信息通信领域产学研合作特色期刊　十佳皖刊
第三届国家期刊奖百种重点期刊　中国科技核心期刊

ISSN 1009-6868
CN 34-1228/TN

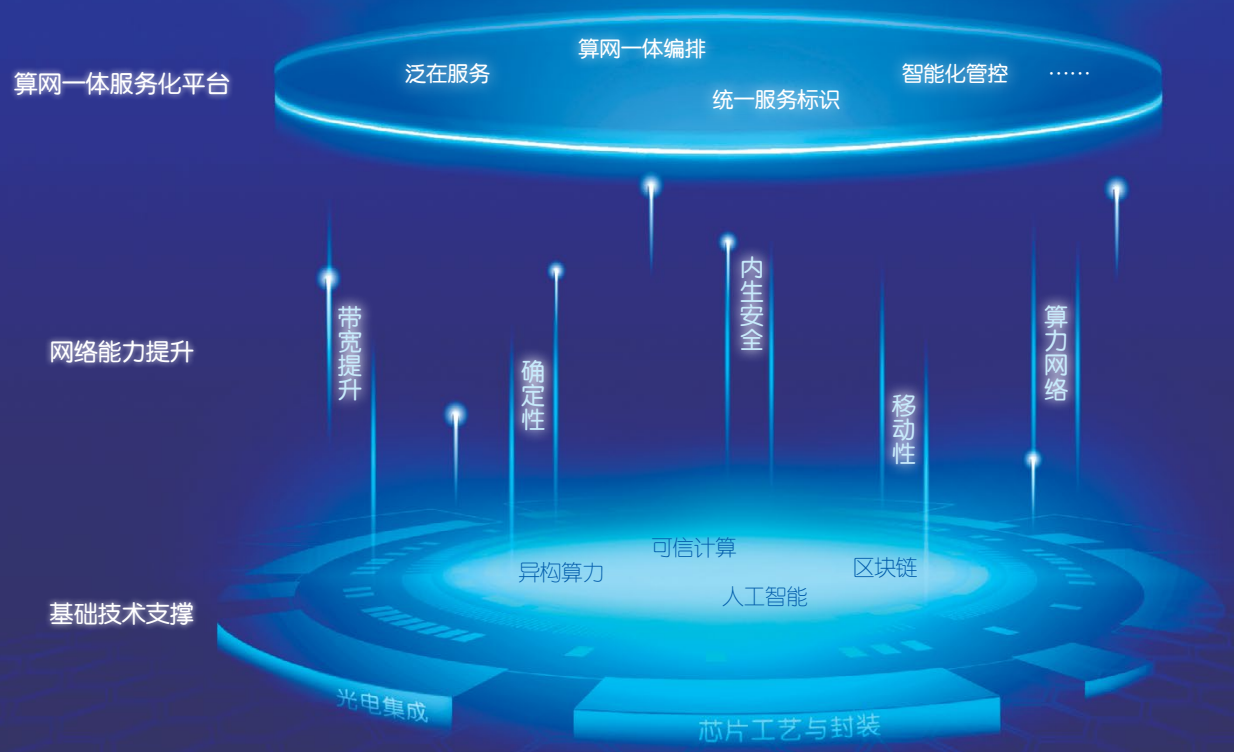
中兴通讯技术

ZTE TECHNOLOGY JOURNAL

http://tech.zte.com.cn

2022 年 2 月 · 第 1 期

专题：新型网络技术



9 771009 686229 02>

《中兴通讯技术》第 9 届编辑委员会成员名单

目　次

中兴通讯技术（ZHONGXING TONGXUN JISHU）
总第 162 期　第 28 卷　第 1 期　2022 年 2 月

信息通信领域产学研合作特色期刊　第三届国家期刊奖百种重点期刊　中国科技核心期刊　工信部优秀科技期刊　十佳皖刊　中国五大文献数据库收录期刊　1995 年创刊

卷首特稿 ▶

01　直面真问题 服务大产业 …… 陆建华

热点专题 ▶

新型网络技术

02　专题导读 …… 唐雄燕
03　IPv6+网络创新体系发展布局 …… 田辉，关旭迎，邬贺铨
08　云网络：云网融合的新型网络发展趋势 …… 史凡
11　基于 SRv6 的算力网络技术体系研究 …… 张帅，曹畅，唐雄燕
16　存转算一体的多模态网络共性平台技术研究 …… 董永吉，胡宇翔，崔鹏帅
21　时间敏感网络中基于网络演算的队列分析与优化 …… 尹淑文，汪硕，黄韬
29　数字孪生网络接口设计及其协议分析 …… 陈丹阳，陆璐，孙滔
34　多样化业务需求与全维网络能力的映射 …… 范琮珊，周旭，任勇毛
41　一种轻量化传输模拟器设计与实现 …… 叶洪波，潘俊臣，崔勇
47　ODICT 融合的网络 2030 …… 王卫斌，周建锋，黄兵

专家论坛 ▶

57　大规模网络向 IPv6 单栈演进的技术方案 …… 解冲锋，李星，李震，余勇志

企业视界 ▶

62　大容量、智能化光传输系统：机遇、挑战与应对策略 …… 冯振华，方瑜，施鹄

技术广角 ▶

70　微服务架构下的算力路由技术 …… 陈晓，黄光平

2022 年第 1—6 期专题计划及策划人

1. 新型网络技术
中国联通研究院副院长 唐雄燕

2. 自然语言预处理模型
中国工程院院士 郑纬民

3. 智能超表面技术
中兴通讯技术预研总工 赵亚军
北京理工大学教授 费泽松

4. 多频段协同通信
电子科技大学教授 李少谦

5. 通信感知一体化
中国科学技术大学教授 卫国

6. 网络内生安全
北京航空航天大学教授 刘建伟

CONTENTS

ZTE TECHNOLOGY JOURNAL
Vol. 28 No. 1 Feb. 2022

Guest Paper

01 Face Real Issues and Serve for Big Industries ····· LU Jianhua

Special Topic

New Network Technology

02 Editorial ····· TANG Xiongyan

03 Development Layout of IPv6+ Network Innovation ····· TIAN Hui, GUAN Xuying, WU Hequan

08 Cloud Network: New Network Development Trend of Cloud Network Convergence ····· SHI Fan

11 Computing Power Network Technology Architecture Based on SRv6 ····· ZHANG Shuai, CAO Chang, TANG Xiongyan

16 PINet Data Plane Platform Technology for Storage, Forwarding and Computing Integration ····· DONG Yongji, HU Yuxiang, CUI Pengshuai

21 Analysis and Optimization of Queues Based on Network Calculus in Time-Sensitive Networking ····· YIN Shuwen, WANG Shuo, HUANG Tao

29 Multi-Protocol Cooperative Interface for Digital Twin Network ····· CHEN Danyang, LU Lu, SUN Tao

34 Mapping Between Diversified Service Requirements and Full-Dimensional Network Capabilities ····· FAN Congshan, ZHOU Xu, REN Yongmao

41 Design and Implementation of a Lightweight Transmission Simulator ····· YE Hongbo, PAN Junchen, CUI Yong

47 ODICT Integrated Network 2030 ····· WANG Weibin, ZHOU Jianfeng, HUANG Bing

Expert Forum

57 Technical Solution of Transition to Large-Scale IPv6-Only Networks ····· XIE Chongfeng, LI Xing, LI Zhen, YU Yongzhi

Enterprise View

62 Towards Large-Capacity and Intelligent Optical Transmission Systems: Opportunities, Challenges, and Solutions ····· FENG Zhenhua, FANG Yu, SHI Hu

Technology Perspective

70 Computing-Power Routing Technologies Under Architecture of Micro-Service ····· CHEN Xiao, HUANG Guangping

期刊基本参数：CN 34-1228/TN*1995*b*16*74*zh*P*¥20.00*6500*14*2022-02

卷首语 >>>

直面真问题 服务大产业

◎ 陆建华 / 中国科学院院士、本刊编委会主任

当前，中国 5G 已大规模商用，6G 尚处于早期研究阶段。6G 的发展须坚持需求导向、问题导向，谨防路径依赖带来的负面效应。《中兴通讯技术》秉承“迎接挑战，把握世界通信技术动态；立即行动，求解通信发展疑难课题；励精图治，促进民族信息产业崛起”的办刊宗旨，倡导“以人为本，荟萃通信技术领域精英”，为 6G 研究发展及产学研用交融提供学术交流平台。

2022 年将是 6G 研究的关键年，《中兴通讯技术》鼓励并欢迎业界同人为共同打造 6G 中国方案贡献真知灼见，内容聚焦但不限于以下几个方面：

（1）从需求定义 6G

习近平总书记在 2020 年 9 月 11 日的科学家座谈会上强调，研究方向的选择应坚持需求导向和问题导向，从国家急迫需要和长远需求出发，真正解决实际问题。6G 需求应来源于未来的市场，6G 研究需要到创新的主战场寻找问题，不能简单地用技术趋势代替应用需求。6G 研究更要关注长远需求，可通过战略研判、发展趋势预测，以明确 6G 愿景需求。6G 研究既要避免空中楼阁，防止脱离社会经济的基本面，也要主动寻求突破，避免受困于技术演进的路径依赖。当前，中国新疆、西藏等边远地区仍普遍缺乏宽带覆盖；北极航道、中国近海等区域宽带覆盖仍然不足，制约了绿色航运和国家海洋战略发展。填补这些“数字鸿沟”理应是 6G 的重要需求。

（2）以系统方法研究 6G

3G、4G 主要服务消费领域，5G 开始关注生产领域。可以预见，6G 将以服务生产领域为重心。消费领域的服务对象是人，这具有一定的统一性，因此 3G/4G/5G 网络可以统一服务框架。而生产领域则不同，其服务对象千差万别，难有统一的框架。以系统方法研究 6G，就要尊重特殊性，客观认识多样性、差异性，变技术统一为实现方法统一。比如，未来 6G 可以探索基站和终端白盒化，仅仅在网络架构和资源组织层面统一规范，从而纲举目张、以简生繁，适应多样化需求。为此，6G 宜重点研究白盒基站、模组化开放终端、开源软件等使能技术，同时结合天地融合网络研究，实现全球智简、开放互联，并以此形成开放共融的 6G 研究新生态。

（3）以务实精神推进 6G

6G 涉及面宽，其应用几乎包罗万象。从哪儿入手，如何入手，如何以点带面循序渐进，都是需要仔细琢磨、认真推敲的。一方面，基础研究要实、要有针对性，需要努力营造基础研究、实验研究、应用研究交叉互动的创新生态。面向“智简”6G，需要强化“大”系统论证，凝练“真”问题，力争实现信息领域基础理论、关键核心技术新突破，牵引范式变革。另一方面，产业推进要稳、要有前瞻性，需要强化产业链各环节的实验验证，积极布局面向 6G 的各类测试仪器、测试系统研究，建立 6G 技术、产品、系统、应用等系列测试床，并充分考虑到“双碳”等国家战略目标，科学、务实地推进 6G 产业发展。同时，面向 6G 创新主战场，还需要积极探索综合性、战略性人才培养新模式。

中国 5G 网络建设规模处于国际领先地位，这体现了国家发展信息通信产业的坚定决心。谋求保持领先的长久大计、发展好 6G 是信息通信人共同的责任。《中兴通讯技术》愿与专家学者们一路同行，以“十年磨一剑”的韧性和毅力，直面真问题，服务大产业，攀登新高峰。

DOI：10.12142/ZTETJ.202201001
收稿日期：2021-12-18

新型网络技术专题导读

专题策划人 >>>

唐雄燕

中国联通研究院副院长、首席科学家，“新世纪百千万人才工程”国家级人选，北京邮电大学兼职教授、博士生导师，工业和信息化部通信科技委委员兼传送与接入专家咨询组副组长，中国通信学会理事兼信息通信网络技术委员会副主任，中国光学工程学会常务理事兼光通信与信息网络专家委员会主任，开放网络基金会（ONF）董事；长期从事通信新技术研发和管理，主要专业领域为宽带通信、光纤传输、互联网/物联网、新一代网络等。

信息通信网络是数字经济的基石，也是新一轮科技革命和产业变革的关键领域。经济社会数字化转型对网络的高质量、确定性、可靠性、安全性、泛在性、智能化等方面不断提出新要求和新挑战，推进网络技术的进步。近20年，新型网络技术研究一直是学术界和产业界关注的焦点，特别是基于开放架构的软件定义网络（SDN）和网络功能虚拟化（NFV）等技术得到了广泛应用，并已成为5G时代的重要网络技术特征。同时，人工智能技术逐步融入网络领域，驱动网络智能化转型。作为网络基础协议的互联网协议（IP）技术也在不断演进，目前已发展到了互联网协议第6版（IPv6）阶段，并开启了IPv6+技术创新。通信技术（CT）、信息技术（IT）、数据技术（DT）和运营技术（OT）日趋融合，传统通信网络逐步迈向智能化综合性数字信息基础设施阶段，实现云网融合与算网一体服务。

网络技术创新呈现多元化趋势，因此本专题从多维度介绍了新型网络技术。《IPv6+网络创新体系发展布局》提出了IPv6+网络创新体系的发展目标，明确了面向2030年的IPv6+网络创新发展路径。结合云网融合发展趋势，《云网络：云网融合的新型网络发展趋势》和《基于SRv6的算力网络技术体系研究》两篇文章，探讨了云网络架构组成及关键技术，并提出了基于SRv6的算力网络技术体系。面向多模态网络需求，《存转算一体的多模态网络共性平台技术研究》提出了一种存转算一体化的数据平面共性网络平台。《时间敏感网络中基于网络演算的队列分析与优化》提出了基于网络演算的循环队列转发（CQF）性能分析方法，实现时间敏感网络中性能和成本的优化。《数字孪生网络接口设计及其协议分析》给出了数字孪生网络构建的通用接口适用性建议，提出了实现孪生层内部接口的多协议协同方法。《多样化业务需求与全维网络能力的映射》提出了全维可定义网络能力模型，从多个维度实现网络能力开放可定义和动态演进发展。《一种轻量化传输模拟器设计与实现》提出了一种支持多种传输功能的轻量化传输模拟器，并利用算法接口抽象提升模拟器的易用性。《ODICT融合的网络2030》阐明了面向2030年的下一代网络是ODICT融合的网络，并介绍了未来网络相关支撑技术。

本专题文章分别由来自中国的高校、科研院所、设备制造商和电信运营商中从事新型网络技术研究的专家学者撰写，较为全面地反映了中国网络技术创新的最新成果。在此，对各位作者的大力支持和精心撰稿表示衷心的感谢！希望本专题能对新型网络技术的进一步研究和发展起到重要参考和积极推动作用。

唐雄燕

2022年2月8日

DOI：10.12142/ZTETJ.202201002
收稿日期：2022-02-09

IPv6+网络创新体系发展布局

Development Layout of IPv6+ Network Innovation

田辉/TIAN Hui[1]，关旭迎/GUAN Xuying[2]，
邬贺铨/WU Hequan[3]

（1. 中国信息通信研究院，中国 北京 100191；
2. 中国移动通信集团有限公司，中国 北京 100007；
3. 中国工程院，中国 北京 100088）
(1. China Academy of Information and Communications Technology, Beijing 100191, China;
2. China Mobile Communications Group Co., Ltd., Beijing 100007, China;
3. Chinese Academy of Engineering, Beijing 100088, China)

DOI：10.12142/ZTETJ.202201003
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20220223.0922.002.html
网络出版日期：2022-02-23
收稿日期：2021-12-20

摘要：在中国互联网协议第6版（IPv6）规模部署全面实施的背景下，提出了IPv6+网络创新体系的发展目标，厘清了其概念的内涵与外沿，形成关键技术规划布局，并据此构建面向2030年的IPv6+网络创新体系发展路径。研究表明，中国应布局网络编程、网络层切片、网络确定性、网络随路测量、新型组播、网络自治以及可信安全等关键技术研究，推动核心设备、系统以及解决方案研发来引导产业发展，实施面向基础电信网络、行业信息网络的试点示范，加快IPv6+国家/国际标准体系的构建。

关键词：IPv6+；网络创新体系；技术布局；发展路径；标准体系

Abstract: China´s Internet Protocal Version 6 (IPv6) has achieved large-scale deployment. The development goal of the IPv6+ network innovation system is put forward, the connotation and external edge of the concept is clarified, the planning of key technologies has been formed, and the development path of the IPv6+ network innovation system for 2030 is constructed accordingly. The research of key technologies including network programming, network slicing, certainty network, with-road measurement, new multicast routing, network autonomy, and credible safety network is arranged. The research and development of core products, systems, and solutions are promoted to guide industrial development, and the pilot project and demonstrations for basic telecommunications networks and industrial information networks are implemented, which accelerates the construction of an IPv6 national/international standard system.

Keywords: IPv6+; network innovation system; technical layout; development path; standard system

互联网是国民经济和社会发展的重要基础设施。当前，在全球范围内，以5G、云计算为代表的新一轮科技革命和产业变革蓬勃兴起，推动了互联网通信模式从人与人通信向物与物通信以及人机交互模式转变，这需要互联网更加弹性、高效、可靠、安全。具体来说，互联网应满足以下需求：

（1）海量连接扩展需求。随着移动互联网、工业互联网、物联网等业务发展，海量异构终端将会接入互联网，这要求网络在具备海量接入能力的同时，还能够保证带宽、时延、抖动等指标要求，并尽量减少与业务特性无关的限制。

（2）灵活流量疏导需求。一方面，在互联网时代，业务流量的爆发式增长已是必然趋势；另一方面，云网融合推动网络流量从南北向传输向东西向流量传输发展，这需要网络具备灵活疏导、智能调度等能力。

（3）便捷网络服务需求。从云的视角来看，计算、存储、网络等功能都要实现便捷的服务化。网络服务化是云网融合对网络联接能力的内在要求，基本内涵包括简化接口、自动化部署、路由可编程、故障快速闭环等。

（4）个性化服务质量需求。智能制造、交通、物流等垂直行业数字化转型，对承载网络提出毫秒级时延和100%可靠性保障等极致服务要求。而传统网络只提供尽力而为服务，不能满足行业差异化和定制化需求。

（5）可信安全保障需求。产业在实现数字化发展的同时也存在新的安全风险，如云服务的虚拟化、数据开放化、松耦合的架构等。而原有的安全防护手段已不适用，因此有必要重建多维度、多领域的信任网络安全架构。

1 IPv6+网络创新体系

近年来，世界主要国家纷纷加强对网络演进创新领域的战略部署，力争在新一轮技术和产业竞争中占据优势。以互联网工程任务组（IETF）、欧洲电信标准化协会（ETSI）、国际电信联盟（ITU）为代表的国际标准组织，不断扩大研

基金项目：国家重点研发计划（2018YFB1800100）

究范围，持续开展新型网络技术的研究和探索。中国也高度关注并重视网络创新演进发展，2017年中共中央办公厅、国务院办公厅联合发布《推进互联网协议第六版（IPv6）规模部署行动计划》[1]，明确了中国IPv6规模部署的总体目标、路线图、时间表和重点任务，提出了“强化网络前沿技术创新”“布局下一代互联网顶层设计”“构建自主技术产业生态”等重点任务。经过4年多的发展，中国建成了全球最大规模的IPv6网络，典型应用和特色应用不断增多，IPv6规模部署取得了显著成效，已具备开展网络创新的坚实基础。2019年，推进IPv6规模部署专家委员会指导成立了IPv6+创新推进组，提出打造IPv6+网络创新体系的战略发展目标，确定用10年左右的时间，以推进IPv6规模部署国家战略为契机，建立可演进创新、可增量部署的IPv6+网络技术创新体系[2]，引领中国IPv6+核心技术、产业能力及应用生态实现突破性发展，并提供以IPv6+系列标准为代表的网络演进创新中国方案，打造赋能数字化转型发展的新型基础设施。

1.1 概念内涵

IPv6+是基于IPv6的下一代互联网的升级，是对现有IPv6技术的增强，是推动技术进步、效率提升，面向新一轮科技革命和产业变革的互联网创新技术体系。基于IPv6技术体系再完善、核心技术再创新、网络能力再提升、产业生态再升级，IPv6+可以实现更加开放活跃的技术与业务创新、更加高效灵活的组网与业务服务提供、更加优异的性能与用户体验，以及更加智能可靠的运维与安全保障。

IPv6+核心技术创新内容包括3个方面：一是在IPv6基础上进行路由转发协议及其功能的增强、完善，例如IPv6分段路由、新型组播技术等；二是IPv6与其他技术的融合应用，例如IPv6与人工智能、软件定义网络等技术融合形成的网络层切片、确定性转发等；三是基于IPv6开展的网络技术体系创新，例如确定性转发、随流检测和应用感知网络等。

除了上述核心技术创新之外，IPv6+还将以网络故障发现、故障识别、网络自愈、自动调优等为代表的智能运维创新作为发展目标，同时将以5G面向企业用户（ToB）、云网融合、用户上云、网安联动等为代表的商业模式创新作为典型融合应用场景。

1.2 外沿关系

业界普遍认为IPv6不是下一代互联网的全部，而是下一代互联网创新发展的起点和平台。IPv6+正是基于IPv6网络技术体系的全面能力升级。借助海量地址和其他重要特性，IPv6成为万物互联的网络基础。IPv6+的技术体系得到了全面升级，可以满足数字化转型的多样化承载需求，它必将推动万物互联走向万物智联。当前以IPv4/IPv6为代表的网络技术体系促进了消费互联网的繁荣，下一步IPv6+可以全面升级网络信息基础设施，必将满足千行百业数字化、网络化和智能化转型发展需求。

从代际演进的角度来看，IPv6+是面向5G和云时代的网络体系创新，是数字化时代的信息基础设施“底座”。未来网络则是以人类可持续发展为目标，解决社会、经济和环境可持续发展问题的信息基础设施，是未来人类社会的“基石”。如果说IPv6+着重关注中近期网络演进创新，那么未来网络的目标则被定位为远期发展。可以预见，IPv6+与未来网络将持续接力，不断提升网络服务能力，全面支撑社会可持续发展。

1.3 关键技术布局

网络技术体系方面的创新具体包括：面向5G承载、云网融合以及产业互联网提出的泛在、多元、弹性、高效、可靠、可信的承载需求，并基于IPv6开展协议创新，研究多样灵活的分段路由控制机制[3]，实现业务的快速开通、跨域互通、业务隔离、可靠保护；研究简化控制的网络编程机制[4]，以提高协议运行效率，减少协议开销[5]，降低维护复杂度；研究泛在连接的差异化服务级别协议（SLA）技术，提供有时延、抖动、丢包边界保障的确定性能力[6]；研究大规模网络层切片技术[7]，提供可交付、可测量、可度量、可计费的切片服务；研究带内遥测的随路测量技术[8]，支持异构网络扩展、轻量开销、协议健壮的网络状态数据采集；研究应用特征的网络感知机制[9]，根据用户、业务以及性能参数要求，进行无缝融合、后向兼容、可扩展性、无状态依赖的精细化运营。IPv6+网络创新体系技术布局如图1所示。

智能运维体系方面的创新具体包括：针对网络长期处于人工为主、半自动运维为辅，且对网络运行状态缺乏感知的现状，研究在IPv6+网络体系中以数据为核心、构建物理网络的数字孪生，支持基于模型驱动的网络服务创新；研究网络能力开放编程技术，将高度抽象的网络服务接口化，向用户业务开放，使用户能像调用计算和存储资源一样方便地调用网络资源；研究在运维体系中引入人工智能技术，开展智能网络资源编排、流量预测分析、网络信息安全、用户行为分析，实现被动运维到主动运维的转变；研究网络故障智能发现、识别、定界的优化闭环技术，使能自动、自优、自愈、自治的自动驾驶网络。

网络商业模式方面的创新具体包括：借助IPv6+路径可规划、业务速开通、运维自动化、质量可视化、SLA可保

障、应用可感知的特性，开展5G、云计算及产业互联网融合应用场景创新，并研究5G园区海量终端延伸到云端的场景，实现业务的高速接入、分片隔离、快速上云和业务质量保障；研究企业不同业务使用多云联接的场景，根据业务时延、带宽、可靠性等要求灵活选择网络路径，实现云网资源的统一调度；研究工业互联网全IP化场景需求，构建联接工业园区、工业云平台、工业内网的高质量网络设施，确保业务不绕路、不断网、不丢包、不延误，满足确定性服务需求。IPv6+典型融合应用场景如图2所示。

1.4 能力纬度

基于IPv6技术体系的全面演进与创新，IPv6+从超宽、广联接、确定性、低时延、自动化、安全可信等6个维度，大幅提升信息网络基础设施的整体服务能力，这必将有力支撑千行百业数字化转型与创新。

超宽能力持续释放宽带能力以应对未来业务不确定性的挑战。端到端高速连接覆盖从接入网络、骨干网到数据中心网络，承载千亿联接和万物上云的数字洪流。

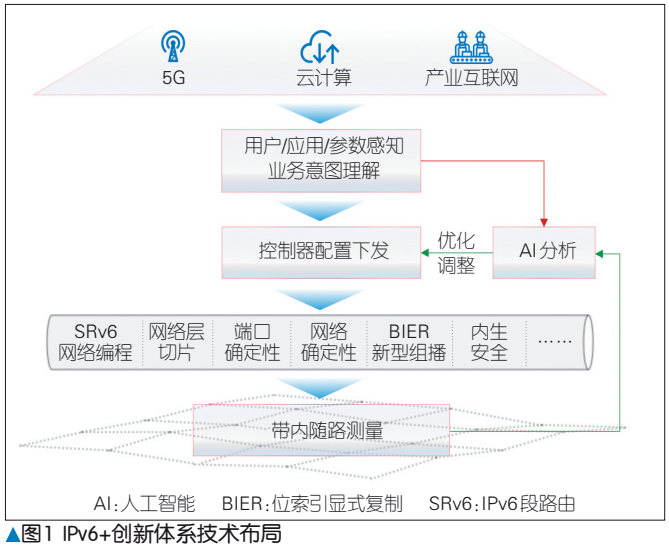
广联接能力提供灵活多业务承载和网络服务化能力。利用网络编程技术，该能力实现端到端流量调度、协议简化和用户体验保障，满足多业务融合承载体验需求。

确定性能力为网络提供可预期的确定性体验。该能力可以利用网络切片技术提供高安全、高可靠、可预期的网络环境，实现微秒级抖动，并可利用无损网络技术实现数据中心零丢包。

低时延能力提供人与虚拟世界实时交互的沉浸式体验。在该能力的支持下，园区网络端到端时延达到毫秒级，数据中心网络端网协同时延达到微秒级。

自动化能力使能自动、自愈、自优、自治的自动驾驶网络。该技术结合人工智能、随流检测、知识图谱等关键技术可以实现异常智能分析，将故障恢复时间从小时级缩短到分钟级。

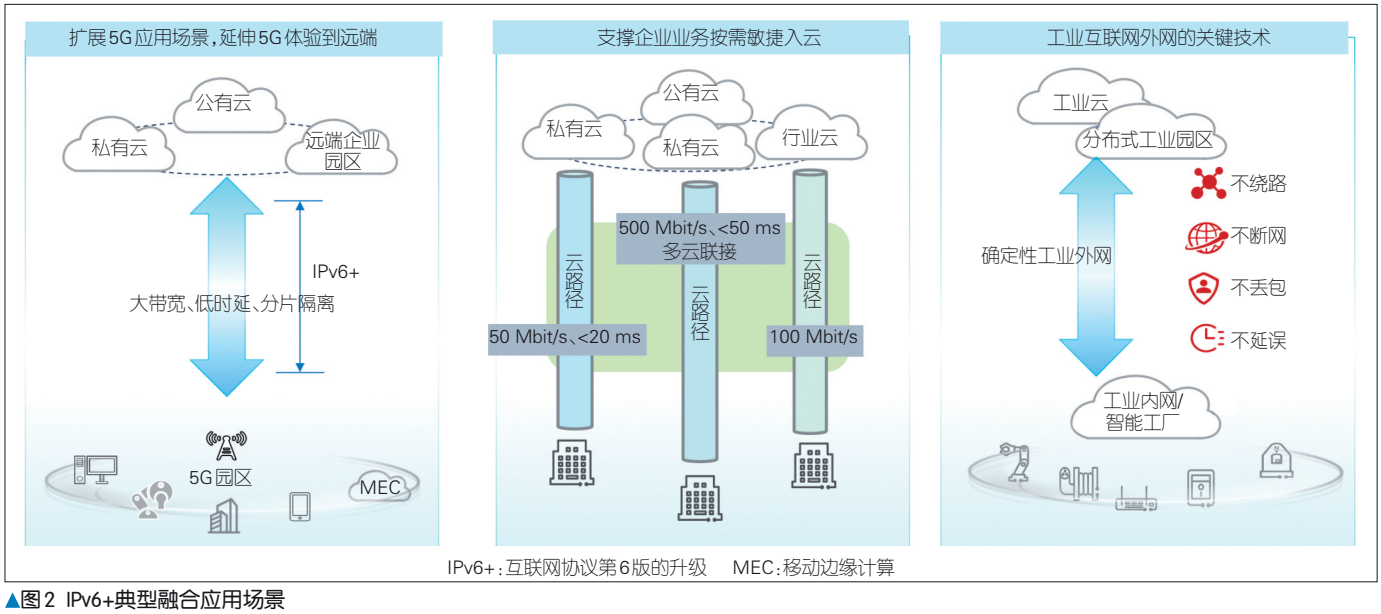
安全可信能力为网络打造内生安全体验。对所有访问进行认证和鉴权，限制最小访问权限。基于云网一体威胁协同处置，实现从小时级到分钟级的威胁遏制。



▲图1 IPv6+创新体系技术布局

1.5 发展路径

IPv6+技术体系演进大致划分为3个发展阶段，具体如



▲图2 IPv6+典型融合应用场景

图3所示。

IPv6+1.0：重点开展技术体系创新，构建网络开放编程能力，通过发展基于IPv6转发平面的段路由（SRv6）实现对传统多协议标记交换（MPLS）、网络基础特性（虚拟专用网）、尽力而为业务（BE）、流量工程（TE）和快速重路由（FRR）等[10]的替代，实现业务快速发放、灵活路径控制，利用自身的优势来简化IPv6网络的业务部署。

IPv6+2.0：重点通过智能运维创新，提升用户体验，并通过发展网络切片/随流检测/新型组播/无损网络等技术，提升算力，优化体验。该阶段需要发展面向5G和云的新应用，如面向5G ToB的行业使能、云虚拟现实（VR）/增强现实（AR）、工业互联网，以及基于数据/计算密集型业务，如大数据、高性能计算、人工智能计算等。这些应用体验的提升需要引入一系列新的创新，包括但不局限于网络切片、随流检测、新型组播[11]和无损网络等。

IPv6+3.0：重点通过商业模式创新，发展应用驱动网络。一方面，随着云和网络的进一步融合，需要在两者之间设置更多的信息交互，也需要将网络能力更加开放地提供给云来实现应用感知和即时调用；另一方面，随着多云的部署加速，网络需要更加开放的多云服务化架构来实现跨云协同和业务的快速统一发放和智能运维。

2 IPv6+标准工作进展

2.1 国家/行业标准体系

2021年，中国通信标准化协会牵头成立了IPv6标准工作组，汇聚各方力量，统筹推进IPv6国家标准、行业标准和团体标准的制定。工作组计划用5年时间形成较为完善的IPv6+标准体系，并持续提升标准对细分行业及领域的覆盖程度，提高跨行业网络应用水平，保障数字经济快速发展。规划中的IPv6+标准体系包括基础创新类、网络安全类、行业应用类、监测评价类标准。

基础创新类标准是IPv6+网络适应5G、云等应用发展，发挥价值的基础性、指导性和通用性标准，包括总体、基础特性与增强的技术规范、关键业务的技术规范、操作维护管理（OAM）与保护技术规范、传统承载与云网融合技术规范、网络应用感知技术规范等。

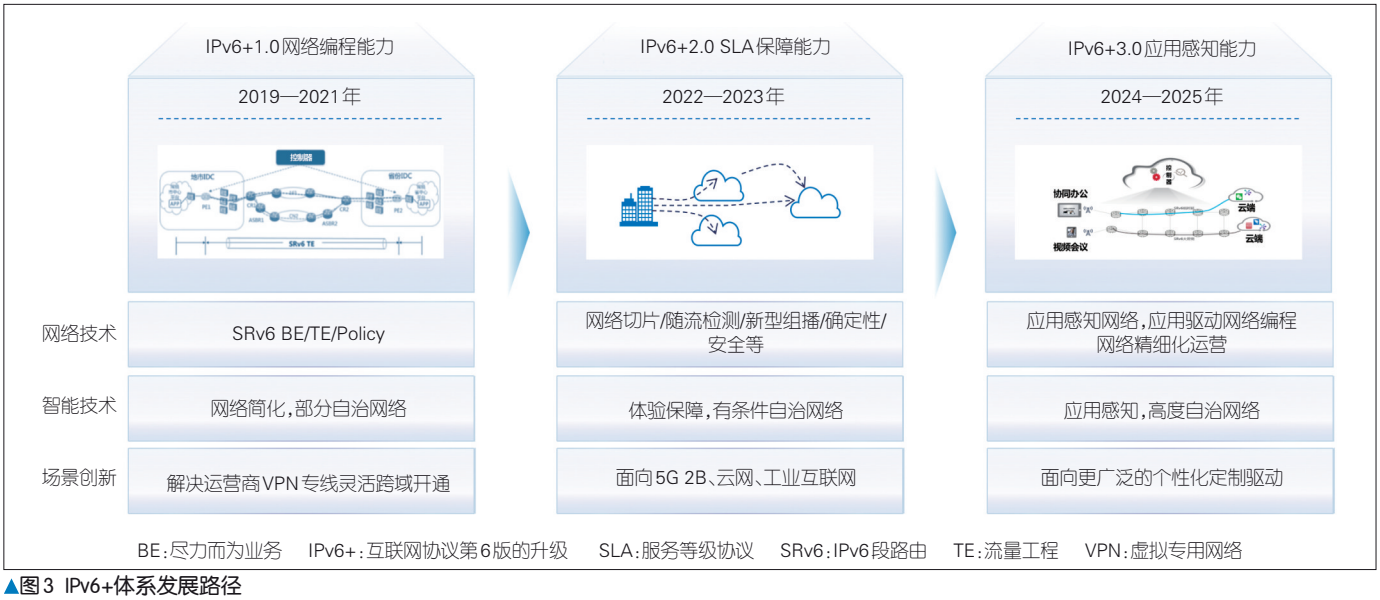
网络安全类标准是IPv6+网络基础的安全基石，包括网络设备通用安全技术要求、骨干/边缘路由器设备网络安全技术要求、数据中心/园区交换机设备网络安全技术要求、网络安全设备IPv6网络安全技术要求等。

行业应用类标准是IPv6/IPv6+网络在主要产业网络部署落地的指南和规范，主要包括金融行业应用标准、能源行业应用标准、交通行业应用标准、教育行业应用标准、政务行业应用标准等。

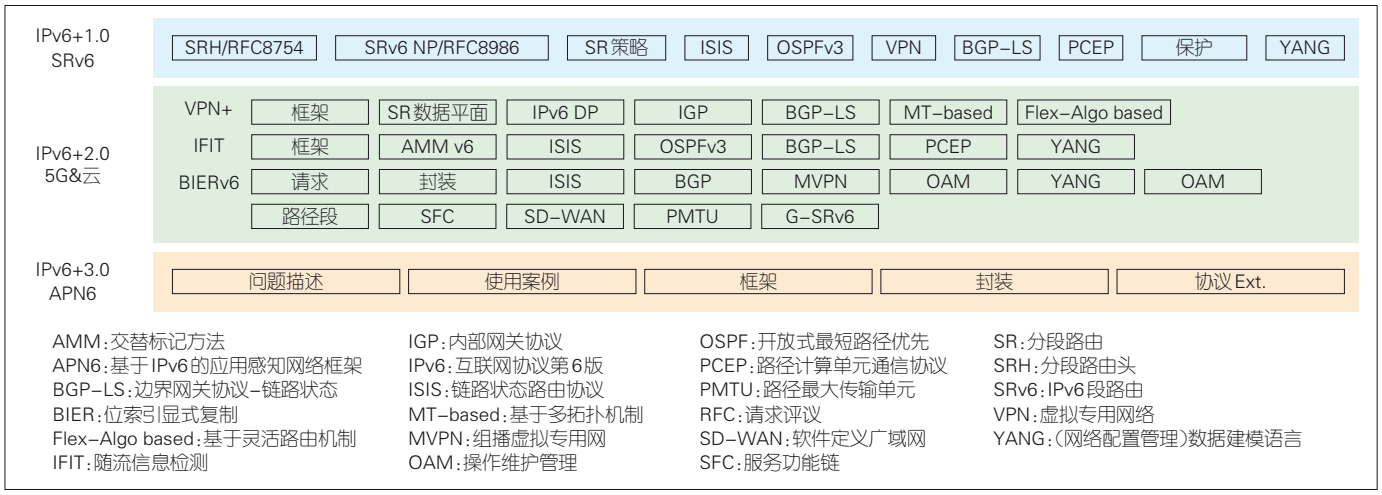
监测评价类标准是IPv6/IPv6+网络服务质量的统一评价规范，指导着IPv6网络建设、运行、维护，主要包括用户、流量、网络浓度标准测试方法、应用浓度标准测试方法、终端浓度标准测试方法等。

2.2 国际标准分布

IPv6+标准的相关工作正在互联网工程任务组



▲图3 IPv6+体系发展路径



▲图4 IPv6+国际标准分布

(IETF)[12]、电气与电子工程师协会(IEEE)、欧洲电信标准化协会(ETSI)等标准组织中有条不紊地展开，国际标准分布如图4所示。在多个技术方向上，中国标准已经与国际标准呈现齐头并进的态势，特别是一些与新应用、新场景结合紧密的方向上，中国标准创新已经走在世界前沿。

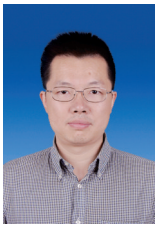
4 结束语

加快IPv6+技术创新、产业发展和应用部署，有利于重塑中国互联网创新体系，激发创新活力，培育新兴业态。这对打造IPv6规模部署、应用高质量发展新优势、加快互联网演进升级、助力经济提质增效具有重要意义。下一步，建议相关部门强化政策引导，统筹各方力量，完善IPv6+技术体系顶层设计，并围绕IPv6+关键技术、核心产品及解决方案等加强测试验证、试点示范，提升创新成果转化水平，增强自主创新能力，形成中国在网络技术演进创新领域的先发优势。

参考文献

[1] 推进互联网协议第六版(IPv6)规模部署行动划 [EB/OL].(2017-11-26)[2021-12-10]. http://www.gov.cn/zhengce/2017-11/26/content_5242389.htm
[2] 田辉, 魏征. "IPv6+"互联网创新体系 [J]. 电信科学, 2020, 36(8): 2-10. DOI: 10.11959/j.issn.1000-0801.2020256
[3] FILSFILS C, DUKES D, PREVIDI S, et al. IPv6 segment routing header: RFC8754 [S]. IETF, 2020
[4] FILSFILS C, CAMARILLO P, LEDDY J, et al. Segment routing over IPv6 (SRv6) network programming: RFC 8986 [S]. IETF, 2021
[5] CHENG W, LI Z B, LI C, et al. Generalized SRv6 network programming for SRv6 compression [EB/OL]. (2021-10-25) [2021-12-02]. https://datatracker.ietf.org/doc/draft-cl-spring-generalized-srv6-for-cmpr/
[6] FINN N, THUBERT P, VARGA B, et al. Deterministic networking architecture: RFC8655 [S]. IETF, 2019
[7] DONG J, BRYANT S, LI Z Q, et al. A framework for enhanced virtual private networks services [EB/OL]. (2018-12-17)[2021-12-10]. https://datatracker.ietf.org/doc/draft-ietf-teas-enhanced-vpn/
[8] SONG H, QIN F, CHEN H, et al. In-situ flow information telemetry [EB/OL]. (2021-10-21) [2021-12-05]. https://datatracker.ietf.org/doc/draft-song-opsawg-ifit-framework/
[9] LI Z, PENG S P, VOYER D, et al. Application-aware networking framework [EB/OL]. (2021-10-25)[2021-12-10]. https://datatracker.ietf.org/doc/draft-li-apn-framework/
[10] CHEN H, HU Z, CHEN H, et al. SRv6 midpoint protection [EB/OL]. (2021-12-19) [2021-12-20]. https://datatracker.ietf.org/doc/draft-chen-rtgwg-srv6-midpoint-protection/
[11] MCBRIDE M, XIE J, DHANARAJ S, et al. BIER IPv6 requirements [EB/OL]. (2021-04-01) [2021-12-10]. https://datatracker.ietf.org/doc/draft-ietf-bier-ipv6-requirements/
[12] 李振斌, 赵锋. "IPv6+"技术标准体系 [J]. 电信科学, 2020, 36(8): 11-20. DOI: 10.11959/j.issn.1000-0801.2020258

作 者 简 介



田辉，中国信息通信研究院技术与标准所互联网中心主任；从事数据通信领域标准和测评技术研究；主持多家基础电信运营企业的互联网、IP承载网建设工程咨询工作；负责20余项数据通信领域国家/行业标准的制定工作，主持5项国家专项研究工作。

关旭迎，中国移动通信集团有限公司研究院物联网技术与应用研究所研究员；从事物联网、车联网、4G/5G无线通信网络等关键技术及应用体系研究；先后负责、参与国家发展改革委、工业和信息化部等国家重大专项研究工作。

邬贺铨，中国工程院院士，曾任中国工程院副院长，现任推进IPv6规模部署专家委员会主任；长期从事数字和光纤通信系统的研发工作，10余年来先后负责"中国下一代互联网示范工程""新一代宽带无线移动通信网"重大专项等国家重大科技战略的咨询工作；曾获全国科学大会奖、国家科技进步奖等多个奖项；出版专著1部。

云网络：云网融合的新型网络发展趋势

Cloud Network: New Network Development Trend of Cloud Network Convergence

史凡/SHI Fan
（中国电信集团有限公司，中国 北京 100033）
(China Telecom Co. , Ltd. , Beijing 100033, China)

DOI：10.12142/ZTETJ.202201004
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20220217.1742.008.html
网络出版日期：2022-02-18
收稿日期：2021-12-25

摘要：云网融合已经成为信息基础设施的核心特征，其关键是打破云和网的边界，实现多个层面的一体化。在信息世界以云为核心的背景下，改变传统的网络组织模式，构建云网络的形态，将是必然选择。重点阐述了对云网络的需求、云网络的架构组成及关键技术，并分析了云网络的表现形式和未来发展前景。

关键词：云网络；云网融合；软件定义广域网

Abstract: As cloud network convergence is the core feature of the information infrastructure, the key challenge is to break the border of cloud and network and bring them together on several levels. It is time for us to change the traditional networking mode, and build up the cloud network, as the core of the information world is cloud. The requirement, the architecture, and the key technologies of the cloud network are highlighted, as well as its existed service type and promising future.

Keywords: cloud network; cloud network convergence; software-defined wide area network

2021年10月18日，习近平总书记在主持中共中央政治局第三十四次集体学习时指出，要加快新型基础设施建设，加强战略布局，加快建设高速泛在、天地一体、云网融合、智能敏捷、绿色低碳、安全可控的智能化综合性数字信息基础设施。这标志着云网融合正式成为数字信息基础设施建设的重要内容。

云网融合的核心特征在于“融”，即打破现有云和网相对独立和隔离的局面，在基础架构、底层设施和资源调度等方面趋于一体化。该特征在反映到现实世界中时，将演变为“云网络”这一新型网络形态。这种网络形态将使得云和网的传统边界变得模糊，使云和网在连接层面率先实现融合。

1 云网络的产生背景

过去的网络组织形式，本质上是一种以基础网络和物理连接为核心的模式，即“先修路”。这种模式根据行政地域层级和人口分布，构建多层级、中心化的网络，然后再将对应的业务和应用资源挂接到网络上。在过去信息通信的源头和形态相对单一、通信流量和数据总量较少的情况下，这种组网形式是便于管理、行之有效的。但是随着互联网的兴起和发展，信息服务的内容逐渐变得丰富多样，网络流量和应用数据呈现爆发式增长。近年来，以云计算技术和云服务载体为代表的新型业务占据主导。业务和应用所需的资源逐步集中到平台化的云上。原先的模式难以满足新形势的需求。以电信运营商为代表的基础网络服务商，虽然逐年提升网络带宽，但仍然难以满足业务流量动态变化（特别是云）的要求。这是因为，一方面现有网络的拥堵点难以消除，另一方面网络也无法结合应用的需求来提供相应合理的服务质量（QoS）/服务等级协议（SLA）。以互联网云商为代表的OTT（指互联网公司越过运营商开展业务）企业，虽然努力构建以Overlay网络为主的网络体系，但是由于其Underlay网络较为封闭，无法和Overlay网络形成高效的协同，因此无法提供与云资源动态弹性、按需服务、按量计费等相匹配的网络能力。

在未来相当长的一段时间内，业务和应用所需的资源主要以云的形态存在，并且该形态是分布式的（在某些情况下甚至是去中心化的）、高频度动态变化的。传统的“先修路再部仓库”的组网模式必须适应新的变化。因此，网络的组织和构成模式需要调整为“随云而动、应云而生”，即需要构建一个云网络。这种云网络的本质特征主要有以下4个：

（1）网络组织以云为核心。网络的布局和架构充分匹配云计算和云业务所需的灵活性，具有高度的弹性，能够提供不同等级应用的QoS。

（2）网络资源云化部署。网络的节点、带宽、流量等打破了与地理位置和物理形态绑定的局面。以网元为主要载体可实现虚拟化、云化部署，能够实现按需提供和调整。

（3）网络边界深入云内。网络连接的端点需要与云业务相关联，以实现信息传送的深度直达，即让信息的“包裹”能够送到用户手中。

（4）网络服务与云融合。从最终用户的角度看，今后大量的应用直接调用的是数据和算力。承载这些数据和算力的载体是云计算和云资源。而网络是作为更加底层的连接支撑存在的。从感知的角度看，理想的网络模式应该是“见云不见网”。网络自动化可提供服务，但它“隐藏”在云的后面。

2 云网络的构成和关键技术

新型云网络的架构包括3个层面。

（1）基础网络层。这一层的作用是构筑一个泛在的高速连接基础。它类似于我们日常生活中的高速公路，但需要具备一定的弹性能力和快速调整能力，能提供一定的差异化QoS和网络开放能力。基础网络层的布局要实现去行政区域化，并根据云资源的布局来设计。基础网络层主要解决网络组织以云为核心、网络资源云化部署的问题，需要从网络拓扑、路由组织、协议选择等维度更新传统网络的设计理念。从传统运营商的视角看，基础网络层又是Underlay网络层。

（2）业务网络层。在基础网络层上有一个业务网络层。业务网络层的作用是根据云计算的需要，实时建立或拆除网络连接，按需提供网络带宽和质量保障。同时这种连接是深入到应用和最终用户的，是真正端到端的连接。相对于基础网络层提供的连接能力，业务网络层实现的网络连接更加细粒度、精准化，并与日常生活中的物流快递类似，一般具有高并发、高时效性。业务网络层主要解决的是网络资源云化部署、网络边界深入云内的问题。简化网络的层级和拓扑、路由组织和接续，并且让网关等节点实现虚拟化甚至云化[1]，有助于实现按需扩（缩）容和随云部署。从传统运营商的视角看，业务网络层也被称为Overlay网络层。

（3）网络导航层。除了上述两个网络层外，系统还需要一个集中控制的网络导航层。这一层的作用在于让云网连接的准确性和可靠性得到提升，使信息物流的效率达到最大。网络导航层主要解决的是网络服务与云融合的问题，可形成多维全域资源视图，为不同的应用和业务设计相应的网络策略，并结合实际资源效能形成最优的调度和配置。该层是新型网络的核心智能所在，具有统一调度和管理的职责。从传统运营商的视角看，网络导航层也可以被称为控制编排层。

上述不同的层面需要引入不同的技术来发挥各自的作用，以实现各层的功能定位。

具体来说，基础网络层要实现比传统的Underlay网络更好的灵活性和差异化性能，就需要在协议层进一步简化，方便基础网络实现端到端的可管可控，改变过去“铁路警察各管一段”的协议跨域拼接状况。目前，最有代表性的是基于IPv6的段路由（SRv6）协议[2]和以太网虚拟专用网（EVPN）技术。SRv6可以让广域网过去的多个协议简化为一个协议，有助于实现“高速路的一站直达”；EVPN技术可以让适合不同客户的L2虚拟专用网（VPN）和L3 VPN实现统一发放，大大简化网络的开通配置。同时，在该层面上还存在着光传输和互联网协议（IP）两个承载层。在现实网络中，出于不同质量、经济性、安全性的考虑，光传输和IP网络各有优势，两者将长期并存。但是目前最缺乏的是两者之间的高效协同，特别是在连接组织方面。人为的设定超出了光网络和IP网络的界限。从云网络承载的角度看，只有让IP+光传输成为一个整体网络，才能发挥网络最大价值。因此，业界争论多年的“彩光”技术将为IP和光网络的融合带来新的机遇。由于云资源的一大功能是实现对算力的承载，特别是考虑到未来大量应用需要同时使用存储和计算两种能力，因此基础网络层尤其需要匹配这一能力需求，需要能够形成“服务器（计算和存储）+传输+路由”的统一调度能力。

业务网络层拥有一端入云、另一端连接最终用户的海量信息分发能力。这是传统Underlay网络所不擅长的。此外，该层还需要具有面向应用的高度定制化能力和快速连接处理能力。为此，SRv6和快速用户数据报协议（UDP）网络连接（QUIC）协议[4]可用来满足相关要求。其中，SRv6提供的是Overlay层面的精细化连接能力。与Underlay网络采用的SRv6协议不同，该层面可以根据网络端点的能力，采用进一步简化的SRv6协议栈，并通过开源方式来部署。如果诸如Sonic、FRRouting这样的开源项目能够满足需求的话，系统还需要专用的网络设备。这将大大降低SRv6部署到云内端点的要求。而随着基础网络质量的不断提升，传统的传输控制协议（TCP）3次握手过程的必要性已经很小，大量基于UDP的应用将应运而生。为此，QUIC协议的高效性愈发明显，将有助于入云数据“包裹”的及时送达。当然，需要指出的是，目前QUIC协议主要应用在Client–Server模式中，其对传统网络的潜在影响也是巨大的。由于不同网元还处在逐步云化的过程中，各个网络节点的传输层协议完全可以基于QUIC协议来简化和优化。比如，当基础网络层采用移动网络承载并遇到频繁跨区切换时，采用QUIC来构建业务网络层就能够规避TCP慢启动对业务和应用产生的问题。

网络导航层需要借助软件定义网络（SDN）、Telemetry

等技术来实现对各种网络资源信息的实时采集，并结合人工智能（AI）和大数据技术，利用机器学习等手段实现智能化的信息处理和闭环控制，让云网络真正具有大脑功能。需要指出的是，网络自动驾驶技术并非这一层面的最大挑战。在云计算领域中，相关的自治技术已经非常成熟，完全可以借鉴到网络中来。网络导航层的技术难点是如何引入语义化的信息交互和数据处理，以便让云网络更加开放，让使用者方便调用。对此，目前业界还没有现成的技术可用。

需要特别指出的是，这3个层面是逻辑功能的不同分工，在物理实现上，不是绝对隔离的，而是彼此嵌入的。以入云专线为例，在城域网范围内，基础网络层可以是无源光网络（PON）、光传输网（OTN），也可以是5G网络。为便于调度，可以在此层叠加一个软件定义广域网（SD-WAN）作为业务网络层，以方便客户侧与云资源池按需连接的建立和拆除。在骨干网范围内，可构建一个需要跨越不同网络域的数据中心互联（DCI）。从互联网数据中心（IDC）互联的角度看，该DCI属于一个基础网络；但是从云资源池互联和互通的角度看，该DCI又可以起到业务网络层的作用。与此同时，位于网络导航层的平台系统，比如编排和控制系统，可能需要连接上述不同网络域内的多个控制器、云管平台等，以实现端到端的协同和能力开放。

3 云网络的现状和发展前景

其实，从业界的应用部署来看，SD-WAN[5]就是一种云网络形态。SD-WAN具备上述特征并包含上述3个层面。

从网络节点的形态来看，SD-WAN在局端和用户侧都可以采用云化部署的方式，不局限于传统的专用物理设备，它采用的组网拓扑和局端部署都呈现出典型的去中心化特征。同时，与云的结合是SD-WAN高速发展的一大驱动力，也是企业实现上云、云互联的重要手段。如果将云比作电商的平台，那么SD-WAN就像物流服务一样。SD-WAN与云的深度融合没有明确业务边界。近来，随着安全访问服务边缘（SASE）、软件定义分支（SD-Branch）的加持，SD-WAN被进一步应用在信息通信技术（ICT）服务中。换句话说，SD-WAN已经从单一的组网连接演变为综合性信息服务。因此，我们也可以认为它是新型云服务的一种。

SD-WAN在实际部署中，并不是一种新型的物理网络形态，而是基于各种现有的接入手段和城域网、骨干网衍生而来的。SD-WAN的Underlay部分可以是目前的互联网专线、互联网宽带，也可以是传输专线，甚至是5G无线接入。所以我们不能认为SD-WAN的网络带宽等同于互联网带宽。在SD-WAN的Overlay部分，通过在客户侧网关和服务侧的网关建立安全封装后的隧道，可实现点到点、点到多点的灵活组网。结合SDN控制器和编排器，SD-WAN能够实时感知底层资源的变化，根据业务的QoS要求来智能选择最佳路径并实施差异化保障手段，使得业务的开通与变更以及服务保障更加灵活、更有性价比。相对于传统的专线，SD-WAN如果选择各种专线类技术，例如将OTN、多协议标签交换（MPLS）VPN作为其Underlay部分（也就是基础网络层），就可以实现很高的网络带宽和QoS保障。出于性价比和可靠性的综合考虑，在党政军和大企业应用中，SD-WAN的基础网络层经常出现专线+互联网互为备份的情况。

当然，除了SD-WAN以外，云网络还有不同的表现形式。比如，多家运营商推出的多云互联服务，多家OTT提供的多网接入服务。这些不同的表现形式在实现云网络的能力上存在一定的差异。

4 结束语

在云网融合的大势趋下，传统网络需要升级为云网络。这种云网络不仅是一种新型的网络形态或者云网载体，还是一种云网能力服务化的创新模式。在技术上，云网络本质上既打破了传统网络和业务的界限，又打破了传统通信技术（CT）和信息技术（IT）的界限，是云网融合能力和服务的供给侧结构性改革体现。当然，云网络的实现方案和可选技术是多样化的，相关的具体应用还处于比较初期的阶段。目前业界还没有找到绝对理想的标准方案和部署模式。正因为如此，业界同人需要共同关注，一起携手，才能解决相关问题，让云网络真正成为助推云网融合的新动能。

参考文献

[1] 周家荣. 基于云网PoP的弹性边缘网络设计思路探讨 [J]. 电信工程技术与标准化, 2021, 34(9): 31-38, 68. DOI: 10.13992/j.cnki.tetas.2021.09.006
[2] 王巍, 王鹏, 赵晓宇, 等. 基于SRv6的云网融合承载方案 [J]. 电信科学, 2021, 37(8): 111-121. DOI: 10.11959/j.issn.1000-0801.2021205
[3] 黄光平, 史伟强, 谭斌. 基于SRv6的算力网络资源和服务编排调度 [J]. 中兴通讯技术, 2021, 27(3): 23-28. DOI: 10.12142/ZTETJ.202103006
[4] 李学兵, 陈阳, 周孟莹, 等. 互联网数据传输协议QUIC研究综述 [J]. 计算机研究与发展, 2020, 57(9): 1864-1876. DOI: 10.7544/issn1000-1239.2020.20190693
[5] 史凡. SD-WAN服务持续升级 带来云网络新业态 [J]. 通信世界, 2021(14): 36-38. DOI: 10.13571/j.cnki.cww.2021.14.013

作　者　简　介

史凡，中国电信集团有限公司科技创新部技术战略处处长，高级工程师，城域以太网论坛（MEF）中国工作组主席，中国通信学会首届青年科技奖获得者；从事电信网络技术科研工作20余年；发布全球标准20余项，拥有专利30余项，出版专著4部。

基于SRv6的算力网络技术体系研究

Computing Power Network Technology Architecture Based on SRv6

张帅/ZHANG Shuai， 曹畅/CAO Chang，
唐雄燕/TANG Xiongyan
（中国联合网络通信有限公司研究院，中国 北京 100048）
(The Research Institute of China Unicom, Beijing 100048, China)

DOI：10.12142/ZTETJ.202201005
网络出版地址：https://kns. cnki. net/kcms/detail/34.1228. TN. 20220224.0944.002.html
网络出版日期：2022-02-25
收稿日期：2021-12-20

摘要：计算形态的云-边-端泛在化分布的演变，以及进入网络内部的计算，都推动承载网络从“云网融合”向“算网一体”演进。面向算力网络时代，提出基于互联网协议第6版段路由（SRv6）的算力网络技术体系，并详细介绍其内涵——“1+*N*+*X*”，即1个算网能力底座，*N*种算力网络能力，*X*种商业场景。通过对此技术体系的研究，构建新一代数字技术设施，增强“联接+计算+智能”的网络内生能力。

关键词：云网融合；算力网络；技术体系；SRv6

Abstract: The evolution of the cloud-edge-end ubiquitous distribution of computing forms and computing power embedded in the network promotes the evolution of carrier networks from "cloud and network convergence" to "computing power and network integration". Facing the computing power network era, computing power network technology architecture based on segment routing IPv6 (SRv6) is put forward, and its contents——“1+*N*+*X* ” are introduced in detail. “1+*N*+*X* ” means 1 computing power network capacity base, *N* kinds of computing power network capacity, and *X* kinds of business scenarios. Through the research of the technology architecture, the new generation of digital technology facilities is established, enhancing the network endogenous capacity of “Connectivity + Computing + Intelligence”.

Keywords: cloud network convergence; computing power network; technology architecture; SRv6

1 数字经济时代，“联接+计算+智能”成为数字新业态的基石

在“十四五”时期，互联网、大数据、人工智能（AI）同各行各业深度融合，产业数字化迅速发展，这对经济社会发展起到重要的战略作用。在移动互联网、大数据、超级计算、传感网、脑科学等新理论新技术的驱动下，AI加速发展，成为新一轮科技革命和产业变革的重要驱动力。加快传统产业数字化改造，就要推进“上云、用数、赋智”。这既是提高国家竞争优势的战略选择，也是构建新发展格局的现实需要，更是实施创新驱动发展战略的主要抓手。

2021年3月，中国联通正式发布CUBE-Net 3.0网络转型计划[1]，旨在携手合作伙伴共同构建面向数字经济新需求，增强网络内生能力，实现“联接+计算+智能”融合服务的新一代数字基础设施，积极推动网络从“云网融合”迈向“算网一体”。

在新架构中，通过支持“联接+计算+智能”的融合服务，结合互联网协议第6版升级（IPv6+）创新技术，网络具备了感知、智能、体验保障等新能力，这些能力组合后形成智能联接。网络不仅可以提供透明的数据交换，还可以为行业客户提供边缘计算服务，为家庭提供智能业务体验保障，为个人提供高精度定位等融合感知服务。

面向算力网络时代，中国联通提出了基于IPv6段路由（SRv6）的算力网络技术体系，提供算网一体服务，即“1+*N*+*X*”：1个算网能力底座，*N*种算力网络能力，*X*种商业场景，构建新一代数字技术设施，增强“联接+计算+智能”的网络内生能力[2]。

2 1个算网能力底座，实现“一网联多云，一键网调云”

基于“1+*N*+*X*”的技术体系，网络是以用户为中心的。从用户的视角看，一网多云需要网络支持低时延、安全可信的通信，以及高确定性的服务质量，因此网络成为价值中心。算网能力底座的构建，需要结合源路由IPv6网络可编程能力，通过将Service ID和SRv6段识别信息（SID）关联，进行联合注册和统一编排。

SRv6是一项新兴的互联网协议（IP）。SRv6通过灵活的

Segment组合、Segment字段，以及类型长度值（TLV）组合实现3层编程空间，更好地满足新网络的业务需求，而其兼容IPv6的特性也使得网络业务的部署更为简便。

在算力网络中，SRv6技术可以简化网络结构，实现网络之间的无缝衔接。同时，路由能力不再割裂，并可以结合网络软件定义网络（SDN）将能力开放出来，实现资源互调。这样可以大幅降低业务连接和业务部署的复杂度，使得网络服务化具备必要条件，即将网络的能力和数据进行开放，支撑业务实现实时、按需、动态部署。网络服务化能力的提供，首先需要对网络能力进行抽象，以服务的形式对外提供，这可以包括网络资源连接服务、切片服务、服务级别协议（SLA）可视等多种服务类型；其次，网络服务化需要能够灵活定义、敏捷编程，并能根据业务场景对网络的各项原子能力进行组装，对运维活动进行灵活编排，实现流程自动化；最后，网络服务化能力需要自闭环，实现服务状态实时感知。

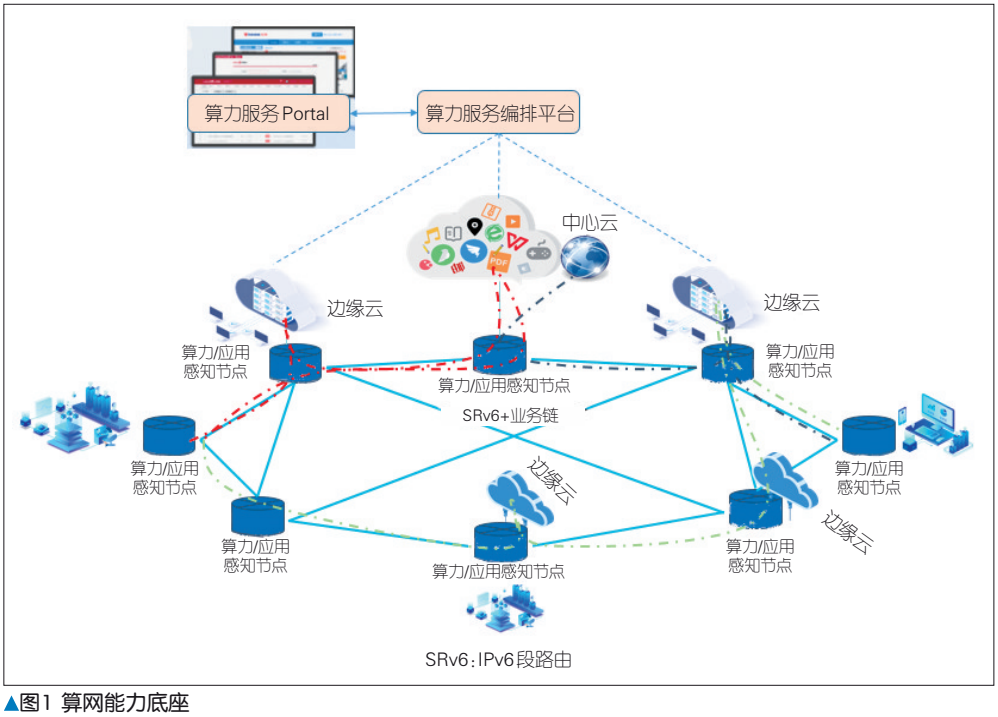
算网能力底座具体如图1所示，该底座对外统一客户入口，提供服务目录；对内实现算网的智能化，包括资源管理、服务管理等，以支持算网服务一体化自动开通、全流程可视，并可利用算网的敏捷性、灵活性和快速响应能力不断提升和优化客户体验。

随着网络、云、终端技术和能力的各自持续演进，云、网、边、端、业的协同需要进一步加强和延伸，并需要根据业务特点、网络特征、设备能力及运行环境等，智能选择原来由终端执行的非实时复杂计算和存储任务，并将其转移至云端或边缘计算节点处理，再将运算结果返回终端执行。云与网的深度融合、相互协同，可以提供云网一体化的综合服务。这就需要云和网的资源能够无缝对接，网络设备与云网元统一纳管，以形成统一的资源视图，从而使得网络的拓扑、带宽、流量和云的计算、存储能力等实时呈现[3]。多云协同支撑业务融合创新，有效地控制了负载和成本，并整合多云资源，从而提升数据的可移植性和互操作性，实现精细化管理，助力企业业务创新，提升云服务的协同能力，丰富云服务生态。充分利用不同云服务提供商的能力，可以为企业提供一致的管理、运营和安全体验。

3 *N*种算力网络能力，使能新业态算网一体

在“1+*N*+*X*”技术体系中，IPv6+技术（具体包含网络切片、确定性传输、新型多播和应用感知等协议创新[4]）使网络具备业务快速发放、灵活路径控制、业务体验保障等能力，以实现服务应用感知、资源及时调用与网络能力开放之间的协调机制。

当前，算力网络已具备SRv6策略路由和网络切片等能力，可以将时延、带宽等网因子，和成本、算力等云因子纳入路由算法，计算出满足SLA要求且云资源利用率最优的业务路径。同时，通过切片实现多种业务的差异化承载，保障了算力传输的确定性[5]。面向未来，算力网络将引入应用感知技术与算力感知技术，使网络及时感知应用类型及需求，并将服务所需的异构算力资源信息结合路由机制在网络发布，以作为业务路径选择的关键依据[6]。



▲图1 算网能力底座

• 智能选路：业务运行时，不同类型的业务对计算能力、存储能力、网络服务的需求是不同的。从现有业务上看，超算类应用、大型渲染类业务对算力的需求是最高的，可达到P级。其次是AI类的训练类应用，根据算法的不同以及训练数据的类型和大小的不同，这类应用所需的算力从T级到P级不等。AI推理类业务则大多部署在终端边缘，对算力的需求稍微减弱，在几百G到T级别不等。同时，数据处理过程

中的存储起到至关重要的作用，内存与显存的数量可以作为关键指标来衡量计算存储的能力。所以，我们需要将这些因素综合起来，量化成度量值，并嵌入路由算法中。同时，基于服务类型及应用场景，对指标因素及权重进行调整，生成新的度量值，可以指导业务转发。

• 网络切片：对5G承载网进行精细化的切片处理，可实现多种业务的差异化承载，从而保障算力传输确定性体验，并最大化网络价值。在SRv6网络中，网络切片将物理网络分成多个网络切片平面，各网络切片由不同的路由分段（SR）Locator以及对应的SID集合组成。各网络切片通过对SID进行编程及组合，约束业务在特定的切片拓扑内使用为切片预留的资源进行转发。在承载网络中，根据业务需求使用灵活以太网（FlexE）技术在承载网设备物理接口上划分出不同的逻辑接口，为切片分配独立的网络资源，以实现不同切片之间的资源隔离。

• 应用感知：将应用信息带入网络中，使网络及时感知应用类型及需求，以提供智能化和定制化的服务。结合应用感知技术，可利用IPv6扩展头将应用信息及其需求传递给网络，有效衔接网络与应用以适应服务的需求。通过业务的部署和资源调整来保证应用的SLA要求，可以将流量引向满足其要求的网络路径，从而充分发挥网络节点尤其是边缘节点的优势。

• 算力感知：通过整合计算资源，以服务的形式为用户提供算力。在电信网络中，承载计算资源信息的通信协议可以位于网络层之上（包括网络层）的任意层，并以网络层协议为基础，将服务所需的异构算力资源信息和路由机制结合并在网络发布，从而作为服务寻址的关键依据[7]。目前计算优先网络协议（CFN）主要通过在路由协议的边界网关协议（BGP）报文头中以扩展字节信息的方式携带算力信息，将网络中计算节点的负载情况实时向全网扩散。CFN协议与基于链路度量值进行路径计算的网络路由协议类似，都是基于算力度量值来完成路径的计算。而算力度量值来源于全网计算资源信息及网络链路的带宽、时延、抖动等指标。

4 X种商业场景探索实践，构建“网络路径编程即服务（SaaS）”生态运营体系

4.1 云网专线升级，赋能千行百业

面向专线应用，结合不同地区、不同行业的实际情况，因地制宜地制定符合客户需求的最佳方案，可以实现敏捷转发、分钟级开通，解决客户灵活组网需求。

在广东省开展的围绕智能城域网和省内云骨干网的云网架构创新活动中，通过SRv6，基于IP可达（即业务可达），可实现新业务的快速开通。IPv6可延展到网络的各个层级，从传统的城域网、骨干网，到接入园区、数据中心，并且通过SRv6可以拉通端、网、云，可提供智能时代多点之间的任意连接。同时，管控体系实现云网自动化对接、业务快速开通，达到了电商化体验效果。

电商化是提升客户体验的一个重要手段，解决了运营商连接客户“最后一米”问题。电商化在实践中又可分为售前、售中和售后3个环节：售前能力主要提供客户直观销售，如客户无须到营业厅，在线即可选择套餐，并可以基于自身需求，快速获取理想的资源，随时随地完成下单操作；售中能力主要提供交付流程可视化，包括实时资源可视可查，时长可预估、可承诺，网络能力开放，新型、叠加型业务的快速研发等，满足了客户定制化的需求；售后能力主要提供在线网络质量可视可查、路径调优等，包括专线质量实时可视、可看可追溯，降低客户误报率，使得网络隐患提前发现，防患于未然。

4.2 算力网络业务链，打造云网安一体服务

传统企业核心业务部署在企业数据中心（DC）。分支企业则通过组网专线访问总部核心业务，并通过企业总部统一的互联网出口访问互联网，所有的安全防护均部署在总部。随着核心系统上云，分支企业经过总部访问的模式不再具有经济效益，分支企业存在直接访问云侧服务的需求。因此，安全策略也需要分布式部署，且策略更加复杂化，要求企业分支和总部一起联动防护，并保持安全策略的一致性。

面向安全服务，我们在大湾区进行了一系列的尝试，在多个资源池分别部署网站应用级入侵防御系统（WAF）、防火墙（FW）等安全服务。根据业务的诉求，客户可以选购不同的安全服务。通过基于SRv6的业务功能链（SFC）技术，网络将按需灵活地串接安全服务，使企业分支通过安全服务后访问云业务，以提供个性化安全保障。业务链是一种业务功能的有序集，可以使指定的业务流按照指定的顺序依次经过指定的增值业务设备，以使业务流量获取一种或多种增值服务。这种增值业务设备，可以是物理设备上的一个模块或者虚拟化的实例。依据客户的意图，结合SRv6 SID即服务，通过业务链将算力服务连接，可以将服务提供给客户。如图2所示，基于服务链的服务灵活调度场景具有两大算网能力：算网技术能力和算网生态服务能力。算网技术能力包括网调应用能力和网随算动能力。网调应用实现了极致灵活网络，使算力云池内的应用按需调度；网随算动使得网络具备基于时延、带宽、丢包、指定路径等选路能力，能够

满足算力的不同诉求。针对算网生态服务能力，中国联通作为服务平台将支持集成第三方服务，并按需调度网络侧差异化承载服务能力，与应用合作伙伴共同构筑产业生态。

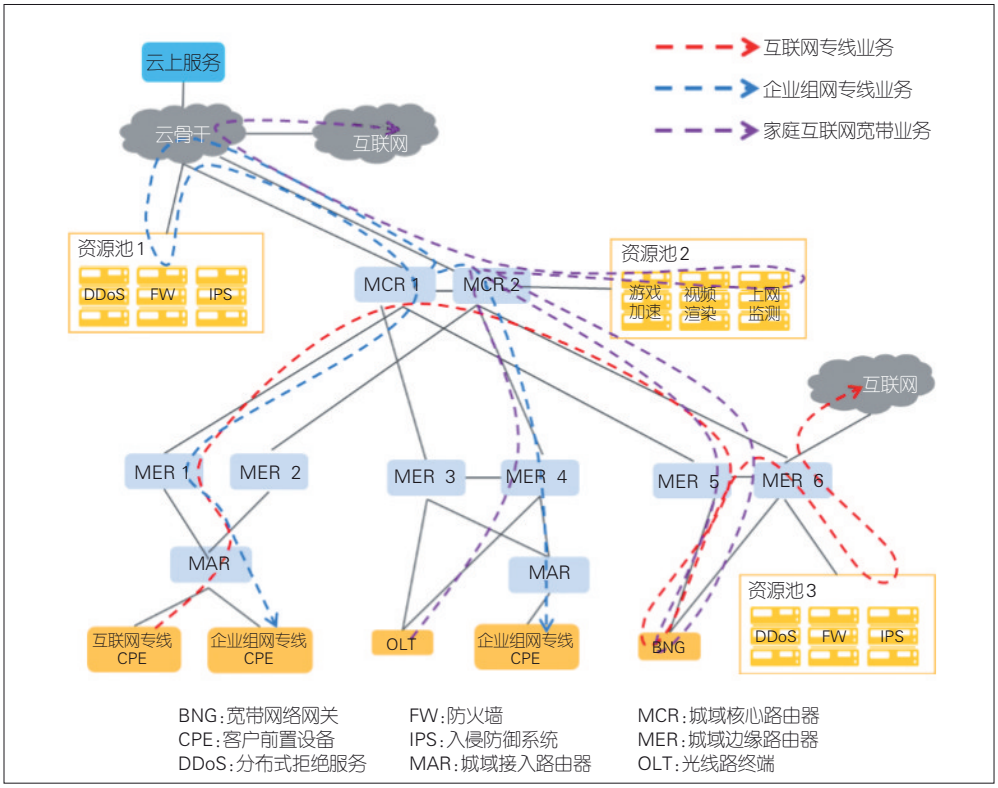
4.3 算力网络新型组播，打造大流量高品质服务

面向视频类业务高清视觉、全新交互需求，Cloud*X*业务正逐渐成为5G市场主流。除了交互式网络电视（IPTV）直播，大部分业务都采用单播技术。随着用户数量急剧增长，网络将面临着极大的带宽压力和挑战，因此推广新型组播技术已是大势所趋。我们提出基于IPv6位索引显式复制（BIER6）视频分发解决方案，以解决传统组播网络面临的资源消耗大、运维复杂度高和网络演进受限等问题。我们在北京进行了BIER6组播场景和酒店行业的场景落地，实现专线和IPTV混合场景下的简化组网、业务解耦和业务质量提升。基于BIER6组播的IPTV场景示意如图3所示。

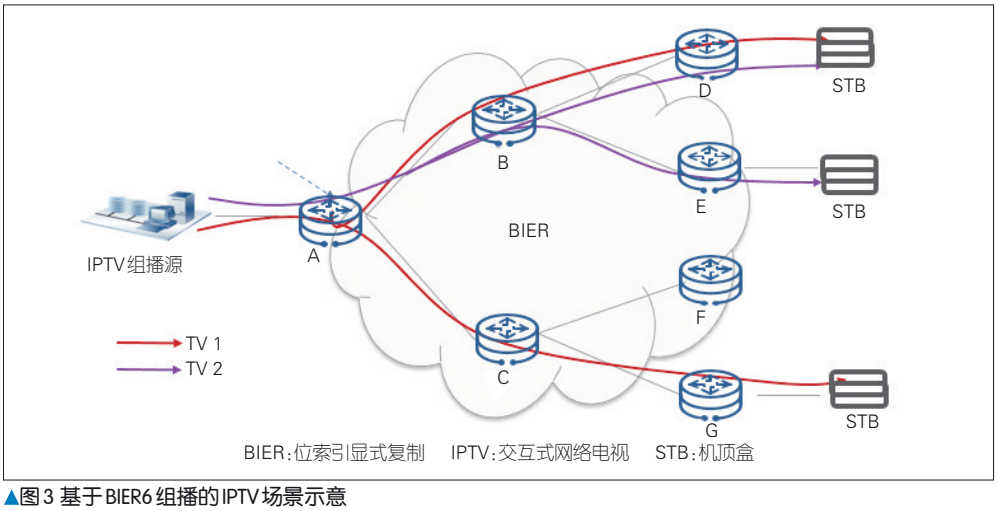
在BIER6组播技术中，上层组播应用和底层路由环境没有耦合关系，因此能够灵活适配各种路由环境，以及各种上层组播应用。现有网络只须在路由层控制面协议增加BIER6能力扩展，在组播叠加层控制面协议增加BIER6隧道能力扩展，在数据面增加BIER6封装/解封和BIER6转发能力，就可以向BIER6平滑演进。在实践过程中，通过端到端IPTV视频业务的验证，我们实现了单自治系统（AS）域及跨AS域间的移动虚拟专用网业务（MVPN）的业务承载及多业务隔离、MVPN可靠性保护（含接入侧保护、中间节点保护和根节点保护），以及BIER6的操作、管理与维护（OAM）和虚拟专用网（VPN）内单播组播共存能力。对承载网而言，点到多点转发能有效减少网络冗余流量，降低网络负载；对应用平台而言，点到多点的应用能减轻服务器和中央处理器（CPU）负荷，降低用户数增长对组播源的影响，为云间多点互联业务带来更大的发展空间。

5 结束语

基于SRv6的算力网络技术体系，运营商将5G的通信技术（CT）优势与云计算、大数据、AI等信息技术（IT）能力相结合，打造算网一体、智能安全的新一代数字基础设



▲图2 基于服务链的服务灵活调度场景示意



▲图3 基于BIER6组播的IPTV场景示意

施，持续赋能千行百业数字化转型，驱动治理方式、生产方式和生活方式变革。

算力网络是中国信息通信业倡导的新兴技术概念，反映了通信与计算服务融合的现实需求和演进趋势，其技术标准和内涵外延还需要通过实践不断丰富和发展。从网络角度看，算力网络是面向计算和智能服务的新型网络体系，IPv6+是其技术基石，增强网络内生算力是其演进的重要方向；从算力角度看，算力网络是网络化的算力基础设施，是依托网络构建的多样化算力资源调度和服务体系；从服务角度看，算力网络的目标是提供算网一体服务，是云网融合服务的新阶段，是数字基础设施服务的新形态[8]。

运营商的算网一体服务演进，须顺应千行百业数字化转型要求，将CT、IT和数据技术（DT）的能力打包提供。“算力即服务”的目标涉及IT产业、CT产业、DT产业的超级融合，每个产业的能力均需做强，并使其产生化学效应，实现“1+1+1>3”的效果[9]。立足5G、放眼6G，我们需要站在建设网络强国、数字中国和智慧社会的高度开启新一轮网络转型。同时，坚持技术和商业双轮驱动，打造泛在、柔性、协同、智能、安全和可定制的新一代数字基础设施，以数据要素支持数字化改造，加快数据要素在东部与中西部各大区域之间流通，实现东部互联网、大数据、AI等先进产业链延伸。

参考文献

[1] 中国联通.中国联通CUBE-Net 3.0网络创新体系白皮书 [R]. 北京: 中国联通研究院, 2021
[2] 中国联通.中国联通算力网络实践案例 [R] . 北京: 中国联通研究院, 2021
[3] TANG X Y, CAO C, WANG Y, et al. Computing power network: the architecture of convergence of computing and networking towards 6G requirement [J]. China communications, 2021, 18(2):175-185
[4] 田辉,魏征 . “IPv6+”互联网创新体系 [J]. 电信科学, 2020, 36(8):3-10. DOI: 10.11959/j.issn.1000-0801.2020256
[5] 曹畅, 唐雄燕 . 算力网络关键技术及发展挑战分析 [J]. 信息通信技术与政策, 2021, 47(3): 6-11. DOI: 10.11959/j.issn.1000-0801.2020256
[6] 曹畅, 张帅, 刘莹,等 . 云网向算网演进中的若干关键技术问题 [J]. 电信科学, 2021, 37(10): 93-101
[7] 中国联通 . 算力网络架构与技术体系白皮书 [R] . 北京:中国联通研究院, 2020
[8] 中国联通 . 云网融合向算网一体技术演进白皮书 [R]. 北京:中国联通研究院, 2021
[9] 唐雄燕, 张帅, 曹畅 . 夯实云网融合, 迈向算网一体 [J]. 中兴通讯技术, 2021, 27(3): 42-46. DOI: 10.12142/ZTETJ.202103009

作 者 简 介

张帅，中国联通研究院未来网络研究部工程师、中国通信标准化协会“网络5.0技术标准推进委员会”需求组副组长；主要研究方向为IP网络宽带通信、SDN/NFV、新一代网络编排技术等。

曹畅，中国联通研究院未来网络研究部高级专家、智能云网技术研究室主任、“算力网络架构与关键技术”院级重点项目攻关经理、第七届中国通信学会信息通信网络技术委员会委员、中国通信标准化协会“网络5.0技术标准推进委员会”架构组副组长、边缘计算网络基础设施联合工作组（ECNI）技术规范组组长；主要研究方向为IP网络宽带通信、SDN/NFV、新一代网络编排技术等；获中国联通科技进步奖2项；已发表论文30余篇，获授权专利10余项。

唐雄燕，中国联通研究院副院长、首席科学家，“新世纪百千万人才工程”国家级人选，北京邮电大学兼职教授、博士生导师，工业和信息化部通信科技委委员兼传送与接入专家咨询组副组长，中国通信学会理事兼信息通信网络技术委员会副主任，中国光学工程学会常务理事兼光通信与信息网络专家委员会主任，开放网络基金会（ONF）董事；长期从事通信新技术的研发和管理，主要专业领域为宽带通信、光纤传输、互联网/物联网、新一代网络等。

存转算一体的多模态网络共性平台技术研究

PINet Data Plane Platform Technology for Storage, Forwarding and Computing Integration

董永吉/DONG Yongji， 胡宇翔/HU Yuxiang，
崔鹏帅/CUI Pengshuai
（解放军战略支援部队信息工程大学，中国 郑州 450002）
(PLA Strategic Support Force Information Engineering University, Zhengzhou 450002, China)

DOI:10.12142/ZTETJ.202201006
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20220223.2024.004.html
网络出版日期：2022-02-25
收稿日期：2021-12-19

摘要：面向多模态网络需求，提出了一种存转算一体化的数据平面共性网络平台，并介绍了该平台的框架组成及关键技术。通过软硬件协同、异构融合的设计优势，该平台充分发挥了数据平面的灵活性与可扩展性，支持多样化业务可定义的转发和处理，满足多模态网络发展中存储、转发与计算融合的需求。

关键词：多模态网络；可编程数据平面；异构融合

Abstract: For the multimodal network requirements, a data plane platform integrating storage, forwarding, and computing is proposed, and the framework composition and key technology are introduced. The platform further releases the flexibility and scalability of the data plane through the coordination of software processing and hardware processing. The advantages of heterogeneousness and integration support diversified network applications with definable forwarding and processing, which can cope with the urgent needs of integration of storage, forwarding, and computing in the polymorphic networks.

Keywords: polymorphic network; programmable data plane; heterogeneous fusion

当前互联网发展进入了新时代，随路计算、确定性网络、低时延网络等新型网络技术的出现加速了工业生产、社会生活与网络的深层次融合。当前互联网的基础架构与技术体系，在智慧化、多元化、个性化、高鲁棒、高效能等方面面临重大挑战，亟须变革网络基础架构并构建全维可定义的多模态智慧网络[1]。

然而，传统单一面向尽力而为转发功能的网元设备结构，既无法满足5G、物联网等的业务流量多样化处理需求，也无法满足低时延网络、确定时延网络等细粒度服务需求，以及网络新业务的快速部署，更无法支持网络体系创新，所以网元设备的困局成为网络发展的瓶颈[2]。

为了应对这一挑战，学术界和工业界不断推出新技术和新理念。2008年，斯坦福大学提出了软件定义网络（SDN）及OpenFlow技术[3]，将传统刚性封闭网元设备结构中的数据平面与控制平面解耦。分离出的应用平面和控制平面通过编程化来应对不同场景的需求，并采用标准化的OpenFlow协议对网元设备进行配置和管理，提升了网络管控的灵活性，一定程度上增强了部署新业务的能力。但由于OpenFlow协议定义能力有限，导致为了支持新协议或新业务，该协议需要向下兼容地扩展内容。于是该协议从最早的1.0版本的12个匹配域，不断扩充到1.5版本的45个匹配域。这种刚性补丁扩容式的协议支持方式，无法满足灵活可定义的新协议需求。同时，每一次协议版本的升级和扩展，都会导致数据平面和控制平面设备的重新研制。这样一来，新技术的应用时间和开发成本则无法有效缩减。

针对SDN数据平面网元设备协议处理可扩展性差的问题，协议无关转发（POF）研究引起业内的广泛关注，POF[4]、P4语言及协议无关的交换机架构（PISA）[5]先后被提出。其中，P4具有与协议无关、可重配置性和平台无关性三大特点，缓解了SDN编程能力不足以及可拓展性差的状况。尤其是数据平面的可编程性，推动了虚拟扩展局域网（VxLAN）、基于用户数据报协议（UDP）的低时延的互联网

基金项目：国家重点研发计划项目（2019YFB1802502）

传输层协议（QUIC）、网络带内遥测等新协议与功能的快速定制与部署，但P4技术规范[6]仅仅描述了数据转发的模型。虽然支持面向转发的功能可定义，但仍无法满足数据包的存储、转发、计算协同处理复杂场景应用需求。因此，为了提供更好的网络传输服务，业界开始在数据平面上整合现场可编程门阵列（FPGA）、图形处理器（GPU）等异构加速单元[7]，通过本地的硬件加速技术实现业务功能的卸载，通过网元实现数据的加速，从而实现整个网络性能和服务质量。

虽然数据平面在硬件可编程方面不断发展，但网元设备的设计却依旧停留在封闭结构上，没有发挥可编程硬件灵活可扩展的能力。针对此，底层可编程硬件与上层软件功能和管理配置解耦的白盒结构被广泛采纳。白盒设备的硬件兼容多种转发芯片的接口，并遵循开放计算项目（OCP）标准化规范设计。同时，软件的平台操作系统及承载的各种网络应用也采用规范化的接口，实现底层硬件与软件的独立设计。例如，微软打造的开源网络交换机操作系统SONIC[8]，采用的交换机抽象接口（SAI），解耦了软硬件的实现细节，但是该系统接口与协议紧耦合，无法适配和部署新的业务。为了解决这个问题，谷歌提出了Stratum系统[9]，旨在配合硬件的可编程特性，实现一个完全可编程的数据平面网元设备。为此，该系统集成了一个协议无关转发接口P4Runtime，在开放网络操作系统（ONOS）控制器的管理下，可以灵活转发部署新型协议；但是当前网元操作系统仅可以实现对交换芯片的抽象，硬件异构的加速单元没有纳入其管理范畴。

与此同时，为了更好地支持数据平面的可编程，业界开展了多种编译器的研究工作，尤其针对交换芯片、FPGA或软件交换机的P4语言编程，以及多种异构芯片的整体硬件编译方法设计[10]。Intel公司的tonifo芯片[11]支持基于商用的软件开发环境（SDE）开发，但是由于其不开源的特性，无法支持用户可扩展的、定制化的功能编译。在基于FPGA的编译器及结构设计方面，业界提出了多种解决方案[12-13]，但这些方案主要集中在转发功能的实现上，没有充分利用FPGA的可编程特性。另外，面向交换机软件模型（BMv2）的软件编译结果限于性能瓶颈[14]，无法直接应用于高速流量场景。而当一个网元设备同时包含多样的可编程目标器件时，由于缺乏一个统一调度的协同编译环境，编译器无法有效糅合异构资源的优势，因此不能达到高效、灵活的编译效果。

1 多模态网络的提出

针对现有网络架构存在的结构僵化、IP单一承载、难以抑制未知威胁等问题，我们从网络构造的角度来提升网络的功能、性能、效能、安全等，将“结构可定义”贯穿网络的各个层面。邬江兴院士提出了一种网络各层功能多模态呈现的网络架构——全维可定义的多模态智慧网络（PINet）[2]。PINet是一种技术体制与物理平台分离的网络发展范式，如图1所示，PINet将各种网络技术体制以模态的形式在多模态网络环境上动态加载并运行，按照模态自定义的报文格式、路由协议、交换方式、转发逻辑等进行处理，实现多种模态在同一物理网络平台上的共存、演进或变革发展。

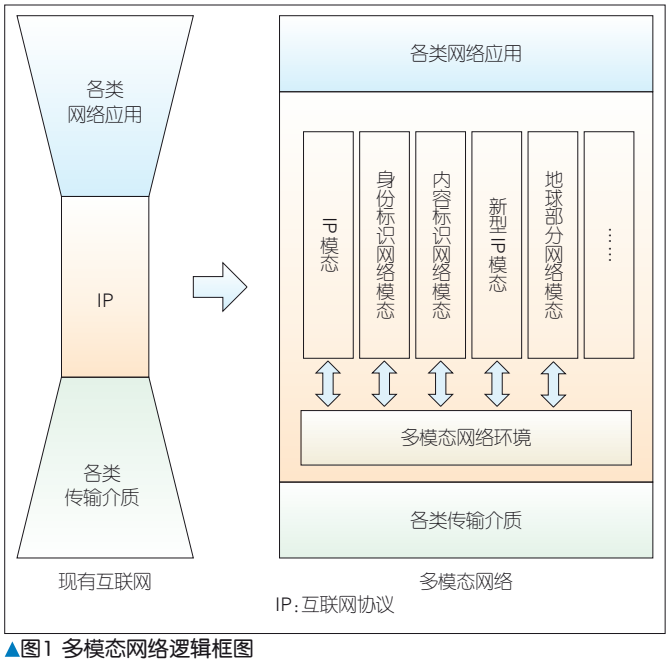
在多模态网络逻辑框架中，多模态网络环境是多模态网络的核心，而网元设备是支撑多模态网络环境的基石。然而，当前网元设备却无法满足多模态网络环境的构建需求。因此，我们需要一个共性平台来重新整合数据平面中存储、转发和计算，支持自定义模态报文的解析、转发及交换处理。

2 多模态网络共性平台的技术特征

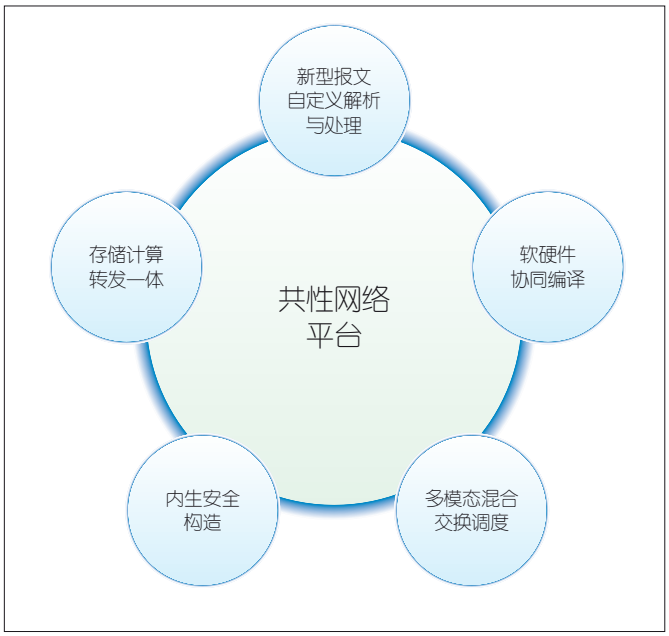
为应对新协议和应用部署的挑战，推动新型网络结构的演进发展，网元设备需要提供灵活可扩展、高性能和安全的服务支撑。本文梳理出面向存转算一体化的数据平面共性网络平台的技术特征，如图2所示。

（1）新型标识和报文的自定义解析与处理

面向定制化、个性化服务承载的需求，共性网络平台需要支持用户自定义接入标识结构和数据报文结构，并按照自定义逻辑进行报文处理，进而支撑传输协议、寻址路由等的全维度可定义。其中，寻址路由体现为基于互联网协议



▲图1 多模态网络逻辑框图



▲图2 共性网络平台的技术特征

(IP)、身份、内容、地理空间等标识的多种寻址路由方式与机制，传输协议体现为面向功能、场景、业务等需求的各种网络协议。通过支持各种新型网络机制的互联互通、协同组合，可以提高网络服务的多元化能力和对于用户需求的个性化适应能力。

（2）存储、转发与计算一体设计

网络超融合、边缘计算、随路计算等日益多样化的业务功能需求对共性网络平台提出了更大的挑战。在满足共性网络平台支持灵活转发能力的基础上，在数据平面引入存储、转发与计算一体的、可定义的复合流水线结构设计，实现存储、转发、计算3种资源在数据转发的过程中灵活组合、调用，适配不同应用场景下特质化的处理需求。

（3）内生安全构造

在由内生安全构造的共性网络平台中，“动态异构冗余”的设计思想已被引入数据平面。该平台以异构处理组件构成的元功能池为基础，生成多种等价异构执行体，并通过多模裁决和负反馈调度机制，实现内生防御的安全机制，以应对平台在软硬件设计过程中不可避免的后门及漏洞等安全威胁。该平台能够有效抵御各种病毒/木马等已知或未知的威胁，并通过将网络空间安全能力由“外挂”转变为“内生”，从而实现“高可信、高可用、高可靠”三位一体的网络安全服务。

（4）面向异构软硬件资源的协同编译

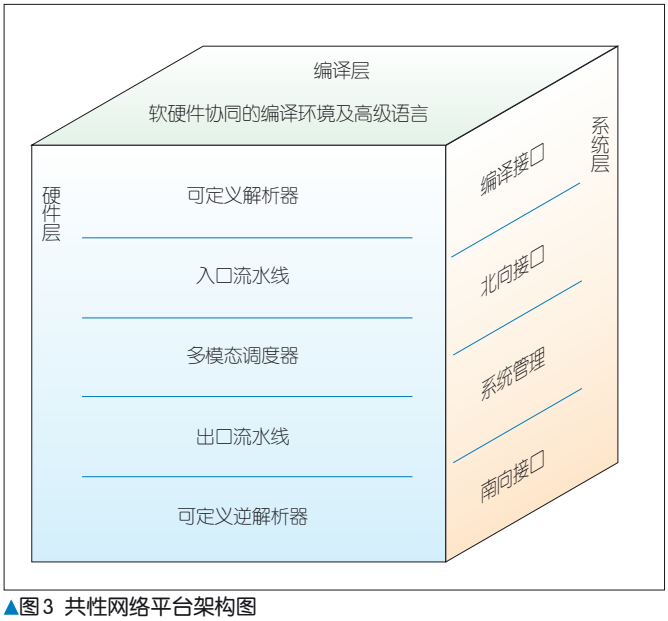
为了提供共性网络平台上多种异构资源的可编程性和通用性，降低设备开发的难度，缩短面向特定场景的应用部署时间，我们需要提供一个“自顶向下”的由高级编程语言和编译器组成的编译环境。该编译环境能够面向多种异构软硬件功能组件构建统一编程模型，提供用户高级语言编程接口，屏蔽底层硬件支持协同调度各类型软硬件资源的细节，并通过单纯的数据包存储、转发、计算等处理逻辑的形式化描述，实现网络处理性能和功能灵活性之间的平衡。

（5）多模态混合交换调度

在差异化的应用场景下，特定业务流的突发性、包长度、流量大小和速率特性有所不同。与此同时，不同的确定时延网络、低时延网络等新型网络技术体制的数据报文交换指标也不尽相同。每一种新型网络技术体制可被视为一种新型的模态。多模态混合调度技术可以为每一种模态提供定制化服务与质量保障，还可以均衡多种模态间的公平/优先级交换的策略，提升交换带宽资源利用率，保障网络交换服务质量。

3 共性网络平台的分层结构

共性网络平台的分层结构模型如图3所示，整体上分为硬件层、系统层和编译层。3个层面相互依存、相互支撑。其中，硬件层是整个平台的底层支撑，通过多种异构资源的叠加，支撑数据包存储、转发和计算的一体化处理；系统层实现异构资源的协同，以及各种网络功能的控制和管理；编译层实现对整个共性网络平台功能的编译，通过将高级语言描述的自定义功能映射到平台上，实现多样化业务的异构接入、交换和传输。



▲图3 共性网络平台架构图

3.1 硬件层逻辑设计

硬件层的处理逻辑如图4所示，主要分为5个组件：可定义解析、入口流水线、调度器、出口流水线和可定义逆解析。5个组件都引入可编程能力，可根据高级语言形式化描述的编译结果，灵活重构出定制化的服务功能。

可定义解析/逆解析组件可通过编译器接口进行配置。可定义解析组件支持按照用户定义的任意协议字段进行解析，并提取对应的字段作为入口/出口流水线中流表的关键词进行多域匹配；可定义逆解析组件支持按照用户自定义的任意数据格式进行封装，实现协议包头内容的增添、修改和转换功能。

入口/出口流水线组件采用融合异构处理资源的多级流水线设计，在支持类似P4可定义的转发流水线的基础上，将计算和存储功能挂载到并行的流水线上，再通过流水线的调用支持多种处理能力的融合，并采用协议无关的配置接口转发流表和动作信息，为不同的模态和协议流分配不同的流水线资源。这样可使得不同的模态和协议相对独立、并行地在平台上运行，而每种流表都可以基于用户定义的关键词进行构建，其匹配的动作集为存储、转发以及计算3种类别的操作集合。

调度器可以通过调度接口配置，实现面向流表的定制化的调度，并通过配置平台流水线上各个队列的属性、带宽保障、优先级和调度策略，实现精细化、差异化的交换能力服务。

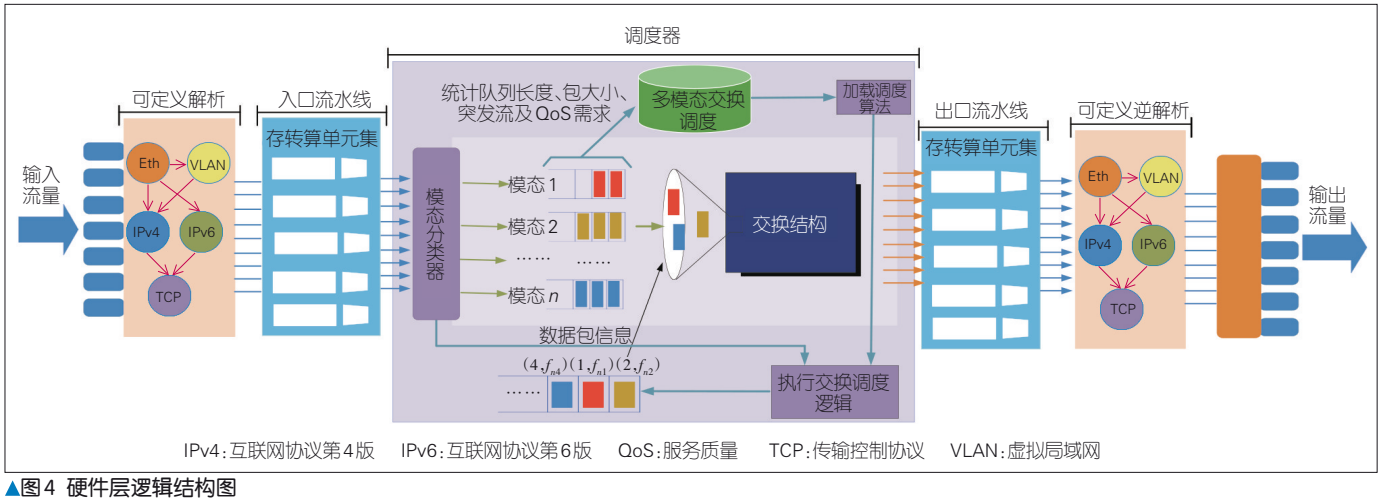
3.2 系统层逻辑设计

系统层实现对硬件层异构资源的管理，以及与控制器间的互通，系统逻辑如图5所示，主要包括编译接口、北向接口、系统管理和南向接口4个部分。编译接口为编译层与系统层的接口，支持编译结果的配置与下发，实现异构资源的可定义重构。北向接口为共性网络平台与控制器之间的配置管理接口，用于接收控制平面的控制器生成的各种流表信息，引导和决定共性网络平台的数据处理行为，并采用协议无关的协议交互机制，支持用户自定义结构和流表信息的传递。系统管理是形成共性网络平台能力的功能集，采用内生安全构造核心节点功能，以动态、随机改变核心功能的静态性、确定性，进而提供内生安全的防护能力。系统管理由运维管理和异构多维资源管理两部分功能组成，通过容器化构建、模块间联动解耦的方式，实现了对硬件层异构多维资源的管理，以及控制器配置信息的维护。南向接口为硬件层和系统层的接口，是对交换芯片、FPGA、多核等异构芯片接口的统一抽象，对上提供一系列标准化的应用程序编程接口（API），使系统层功能不再关心硬件层中异构芯片的硬件细节，并采用统一的方式接口进行管理和配置。

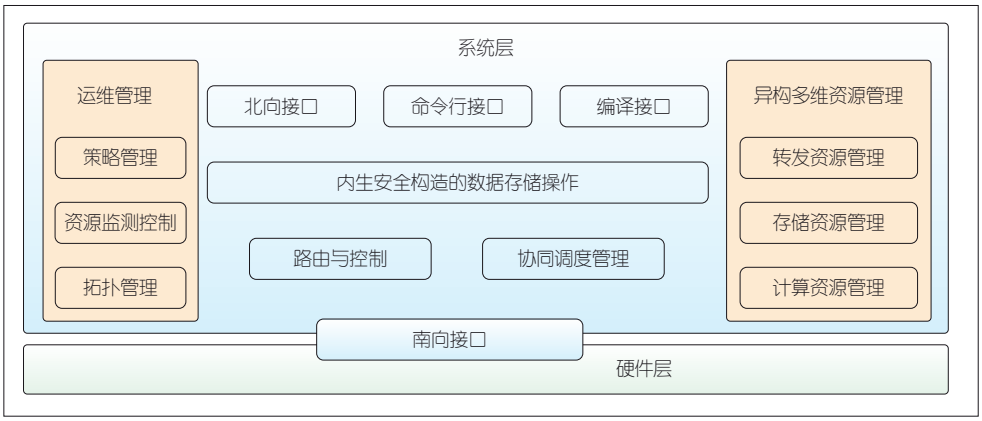
3.3 编译层逻辑设计

编译层逻辑组成如图6所示，主要包括存储、转发、计算一体化的编译器框架和高级可编程语言设计两部分。针对存储、转发、计算一体化的异构特点，编译层对数据平面关键要素进行提取，抽象出一套融合存储、转发和计算的数据平面操作指令集，用于定义和描述数据平面的行为（包括控制原语、存储原语、计算原语、转发原语），更加灵活地描述数据平面，实现数据平面与控制平面的解耦和接口的标准化。

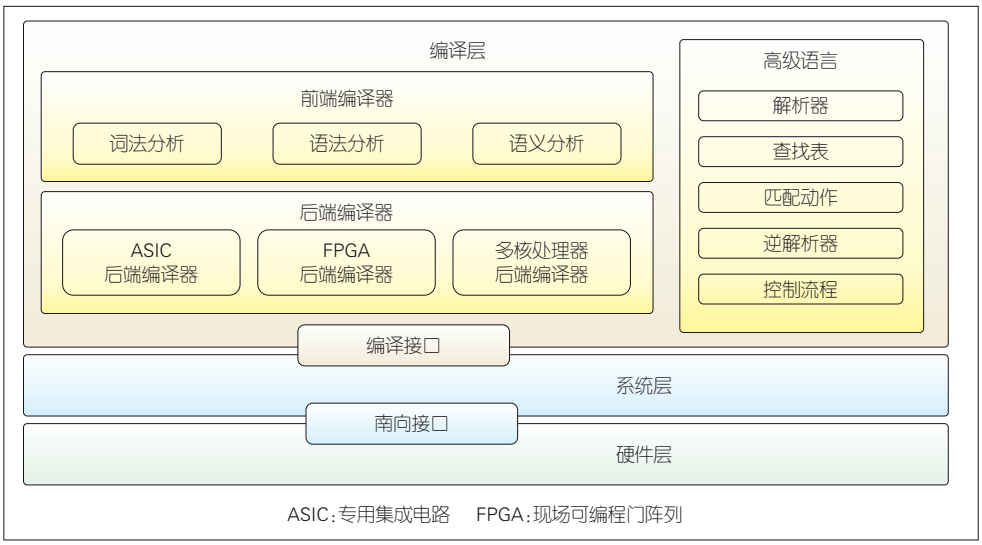
基于开放式结构设计，编译器框架参考了P4编译器前后端解耦分离的设计。前端编译器扩展支持存储、转发和计算功能的统一编译；而后端编译器支持多样化的编译目标，以及可扩展的目标结构，并能设计编译流程与仿真验证环



▲图4 硬件层逻辑结构图



▲图5 系统层逻辑组成图



▲图6 编译层逻辑组成图

境，更好地支持编译器的发展与演进。前端编译器侧重于通用基础的编译功能，主要实现编译过程的结构词法分析、文件语法分析以及段落语义分析3个功能：词法分析主要将待编译的源文件按照语法分割为独立的标记和单词，而源文件中的制表符、空格等编码无效字符会被替换并删除，并根据注释中的相关辅助类语法将语法标记或词组分类；根据语法模板，语法分析可以从语法角度判断不同分组间代码结构的正确性，并生成抽象表达；语义分析针对整个源文件的含义进行分析，排查逻辑漏洞，展开嵌套循环，并生成中间表达形式（IR）。后端编译器面向独立目标器件构建，基于前端生成的中间表达形式，结合具体的芯片属性生成最终的目标文件。后端编译器支持多种芯片类型，如ASIC交换芯片、FPGA芯片、x86多核芯片、ARM多核芯片等。所有类型的后端编译器采用共享通用可扩展的接口，支持新型编程器件的可扩展和现有功能器件的可演进。

面向该框架设计，高级可编程语言统一抽象描述异构编程对象的功能特征，主要包含5个要素：解析器、查找表、匹配动作、逆解析器和控制流程。其中，解析器通过定义并有序描述协议的特征，指导硬件层中可定义解析组件按照一定的逻辑解析数据报文；查找表用于构建硬件层中入口/出口流水线组件流表的匹配关键词和掩码的长度以及匹配模式方法，如最长匹配、精确匹配范围匹配等；匹配动作定义查找表匹配之后的动作模式，分别包括面向数学运算的计算类动作、面向本地存储的存储类动作、面向可定义流转的转发类动作；逆解析器通过定义有序的协议字段组合，指导可定义逆解析组件实现数据报文的协议的重组和封装；控制流程定义数据报文在硬件层的顺序及判断跳转执行逻辑，进而指导硬件层完成对进入系统中数据报文的处理流程。

4 结束语

本文提出了一种支持存储、转发、计算一体的多模态共性网络平台结构。该结构通过融合异构芯片功能构建数据平面的处理能力，采用可扩展的操作系统管理异构资源，从而提供统一的软硬件协同编译环境，形成存储、转发、计算一体的系统平台能力，支撑多模态网络中新型模态、协议和业务的融合、演进与发展。

参考文献
[1] 李军飞，胡宇翔，伊鹏，等. 面向2035的多模态智慧网络技术发展路线图 [J]. 中国工程科学, 2020,22(3): 141-147
[2] 胡宇翔, 伊鹏, 孙鹏浩, 等. 全维可定义的多模态智慧网络体系研究 [J]. 通信学报, 2019, 40(8): 1-12
[3] MCKEOWN N, ANDERSON T, BALAKRISHNAN H, et al. OpenFlow: enabling innovation in campus networks [J]. ACM SIGCOMM computer communication review, 2008, 38(2): 69-74. DOI: 10.1145/1355734.1355746
[4] SONG H. Protocol-oblivious forwarding: unleash the power of SDN through a future-proof forwarding plane [C]//Proceedings of the Second ACM SIGCOMM Workshop on Hot Topics in Software Defined Networking.

下转第74页➡

时间敏感网络中基于网络演算的队列分析与优化

Analysis and Optimization of Queues Based on Network Calculus in Time-Sensitive Networking

尹淑文/YIN Shuwen[1]，汪硕/WANG Shuo[1,2]，
黄韬/HUANG Tao[1,2]
（1. 北京邮电大学网络与交换国家重点实验室，中国 北京 100876；
2. 紫金山实验室，中国 南京 211111）
(1. State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China;
2. Purple Mountain Laboratories, Nanjing 211111, China)

DOI:10.12142/ZTETJ.202201007
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20220217.1553.002.html
网络出版日期：2022-02-18
收稿日期：2021-12-15

摘要：时间敏感网络（TSN）中循环队列转发（CQF）机制保留了时间感知整形中门控调度的转发可控特性，同时又降低了门控列表配置的复杂度，但是缓存队列的长度作为一个关键参数直接影响着网络的调度性能，并且在实现时受到硬件资源的约束。为了寻找到合适的CQF队列长度值以实现网络系统设计的性能和成本的优化，提出了一个基于网络演算的CQF性能分析方法。通过曲线模型的构建和计算，分析流量传输时延和积压的性能上界值，从而选择出合适的队列长度值。通过不同场景的实验，得到了不同流特性参数对队列长度选择的影响。

关键词：时间敏感网络；循环队列转发；网络演算；性能分析；队列长度分析

Abstract: Cyclic queue forwarding (CQF) in time-sensitive networking (TSN) remains the forwarding controllability based on gated scheduling in time-aware shaper and reduces the complexity of the configuration of gate control lists. However, as a key parameter, the length of CQF queues directly affects the performance of network scheduling and is constrained by hardware resources in implementation. In order to find the appropriate value of CQF queue length to realize the performance and cost optimization of network system design, a performance analysis method of CQF-based networks based on network calculus is proposed. Through the construction of curve model and calculation, the upper bounds of delay and backlog of data traffic transmission are analyzed. Then the appropriate length of CQF queues can be selected according to these analysis results. Experiments are conducted in different scenarios, and the influence of different flow characteristic parameters on the selection of queue length is obtained.

Keywords: time-sensitive networking; cyclic queue forwarding; network calculus; performance analysis; queue length analysis

由于越来越多的应用，如工业控制、自动驾驶、实时交互、远程医疗等，对网络端到端传输提出有界低时延的需求，以时间敏感网络（TSN）作为标准的、开放的链路层确定性技术受到了广泛关注[1]。

TSN引入了调度整形机制为时间敏感流量在路径上传输提供确定的传输时隙。其中，基于时钟同步的时间感知整形（TAS）机制[2]通过控制出端口队列上的门开关状态来实现对流量的精确转发控制，是TSN中广泛使用的整形机制。然而，这一机制需要通过复杂的计算为每一个出端口的每一个队列分配一个门控列表。为了简化配置问题，TSN工作组在标准中提出了循环队列转发（CQF）[3]，即一个队列接收数据，一个队列转发缓存数据，两队列采用周期性交替的方式进行传输。

目前，CQF的相关研究目标是实现基于该机制的端到端传输。例如，通过一系列约束条件规划出每条流从终端设备发包的时间，以实现无冲突的有界低时延传输。其中，队列长度仅作为资源约束中一个固定数值参与调度时间规划的计算[4-5]。然而，端口队列长度也对基于CQF机制的网络传输性能产生重要的影响，并与实现难度和成本相关。分析队列长度的选择是基于CQF机制进行交换机设计的重要内容，但当前仍缺乏这类研究。

网络演算作为一种网络性能分析的理论工具，根据流的特性刻画到达曲线以构造数据流量模型，同时根据网络节点转发服务能力刻画服务曲线以构造节点服务模型，提供了一

个从理论上进行网络性能分析的工具[6]。因此，网络演算在TSN领域也得到了关注和应用，如为音视频桥接（AVB）流传输建模提供性能分析和传输保障[7-8]；构造基于传输窗口的模型，并实现TAS机制性能分析和门控列表配置验证等[9]。上述这些研究通常集中在整形机制的时延边界分析方面，很少关注流量积压边界的问题。

针对TSN中的CQF机制，本文提出了一种基于网络演算的队列分析优化方法。该方法首先建立了一个基于CQF的网络系统的性能分析模型，再利用该模型对节点内积压和流量端到端时延进行分析并选择出合适的队列长度值。这一方法能够在理论上整体分析流量在节点上的流量积压边界，不受单条流调度规划的限制，避免了多次约束求解所造成的时间和资源浪费。

1 研究背景与动机

1.1 CQF机制

CQF机制沿用TAS机制的门控方法，当门开启时，队列中的数据被允许转发到下一节点；当门关闭时，进入到队列的数据将被缓存在队列中以等待传输。然而，CQF利用静态门控，并采用两个队列周期交替去取代TAS中一个门控队列的功能，如图1（a）所示。在CQF的奇数循环中，缓存到奇队列中的数据被转发到下一节点，到达该交换机端口的数据则进入到偶队列中；在CQF的偶数循环中,情况则与之相反，即奇队列进行缓存，偶队列进行转发。

数据帧在路径节点上的转发传输过程如图1（b）所示。假设CQF的循环周期，即两队列交替的时间间隔为T_C，流端到端传输路径的总跳数为H，则传输的最小时延$D_{min} = (H-1)\times T_C$，传输最大时延为$D_{max} = (H+1)\times T_C$。

1.2 网络演算理论基础

网络演算以最小加代数为理论工具[10]。其中，最小加代数中的卷积运算如公式（1）所示，最小加反卷积如公式（2）所示[6]：

$$(f \otimes g)(t) = inf_{0\le s\le t}\{f(t-s)+g(s)\}, \quad (1)$$

$$(f \oslash g)(t) = sup_{s\ge 0}\{f(t+s)-g(s)\}。 \quad (2)$$

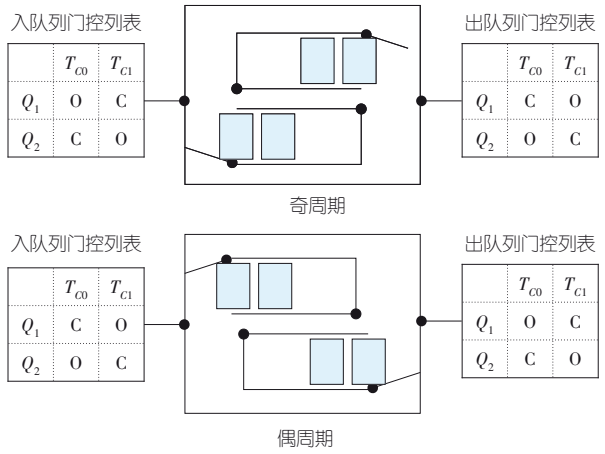
到达曲线通常描述的是到达节点流累积量的上界，为广义增函数。假设累积到达流$R(t)$的到达曲线为$\alpha(t)$，则它们的关系如公式（3）所示：

$$R(t)-R(s)\le \alpha(t-s)。 \quad (3)$$

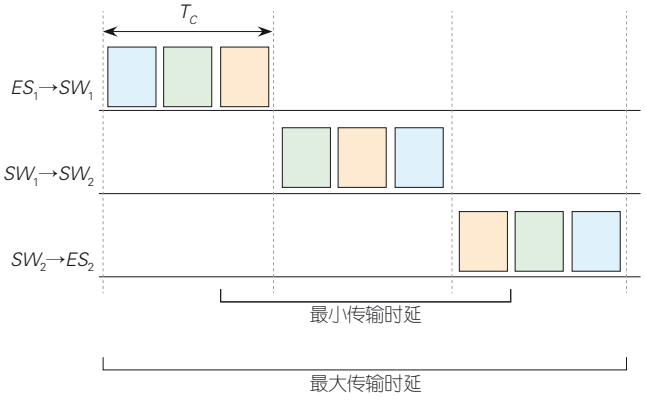
服务曲线通常用来描述节点累计服务量的下界，也是一个广义增函数。假设流量被节点服务后离开的累积量为$R^*(t)$，节点的服务曲线为$\beta(t)$，则其关系如公式（4）所示：

$$R^*(t)\ge R(s)+\beta(t-s)\ge R\otimes\beta(t)。 \quad (4)$$

根据上述所定义的模型，我们可以得到流在该节点的排队时延和该队列的流量积压边界，并进行性能分析。对于一个无损的先入先出系统，在t时刻输入数据的时延上界如式（5）所示，输入数据在该节点的流量积压如式（6）所示。



（a）循环队列转发机制



（b）循环队列转发传输示意图

C：门控状态为关闭，即禁止该队列中数据被转发　O：门控状态为开启，即允许该队列中数据被转发　*SW*：交换设备
ES：终端系统　*Q*：出端口队列　T_C：两队列循环交替的周期，即一个队列连续传输的时间

▲图1 循环队列转发机制传输示意图

$$Delay(t) = \inf\{\tau \geqslant 0 : R(t) \leqslant R^*(t+\tau)\}, \quad (5)$$

$$Backlog(t) = R(t) - R^*(t)。 \quad (6)$$

由于到达曲线和服务曲线表示最差情况下的节点传输状态，因此流在节点的传输时延上边界为两条曲线的最大水平距离，即$H(\alpha,\beta)$，如公式（7）所示。流量积压上边界为两曲线最大垂直距离，即$V(\alpha,\beta)$，如公式（8）所示：

$$H(\alpha,\beta) = sup_{t\geqslant 0}\, inf_{s\geqslant 0}\{\tau \geqslant 0 : \alpha(t) \leqslant \beta(t+s)\}, \quad (7)$$

$$V(\alpha,\beta) = sup_{s\geqslant 0}\alpha(s) - \beta(s)。 \quad (8)$$

对于需要控制数据输出的设备或机制，网络演算定义了整形器的概念。整形器具有一条整形曲线$\sigma(t)$。整形器的输出需要遵守该整形曲线，以曲线所定义的输出量为上界。其中，能够将输入数据存放在缓存中，并在满足整形曲线时尽快把数据转发的整形器被称为贪婪整形器。

1.3 队列长度的影响

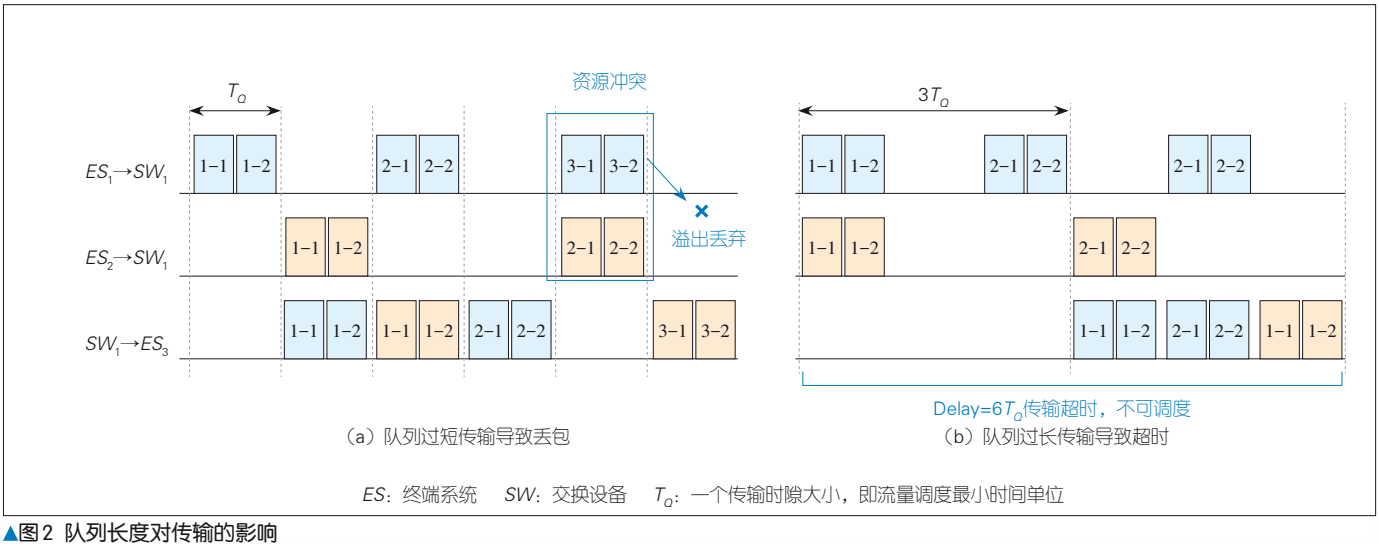
在进行交换机设计或对拓扑抽象建模时，队列长度都是一个重要的设计参数。在进行流量调度规划时，假设基于TAS机制的网络模型的端口队列长度足够使用，不会出现丢包的情况[11]，流在调度约束中不会受到队列资源的限制。然而，在CQF网络中，队列长度与传输时隙的大小有直接关系。一个传输时隙需要保证CQF队列中的所有数据包一跳转发。因此，流传输需要受到队列资源的约束。

实验表明，CQF队列长度的减小能够使其所能容纳的包数量减少，在合理范围内算法的可调度流数量也会随之减小[5]。当CQF队列长度过小时，虽然流量传输的单跳排队时延很小，但是交换机端口内没有足够的空间去缓存更多到达流。如果网络系统中存在发包周期不同的流，就需要对流的发包时间范围进行严格限制，以避免在传输中出现内存溢出导致丢包的现象。这将限制终端设备的发包数量和种类，从而限制网络系统可调度流的数量。

以图2为例，把3个终端设备连接到1台交换机上组成一个简单拓扑。当流f_1从ES_1发送到ES_3时，周期为$2T_Q$，并且一次发两个包；当流f_2从ES_2发送到ES_3时，周期为$3T_Q$。f_1和f_2可容忍的最大时延都为$5T_Q$。两条流在传输时都将经过交换机SW_1连接ES_3的端口，在调度规划时需要避免队列溢出丢包的情况。当CQF两个队列功能切换的交替周期T_C为T_Q时，队列长度可容纳两个最大传输单元大小的包。然而，由于流f_1和f_2的发包周期互为质数，在调度的超周期（两周期的最小公倍数）内系统无法实现无溢出的调度规划，如图2（a）所示。因此，数据包一定会因为队列资源不足而溢出，进而造成丢包。

当CQF队列过长时，流传输的单条排队时延会同样变得过大。这就使流的端到端传输时延过长，从而可能导致流的不可调度，造成网络系统的可调度性能变差。同样以图2为例，网络和流特性参数同上，在图2（b）中CQF两队列交替周期T_C为$3T_Q$。根据该机制包传输时延计算，其时延最大可达到$6T_Q$。可以看出，该时延边界超出了流可容忍时延的范围。

此外，由于底层硬件资源有限，内存越大其实现的难度和成本也就越高。因此，队列长度在满足传输需求的同时应尽可能地小，以减少所需资源。



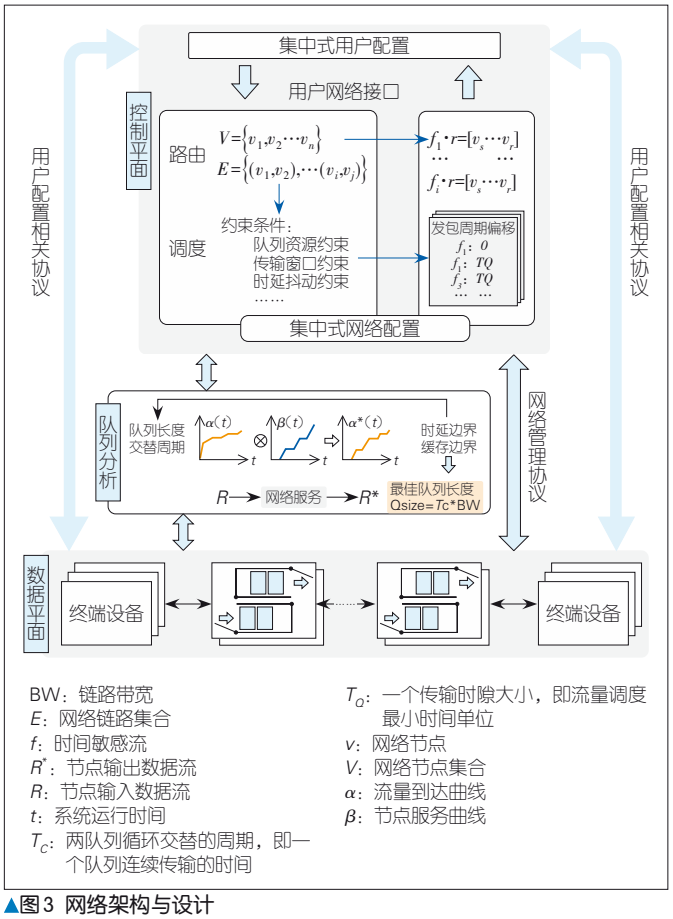
▲图2 队列长度对传输的影响

队列长度作为流量调度时的一个资源约束条件的同时，也对流量调度结果有着重要的影响。基于CQF机制的调度算法设计的重点在于对流发包时间的规划。将时间敏感流量映射到基于TSN的底层硬件资源上时，队列长度被赋予一个固定值来调度并进行约束[4]。目前，有关队列长度对调度影响的研究[5]还比较少，对于队列长度如何进行选择的研究更是缺乏。因此，我们需要一种能够根据场景进行队列分析优化的方法。

2 模型与分析

2.1 模型概述

根据IEEE 802.1 Qcc（电气与电子工程师协会标准）所提出的集中式控制架构，基于CQF的TSN网络架构可以分为数据面和控制面两个部分，如图3所示。数据面的拓扑、设备和流特性抽象为一个全局的资源视图，并作为控制面的输入数据。控制面采用设计的调度算法，根据输入进行调度计算，并下发给数据面设备。随后，数据面设备根据收到的配置信息来控制数据包的发送。



BW：链路带宽
E：网络链路集合
f：时间敏感流
R^*：节点输出数据流
R：节点输入数据流
t：系统运行时间
T_C：两队列循环交替的周期，即一个队列连续传输的时间
T_Q：一个传输时隙大小，即流量调度最小时间单位
v：网络节点
V：网络节点集合
α：流量到达曲线
β：节点服务曲线

▲图3 网络架构与设计

在从数据面信息抽象时，设备的队列长度将作为一个既定的信息参数图输入至控制面。在控制面中，调度器利用一个确定的队列长度值来进行资源约束。本文所提出的分析方法应用在整个网络架构之外，如图3所示。在理论层面，该方法利用TSN网络抽象模型中数据面的全局资源信息和控制面路由结果，分析出最佳队列长度参数，并将该参数反馈至数据面，调节网络模型的队列长度，为可编程交换机参数设置[12]或者TSN交换机设计和型号选择提供参考。

该方法不仅能根据当前场景的终端设备和流量特性构造出数据流量模型，还可根据交换机信息构造队列服务模型，基于网络演算理论计算流量端到端传输性能参数，并反馈调节队列参数，最终可得到最佳队列长度值。

2.2 队列服务模型

CQF机制利用两个相同的队列进行乒乓交替传输。从节点传输上来看，该端口节点始终为线速转发。然而，在数据传输方面，数据帧先进入端口的一个队列等待，再以线速传输转发。在对端口节点进行服务模型分析时，从单独CQF队列出发，我们把CQF服务模型分为奇队列和偶队列两个部分。其中，奇队列在奇数交替周期转发，偶队列在偶数周期进行转发。

根据网络演算理论，CQF服务队列可建模为贪婪整形器[6]。在整形曲线$\sigma(t)$满足次加性并且初始时值为0的情况下，缓冲区初始状态为空且足够大的贪婪整形器的输入输出特性满足公式（9），即节点为流提供了一条等于σ的服务曲线。

$$R^* = R \otimes \sigma。 \tag{9}$$

为了获得更细粒度的CQF队列服务模型，整形曲线的构造可根据时分多址（TDMA）总线协议的经典服务模型[13]来完成，进而分阶段描述出转发服务和缓存等待两个过程。在转发服务阶段，整形曲线以交换机传输速率为斜率递增；在缓存等待阶段，整形曲线累积服务量不随时间的增加而变化。根据两队列工作机制，奇队列在一个调度超周期初始时刻即开始进行转发，其整形曲线表达如公式（10）所示。偶队列等同于在奇队列前再加入一个恒定突发延迟函数δ_{T_Q}。该函数的值在$t \leqslant 0$时为0，其他情况为∞。偶队列整形曲线表达满足公式（11）。其中，T_Q为队列转发的交替周期，C为端口转发速率。

$$\beta^1(t) = \sigma_{\text{odd}}(t) = C \cdot \min\left(\left\lceil \frac{t}{2T_Q} \right\rceil \cdot T_Q, t - \left\lfloor \frac{t}{2T_Q} \right\rfloor \cdot T_Q\right), \tag{10}$$

$$\beta^0(t)=\sigma_{\text{even}}(t)=C\cdot\max\left(\left\lfloor\frac{t}{2T_Q}\right\rfloor\cdot T_Q,t-\left\lceil\frac{t}{2T_Q}\right\rceil\cdot T_Q\right)。\tag{11}$$

2.3 数据流量模型

在构造数据流量模型时，模型仅对流特性已知，对流量发包时间的调度未知。因此，流量到达曲线的构造以经典的漏桶模型为基础，相关定义为$\alpha(t)=b+rt$。对于终端设备输出数据量曲线，参数恒定速率r和瞬时突发b可以根据流的发包周期*f.period*、数据包大小*f.size*和一次发包数量*f.num*求出。一台终端设备累积输出数据量上界的曲线参数r和b的计算公式如（12）和（13）所示：

$$r=\frac{f.size\cdot f.num}{f.period},\tag{12}$$

$$b=f.size\cdot f.num\cdot(1-\frac{r}{C})。\tag{13}$$

基于上述漏桶曲线，根据CQF机制工作模式得到终端设备在奇偶周期流量发送曲线。下一节点的偶队列的到达曲线，即终端设备在奇周期的累积输出数据量$\alpha_{\text{odd}}(t)$如公式（14）所示；奇队列的到达曲线，即终端设备在偶周期的累积发送数据量$\alpha_{\text{even}}(t)$为$\alpha_{odd}(t)$延迟一个交替周期T后的结果，如公式（15）所示。

$$\alpha_{\text{odd}}(t)=\begin{cases}0 & t=0\\ b+r\cdot\min\left(\left\lfloor\frac{t}{2T_Q}\right\rfloor\cdot T_Q,t-\left\lceil\frac{t}{2T_Q}\right\rceil\cdot T_Q\right) & t>0\end{cases},\tag{14}$$

$$\alpha_{\text{even}}(t)=(\alpha_{\text{odd}}\otimes\delta_{T_Q})(t)。\tag{15}$$

2.4 性能与参数分析

流从发送端输出后将根据传输路径经过多个串联的网络节点，然后到达接收端。根据串联等效定理[10]，这些串联的节点为一条流所提供的服务量可以使用一个服务模型来代替。假设这些串联节点提供的服务曲线依次为$\beta_1(t),\beta_2(t),\cdots,\beta_n(t)$，则该串联等效模型服务曲线为$\beta(t)=\beta_1\otimes\beta_2\cdots\otimes\beta_n(t)$。

当时间敏感流在基于CQF的网络系统中传输时，缓存在奇队列中的数据将转发至偶队列中，同时偶队列中的数据将转发至奇队列中。根据CQF传输时延边界公式、串联等效定理和贪婪整形器输入输出特性，奇数周期从发送端输出的到达曲线为$\alpha_{\text{odd}}(t)$的数据流$R_{\text{odd}}(t)$。在经过路径节点$v_0,v_1,\cdots,v_n$的服务后，接收端所接收到的输出数据流$R^*_{\text{odd}}(t)$满足公式（16）。同理，偶数周期发送流的输出$R^*_{\text{even}}(t)$满足公式（17）。

$$R^*_{\text{odd}}(t)\geqslant\alpha_{\text{odd}}\otimes\beta_0^0\otimes\beta_1^1\cdots\otimes\beta_n^{n\%2}\otimes\delta_T(t),\tag{16}$$

$$R^*_{\text{even}}(t)\geqslant\alpha_{\text{even}}\otimes\beta_0^1\otimes\beta_1^0\cdots\otimes\beta_n^{(n+1)\%2}\otimes\delta_T(t)。\tag{17}$$

根据传输时延公式，CQF的网络系统中到达曲线为$\alpha(t)$数据流的端到端传输时延边界，如公式（18）所示：

$$Delay(t)=inf\{\tau\geqslant 0:R(t)\leqslant R^*(t+\tau)\}\leqslant$$

$$inf\{\tau\geqslant 0:\alpha(t)\leqslant\alpha\otimes\beta_0\cdots\otimes\beta_n\otimes\delta_T(t)\}。\tag{18}$$

根据剩余服务定理[10]，当多条流同时到达同一网络节点并竞争使用该节点提供的服务时，假设这些流的到达曲线分别为$\alpha_1,\alpha_2,\cdots,\alpha_n$，节点提供给所有流的总服务曲线为$\beta(t)$，则节点提供给到达曲线$\alpha_n$的数据流的服务曲线为$\beta^n(t)=\max(0,\beta-\alpha_1-\alpha_2\cdots-\alpha_{n-1})$。

在CQF机制下，流量到达队列先缓存后进行转发。相对于流到达，CQF队列对流服务有一个周期的延迟。节点的剩余服务为$\beta^n(t)=\max(0,\beta-\alpha_1\otimes\delta_T-\alpha_2\otimes\delta_T\cdots-\alpha_{n-1}\otimes\delta_T)$。在进行性能分析时，系统按照最大可容忍时延从小到大的顺序对流进行逐一端到端传输分析，并沿路由更新交换机端口剩余服务曲线，以用于下一条到达流量服务分析。端口剩余服务曲线不足以服务的流则滞留在该端口队列中。根据流量积压公式，端口队列中流量积压量如公式（19）所示：

$$Backlog(t)=R(t)-R^*(t)\leqslant\sum_{i=1}^{n}[\alpha_i(t)-\alpha_i\otimes\beta^i(t)]。\tag{19}$$

如图3中队列分析部分所示，在利用上述分析模型进行队列参数的选择时，首先将队列长度的初始状态设置为一个最大传输单元，再构建该系统的流量和服务模型，以便得到流量端到端传输时延和各个端口的流量积压值。

当队列过小时，队列长度不足以容纳端口流量积压量，数据包将被丢弃。此时，系统会根据当前队列长度和流量积压参数将队列长度调大，队列长度增量值如公式（20）所示。其中，$Backlog_{\max}$为超周期内所有队列流量积压量的最大值，*AdjustNum*参数控制节幅度随着调节次数的增加而减小，队列长度调节单位为最大传输单元1 500 B。

$$\Delta Q_{\text{size}}=\max\left(1,\left\lceil\frac{Backlog_{\max}-Q_{\text{size}}}{1\,500}\right\rceil/AdjustNum\right)\times 1\,500。\tag{20}$$

当队列过大时，传输时延不满足可容忍最大时延约束条件。此时系统会根据当前时延和流的可容忍最大时延差值将队列长度调小，直至传输时延和队列缓存都满足相关条件，相关计算如公式（21）所示。其中，$Delay_i$为流f_i性能分析得到的传输时延，$delay_i$为流f_i最大可容忍传输时延，Hop_i为流f_i跳数。

$$\Delta Q_{size} = \max\left(1, \left[\max\left(\frac{Delay_i - delay_i}{Hop_i}\right)\right] / AdjustNum\right) \times 1\,500。\quad (21)$$

随后，系统会返回该理想的队列长度值并将其作为最佳队列长度参数，或者达到最大调节次数后退出。如果选择失败，则说明当前场景无法选择出理想的队列长度值。

3 实验与结果

3.1 实验设置

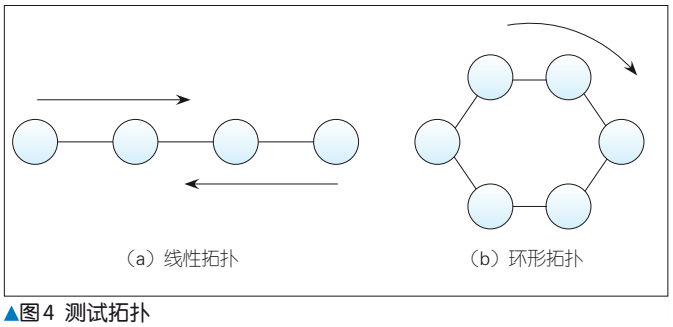
3.1.1 网络拓扑

本文实验所采用的拓扑为常用于工业控制网络的线性拓扑和环形拓扑，如图4所示。线性拓扑的流量数据在交换机节点中可以双向传输，环形拓扑的流量数据在交换机节点之间只能沿着一个方向进行传输。

由于队列分析方法的应用不受网络规模限制，实验时线性和环形拓扑中的交换机数量固定为10个，网络带宽都设置为1 Gbit/s。CQF队列长度初始值被设置为一个最大传输单元1 500 B，系统以1 500 B的幅度进行调节，以保证数据传输的完整性。实验在Intel(R) Core(TM) i7–6700 RAM 16GB的Windows设备上，基于Python开发的分析系统进行测试。

3.1.2 流量特性

由于没有可以直接应用于实验测试的标准TSN流量集，本实验参考国际电工委员会（IEC）/IEEE 60802标准[14]中描述的工业自动化网络流量特性进行参数设置，从不连接同一交换机的终端设备中随机选取一组作为流的发送和接收设备，并采用最短路径计算出相应的传输路径。流的发包周期以毫秒为单位，并且该发包周期从集合{2，4，8} ms中选取。流的最大可容忍时延一般为时延系数集合中随机选取的一个系数与其发包周期的乘积，流的一个数据帧长度范围为64～1 500 B。



▲图4 测试拓扑

在该实验中，时间敏感流量类型主要分为循环流、同步流和视频流3种[15]，相关参数如表1所示。循环流主要用于设备之间的周期性通信，它的可容忍最大时延与发包周期相关，一般不超过其周期值。同步流主要用于控制器或设备之间的同步交互，它的最大时延通常在一个周期以内，数据帧通常很小。视频流是终端直接传输的视频数据流，面向用户的视频流的性能相对较低，其特征是延迟小于10 ms，以保障用户体验。视频流的帧长度一般为1 000～1 500 B。根据码率可以近似得到等价发包周期。

3.2 实验结果

实验分别分析了流特性中的发包周期、最大可容忍时间和网络拓扑类型对队列分析结果的影响，并对具有明显差异的同步流和视频流进行队列分析，将其作为队列分析场景应用示例。

3.2.1 发包周期

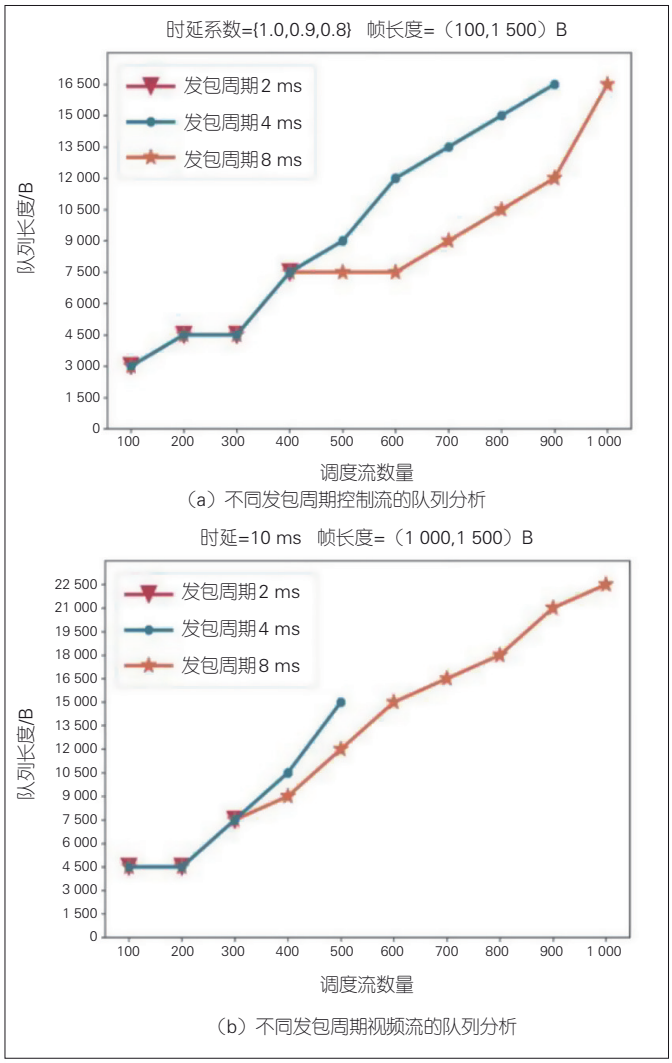
在不同发包周期下进行队列分析的实验中，测试采用线性拓扑并分别采用具有固定时延系数的循环流和固定时延的视频流，测试结果分别如图5（a）和图5（b）所示。随着发包间隔的增大，流从终端设备注入的时间有更大的选择范围，流量更加不易聚集，因此可调度的流数量随之增加。随着调度流数量的增加，发包周期大的流所需要的队列长度逐渐小于发包周期小的流量。

3.2.2 最大时延

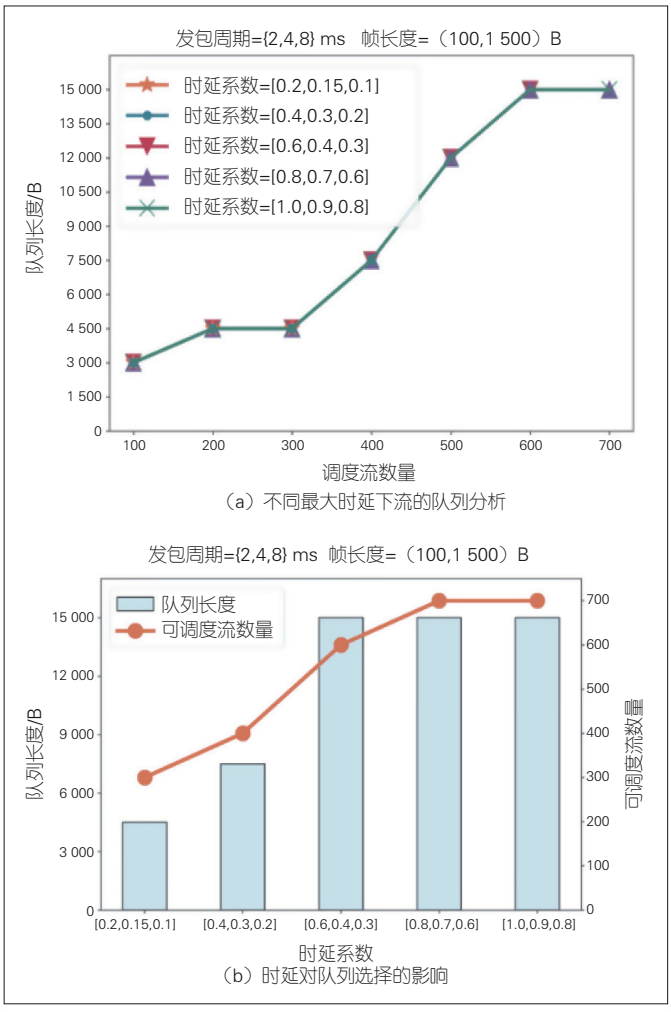
在不同可容忍最大时延下进行队列分析的实验中，测试采用线性拓扑和具有不同时延系数的循环流。该循环流单条流的周期从周期集合为{2，4，8} ms中进行选取，帧长度从100～1 500 B中选取。实验采用5组时延系数进行分析，实

▼表1 测试流参数

流量类型	发包周期/ms	最大时延/ms	帧长度/B
循环流	{2,4,8}	不超过周期值	64～1 500
同步流	{2,4,8}	不超过周期值	64～300
视频流	{2,4,8}	小于10	1 000～1 500



▲图5 不同发包周期下的队列分析



▲图6 不同可容忍最大时延下的队列分析

验结果如图6所示。从图6（a）中不同组队列长度曲线重合可知，在可调度流数量范围内，流的可容忍最大时延并不影响队列长度的选择；然而，流可容忍最大时延会影响队列选择范围和与之对应的流的最大可调度数量，即随着可容忍最大时延的增大，可选择队列长度会增大，可调度流数量也会增多，如图6（b）所示。

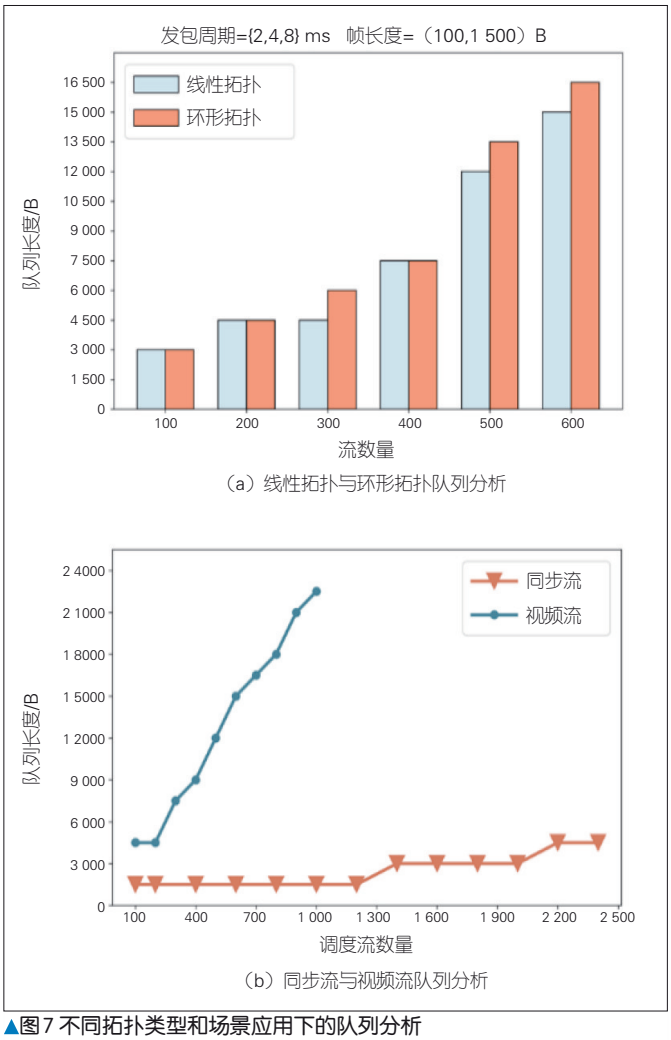
3.2.3 拓扑类型与应用

在不同网络拓扑类型下进行队列分析的实验中，我们分别采用线性拓扑和环形拓扑传输循环流，如图7（a）所示。由于环形拓扑中所有交换机之间只能够进行单向传输，相对于线性拓扑，环形拓扑队列缓存需求更大。这说明在相同队列资源的条件下，线性拓扑可映射的流数量比环形拓扑更多[4]。

图7（b）展示了同步流和视频流的队列分析应用场景。对时延要求高、轻负载的同步流更适合短队列，可实现快速转发；对时延要求低、数据量大的视频流更适合长队列，可避免丢包。由此可见，不同类型流量和应用场景对队列长度的要求具有一定的差别，有必要根据网络流量的特性进行分析。

4 结束语

随着实时性和交互性网络应用的发展，端到端有界低时延的确定性传输需求给当前网络提出了挑战，同时也为网络革新带来了机遇。TSN作为链路层上的确定性技术正趋于完善。在本文中，我们针对TSN循环队列转发机制的应用进行分析研究，提出了一种队列分析优化方法，并对基于该方法的网络进行性能分析和转发队列长度选择，以优化不同场景下流的可调度性和队列资源成本，为交换机参数设置或选择提出了理论层面的建议。



（a）线性拓扑与环形拓扑队列分析

（b）同步流与视频流队列分析

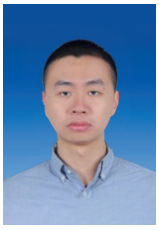
▲图7 不同拓扑类型和场景应用下的队列分析

参考文献

[1] 黄韬, 汪硕, 黄玉栋, 等. 确定性网络研究综述 [J]. 通信学报, 2019, 40(6): 160–176. DOI: 10.11959/j.issn.1000–436x.2019119
[2] IEEE. IEEE standard for local and metropolitan area networks – bridges and bridged networks – amendment 25: enhancements for scheduled traffic: IEEE [S]. 2015
[3] IEEE. IEEE Standard for local and metropolitan area networks – bridges and bridged networks – amendment 29: cyclic queuing and forwarding: IEEE [S]. 2017
[4] YAN J L, QUAN W, JIANG X Y, et al. Injection time planning: making CQF practical in time–sensitive networking [C]//IEEE INFOCOM 2020 – IEEE Conference on Computer Communications. IEEE, 2020: 616–625. DOI: 10.1109/INFOCOM41043.2020.9155434
[5] QUAN W, YAN J L, JIANG X Y, et al. On–line traffic scheduling optimization in IEEE 802.1Qch based time–sensitive networks [C]//2020 IEEE 22nd International Conference on High Performance Computing and Communications; IEEE 18th International Conference on Smart City; IEEE 6th International Conference on Data Science and Systems. IEEE, 2020: 369–376. DOI: 10.1109/HPCC–SmartCity–DSS50907.2020.00045
[6] BOUDEC J, THIRAN P. Network calculus: a theory of deterministic queuing systems for the Internet [M]. Berlin: Springer–Verlag, 2001
[7] DE AZUA J A R, BOYER M. Complete modelling of AVB in network calculus framework [C]//Proceedings of RTNS '14: Proceedings of the 22nd International Conference on Real–Time Networks and Systems. ACM, 2014: 55–64. DOI: 10.1145/2659787.2659810
[8] ZHAO L X, POP P, ZHENG Z, et al. Latency analysis of multiple classes of AVB traffic in TSN with standard credit behavior using network calculus [J]. IEEE transactions on industrial electronics, 2021, 68(10): 10291–10302. DOI: 10.1109/TIE.2020.3021638
[9] ZHAO L X, POP P, GONG Z J, et al. Improving latency analysis for flexible window–based GCL scheduling in TSN networks by integration of consecutive nodes offsets [J]. IEEE Internet of Things journal, 2021, 8(7): 5574–5584. DOI: 10.1109/JIOT.2020.3031932
[10] 李焕忠. 基于随机网络演算的性能分析技术研究 [D]. 长沙: 国防科学技术大学, 2011
[11] CRACIUNAS S S, OLIVER R S, CHMELÍK M, et al. Scheduling real–time communication in IEEE 802.1Qbv time sensitive networks [C]//RTNS' 16: Proceedings of the 24th International Conference on Real–Time Networks and Systems. ACM, 2016: 183–192. DOI: 10.1145/2997465.2997470
[12] YAN J L, QUAN W, YANG X R, et al. TSN–builder: enabling rapid customization of resource–efficient switches for time–sensitive networking [C]//2020 57th ACM/IEEE Design Automation Conference. IEEE, 2020: 1–6. DOI: 10.1109/DAC18072.2020.9218753
[13] WANDELER E, THIELE L. Optimal TDMA time slot and cycle length allocation for hard real–time systems [C]//Asia and South Pacific Conference on Design Automation, 2006. IEEE, 2006: 479–484. DOI: 10.1109/ASPDAC.2006.1594731
[14] IEC, IEEE. Time–sensitive networking profile for industrial automation: IEC/IEEE 60802 [S]. 2018.
[15] Industrial Internet Consortium. Time sensitive networks for flexible manufacturing testbed characterization and mapping of converged traffic types [R]. 2019

作 者 简 介

尹淑文，北京邮电大学在读硕士研究生；主要研究方向为确定性网络、时间敏感网络等。



汪硕，北京邮电大学讲师；主要研究方向为确定性网络、数据中心网络、软件定义网络、网络流量调度等；获“青年人才托举工程”项目资助；发表论文20余篇，申请发明专利10余项。

黄韬，北京邮电大学教授、江苏省未来网络创新研究院副院长、紫金山实验室未来网络中心主任；主要研究方向为路由与交换、软件定义网络、内容分发网络等；先后主持科技部、工信部等重大项目10余项；获中国通信学会技术发明奖一等奖（排名第1）1次；发表论文65篇，申请专利64项，牵头完成行业标准6项，提交国际标准提案25个，出版英文专著1部、中文专著10部。

数字孪生网络接口设计及其协议分析

Multi-Protocol Cooperative Interface for Digital Twin Network

陈丹阳/CHEN Danyang， 陆璐/LU Lu， 孙滔/SUN Tao
（中国移动通信有限公司研究院，中国 北京 100053）
(Research Institute of China Mobile, Beijing 100053, China)

DOI：10.12142/ZTETJ.202201008
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20220222.1905.004.html
网络出版日期：2022-02-23
收稿日期：2021-12-22

摘要：探讨了数字孪生网络南北向接口和孪生层内部接口应具备的不同特性，并给出了当前一些通用接口在数字孪生网络中的适用性建议。同时，针对数字孪生网络面临的多协议问题，提出在孪生网络层和物理网络层间引入南向接口协议适配功能，在孪生网络层和网络应用层间引入北向接口协议适配功能。借助南向和北向的接口协议适配功能，分别提出南北向多协议的识别、解析、转换方法，实现孪生层内部接口的多协议协同，降低构建数字孪生网络的协议处理的复杂度。

关键词：数字孪生网络；网络孪生；多协议协同；闭环控制

Abstract: The different characteristics of the north-south interface and the internal interface of the twin layer in the digital twin network and the applicability suggestions of some general interfaces for the digital twin network in the current network are given. At the same time, aiming at the multi-protocol problem of digital twin network, the southbound interface protocol adaptation between the twin network layer and physical network layer, and northbound interface protocol adaptation between the twin network layer and network application layer are introduced. Based on the protocol adaptation function of southbound and northbound interfaces, the identification, parsing, and transformation methods of southbound and northbound multi-protocol are proposed to realize the multi-protocol collaboration of interfaces in the twin-layer and reduce the complexity of protocol processing in constructing digital twin network.

Keywords: digital twin network; network twin; multi-protocol collaboration; closed-loop control

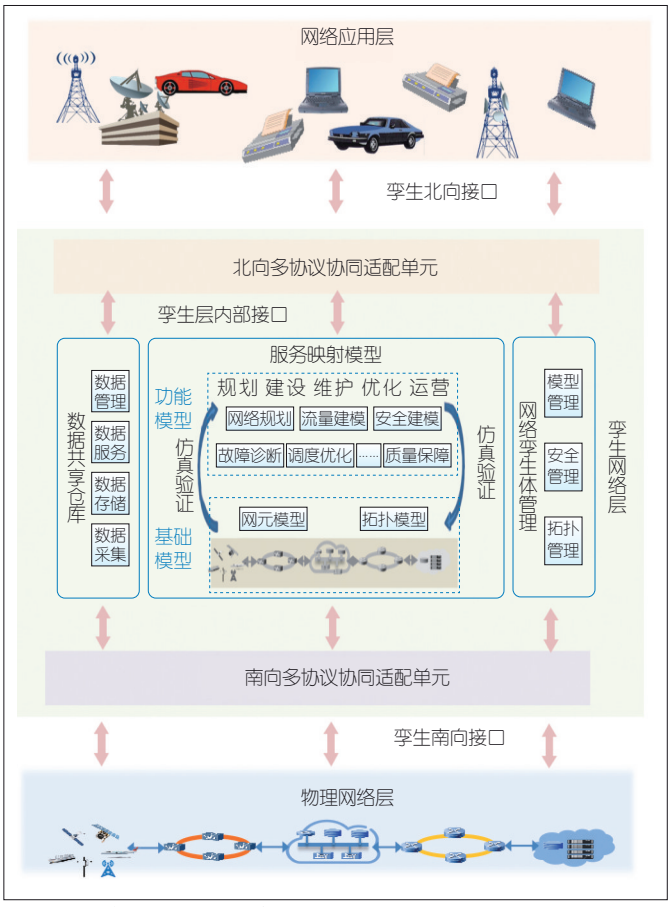
1 数字孪生网络（DTN）发展概述

2003年，美国密歇根大学的M. GRIEVES教授于产品全生命周期管理课程上提出了数字孪生的概念[1]，并将其定义为包括实体产品、虚拟产品以及二者间联系的三维模型。到2012年，美国空军研究实验室和美国国家航空航天局（NASA）合作并提出了构建未来飞行器的数字孪生体，同时将数字孪生定义为高度集成的多物理场、多尺度、多概率的仿真模型，并且能够利用物理模型、传感器数据和历史数据等反映与该模型对应的实体的功能、实时状态以及演变趋势等。近年来，随着建模仿真技术的迅猛发展，数字孪生技术逐渐开始应用在诸如卫星、医疗、能源、交通等各行各业中[2]。

随着数字孪生技术在各行各业的应用，如何在通信网络领域中引入数字孪生技术并构建DTN成为了新的研究热点之一。同时，DTN也逐渐被认为是6G网络的关键技术之一。文献[3]中，DTN被定义为“一个具有物理网络实体及虚拟孪生体，且二者可进行实时交互映射的网络系统”，其应当具备4个核心要素：数据、模型、映射和交互，并相应设计了“三层三域双闭环”架构。

基于如上所述的架构定义，数字孪生网络的各层接口及所处位置如图1所示。物理实体网络中的网元通过孪生南向接口同孪生网络层交互网络数据和网络控制信息。孪生网络层中含有数据共享仓库、服务映射模型和数字孪生体管理3个关键子系统，也通过相应的接口协议，满足其构建与交互需求，并通过孪生层内部接口实现3个关键子系统之间以及与物理网络层和网络应用层间的交互。网络应用通过孪生北向接口向孪生网络层输入需求，并通过模型化实例在孪生网络层进行业务部署。综上所述，DTN不同层间，以及孪生层内部的接口协议需求存在差异性。此外，物理网络层中的不同设备所支持的协议也多有不同，因此DTN的构建也需要考虑如何实现不同协议间的高效协同。

基金项目：基于服务的6G核心网关键技术研究（62032003）



▲图1 数字孪生网络接口示意

2 数字孪生网络的接口及协议

本章依据文献[3]所提架构，进一步提出了孪生北向接口、孪生层内部接口和孪生南向接口需具备的能力，总结了各接口应具备的特性，并在此基础上总结分析了一些现有通信协议在DTN构建中的适用性。

2.1 孪生北向接口

孪生北向接口是网络应用层与孪生网络层间的接口，网络应用需求由孪生北向接口输入到孪生网络层。孪生北向接口能够支持网络运维和优化、网络可视化、意图验证、网络自动驾驶等网络应用以更低的成本、更高的效率和更小的现网业务影响快速部署。因此，孪生北向接口应具备4个方面的特征。

（1）开放性：孪生北向接口须让不同网络应用的业务需求输入到孪生网络层，因此它需要具备良好的开放性与兼容性。

（2）可扩展：网络应用层内部存在着多种网络应用，这必然会导致更新换代的发生。同时，网络的不断发展势必引入新的网络应用。随着网络应用的升级和新应用的产生，孪生北向接口应能及时地扩展，以满足新网络的应用需求。

（3）可移植：孪生网络层中存在大小不一、功能不同的孪生体，网络应用层中各类应用的相同或相似需求可能部署在不同的孪生体上，因此，孪生北向接口应能够较方便地移植、部署到不同的孪生体上。

（4）灵活易部署：为减少部署时间，降低部署成本，孪生北向接口须能灵活部署。

2.2 孪生层内部接口

孪生网络层内部包含数据共享仓库、服务映射模型和数字孪生体管理3个关键子系统，它们是数字孪生网络最关键的部分。孪生层内部接口指的是这3个子系统内部及其之间的接口。为支持这3个子系统各自的功能及其交互，孪生层内部接口应具备如下4个功能。

（1）统一性：孪生网络层中任一子系统应能通过孪生层内部接口为其他子系统提供统一的数据格式和数据服务，即接口应具备统一性。

（2）适配性：孪生网络层须与网络应用层和物理网络层交互，应能很好地适应于各种网络设备，并与多种接口适配，因此孪生层内部接口也需要具备适配性。

（3）可移植：服务映射模型子系统为不同应用提供的数据模型实例可能存在极高的相似度，为提高效率，数据模型实例须能通过不同的孪生层内部接口提供及部署。

（4）灵活可扩展：孪生网络层须能对不同的网络新业务进行验证，为缩短功能实现时间，孪生层内部的功能实现应尽可能简化，因此孪生网络层内部接口须做到灵活可扩展。

2.3 孪生南向接口

孪生南向接口是孪生网络层与物理实体网络间的接口。控制更新由孪生南向接口下发至物理实体网络，同时物理实体网络中的各种网元通过孪生南向接口同孪生网络层交互网络数据和网络控制信息。因此，孪生南向接口应具备3个功能。

（1）信息交互能力：孪生南向接口应能收集不同物理网元或网络设备的信息，同时将孪生网络中的配置信息下发给物理网络来执行，即能实现孪生网络层与物理实体网络间的信息交互。

（2）实时性：孪生网络配置验证等功能的实现须具备一定的实时性，因此从物理实体网络中采集并上传的信息以及从孪生网络下发给物理网络的配置信息均须具备一定的实时

性，以满足数字孪生网络的实时性要求。

（3）可兼容：不同厂商生产的网络设备及网元所使用的接口及协议千差万别，孪生南向接口应具备良好的兼容性，保证信息采集与配置下发的可靠性。

2.4 通用协议

目前网络中存在多种多样的南北向及网络内部协议，如RESTCONF[4]、NETCONF[5]、OpenFlow[6]、可扩展消息处理线程协议（XMPP）[7]、East-West Bridge[8]等。不同协议适用于不同的孪生网络接口，如表1所示。表1给出了目前适合DTN的一些通用协议的适用性建议。

3 多协议协同的接口实现机制

如上所述，DTN中的物理网络涵盖移动接入网、核心网、数据中心网等多种网络类型，因此网元设备种类繁多，各厂家设备所支持的协议存在差异性。同时，DTN中的网络应用层也须支持不同网络应用的多样化协议。因此，孪生层内部接口须能实现多协议协同，以满足不同厂家的网元或网络设备所支持的多样化协议，以及差异化的数据格式。此外，孪生层内部接口也须支持不同应用、应用升级等带来的需求改变和接口协议的适配变化。同时，由于孪生网络层的构建并非简单的、1∶1的完全复制物理网络，而是通过模型抽象的方式实现物理网络的映射，因此在孪生层内部通过多协议协同实现协议转换等处理，既能实现孪生层内部协议简单化，又不会影响DTN原系统构建。

当前，针对网络中存在的协议类别多的问题，业界也展开了相关研究。例如，文献[10]对窄带物联网中网关处不同协议的转换问题进行了探讨；文献[11]研究了基于OpenFlow协议实现的支持多业务融合和多协议转发的网络架构设计；文献[12]结合了简单网络管理协议（SNMP）、超文本传输协议（HTTP）等多协议，用于网络拓扑发现；文献[13]提出了一种使用元数据以抽象方式指定变量和消息的多代理协议组合方法，实现相同协议配置的多次调用，提升了协议组合的效率。由此可见，多协议转换、融合的研究已有一定基础，但在DTN中如何实现多协议协同仍有待研究。另外，由于孪生北向接口与孪生南向接口所需处理的协议不同，南北向接口协议适配功能存在一定的差异性。

3.1 孪生南向接口协议适配功能

基于上述相关协议融合与协议转换的研究，为实现孪生网络层与物理网络层间协议的转换，保证协议处理的高效性，配置信息分发的准确可执行，并尽可能地降低协议处理的复杂度，本文在孪生网络层与物理网络层交互处引入了南向接口协议适配功能。如图2所示，南向接口协议适配功能由协议配置管理、协议解析及转换、协议识别及匹配和数据管理4个模块组成。

▼表1 通用协议的适用性建议

协议	特性	孪生北向接口	孪生层内部接口	孪生南向接口
RESTCONF	RESTCONF以HTTP作为传输协议，用XML/JSON作为消息交换格式，允许WEB应用以模块化和可扩展的方式访问网络设备的配置和操作数据。	√		
NETCONF	NETCONF使用基于RPC的机制为客户机和服务器之间提供一套能够新增、修改、删除网络设备配置，查询配置、状态和统计信息的框架机制，可以作为网络管理员或网络配置应用程序与网络设备之间进行逻辑连接。NETCONF可传输配置数据和状态数据两类信息。	√		√
OpenFlow	用于OpenFlow交换机与控制器的信息交互。			√
XMPP	用于即时消息传递、多方聊天、语音和视频呼叫、协作、内容联合以及通用的XML数据路由的开放技术。		√	√
I2RS	可基于拓扑变化、流量统计等信息动态下发路由状态和策略，能支持外部应用或控制实体读取路由器中的信息。		√	√
East-West Bridge协议	East-West Bridge是基于TCP/SSL的一种应用层协议，具有良好的移植性和可扩展性。可将网元抽象为节点、链路、端口、流等概念，通过扩展的链路层发现协议获取域内各网元的标识、容量、状态等信息。		√	
SNMP[9]	专门设计用于在IP网络管理网络节点的一种标准协议。网络管理员能够使用SNMP管理网络效能，发现并解决网络问题以及规划网络增长。	√	√	

HTTP：超文本传输协议
I2RS：路由系统接口协议
JSON：一种轻量级的数据交换格式
RPC：远程过程调用
SNMP：简单网络管理协议
SSL：安全套接字协议
TCP：传输控制协议
XML：可扩展标记语言
XMPP：可扩展消息处理线程协议

（1）协议配置管理模块：对物理网络层发给孪生网络层的所有数据包进行处理并得到相应的配置信息，为协议识别及匹配模块和协议解析及转换模块提供所需的配置信息。

（2）协议识别及匹配模块：通过孪生南向接口实现与物理网络层的交互，根据网元设备标识、终端设备信息、接入控制携带的网元信息等，识别并记录物理网络层设备所支持的协议类别，形成相应的终端协议表，具体形式如表2所示。此外，协议识别及匹配模块在孪生网络层完成相应的功能验证并生成相应的网络配置，再将网络配置信息下发至物理网络的具体设备。这时协议识别及匹配模块会根据终端协议表，确保命令传输协议为相应设备所支持的协议类型。

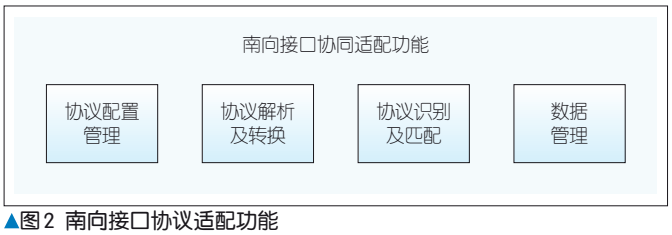
（3）协议解析及转换模块：对从物理网络上传的信息等进行解析并转换为孪生网络层内部数据共享仓库、服务映射模型和数字孪生体管理3个子系统支持的协议类型；或做逆处理，即将孪生层内部3个子系统的数据信息、模型信息、配置信息等解析并转换为外部应用和物理设备支持的协议类型。同时，孪生网络层内部不同子系统功能各不相同，协议解析及转换模块须在孪生层内部将协议转换为统一的且孪生网络层3个子系统均支持的协议格式，以简化孪生网络层内部协议转发及信息交互流程。

（4）数据管理模块：将物理网络层和网络应用层所用的不同协议的不同数据格式转换成孪生网络层内部所用协议适用的数据格式。

孪生层内部南向接口协议适配功能对物理网络层中不同终端、设备用到的多种协议进行识别、解析和转换等操作，简化了孪生网络层内部3个子系统间以及孪生网络层与物理网络层间的信息交互，实现了孪生层内部协议无关的信息处理和数据转发等功能。南向接口协议适配功能简化流程如图3所示。通过引入南向接口协议适配单元，无须过多修改底层物理网络中的网络设备，协议转换、适配等工作就可以全部由南向多协议适配单元完成。这使得孪生网络层的功能更容易实现，数字孪生网络的构建复杂度进一步降低。

3.2 孪生北向接口协议适配功能

与物理层网元设备等支持的协议种类繁多相比，在当前网络应用层中，应用所使用的协议种类数量较少，采用基于Rest应用程序编程接口（API）实现方式的应用占绝大多数。因此，相比于南向接口协议适配功能，孪生北向接口协议适配功能相对要简单一些。类似于南向接口协议适配功能，北向接口协议适配功能同样需要有协议解析及转换模块，实现将基于Rest API接口的网络应用的业务需求转换成网络孪生层可执行的协议类型。

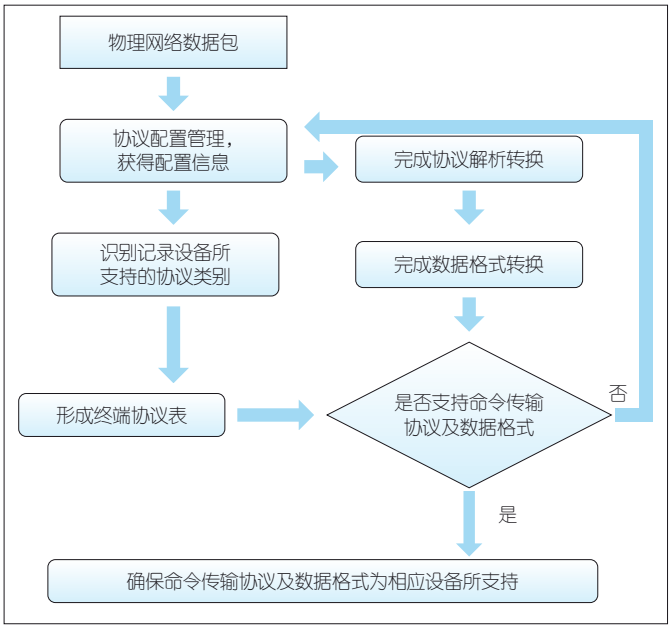


▲图2 南向接口协议适配功能

▼表2 终端及应用协议表

终端/应用	协议类别
交换机1	SNMP
交换机2	OpenFlow
服务器1	XMPP
……	……

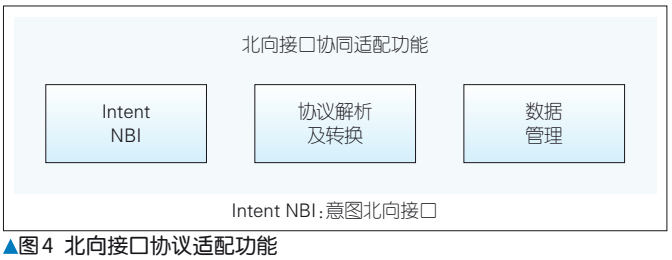
SNMP：简单网络管理协议 XMPP：可扩展消息处理线程协议



▲图3 南向接口协议适配功能流程图

此外，近年来随着意图网络[14]的发展，意图北向接口（Intent NBI）概念及部署逐渐兴起。Intent NBI作为一种与具体网络实现无关的北向接口，仅关注应用相关的内容，而不关注具体的网络协议及网络技术。不同于Rest API类接口的实现方式，Intent NBI采用声明式的表达形式，仅关注实现结果而不指定具体的操作方式。独特的设计理念使Intent NBI在一定程度上具备了协议处理功能，与北向接口协议适配功能高度吻合。目前业界已有关于Intent NBI的一些设计实现工作。因此，我们考虑将Intent NBI作为孪生北向接口协议适配功能的一部分，以简化部分孪生北向接口多协议协同的实现。孪生北向接口协议适配功能的设计如图4所示。

如上所述，网络应用层目前主流的北向接口为基于



▲图4 北向接口协议适配功能

Rest API的实现方式。该方式多采用RESTCONF协议，其他协议占比较少。同时，考虑到随着意图网络的发展，以及Intent NBI设计理念的契合性，在北向接口协议适配功能中，设计引入Intent NBI，同时面向Rest API等接口设置协议解析及转换功能模块和数据管理功能模块，在降低协议处理的复杂度的同时，完成北向接口和孪生层内部接口间的协议转换和统一数据格式等工作。

4 结束语

数字孪生网络作为网络智能化、自治化、6G网络演进的技术支撑，越来越多地受到专家学者的关注。但由于通信网络与其他行业存在明显不同，数字孪生网络的构建不能单纯照搬其他行业中的应用方案，必须考虑通信网络本身的各种特性。本文中，我们从数字孪生网络构建的不同接口应具备的特性出发，给出了当前一些协议在数字孪生网络中的适用性建议。同时，针对孪生网络层如何实现多协议共存和协调的问题，我们提出在孪生网络层面向物理网络层和网络应用层处分别引入南北向接口协议适配功能，以实现南北向多协议的识别、解析与转换，简化孪生网络层内部信息传递、数据转发等功能，并降低对网络现有设备的影响和数字孪生网络构建的复杂度。多协议协调的接口实现只是构建数字孪生网络所面临的诸多问题中的一个，未来我们将继续探讨并研究数字孪生网络构建中所面临的其他问题与挑战。

参考文献

[1] GRIEVES M. Digital twin: manufacturing excellence through virtual factory replication [EB/OL]. [2021-12-12]. https://www.researchgate.net/profile/Michael-Grieves/publication/275211047_Digital_Twin_Manufacturing_Excellence_through_Virtual_Factory_Replication/links/5535186a0cf23947bc0b17fa/Digital-Twin-Manufacturing-Excellence-through-Virtual-Factory-Replication.pdf
[2] TAO F, ZHANG H, LIU A, et al. Digital twin in industry: state-of-the-art [J]. IEEE transactions on industrial informatics, 2019, 15(4): 2405-2415. DOI: 10.1109/TII.2018.2873186
[3] 孙滔, 周铖, 段晓东, 等. 数字孪生网络(DTN): 概念、架构及关键技术 [J]. 自动化学报, 2021, 47(3): 569-582. DOI: 10.16383/j.aas.c210097
[4] BJORKLUND M, SCHONWALDER J, SHAFER P, et al. RESTCONF extensions to support the network management datastore architecture, RFC 8527 [EB/OL]. (2019-03-06)[2021-12-10]. https://datatracker.ietf.org/doc/rfc8527/
[5] BJORKLUND M, SCHONWALDER J, SHAFER P, et al. NETCONF extensions to support the network management datastore architecture, RFC 8526 [EB/OL]. (2019-03-06)[2021-12-10]. https://datatracker.ietf.org/doc/rfc8526/
[6] MCKEOWN N, ANDERSON T, BALAKRISHNAN H, et al. OpenFlow [J]. ACM SIGCOMM computer communication review, 2008, 38(2): 69-74. DOI: 10.1145/1355734.1355746
[7] ABDRE S P. Extensible messaging and presence protocol (XMPP): address format, RFC 7622 [EB/OL]. (2020-01-21)[2021-12-10]. https://datatracker.ietf.org/doc/rfc7622/
[8] LIN P, BI J, WANG Y Y. East-west bridge for SDN network peering [EB/OL]. [2021-12-12]. https://link.springer.com/chapter/10.1007/978-3-642-53959-6_16
[9] FEDOR M, SCHOFFSTALL M, DAVIN J, et al. Simple network management protocol (SNMP), RFC 1157 [EB/OL]. (2013-03-02)[2021-12-12]. https://datatracker.ietf.org/doc/rfc1157/
[10] 孙握瑜. 基于NB-IoT技术的物联网网关多协议转换研究 [J]. 齐齐哈尔大学学报(自然科学版), 2021, 37(4): 49-53
[11] 冯龙. SDN网络控制器面向多协议虚拟网络接口的研究与实现 [D]. 北京: 北京邮电大学, 2014
[12] 潘爽, 苏亚维. 多协议融合的网络拓扑发现技术研究 [J]. 物联网技术, 2021, 11(9): 14-17. DOI: 10.16667/j.issn.2095-1302.2021.09.005
[13] TAKAHASHI R, TEI K, ISHIKAWA F, et al. A flexible protocol composition for multi-party coordination protocols in multi-agent systems [C]//2008 Sixth Annual IEEE International Conference on Pervasive Computing and Communications (PerCom). IEEE, 2008: 609-614. DOI: 10.1109/PERCOM.2008.94
[14] PANG L, YANG C G, CHEN D Y, et al. A survey on intent-driven networks [J]. IEEE access, 2020, 8: 22862-22873. DOI: 10.1109/ACCESS.2020.2969208

作 者 简 介

陈丹阳，中国移动研究院研究员；研究方向为数字孪生网络和意图网络。

陆璐，中国移动研究院基础网络技术研究所副所长、中国通信标准化协会TC5核心网组组长；长期从事移动核心网策略、演进、标准和技术研究工作，主要涉及未来网络架构、智能管道、边缘计算等领域。

孙滔，中国移动研究院首席专家，正高级工程师，3GPP SA2副主席；主要从事移动通信网络架构、网络融合、网络智能化、5G/6G等网络新技术的研发工作；从2009年开始代表中国移动参加3GPP会议，作为报告人完成5G架构的研究和标准制订的工作。

多样化业务需求与全维网络能力的映射

Mapping Between Diversified Service Requirements and Full-Dimensional Network Capabilities

范琮珊/FAN Congshan， 周旭/ZHOU Xu，
任勇毛/REN Yongmao

（中国科学院计算机网络信息中心，中国 北京 100083）
(Computer Network Information Center, Chinese Academy of Sciences, Beijing 100083, China)

DOI：10.12142/ZTETJ.202201009
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20220218.0922.002.html
网络出版日期：2022-02-18
收稿日期：2021-12-26

摘要：提出一种通用的业务需求与网络能力映射方法，通过网络能力自组织和业务需求自映射灵活适配业务发展需求。构建全维可定义网络能力模型，抽象和分解网络各层能力，并从通信主体、网络功能、网络资源、网络安全4个维度以及30多种具体元素实现网络能力开放可定义和动态演进发展。针对复杂的综合性业务需求，“动态地”选择和组合网络能力，对比业务需求与网络能力间匹配度，采用最优能力组合重构复合型网络服务，设计直观的0-1映射矩阵形式，支撑映射实现。

关键词：业务需求；全维可定义网络能力；映射；网络能力组合

Abstract: A general mapping method between service requirements and network capabilities is proposed, which can flexibly adapt to service developing requirements through network capability self-organization and service requirement self-mapping. By abstracting and decom - posing the capabilities of each layer of the network, a full-dimensional definable network capability model is constructed to realize the opening definition and dynamic evolutionary development of network capabilities from four dimensions, including communication subjects, network functions, network resources, and network security, with more than 30 specific elements. The proposed mapping method "dy - namically" selects and combines network capabilities for complex and comprehensive service requirements. Based on the matching degree calculated between service requirements and network capabilities, the optimal combination of network capabilities is selected to reconstruct composite network functions. An intuitive 0-1 mapping matrix form is designed to support mapping realization.

Keywords: service requirement; full-dimensional definable network capability; mapping; network capabilities combination

近年来，电子、计算机和人工智能等技术的飞速发展催生了大量新型的网络业务，远程医疗、车联网、全息通信、虚拟现实（VR）/增强现实（AR）、智慧家庭等业务不断涌现。全新业务随着技术的成熟和升级将逐渐普及，并将改变社会形态与人们的生活方式。与传统网络业务大不相同，新型业务对未来网络提出更高的要求，是网络发展的一项重要挑战。以远程手术为例，医师在远程操作多个协同的医疗设备对患者进行治疗时，需要与用户的身体直接交互。这对安全性提出极高的要求。治疗过程需要手、眼、耳、鼻等多个器官同时参与控制与反馈。不同类型的信息传输及传输性能也要求精准同步。医疗影像视频传输和手术现场画面的实时观察要求分辨率在4K以上，这对网络的时延及可靠性要求也很高。未来业务包含的信息维度逐步增加，协同性逐渐增强，性能要求也越来越高。因此，只有实现多种网络能力的编排组合，才能够有效支撑更精、更尖、更高的网络业务，提升用户体验质量（QoE）。

与此同时，网络的能力也随着技术的进步和硬件的升级不断完善。在传统互联网协议（IP）网络的尽力而为转发能力基础上，源路由、多标识寻址、智能路由、确定性转发、内生安全等全新网络能力先后出现，打破已有网络能力的单一架构，不断扩展网络的能力维度，丰富网络能力的实现形式，提高网络能力支撑业务的力度。

面向多样化业务需求和差异化网络的能力，如何实现两者之间的匹配，并通过组合优化选择合适的网络能力为业务需求提供高效的支撑是一个关键的问题。运营商通过端到端服务质量（QoS）管理完成对业务关键质量指标（KQI）的

基金项目：国家重点研发计划（2018YFB1800100）

监测和控制，将上层的用户感受折射到业务质量模型，再映射到反映特定网络能力属性的网络关键性能指标（KPI），调整和优化各项指标，满足业务需求，提升QoE[1]。文献[2]定义视频流和长期演进（LTE）语音服务影响QoS和QoE的KPI与KQI，并分析QoS和QoE之间的数学关系，通过QoS和QoE的关联关系预测达到规定QoE级别的概率，以衡量用户的业务体验。文献[3]利用机器学习的方法分析KQI与KPI之间的关系，得到影响KQI的KPI，以及KPI劣化导致KQI劣化的概率，完成KPI劣化小区感知，实现问题定位，提升网络优化的主动化、事先化、自动化。在大型网络中构建业务需求与网络能力映射关系时，我们可以采用数学建模的方式，即通过系统抽象和数学推导得出映射函数，解决实际问题。文献[4]以网络效用最大化为目标构建从服务到连接和从连接到路径的多对多映射数学模型，将路径带宽合理分配给各个服务，使得所有服务的效用之和达到最优。

已有网络映射的研究主要针对特定场景下局部业务需求与网络能力映射，缺乏普遍适用的业务需求与网络能力映射方法。局部业务需求与网络能力映射的方式操作简单，但适用性差，且只考虑单一的业务需求。随着业务形态的丰富和多样化，业务需求迅速增长且不断复杂化。当前有限数量的网络能力形式单一、动态性差、效能低、运维僵化，直接导致业务需求与网络能力之间的差距日益扩大，难以采用灵活的网络能力组合匹配未来多元化业务的需求。因此，人们急需一种普遍适用的映射机制，以全面覆盖单一化和复杂化业务需求的映射。本文提出一种通用的业务需求与网络能力映射模型，通过网络能力自组织和业务需求自映射灵活适配业务发展需求，基于“以网络为核心”的设计理念，抽象和分解网络各层能力，从通信主体、网络功能、网络资源、网络安全4个维度进行细粒度划分，获得30多种网络能力元素，实现全维度网络能力可定义。灵活地扩展网络能力类型能够支持全维度可定义网络能力模型的动态加载和演进发展。根据不同的业务需求“动态地”选择网络能力，灵活地组合网络能力，分析对比业务需求与网络能力之间的匹配度，采用最优能力组合重构复合型网络服务，可以有效支撑未来网络专业化、多样化业务需求，实现业务需求到网络能力的映射。

1 多样化业务需求

随着网络规模的不断扩张以及经济、政治、教育、医疗等专业领域的发展，新型的业务场景开始涌现。新业务场景可以分为消费类业务场景和生产类业务场景，如图1所示。消费类业务场景目标是为用户提供极致的服务体验，满足人类社会智慧化需求，包括AR/VR、远程医疗、智慧家庭、全息通信等；生产类业务场景是传统产业与网络基础设施融合的产物，该场景的目标是促进生产力的大力发展，包括车联网、工业互联网、智能电网等。业务类型的丰富对网络提出了多样化功能性需求和性能性需求，具体反映为不同维度、类型的网络能力。新型的业务需求推动技术发展，促进了网络能力的动态演进。

新型业务对网络的功能性需求不断增加，不仅体现在已有功能的全面增强，同时也体现在新的功能性需求。大量的人、手机、传感器、医疗设备，甚至数据、计算作为通信主体接入网络进行通信，网络需要支持数目巨大且类型各异的连接。不同的业务对网络传输质量有不同的需求。例如，安全可靠的远程医疗需要确定性的时延和传输抖动保证，全息通信要求网络支持高通量传输。网络需要根据业务的特性提供定制化、可预测的接入和传输服务，以保证服务质量的确定性和差异化。处于动态变化的业务场景，例如车联网，对移动性支持有超高的要求。网络的发展融入了存储和计算，需要实时感知业务需求和网络状态，进行高效全局的资源管控和编排，优化网络利用率，提升QoE。未来业务场景的复杂化导致更多的安全漏洞，无法通过IP网络“补丁式”的安全方案保障，需要设计内生安全机制，使网络具备内在自免疫、可进化的安全能力，提供高可靠性和隐私性服务。

业务的专业化、智能化使得业务的性能需求更加精准和高效。典型的性能需求包括带宽、时延、抖动及丢包率等。业务超高通量传输需要超大带宽的支持。4K视频的传输需要12 Gbit/s的带宽。大规模科学实验数据传输对带宽的需求已达到100 Gbit/s。抖动是与时延密切相关的业务需求。降低时延、保证有界抖动有助于提供高准确性和高可靠性服务。远程医疗、车联网、工业互联网等业务有明确的端到端时延、抖动的需求：远程手术要求网络传输的基础时延控制在200 ms以内；车联网自动驾驶要求端到端时延小于5 ms；工业互联网的控制业务要求微秒级的时延抖动。精细化控制类业务，比如工业控制、智能电网继电保护等，对丢包率敏感。关键指令的丢失将导致严重后果。因此，精细化控制类业务要求丢包率控制在10^{-3}以下[5]。

2 全维可定义网络能力

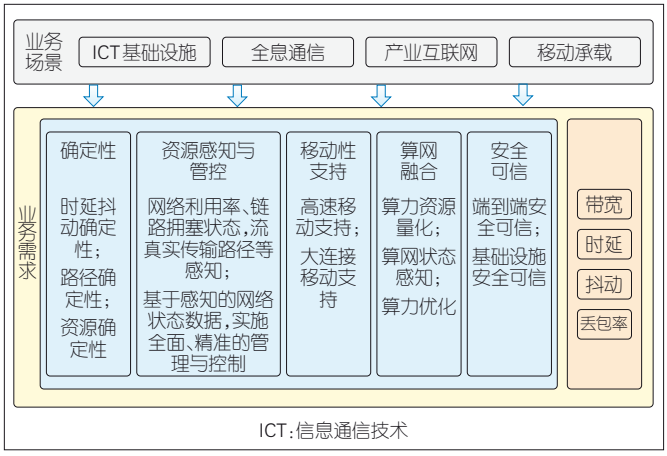
网络技术的发展带来丰富的网络能力。网络维度不断扩展，能力逐步增强。然而，当前网络能力结构僵化、提供方式单一、协调性差，导致网络对新型业务的支持能力低下。为此，本文打破传统面向终端设计网络能力的方式，以网络为中心，构建全维可定义网络能力模型，抽象分解网络各层

能力，细粒度划分网络能力的维度和类别，支持网络能力的灵活扩展，实现多元化网络能力的开放可定义和动态演进发展，为业务需求与网络能力的映射奠定基础[6]。

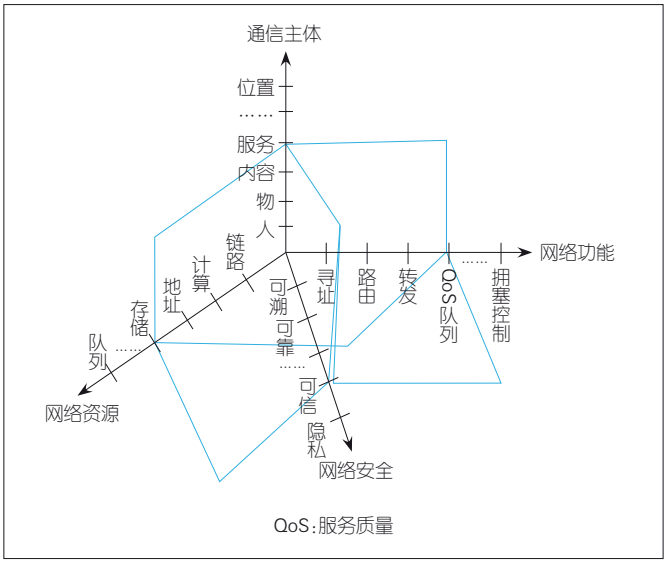
全维可定义网络能力模型构建网络能力空间，如图2所示。整体空间划分通信主体、网络功能、网络资源、网络安全4个维度。每个维度包括不同网络能力类型，一种能力类型支持多种实现形式的能力元素（共30多种）。每个维度的网络能力可以随时更新新型的网络能力，并能及时删除旧网络能力，以保持网络能力模型的动态可扩展性。下面我们对每个维度做具体说明。

（1） 通信主体

通信主体是指网络中参与数据传输行为的主体，包括数据发送方、转发方和接收方。不同的通信主体采用不同的身份标识（ID）进行数据传输。当前网络采用IP地址作为寻址标识。随着工业互联网、卫星网络、车联网等多元化网络融合与互联需求的发展，通信主体的种类不断丰富，支持人（身份）、位置（经纬度、速度、方向）、服务（应用）、物（物联网标签）、内容（视频、文件、图片）等多样化标识并存，实现了大规模网络设备和元素的互通。



▲图1 多样化业务需求



▲图2 全维可定义网络能力

（2） 网络功能

网络功能是指数据在通信主体间完成传递转发设备所承载的功能，包括寻址（定长寻址、变长寻址）、路由（距离矢量路由、链路状态路由）、转发（尽力而为转发、约束路径转发）、QoS队列（先进先出、优先级队列、加权公平队列）、拥塞控制（基于显示拥塞反馈、量化拥塞通知）等能力。随着技术的发展，网络功能不断优化和完善。面对海量的异构通信主体，网络支持多模式接入和连接。为满足专业化、精细化及差异化的业务需求，网络提供可规划、可预期和可定制的数据传输，以保障时延、吞吐量、抖动、丢包率等性能指标，提升用户体验。

（3） 网络资源

网络资源是指数据传输依赖的资源，包括链路（无线、光纤、电缆）、计算（边缘计算、云计算）、存储（内存、硬盘）、地址（互联网协议第4版、互联网协议第6版）等。报文中分配的地址占用地址空间资源，数据在通信主体间传输时占用带宽形式的链路资源。转发设备采用网络处理器及片上内存的计算和存储资源处理报文。硬件升级与软件优化导致网络资源发生了巨大的变化：资源类型不断丰富，多种异构资源并存。资源在网络中的部署位置比较灵活，适合协同调度。资源容量得到提升，体积变小，便于处理。

（4） 网络安全

网络安全维度的网络能力能够保障网络设施、信息和传输的安全，包括可信性（源地址验证、身份验证）、隐私性（加密）、可靠性（完整性校验）和可溯源性（概率包标记法、日志记录）。可信性保证网络信息能够被授权实体访问并合法使用；隐私性保护能够防止信息泄露及非法利用；可靠性通过实时监测，解决网络异常，保证高效正常运行；可溯源性在面对网络攻击时能够快速定位和追踪攻击的源头。随着业务场景的复杂化，外挂式安全技术无法应对网络协议不统一、终端多样化带来的安全隐患，亟须采用内生安全机制[7]。

3 业务需求到网络能力的映射

业务需求与网络能力映射是映射概念在网络服务业务中的具体化。在互联网络中，不同业务面临差异化需求。单个业务实现需要满足的所有条件形成业务需求集合，集合中的每个元素代表一项具体的业务需求。相应地，网络的快速演

进不断丰富网络能力，聚集网络具备的全部能力并构成全维网络能力集合。集合中的每个元素代表一项网络支持能力。业务需求与网络能力映射是为了利用特定网络能力支持业务实现，需要在业务需求集合和网络能力集合两者的元素之间建立对应关系。这种对应关系体现在：一项业务需求需要通过一系列网络能力才能得到满足，同时一种网络能力能够支持多种业务需求。通过选择和组合适当的网络能力，依据合理的顺序执行网络能力，可满足业务需求，实现业务需求到网络能力的映射。

从业务使用者（用户）的角度看，多样化业务需求需要选择合适的网络能力。通过业务需求和网络能力映射，用户从网络运营商提供的可选网络能力方案中选择一种，并为相应的网络能力付费，支持业务实现。从网络运营商的角度，网络具备全维能力。不同的网络能力对业务需求的支持度不同。通过业务需求和网络能力映射，网络运营商为用户提供满足业务需求的网络能力方案和定价，以实现网络能力商品化。

3.1 映射要素

业务需求与网络能力映射的3个要素是原象、象和映射法则。其中，原象是业务实现应该满足的需求元素，象是网络具备的能力元素，映射法则是指原象和象之间对应关系的生成原则。映射法则的产生包括映射形式和映射机制两部分。映射形式包括原象与象一对一、多对一、一对多和多对多4种映射关系。相较于原始的映射定义，映射形式扩展了一对多、多对多两种。映射机制可解决如何将原象和象代表的业务需求元素与网络能力元素进行合理对应的问题。

3.2 映射形式

假设业务实现D具备M项业务需求，采用集合表示为$D=\{d_1,d_2,\cdots,d_M\}$。网络具备N项能力，采用集合表示为$R=\{r_1,r_2,\cdots,r_N\}$。定义矩阵$\boldsymbol{A}=(A_{dr},d\in D,r\in R)$为业务需求和网络能力的映射矩阵。$A_{dr}=1$表示网络能力$r$能够支持业务需求$d$；$A_{dr}=0$表示网络能力$r$不能支持业务需求$d$。业务需求与网络能力间对应关系也可以描述为：$D=\boldsymbol{A}\times R^T$。

图3表示业务需求$D=\{D_1,D_2,D_3\}$与网络能力$R=\{R_1,R_2,R_3,R_4\}$映射关系的示例。对应的0–1映射矩阵为：

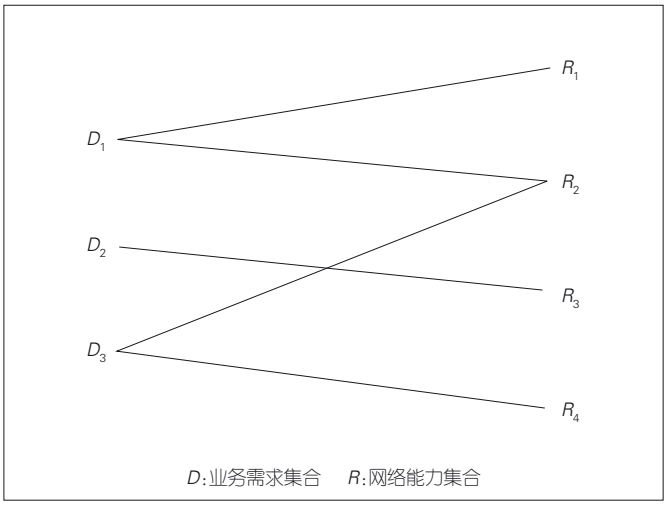
$$\boldsymbol{A}=\begin{bmatrix}1&1&0&0\\0&0&1&0\\0&1&0&1\end{bmatrix}。$$

0–1矩阵形式能够直观地表示业务需求和网络能力之间的复杂映射关系，为灵活的网络能力组合满足业务需求提供有效支撑。矩阵的行代表映射的原象，矩阵的列代表映射的象。矩阵中“1”在行向量和列向量的位置（由下标决定）决定了映射函数的原象和象的对应关系。0–1映射矩阵行向量或列向量中“1”的总数量反映了不同的映射形式。当映射矩阵列向量中“1”的总数量多于一个时，映射形式为多对一，表示一项业务需求需要多种网络能力才能够满足；当映射矩阵行向量中“1”的总数量多于一个时，映射形式为一对多，表示一种网络能力能够支撑多项业务需求；当映射矩阵中所有列向量和行向量的“1”总量都为一个时，映射形式为一对一，表示业务需求与网络能力是一一对应的；当映射矩阵列向量和行向量中的“1”的总数量均多于一个时，映射形式为多对多，表示多项业务需求需要多种网络能力才能够满足。

3.3 映射机制

本文基于业务需求与网络能力间的匹配度设计映射机制。当业务实现规定需求指标时，业务需求与网络能力间的匹配度被定义为：在当前网络状态下，给定网络能力所获得的业务需求指标与规定需求指标间的差值。针对功能性业务需求，当网络执行业务需求规定的网络能力时，定义指标的差值为0，否则差值为1；针对性能性业务需求，定义指标的差值为可达性能指标值与规定性能指标值间的差值。

假设业务D规定的需求指标为$Ind=[Ind_1,Ind_2,Ind_3,\cdots,Ind_M]$。支持业务实现的网络能力集合共有$K$种，表示为$R_k$，$1\leqslant k\leqslant K$。采用$R_k$获得的业务需求指标为$Ind(R_k)=[Ind_1(R_k),Ind_2(R_k),Ind_3(R_k),\cdots,Ind_M(R_k)]$。闵可夫斯基距离可用来衡量两个需求指标向量间的相似度，即业务



▲图3 业务需求与网络能力映射关系示例

需求与网络能力的匹配度：

$G_k = \sqrt[p]{\sum_{l=1}^{M}\alpha_l\left(Ind_l\left(R_k\right)-Ind_l\right)^p}$，其中 $W=\{\alpha_1,\alpha_2,\cdots,\alpha_M\}$ 是业务需求集合 D 对应的优先级权重系数集合。α_l 数值越大，优先级越高，且满足 $\sum_{l=1}^{M}\alpha_l=1$。对比 K 个闵可夫斯基距离值，最小值 $k^*=\min_k G_k$ 对应的网络能力为与业务需求匹配度最高的网络能力 R_{k^*}。利用业务需求集合 D 和网络能力集合 R_{k^*} 生成映射矩阵。

3.4 映射步骤

基于映射三要素，业务需求与网络能力的映射步骤包括原象生成、象生成和映射法则生成3个步骤，如图4所示。

第1步：原象生成

针对用户侧提交的业务需求，拆分复合型业务需求，聚类相似的业务需求，形成业务需求集合 D；根据业务的特性，生成业务需求指标集合 Ind 以及需求权重因子集合 W。

第2步：象生成

网络能力随着业务执行而动态变化，为了更加准确地匹配业务需求，需要提供实时的网络能力状态。利用场景感知等方式，获得动态的网络能力空间 Spa；利用集合标识网络能力 $C=\{C_1;C_2;\cdots;C_I\}$，其中 C_i 表示第 i 种维度的网络能力，包括通信主体、网络功能、网络资源、网络安全等多个维度。每个维度的网络能力具体包括多种网络能力元素。

第3步：映射法则生成

映射法则的生成主要包括网络能力组合、业务需求与网络能力匹配度计算和映射矩阵的生成。

（1）网络能力编排

网络能力组合过程包括：依次分析业务需求 d_m，从网络能力空间的各维度中选择支持业务需求的单个或多个网络能力，形成网络能力组合 $R^m=\{r_j,r_j\in Spa\}$；遍历业务需求集合，将 M 项网络能力组合合并，保留具有关联关系的网络能力，删除具有互斥关系的网络能力，形成全新的网络能力集合 $R_k=\bigcup_{m=1}^{M}R^m$。由于存在不同类型的网络能力均支持同一业务需求的情况，因此网络能力组合 R^m 的数量可能大于1。通过灵活编排，可最终获得 K 个网络能力组合 R_k，$1\leqslant k\leqslant K$。

（2）业务需求与网络能力匹配度计算

假设业务 D 规定的需求指标为 $Ind=\left[Ind_1,Ind_2,\cdots,Ind_M\right]$，网络能力组合 R_k 获得的业务需求指标为 $Ind\left(R_k\right)$，我们利用闵可夫斯基距离 G_k 来衡量需求指标向量 $Ind\left(R_k\right)$ 和 Ind 间的距离。

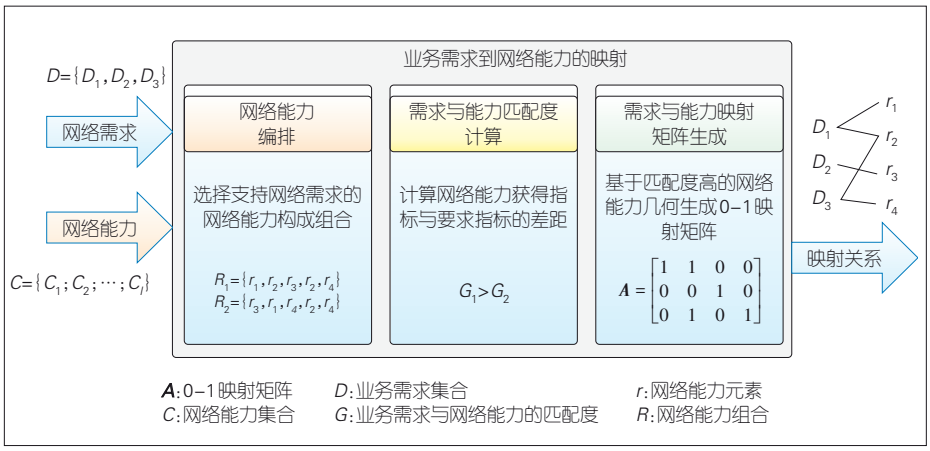
（3）映射矩阵生成

对比 K 个闵可夫斯基距离值，选择最小值 $k^*=\min_k G_k$ 对应的网络能力为与业务需求匹配度最高的网络能力 $R_{k^*}=\{r_1,r_2,\cdots,r_N\}$。利用业务需求集合 D 和网络能力集合 R_{k^*} 生成 M 行、N 列的映射矩阵 $\boldsymbol{A}$。矩阵元素 $\boldsymbol{A}_{m,n}=1$ 表示网络能力 r_n 支持业务需求 d_m，$\boldsymbol{A}_{m,n}=0$ 表示网络能力 r_n 不支持业务需求 d_m。

4 映射实例

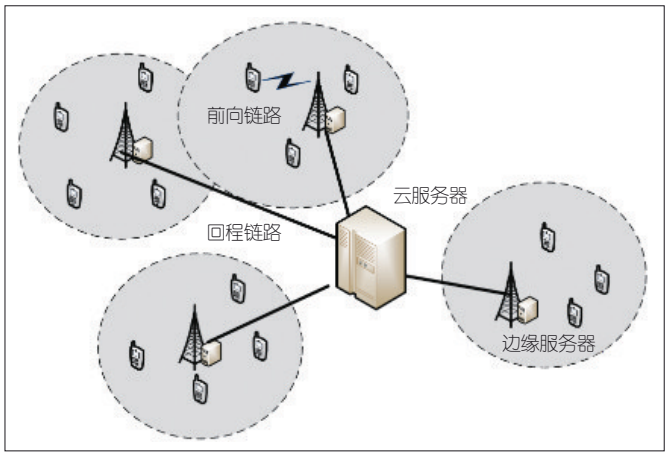
本节以边缘计算场景下任务卸载调用网络资源维度的网络能力来分析业务需求与网络能力映射。任务卸载主要涉及计算资源和链路资源。其中，计算资源支持边缘计算、云计算和云边协同计算3种网络能力元素，链路资源包括无线前向链路和光纤回程链路2种网络能力元素。

如图5所示，部署区域中共有1个云服务中心和4个基站。基站通过光纤回程链路连接到云中心，并且回程容量为 R_{bac}=10 Mbit/s。基站具备边缘计算能力，用户均匀地分布在基站覆盖范围中。由于用户侧计算容量有限，任务将被卸载至边缘服务器或者云中心执行。规定数据包的产生传输均以时隙 $T_S=1\times10^{-3}$ s 为单位。用户在时隙 T_S 以概率 δ 产生数据包，数据包大小为 F=20 kbit。任务到达边缘服务器或云服务器的过程服从泊松分布，到达率为 λ_{mec} 和 λ_C。



▲图4 业务需求与网络能力的映射步骤

边缘服务器和云服务器具备不同



▲图5 边缘计算任务卸载场景

的计算容量。数据包在服务器中的服务时间服从均值为$1/\mu$的指数分布，其中μ表示服务速率，并满足$\mu_C = 2\mu_{\text{mec}} = 3\,\text{packet/slot}$。服务时间的概率密度函数（PDF）表示为：$f_{T_{\text{ser},x}}(t) = \mu_x \exp\left(-\mu_x t\right), x \in \{\text{mec}, C\}$。

任务卸载的开销与数据包到达率有关。设置边缘服务器与云服务器的单位开销系数为$\varepsilon_x, x \in \{\text{mec}, C\}$，并且满足$\varepsilon_C = \varepsilon_{\text{mec}} = 2\,\text{unit/packet}$。用户侧提出业务需求：以最小化时延和最小化开销为目标实现数据包处理，并且业务需求以最小时延为主，规定优先级权重系数$\alpha = 0.7$。

网络运营商为用户的业务需求提供了3种网络能力组合。

（1）边缘计算：用户通过无线链路将数据包传输至边缘服务器，完成数据包处理；

（2）云计算：用户通过无线链路将数据包传输至基站，基站通过回程链路将数据包上传至云服务器，完成数据包处理；

（3）云边协同计算：用户通过无线链路将数据包传输至基站，同时数据包以协同概率θ选择边缘服务器进行处理，并以概率$(1-\theta)$选择通过回程链路上传至云服务器进行处理[8]。

采用随机几何理论和排队论分析任务卸载的时延和开销性能，通过仿真对比3种网络能力组合的匹配度，可实现不同应用场景下业务需求与网络能力的映射。

边缘服务器和云服务器的数据包处理符合$M/M/1$的队列系统。数据包处理时延包括排队时延和服务时延。根据排队论理论[9]，平均处理时延为$1/(\mu-\lambda)$，其中λ和μ分别表示数据包到达率和服务速率。

相应地，第1种网络能力方式的总时延等于边缘计算的平均处理时延：

$$T_1 = T_{\text{mec}} = \frac{1}{\mu_{\text{mec}} - \lambda_{\text{mec}}}。\tag{1}$$

第2种网络能力方式的总时延等于云计算的平均处理时延与回程链路传输时延之和。任务到达云服务器的速率服从泊松过程，满足$\lambda_C = 4\lambda_{\text{mec}}$。总时延表示为：

$$T_2 = T_C + T_{\text{back}} = \frac{1}{\mu_C - \lambda_C} + \frac{F}{R_{\text{back}}}。\tag{2}$$

第3种网络能力方式的总时延包括边缘计算时延和云计算时延两种情况，具体可表示为：

$$T_3 = \theta\left(\frac{1}{\mu_{\text{mec}} - \lambda_{\text{mec}}}\right) + (1-\theta)\left(\frac{1}{\mu_C - \lambda_C} + \frac{F}{R_{\text{back}}}\right)。\tag{3}$$

计算卸载的开销与服务器的数据包到达率有关，第1种网络能力方式的开销等于边缘计算的开销：

$$C_1 = C_{\text{mec}} = \varepsilon_{\text{mec}}\lambda_{\text{mec}}。\tag{4}$$

第2种网络能力方式的开销等于云计算的开销：

$$C_2 = C_C = \varepsilon_C\lambda_C。\tag{5}$$

第3种网络能力方式的开销等于边缘计算开销和云计算开销之和：

$$C_3 = \theta\varepsilon_{\text{mec}}\lambda_{\text{mec}} + (1-\theta)\varepsilon_C\lambda_C。\tag{6}$$

用户的业务需求是保证最小化时延和开销。以时延和开销为指标，基于两者的优先级关系设计业务需求与网络能力的匹配函数$G(z) = \alpha T(z) + (1-\alpha)C(z)$。将公式（1）和（4）、（2）和（5），以及（3）和（6）分别代入匹配函数，以仿真分析3种网络能力方式在不同数据包到达率设置下的匹配函数。

图6对比了时延指标需求优先级权重系数分别为$\alpha = 1$和$\alpha = 0.7$的匹配函数。$\alpha = 1$是传统网络能力满足单一业务需求的方法，只关注时延指标。从图6中可以看出，只有在数据包达到率极低的情况下，选择边缘计算获得的匹配函数最小，即$\lambda_{\text{mec}} \leqslant 0.2\,\text{packet/slot}$。当增加数据包达到率时，$\lambda_{\text{mec}} > 0.2\,\text{packet/slot}$，选择云计算获得的匹配函数最小；由于大量数据包卸载到中心云服务器会增加服务开销，同时导致数据包堆积，因此云计算网络能力组合不适用于业务密集和负载均衡的场景。

综合考虑时延和开销的需求指标，在$\alpha = 0.7$的情况下，当数据包到达率增加时，匹配度值会随着时延和开销的增加而不断增加。当数据包到达率较小时，$\lambda_{\text{mec}} < 0.5\,\text{packet/slot}$，选择边缘计算获得的匹配函数最小。原因在于边缘服务器距离用户近，服务器计算容量足够支撑少量的计算任务，并且边缘服务器的开销远小于云中心服务器。在稀疏业务场景下，用户可以选择第1种网络能力方式，就

近采用边缘服务器卸载计算任务。

当数据包到达率逐渐增大时，即 $\lambda_{mec} \geqslant 0.5$ packet/slot，将数据包卸载到计算容量较小的边缘服务器会导致数据包处理堆积，排队等待时长增加，从而导致总时延增加。当选择将数据包卸载到云服务器时，虽然回程链路传输会带来额外的时延，并且中心云服务器的开销大，但是云服务器的大计算容量能够补偿回程链路带来的时延，减轻开销的影响。在密集业务场景下，用户可以选择第2种网络能力方式，采用中心云服务器卸载计算任务。

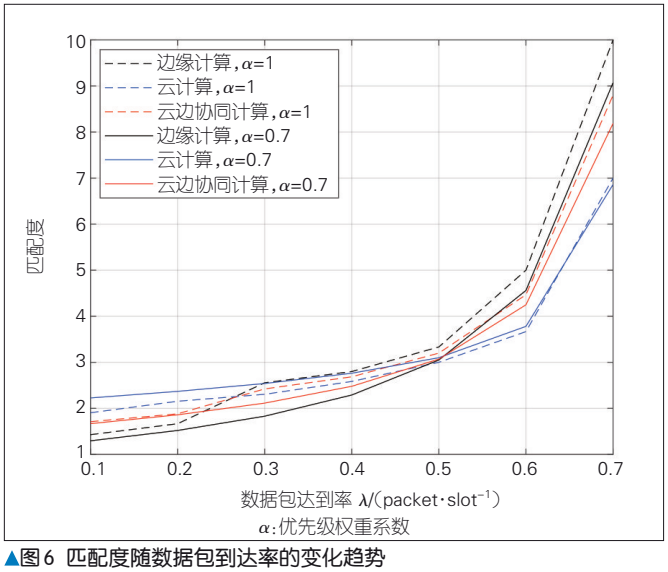
网络中数据包的生成和处理都处于动态变化中，服务器的负载也随之变化。图6中，云边协同计算方式的匹配度位于边缘计算和云计算两者中间。当网络中云中心服务器负载较大，并且数据包的到达率高时，可以选择云边协同计算的方式，满足用户最小化时延和开销的要求，同时通过调节协同概率 θ，实现网络的负载均衡。

以负载均衡场景下的任务卸载为例，业务需求集合表示为 D={时延最小化，开销最小化}，网络能力组合表示为 R={边缘计算，云计算，前向链路，回程链路}，业务需求与网络能力的映射矩阵表示为：

$$\boldsymbol{A}=\begin{bmatrix}1 & 1 & 1 & 1\\ 1 & 1 & 0 & 0\end{bmatrix}。\tag{7}$$

5 结束语

本文提出一种普遍适用的业务需求与网络能力的映射方法。该方法可构建全维可定义网络能力模型，从通信主体、网络功能、网络资源、网络安全等多个维度细粒度划分和定义网络能力，支持网络能力的扩展演进；针对多样化业务需求，通过动态网络能力组合、业务需求与网络能力匹配度计算和映射矩阵生成3个具体步骤形成业务需求到网络能力的映射，以灵活适配业务发展需求。



▲图6 匹配度随数据包到达率的变化趋势

参考文献

[1] 倪萍, 廖建新, 王纯, 等. 一种KPI映射到KQI的通用算法 [J]. 电子与信息学报, 2008, 30(10): 2503-2506

[2] VASER M, FORCONI S. QoS KPI and QoE KQI Relationship for LTE Video Streaming and VoLTE Services[C]//2015 IEEE 9th International Conference on Next Generation Mobile Applications, Services and Technologies. IEEE, 2015: 318-323. DOI: 10.1109/NGMAST.2015.34

[3] 杨磊. 基于FPgrowth机器学习的影响用户感知无线根因问题的快速定位方法研究 [J]. 江苏通信, 2019, 35(2): 56-62. DOI: 10.3969/j. issn. 1007-9513.2019.02.017

[4] 李世勇, 秦雅娟, 张宏科. 基于网络效用最大化的一体化网络服务层映射模型 [J]. 电子学报, 2010, 38(2): 282-289

[5] 蒋林涛. 数据网的现状及发展方向 [J]. 电信科学, 2019, 35(8): 1-15. DOI: 10.11959/j.issn.1000-0801.2019202

[6] 范琮珊, 周旭, 覃毅芳, 等. 全维可定义网络能力 [J]. 电信科学, 2021, 37(10): 31-38. DOI: 10.11959/j.issn.1000-0801.2021235

[7] 郭少勇, 齐芫苑, 代美玲, 等. 面向智能共享的内生可信网络体系架构 [J]. 通信学报, 2020, 41(11): 86-98. DOI: 10.11959/j.issn.1000-436x.2020193

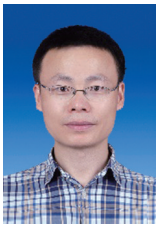
[8] MUKHERJEE S, LEE J. Edge computing-enabled cell-free massive MIMO systems [J]. IEEE transactions on wireless communications, 2020, 19(4): 2884-2899. DOI: 10.1109/TWC.2020.2968897

[9] KLEINROCK L. Queueing systems: theory, vol. 1 [M]. New York: Wiley, 1975

作　者　简　介

范琮珊，中国科学院计算机网络信息中心助理研究员；主要研究领域为未来网络和边缘计算；已发表论文10余篇。

周旭，中国科学院计算机网络信息中心研究员、先进网络与技术发展部主任；主要研究领域为未来网络架构、5G、边缘计算等；主持科技部、工信部、发改委、中科院等重大项目10余项；获得部级科学技术奖一等奖1次、二等奖1次；发表论文60余篇，申请专利50余项，主持或参与制定国际标准和行业标准5项。



任勇毛，中国科学院计算机网络信息中心研究员、先进网络部网络体系结构与系统实验室主任；主要研究领域为计算机网络协议及体系结构；主持国家自然科学基金2项、北京市自然科学基金2项，国家重点研发计划子课题、国家科技重大专项子课题等10余项科研项目；发表论文70余篇，申请发明专利10余项，提交ITU-T国际标准提案2项。

一种轻量化传输模拟器设计与实现

Design and Implementation of a Lightweight Transmission Simulator

叶洪波/YE Hongbo[1]，潘俊臣/PAN Junchen[2]，
崔勇/CUI Yong[2]
（1. 国网上海市电力公司，中国 上海 200122；
2. 清华大学，中国 北京 100084）
(1. State Grid Shanghai Municipal Electric Power Company, Shanghai 200122, China;
2. Tsinghua University, Beijing 100084, China)

DOI：10.12142/ZTETJ.202201010
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20220222.1804.002.html
网络出版日期：2022-02-23
收稿日期：2021-12-20

摘要：传输优化方案的验证依赖于模拟器，但大量基于机器学习的优化方案和端网结合的传输优化方案的出现，使得现有模拟器不适应日益复杂的传输优化方法。为了解决该问题，提出了一种轻量化传输模拟器。该模拟器支持多种传输功能的模拟，并利用算法接口抽象提升模拟器的易用性。通过在所提出的模拟器上构造常见传输场景并对多种算法进行模拟实验，验证了本传输模拟器的有效性和实现的正确性。

关键词：传输模拟器；拥塞控制算法；码率自适应

Abstract: The verification of transmission optimization schemes depends on simulators, but existing simulators are not suitable for increasingly complex transmission optimization methods, including optimization schemes based on machine learning and transmission optimization schemes combined with end networks. To solve this problem, a lightweight transmission simulator is proposed. The simulator supports the simulation of multiple transport functions and uses algorithmic interface abstraction to improve the simulator´s ease of use. Experiments are carried out on the simulator proposed in this paper through common transmission scenarios and different optimization algorithms to verify the effectiveness and correctness of the implementation of the simulator.

Keywords: transmission simulator; congestion control algorithm; adaptive bitrate

随着互联网的发展，直播、线上会议、云游戏等新应用和新场景不断涌现，对网络传输性能都提出了更高的要求，传输优化研究也因此越来越成为业界关注的焦点。为了满足新场景下的传输需求，业界在网络传输过程中的各个细分领域提出了大量的优化方案。码率自适应、拥塞控制算法、队列管理算法等处于网络传输的端侧或网络侧，不同网络层次的优化策略研究不断涌现。这些方案从不同的角度对传输过程进行优化，以期获得更好的服务质量和用户体验。

作为传输优化研究过程中各类方案验证和测试的重要工具，传输模拟器对传输机制的研究有着很重要的意义。相对于在真实设备上进行传输实验，在模拟器上的实验有着部署简单、运行速度快等诸多优点。这些优点使得模拟器成为传输机制研究初期方案验证时的最佳工具。传输模拟器的拟真程度、性能好坏、支持的功能是否全面、是否便捷开发等特性，很大程度上影响新型传输机制的验证效率。

随着传输优化方案设计的日益复杂和机器学习方法的引入，现有模拟器的设计很难满足研究人员的需要。从功能上看，不同细分领域的传输研究对模拟器功能有不同的要求。功能过于单一的模拟器系统不能满足研究人员的需求。从易用性上看，现有传输模拟器的功能不足以支持复杂的方案设计，这带来了使用上的不便和额外改动的成本。一方面，传输机制上的创新方案往往需要对现有模拟器主逻辑进行较大规模的改动，这与模拟器本身作为快速验证手段的目的背道而驰；另一方面，机器学习方法的引入使得模拟器系统同时肩负起快速生成大量学习样本的职责，这对传输模拟器提出了轻量化的要求。

综上所述，现有模拟器已经不能满足日益复杂的传输优化方案快速性能验证需求，因此需要设计轻量化的模拟器来满足传输优化方案的性能验证的新需求。本文首先分析传输

基金项目：国家电网有限公司科技项目（5100-202117386A-0-0-00）

优化研究的几个重要领域的特点，总结这些研究方案对模拟器功能的需求；其次，对现有模拟器存在的问题进行了归纳和总结；以此为基础，提出了一种适应新型传输优化方案的模拟器；最后，通过常见传输场景和算法的模拟效果实验，验证了模拟器的有效性和实现的正确性。

1 多样化传输优化方案对模拟器功能需求

近年来，拥塞控制算法和码率自适应（ABR）算法是网络传输优化研究中的热点。基于启发式策略的算法和新兴的基于机器学习的算法层出不穷。这些算法研究工作大多需要利用模拟器进行方案性能验证。算法复杂度的不断增加，对模拟器的功能也提出了更高的要求。

1.1 常见细分领域研究需求

传输优化领域包含的研究很多，每一种细分领域需要模拟器实现的功能不同。本文以拥塞控制和码率自适应两个领域为例，给出模拟器需要实现的几种重要功能。

拥塞控制算法的研究关注源端速率控制策略，需要模拟传输协议和网络链路。源端拥塞控制算法控制传输速率的方法大致分为两种：基于窗口的策略和基于速率的策略。具体来说，基于窗口的策略通过拥塞窗口及平滑速度控制发送，基于速率的策略则直接通过发送速度来控制发送。窗口策略常见于启发式的传统算法，例如Reno[1]、Cubic[2]只决策拥塞窗口控制发送。瓶颈带宽和位移传播时间（BBR）[3]则以拥塞控制窗口为主，平滑速度控制发包速率为辅。速率策略则常见于机器学习拥塞控制算法中，例如，面向性能的拥塞控制（PCC）[4]直接决策发送速度值来控制发送速度。为了支持不同拥塞控制方案的实施，模拟器需要同时实现上述不同的控制策略。同时，拥塞控制算法研究还需要完整的网络链路功能模拟来支持网络波动模拟，还原真实网络中拥塞控制算法性能。

ABR研究则从应用层角度优化传输过程，并需要模拟流媒体传输的过程。目前常见的流媒体点播协议，例如基于超文本传输协议（HTTP）的动态自适应流媒体传输协议（DASH）[5]，都通过视频分片的形式进行流媒体的传输。由于多种码率的存在，需要码率自适应算法来自动选择适配当前网络质量的视频块。例如，RobustMPC[6]将ABR问题抽象为控制论问题，并利用控制论算法解决视频码率和带宽能力的适配；Pensieve[7]则利用强化学习的方法，利用机器学习模型获取最优策略。总体上，ABR算法需要模拟器实现流媒体点播协议逻辑、ABR算法的统计输入和决策执行逻辑。

1.2 机器学习方法的新需求

随着机器学习方法在传输优化中的大量应用，模拟器的运行效率也成为重要的性能指标。Remy[8]、PCC、PCC-Vivace[9]等拥塞控制算法使用机器学习方法指导决策；Orca[10]则通过传统策略和机器学习的结合，在降低运算开销的同时提升了吞吐和时延表现；Pensieve则首先将强化学习应用于码率自适应研究中，大幅提升了ABR算法在复杂网络状态下的决策性能。虽然两个研究领域的模拟功能需求有所不同，但都依赖于大量实验样本进行机器学习调优，这对模拟器性能提出了更高的要求。机器学习方法往往需要在模拟器系统中进行初步的模拟训练，经过调优后再在真实环境中进行细调。因此，模拟器的性能对机器学习算法的训练效率的影响很大。

2 现有传输模拟器方案缺点

日益复杂的传输优化研究，对模拟器系统的功能提出了更高的要求。但现有的传输模拟器方案往往不能同时满足所有的研究需求，且大部分缺乏易用性设计。这导致学习成本增加、研究效率下降。本文对常见的传输模拟器NS-3[11]和Mahimahi[12]做简要介绍，并分析优劣，指出了当前传输模拟器面临的问题。

2.1 NS-3模拟器

NS-3是一个开源的、可拓展的、基于事件的网络模拟器，旨在服务网络领域的研究和教学工作。NS-3模拟器最大的特点是支持极高的模拟精度。例如，NS-3模拟器可以生成接近系统内核原生网络栈精度的Internet控制报文协议（ICMP）、用户数据报协议（UDP）、传输控制协议（TCP）等多种真实的端到端交互模拟。路由层面支持不同的路由协议，链路层面则支持Wi-Fi、以太网等不同的链路协议。总体上讲，NS-3在设计上着重关注模拟器的拟真度，对网络的模拟非常细致。

但是NS-3模拟器过于复杂且学习成本较高，这导致其易用性较差。另外，高精度的模拟也不利于快速运行：

（1）NS-3基本框架比较复杂，修改起来相当困难，对研究人员并不友好。针对不同协议的模拟在增加模拟精度的同时，也影响了代码的复杂度。

（2）NS-3对链路、路由、传输协议等进行高精度模拟，这提升了模拟器的拟真程度。但对于传输研究来说，NS-3模拟了很多研究者不关心的因素，这增加了开发难度。

（3）NS-3复杂的模拟机制限制了NS-3的性能，这导致其并不适合大规模生成模拟样本。随着机器学习方法的引

入，快速生成大量模拟样本的能力已经逐渐成为模拟器重要功能指标。

2.2 Mahimahi 模拟器

Mahimahi 模拟器最早被用于 HTTP 流量的记录和重放，目前常用于网络传输模拟。Mahimahi 是一个数据包级别的端到端传输模拟器，它能够模拟中间瓶颈链路的缓存变化情况，从而反映带宽和往返时延（RTT）的变化。在 ABR 领域基于机器学习的经典算法 Pensieve 的实验中，底层传输模拟器使用的就是 Mahimahi。Mahimahi 最大的特点是框架较为轻量。模拟器的设计目标并不是模拟复杂的网络设备，因此模拟过程相对粗粒度。这却带来了较好的性能。同时，相比于 NS-3，Mahimahi 框架简单易理解。

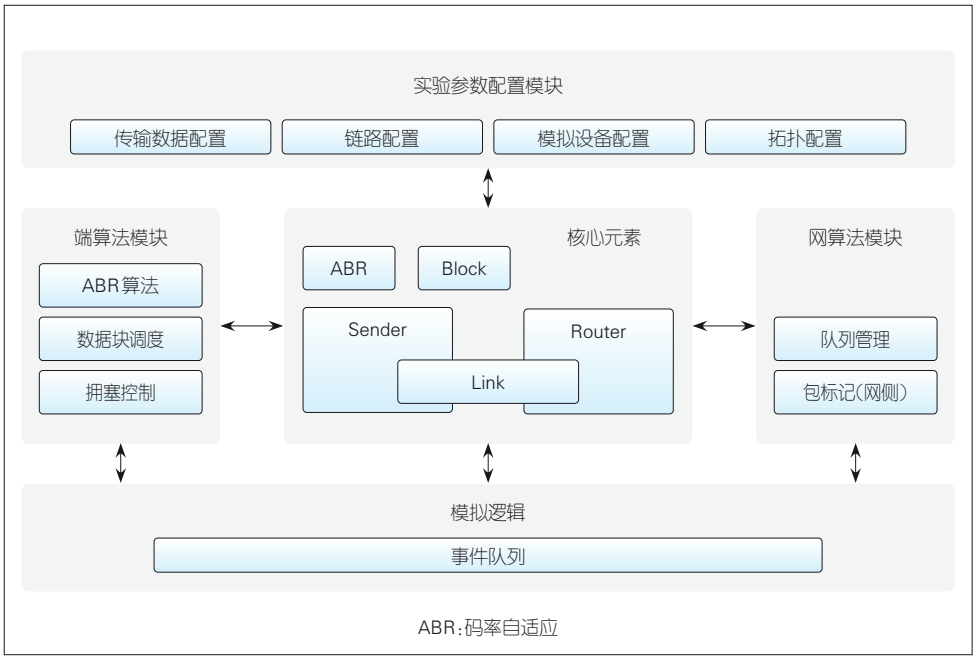
Mahimahi 模拟器由于功能有限，也有着局限性，如 Mahimahi 不支持复杂网络拓扑。对于数据中心这种层次路由的场景，不同流之间不仅有独立的缓存（例如直连的交换机），还有共享的缓存（例如出口路由器）。Mahimahi 不支持这种复杂的缓存模拟，也因此不能模拟复杂拓扑中多对节点之间的同时通信。这就使得 Mahimahi 的模拟功能局限于端到端的场景，不能适配其他传输优化研究。

3 轻量化传输模拟器设计

为了解决现有模拟器不能满足研究需要的问题，本文设计了一种轻量化传输模拟器。本节主要介绍模拟器的总体架构、轻量化设计和易用性设计。

3.1 总体架构设计

模拟器的总体架构包含 4 个模块：算法模块、参数配置模块、核心元素和底层模拟逻辑，具体如图 1 所示。模拟器的底层模拟逻辑是基于事件队列实现的，它利用事件队列将各个功能模块联系起来，实现了模拟设备间复杂的调用关系。在模拟器底层框架之上，还有若干核心元素，构成了模拟设备、链路等模拟要素。具体来说，Block 和 ABR 组件分别负责静态流量重放和 ABR 机制模拟，是模拟器环境中的应用层功能。Sender 元素模拟传输中的端系统以实现模拟器的传输层功能。Sender 元素同时还利用算法接口与数据块调度、拥塞控制等算法模块交互，完成算法的状态更新，并按照决策进行模拟传输。Router 元素模拟中间转发节点的功能，负责模拟缓存和队列调度机制，同时利用算法接口与队列管理算法、包标记算法等模块交互，实现设备算法的自定义。Link 元素负责模拟传输中的链路，代表了模拟设备（Sender 和 Router）之间的连接，其传输特性由链路配置接口指定。



▲图1 模拟器总体架构

3.2 轻量化设计

我们通过简化模拟机制来减少模拟器的运行开销，实现轻量化模拟器设计。具体来说，我们将传输过程中较为复杂但对模拟精度影响不大的设计取消，从而提升了运行效率。

以传输过程简化和 ABR 模拟简化为例，模拟器只模拟其中对精度影响较大的机制。首先是传输过程的简化，模拟器只模拟稳态下的传输流程，不对建立连接等短暂状态进行模拟。虽然可靠的传输协议通常有比较复杂的状态机制，但不同状态下的流量特征有很大的区别，例如 TCP[13]和快速用户数据报协议网络连接（QUIC）[14]的链接建立过程差异就很大。这些可靠传输机制的背后是复杂的网络栈状态机的实现。对于模拟器来说，完整的模拟是非常影响模拟性能和框架简洁性的。但是这些区别于稳态的其他状态往往只占用了传输事件的极短时间，因此单纯模拟稳态的传输过程不会严重影响模拟器的拟真度。所以，模拟器只模拟稳态的传输过

程。其次是ABR过程模拟，模拟器只对比较关键的视频块码率切换逻辑做模拟，忽略了其他的协议细节。例如，完整的点播协议中包含Meta数据的交互、HTTP头参数等，而模拟器则忽略了这些协议细节。服务端和用户端读取trace文件共享Meta信息，只模拟视频传输逻辑。

3.3 易用性设计

为了实现模拟器的易用性，本文设计并实现了多种算法接口。具体来说，模拟器的设计总结了现有传输方案的特点，并设计了4种算法接口，以为不同需求的研究提供支持。4种接口分别是源端ABR算法接口、源端拥塞控制算法接口、中间设备队列调度算法接口和中间设备数据包标记接口。其中，ABR算法接口主要服务于流媒体传输优化研究中ABR算法的实现；拥塞控制算法接口则服务于拥塞控制研究；中间设备队列调度接口控制中间设备队列管理的决策，方便设备缓存管理的研究；包标记接口则用于给数据包标记额外的传输信息，支持端到网和网到端的反馈，可用来实现端网协同的设计方案，例如高精度拥塞控制（HPCC）[15]、数据中心传输控制协议（DCTCP）[16]等。4种接口之间可以独立开发、相互组合，也可以协同设计以完成更复杂的功能，以满足研究者的不同需求。

除了算法接口之外，我们还对模拟器实验场景配置中常用的几种功能进行了接口抽象，具体包括拓扑配置、链路状态配置、设备配置和流配置4种。拓扑配置接口负责表达模拟设备连接关系，用来将模拟设备组成实验拓扑；链路状态配置接口则用来配置设备间链路的状态信息，包括传输物理延时、带宽大小、丢包率等；模拟设备配置接口负责规定模拟设备性能参数以及算法；传输数据配置接口则主要规定了模拟设备之间的静态传输任务，用来配置流量分布。通过这4种接口，研究者能够快速构造目标的实验场景的具体参数，进行验证性实验。

4 验证性实验

我们在模拟器中实现了几种典型传输优化方案，并进行了功能验证实验，用来验证模拟器的功能正确性。我们分别从拥塞控制研究、端网协同算法研究和应用层码率自适应研究角度出发来构造场景，实验验证了多流竞争、DCTCP模拟和ABR模拟的效果。

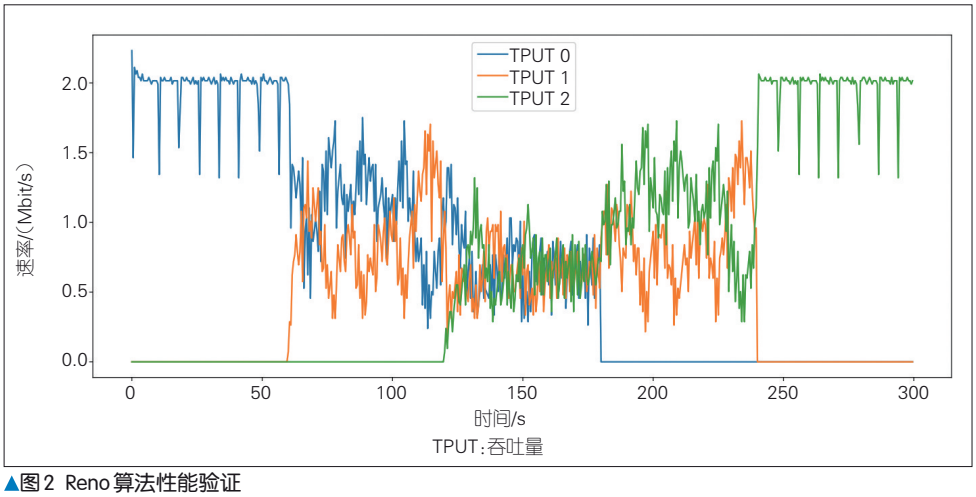
4.1 多流竞争场景

多流竞争实验是拥塞控制算法研究中重要的实验场景，常用于对比算法之间的公平性。Reno是最广泛应用的拥塞控制算法，其包含的慢启动、拥塞避免、快速重传和快速恢复机制是现有的很多新型拥塞控制算法的基础。Reno算法在执行过程中需要维护拥塞控制窗口和阈值。当现有拥塞控制窗口低于阈值时，Reno算法进入慢启动状态，拥塞窗口按照指数增长；反之则进入拥塞避免状态，采用线性增长。拥塞窗口限制了同时能在网络中传输的数据包数量，也就达到了避免网络拥塞的目的。Reno核心的和式增加、积式减少机制自1988年提出以来，就是基于拥塞窗口的算法中最经典的控制机制，后续的很多研究都基于此。因此，我们选用Reno算法验证模拟器对多流复用同一信道的模拟效果。

我们模拟了一个3对1的传输场景，即3条流共享1个转发节点。3个端系统分别运行1个Reno的拥塞控制算法，并在0 s、60 s、120 s依次开始向同一个节点发送数据，分别运行180 s后结束。在此过程中，3个端系统上的Reno会经历单流、双流竞争、三流竞争，再到双流竞争、单流，端系统吞吐结果如图2所示。多流竞争的场景还原了Reno多流竞争时的特点，双流和三流的竞争场景基本上实现了多流的公平性。从吞吐上看，多条Reno流量均分了带宽，模拟的转发节点也维持了2 Mbit/s的转发速度，整体模拟效果符合预期。

4.2 DCTCP模拟场景

作为一种利用网络设备反馈信息优化拥塞控制的经典算法，DCTCP很适合用来验证模拟器在复杂算法设计研究中



▲图2 Reno算法性能验证

的功能。DCTCP使用的显式拥塞通告（ECN）是一种广泛应用的机制，绝大多数商用交换机都已经实现了类似的功能。具体来说，交换机感知实时的队列长度，并在其超过设定阈值时标记数据包，以通知端系统主动退让。DCTCP利用ECN机制，实现了算法收敛快和传输时延低。DCTCP使用与ECN信号相关的alpha值来调整拥塞窗口（CWND），这在加速了收敛的同时使得链路缓存维持在特定阈值。区别于Reno倾向于填满缓存，DCTCP会主动降低队列占用，从而降低了时延。

我们构造了1对1的传输场景，用来验证DCTCP的模拟效果。链路的传输速度在20 s时折半，转发节点的ECN标记阈值被设置为20个包，测试的结果如图3所示。我们将吞吐表现、转发节点缓冲长度和DCTCP的端系统窗口大小绘制在了图中。模拟器中的DCTCP在队列累积到阈值时，就会主动降低传输速度避免拥塞丢包缓存，这一点充分体现出了DCTCP的方案特点。与Reno只基于丢包的控制逻辑不同，DCTCP利用alpha状态量和ECN标记，在队列还未因完全填满导致丢包时就主动消去了部分CWND来降低发送速度。在图3中，DCTCP的窗口大小在一定区间内小幅波动，而不是类似于Reno的折半再增长。

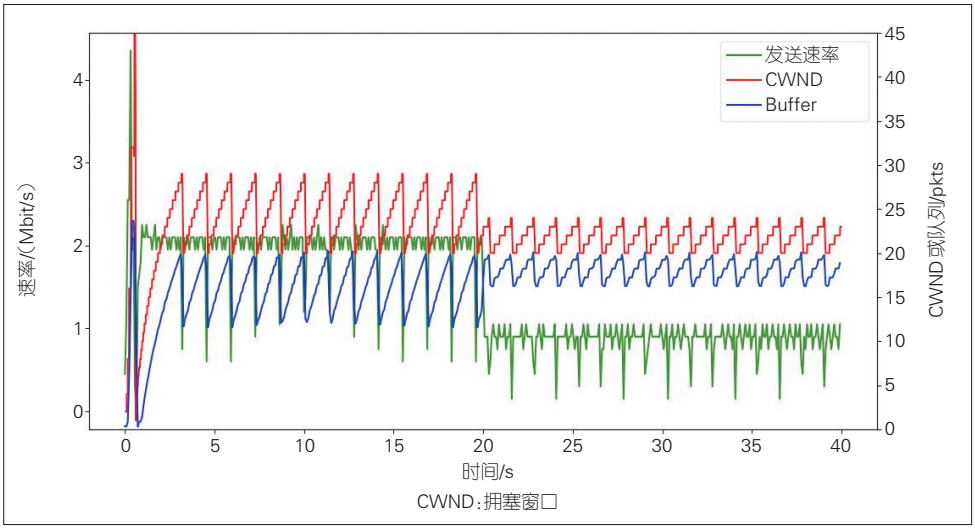
4.3 ABR模拟场景

为了验证模拟器在流媒体码率自适应研究中的功能，我们构造了基于Buffer的码率自适应算法（BBA–0）[17]模拟实验。BBA–0算法是一种基于Buffer的ABR算法，以当前缓冲量作为输入来决定目标比特率。BBA算法有低阈值和高阈值两个参数，当Buffer量低于低阈值时只采用最低清视频流快速填充，高于高阈值时只使用最高清视频流，介于两者之间的则按照比例线性选择中间的清晰度。总体上BBA–0会展示出优先免卡顿的特点，避免Buffer量下降。

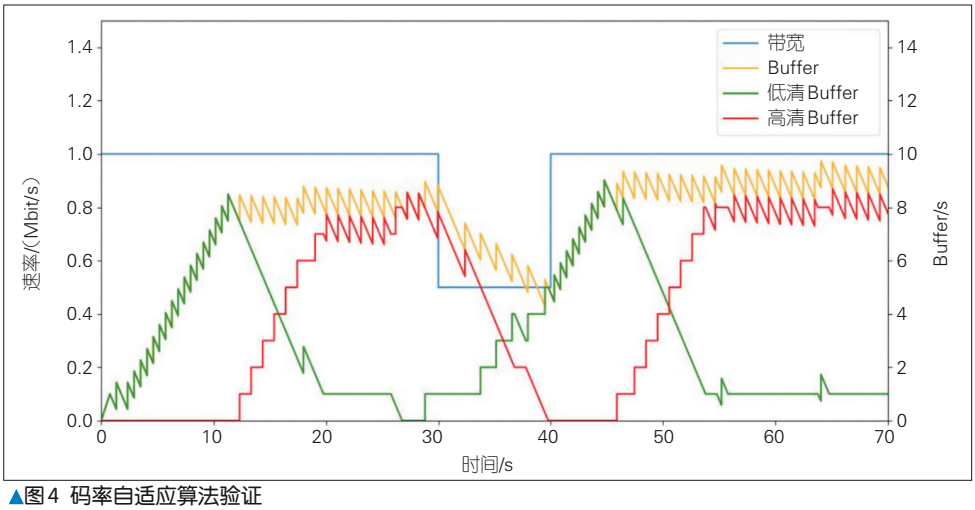
我们设置了简化的流媒体点播场景来对码率自适应功能进行测试。视频流量只有低清和高清两种，网络带宽在30～40 s折半。我们将总Buffer量、低清Buffer量和高清Buffer量的变化绘制如图4。启动阶段，我们优先选择低清流量建立Buffer，当Buffer量较为稳定后再切换至高清流量,以获取更好的用户体验。在带宽骤降阶段，Buffer被消耗，BBA–0切换回低清流量以重新填满Buffer。从模拟结果来看，模拟器中实现的BBA–0算法符合其功能特点，很好地展示了模拟器对码率自适应功能的模拟。

5 结束语

模拟器验证是传输优化研究过程的重要组成部分，但现有模拟器并不适应日益复杂的研究需求。本文分析了传输优化研究的几个重要细分领域对传输模拟器的需求，并总结了现有模拟器存在的问题。以此为基础，提出了一种轻量化传



▲图3 数据中心传输控制协议性能验证



▲图4 码率自适应算法验证

输模拟器系统。本传输模拟器在提供算法接口保证易用性的同时实现轻量化，并取得了较好的模拟效果，能够支持机器学习及非机器学习的各类传输优化方案的性能验证。

模拟器的功能拓展和性能优化都需要更多的研究。一方面，传输优化研究不局限于文中提到的拥塞控制算法和码率自适应算法。模拟器需要进一步设计并实现更多的算法接口以支持更多复杂的传输优化研究。另一方面，模拟器需要进一步优化运行效率，最大限度地发挥轻量化设计框架带来的效率优势。

参考文献

[1] FLOYD S, HENDERSON T, GURTOV A. The newReno modification to TCP's fast recovery algorithm [S]: Rfc3782. IETF, 2004

[2] HA S, RHEE I, XU L. Cubic: a new TCP-friendly high-speed TCP variant [J]. ACM SIGOPS operating systems review, 2008, 42(5): 64-74

[3] CARDWELL N, CHENG Y, GUNN C S, et al. Bbr [J]. Communications of the ACM, 2017, 60(2): 58-66. DOI: 10.1145/3009824

[4] DONG M, LI Q X, ZARCHY D, et al. PCC: re-architecting congestion control for consistent high performance [EB/OL]. (2014-09-24)[2021-12-10].https://arxiv.org/abs/1409.7092

[5] AKAMAI. Dash. js [EB/OL]. [2021-12-12]. https://github. com/Dash-Industry-Forum/dash.js/

[6] YIN X Q, JINDAL A, SEKAR V, et al. A control-theoretic approach for dynamic adaptive video streaming over HTTP [C]//Proceedings of the 2015 ACM Conference on Special Interest Group on Data Communication. ACM, 2015: 325-338. DOI: 10.1145/2785956.2787486

[7] MAO H Z, NETRAVALI R, ALIZADEH M. Neural adaptive video streaming with pensieve [C]//Proceedings of the Conference of the ACM Special Interest Group on Data Communication. ACM, 2017: 197-210. DOI: 10.1145/3098822.3098843

[8] WINSTEIN K, BALAKRISHNAN H. TCP ex machina [J]. ACM SIGCOMM computer communication review, 2013, 43(4): 123-134. DOI: 10.1145/2534169.2486020

[9] DONG M, MENG T, ZARCHY D, et al. PCC vivace: online-learning congestion control [EB/OL].[2021-12-10]. https://www.usenix.org/system/files/conference/nsdi18/nsdi18-dong.pdf

[10] ABBASLOO S, YEN C Y, CHAO H J. Classic meets modern: a pragmatic learning-based congestion control for the Internet [C]//Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication on the Applications, Technologies, Architectures, and Protocols for Computer Communication. ACM, 2020: 632-647. DOI: 10.1145/3387514.3405892

[11] HENDERSON R T, LACAGE M, RILEY F G. Network simulations with the ns-3 simulator [EB/OL]. [2021-12-10]. http://conferences. sigcomm. org/sigcomm/2008/papers/p527-hendersonA.pdf

[12] NETRAVALI R, SIVARAMAN A, DAS S, et al. Mahimahi: accurate record-and-replay for HTTP. [EB/OL]. [2021-12-10]. http://mahimahi. mit. edu/mahimahi_atc.pdf

[13] TANTAWY M M. Effect of transmission control protocol on limited buffer cognitive radio relay node [J]. Communications and network, 2015, 7(3): 20-27. DOI: 10.4236/cn.2015.73013

[14] LANGLEY A, RIDDOCH A, WILK A, et al. The QUIC transport protocol: design and Internet-scale deployment [C]//Proceedings of the ACM Special Interest Group on Data Communication. ACM, 2017: 183-196. DOI: 10.1145/3098822.3098842

[15] LI Y L, MIAO R, LIU H H, et al. HPCC: high precision congestion control [C]//Proceedings of the ACM Special Interest Group on Data Communication. ACM, 2019: 44-58. DOI: 10.1145/3341302.3342085

[16] ALIZADEH M, GREENBERG A, MALTZ D A, et al. Data center TCP (DCTCP) [C]//Proceedings of the ACM SIGCOMM 2010 Conference on SIGCOMM-SIGCOMM '10. ACM, 2010: 63-74. DOI: 10.1145/1851182.1851192

[17] HUANG T Y, JOHARI R, MCKEOWN N, et al. A buffer-based approach to rate adaptation: evidence from a large video streaming service [C]//Proceedings of the 2014 ACM Conference on SIGCOMM. ACM, 2014: 187-198. DOI: 10.1145/2619239.2626296

作 者 简 介

叶洪波，国网上海市电力公司副总工程师兼调度控制中心主任，高级工程师；主要从事电网调度运行管理、设备运维检修管理等工作；负责并参与了多项国网公司和上海公司科技项目，获得省部级科技奖5项，被评为国家电网公司劳动模范；发表论文10余篇。

潘俊臣，清华大学在读博士研究生；主要研究领域为恶意流量检测和可编程网络设备应用。

崔勇，清华大学计算机系教授、博士生导师、网络技术研究所所长，教育部青年长江学者，国家优秀青年科学基金获得者，教育部新世纪人才和中创软件人才奖获得者，中国通信标准化协会理事，国际互联网标准化组织IETF IPv6过渡工作组主席；获得国家技术发明奖二等奖1次、国家科学技术进步奖二等奖1次、省部级科技进步一等奖4次以及国家信息产业重大发明2次，所提出的IPv6过渡技术被国际互联网标准化组织IETF制定为9项RFC；发表论文100余篇，获国家发明专利40余项，出版学术著作4部。

ODICT融合的网络2030

ODICT Integrated Network 2030

王卫斌/WANG Weibin， 周建锋/ZHOU Jianfeng，
黄兵/HUANG Bing
（中兴通讯股份有限公司，中国 深圳518057）
(ZTE Corporation, Shenzhen 518057, China)

DOI：10.12142/ZTETJ.202201011
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20220224.1620.004.html
网络出版日期：2022-02-25
收稿日期：2021-12-23

摘要：围绕下一代网络场景、愿景、架构、技术发展等进行阐述，认为下一代网络是运营、数据、信息、通信技术（ODICT）融合的网络。该网络可支撑碎片化新业务，提供新的网络服务模式，是智能高效的自治网络。未来网络拥有算网一体化的能力平台，具备高带宽确定性能力，可应对增强现实（AR）/虚拟现实（VR）/扩展现实（XR）及元宇宙演进、网络安全内生挑战，以及高速移动场景的网络服务无缝迁移衔接能力要求。算网一体的新型网络能够满足未来业务对于算力资源和网络连接的极致需求。

关键词：未来网络；网络2030；算力网络；网络智能化；网络确定性；数字孪生；联邦学习；安全内生

Abstract: The scenario, vision, architecture, and technical development of the next generation network are highlighted. The next generation network is a network integrating operation, data, information and communication technology (ODICT), which supports new fragmented services, provides new network service modes, and is an intelligent and efficient autonomous network. The future network has the capability of integrated network calculation platform, and has high bandwidth deterministic capability to cope with augmented reality (AR)/virtual reality (VR)/extended reality (XR) and the evolution of the Metaverse, network security endogenous, and seamless migration and connection of network services in high-speed mobile scenarios. The new network integrating network calculations meets the extreme requirements of future services for computing power resources and network connections.

Keywords: future network; network 2030; computing power network; network intelligence; network determinacy; digital twin; federal learning; security endogenous

1 未来网络的发展方向

对于未来网络的发展，中国政府、运营商、设备商都提出了相应的要求或发展方向。《工业互联网创新发展行动计划（2021—2023年）》提出，需要加快工业设备网络化改造，推进企业内网升级，推动信息技术（IT）网络与运营技术（OT）网络的融合，建设工业互联网园区网络；中国移动提出“5G+AICDE”（5G与人工智能、物联网、云计算、大数据、边缘计算的融合）发展战略；中国电信构建2030云网一体的融合网络架构；中国联通发布《CUBE-Net 3.0》，确定“联接+计算+智能”的发展方向。

中兴通讯对未来网络提出的发展愿景是：运营、数据、信息、通信技术（ODICT）融合的网络2030，使能网络绿色、智能、安全、确定、可管可控，最终实现万物智联。

2 未来网络的场景和愿景

随着网络规模不断扩大，网络带宽速率持续提高，新应用层出不穷。当前，产业和社会正在进行数字化转型，人类经济形态由工业经济向信息经济、智慧经济转化，这使得社会交易成本极大降低，资源优化配置效率得到提高。

为满足未来5~10年生产、生活、社会管理的需求，网络基础设施也需要不断演进。网络演进的驱动力共有3个：新型业务形态、新型服务模式和新型智能网络，如图1所示。

（1）新型业务形态

未来5~10年，从消费互联网向产业互联网的演进最令人期待。未来网络将融合万网、万物和万业，将各种异构异质的垂直行业网络整合成统一的互联网，以支撑工业控制、智能电网、远程医疗、自动驾驶等产业化应用[1]。虚拟现实（VR）、增强现实（AR）、全息等新媒体应用也同样值得关注。元宇宙概念的提出，使得人们对未来虚拟社会和物理社会的无缝衔接充满期待。可穿戴技术、机器人技术、可植入技术、超硅计算与通信技术的快速发展与应用，为业务创新奠定坚实的技术基础。新型业务的快速发展，将创造出新的生活方式、数字经济[2]和社会结构。

（2）新型服务模式

未来网络将提供算网一体服务,将从目前的“管道型”服务向计算、网络、存储一体化的新型基础设施服务演进。未来网络不仅仅是“网络”，还是算网一体的智能化基础设施，将实现“算力无处不在，网络无处不达”的愿景。面对各种行业应用和AR/VR实时业务对算力就近服务的需求，算力资源将从中心云的集中模式逐渐向云、边、端的分布式模式转变。未来网络将把全网的算力资源、网络的精准传输能力更好地结合起来，并实现云、边、端三级算力的分配和协同。同时，未来网络不仅提供“裸资源”的服务，还将成为互联网公共能力的提供者，比如提供人工智能（AI）平台能力、大视频基础能力、内生安全能力等。未来网络提供的能力平台将带动各行各业的业务创新，促进整个社会数字服务的发展，如图2所示。

（3）新型智能网络

未来网络将对自身的网络架构、技术体系、运维模式进行智能化改造，以提高资源利用效率，降低成本和功耗。据统计，信息通信技术（ICT）产业的能耗占到全球总能耗的6%。未来这个比例还将不断增长。未来网络将通过网络架构的优化（比如算网一体等）、资源利用率的提高、新型技术（比如光电集成）的应用，大幅降低流量/能耗比。未来网络还将通过数字孪生等智能技术，促进整个网络的自动化运行，降低运维成本和出错率。

3 网络架构创新

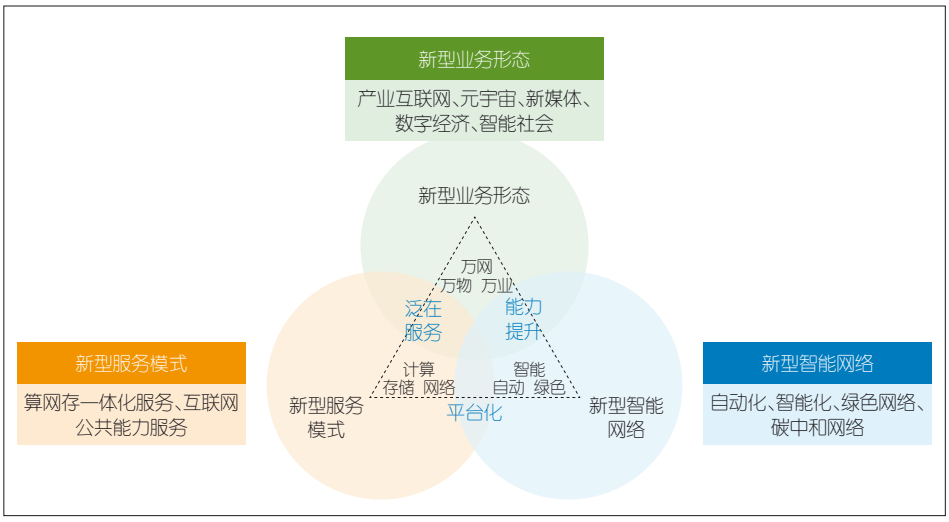
未来网络要达成以上愿景，在网络架构和技术方面需要在以下3个方面进行创新。

3.1 ODICT技术和架构的融合

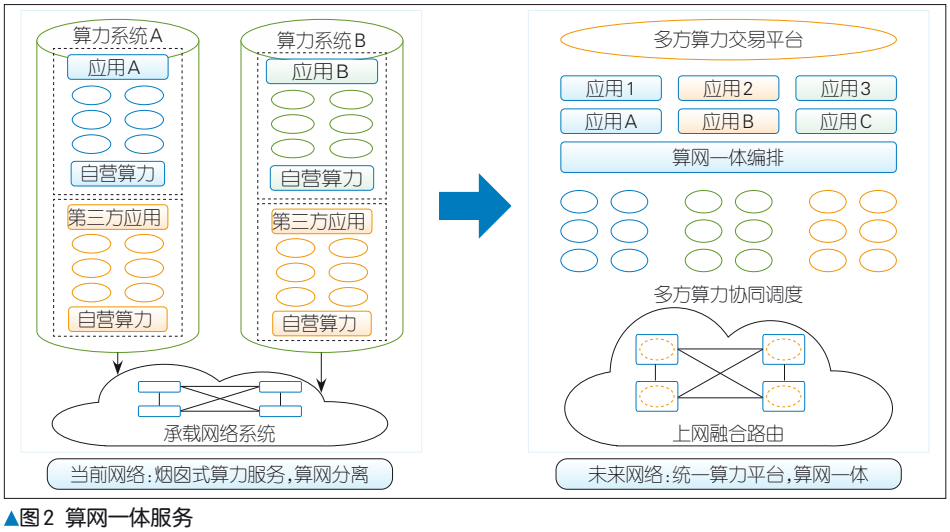
网络的发展历程是业务需求驱动多领域技术不断融合发展壮大的过程。在通信技术（CT）的基础上引入信息技术（IT），能够让网络组网变得更灵活，使上层应用接入网络变得更方便。ICT技术的融合在推动网络发展的同时也推动了IT技术的发展。网络与OT技术的融合，将加快工业设备的网络化改造，深化“5G+工业互联网”[3]，推动企业内网升级和外网建设。

网络和业务的发展相生相随，相互促进。随着网络的发展，行业应用提出了新愿景。高清云游戏、工业视觉、元宇宙等需要网络在满足高带宽的同时也要满足低时延、网络确定性以及边缘高算力等需求。

ICT技术的融合，可使能网络基础更强健，在具备工业领域的低时延、低抖动、高可靠的确定性的同时，也具备满足元宇宙、扩展现实（XR）、工业视觉等领域的上行高带宽网络基础的需求。此时引入智能数据（DT）技术，网络将从能用走向好用，在用户体验优化、高效运维、安全保障等方面发挥巨大作用。随着OT、DT、IT、CT多领域技术的不断融合、相互促进，未来网络架构和技术将推动网络及多



▲图1 未来网络愿景



▲图2 算网一体服务

领域技术共同演进。

3.2 业务和网络的协同

在网络发展的历史上，业务和网络的关系曾经从“耦合”向“去耦合”的方向发展[4]。

传统电信网的网络与业务层紧密耦合，业务功能由网络设备提供。网络提供单一的业务，比如公共交换电话网（PSTN）、数据网、电视网等。这种网络的优点是服务质量好，用户体验好；但缺点是业务体系封闭，不利于技术创新。

随着多媒体业务的发展，业务种类变得越来越丰富，业务和网络耦合的模式已不能满足业务发展的需要。电信行业逐渐将业务功能从网络中解耦出来，形成独立的业务网元，比如智能网、短信、IP多媒体系统（IMS）等。基于传输控制协议（TCP）/互联网协议（IP）的互联网把业务和网络的解耦发挥到了极致。互联网的两大设计原则是端到端原则和分层解耦原则[5]。端到端原则把互联网的复杂性放在两端，使网络层尽可能保持简单；分层解耦原则尽量避免互联网层间的内部交互。这种架构设计使得业务可以脱离网络而独立发展，降低了互联网业务的创新门槛，增加了业务部署的便利。

然而，目前互联网架构把业务和网络过分去耦合，使得两者处于互相割裂的状态。端到端原则隔离了两端和网络，使得终端和云端无法感知网络的状况；分层解耦原则隔离了应用层和网络层，使得上层应用无法向网络传递个性化的需求信息，最终绝大多数业务只能按照“尽力而为”的模式运行。随着互联网业务的纵深演进，尤其是产业互联网的发展，业务和网络的割裂状态越来越不能满足业务的需求。例如，对传输质量有要求的业务希望网络能够提供确定性的传输能力，即带宽、丢包率、时延都是可以预期的，而不仅仅是尽力而为的；对安全性有高要求的垂直行业则希望网络不仅仅提供传输功能，还要提供“有安全保障”的传输，即保持信息传送的完整、可靠、不被非授权访问；此外，还有的业务希望感知网络的状态，如链路利用率、丢包率、缓存队列等，以便调整自身的传输窗口，保持最优的传输效率。

因此，业务和网络的完全耦合或者完全去耦合都不能满足未来的业务需求。在未来网络的架构中，业务和网络必须以某种方式“再耦合”。这既能保持业务的独立性，又使得网络能够感知业务的关键需求，以便于精准匹配相应的服务等级协议（SLA）策略。如何在未来网络的架构和协议方面建立业务和网络之间的桥梁，是未来网络面临的一大挑战。

3.3 服务化平台

网络运营商是商用互联网的主要建设者和运营者。运营商投入巨资拓展网络覆盖范围，提升网络连接速度，极大地促进了互联网的发展。如何在网络发展成功的同时，在业务方面也取得成功，是下一代网络需要考虑的。

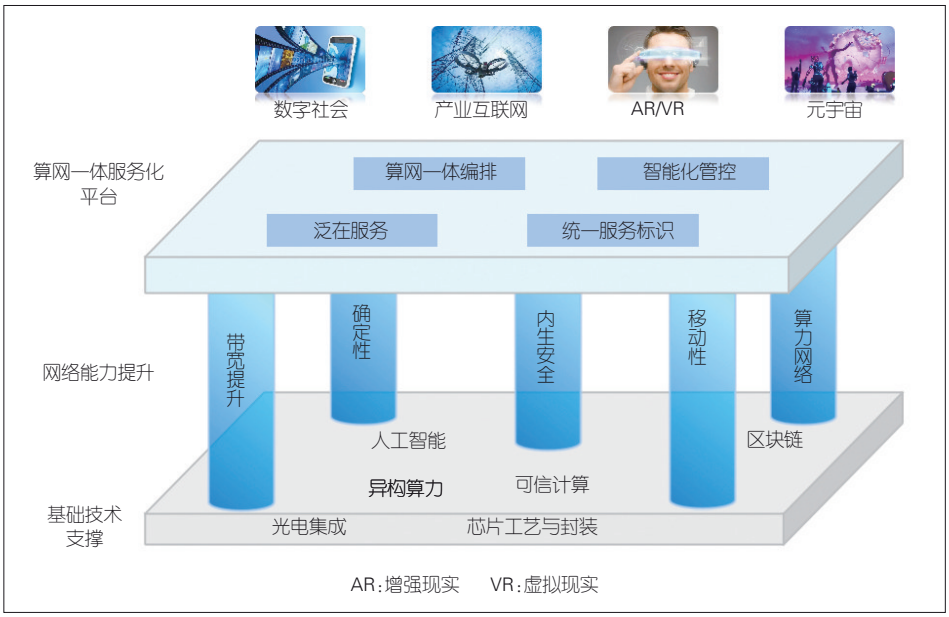
电信行业一直在发展“综合业务数字网”技术，以试图实现网络和业务的综合运营：从20世纪80年代末的综合业务数字网（ISDN）技术，到90年代的异步传输模式（ATM）技术和电信级IP综合承载网技术，再到2000年之后的IP多媒体子系统（IMS）技术。实践表明，电信行业从技术到标准再到应用的发展模式，无法在互联网业务的竞争中取得优势。互联网业务更注重商业模式灵活性、业务创新能力、迭代速度、资本运作等方面，而互联网服务商在这些方面更具优势，比如美国的Facebook、Amazon、Google，中国的BAT（百度、阿里巴巴、腾讯）等。对网络运营商来说，与其在业务创新方面下功夫，不如将自身定位为基础设施和平台的提供者，即从网络运营商扩展为基础设施和平台运营商。

网络运营商曾经对平台运营模式做过尝试。比如，在2010年前后运营商提出“智能管道”的理念，试图把网络功能开放出来供业务调用，但只有短信等少数功能的开放取得成功，而最为重要的网络服务质量的能力未能实现开放。例如，服务质量实现了像DiffServ、IntServ、多协议标签交换流量工程（MPLS-TE）等技术标准的制定，同时新的分段路由流量工程（SR-TE）、SR-Policy等技术标准也在制订之中，但这些技术只在运营商自营业务中得到部署，并没有得到更广泛的应用，尤其是没有和互联网服务商的业务相结合起来。

面对“新型服务模式”的愿景，未来网络应当成为一个服务化平台，不仅能提供网络连接服务，还能提供算、网、存一体化的基础设施服务，甚至通过进一步扩展提供共性能力的服务（比如安全能力、AI能力、大数据等）。

未来网络提供的服务化平台，不同于目前云服务商提供的私有化的“烟囱式”平台。打破“平台垄断”是促进行业竞争、经济健康发展的需要。在这一点上，电信行业有自身的优势，不仅有成熟的标准组织和体系，还有互联互通的文化和传统。因此，未来网络的服务化平台是统一定义的、互联互通的平台。

网络架构创新从ODICT技术和架构融合、业务和网络协同、服务化平台3个方面进行，包含算网一体服务化平台、网络能力提升、基础支撑技术等内容，如图3所示。



▲图3 未来网络架构创新

4 算网一体化服务平台

算网一体化服务平台是未来网络的大脑，是实现算网资源整合、算网能力开放的关键。该平台有如下几条核心要素：

（1）算网一体。未来网络具备感知和调度算、网、存一体化资源的能力。算网一体操作系统将作为资源调度的大脑。

（2）服务化。未来网络解决业务和网络的双向感知问题，通过扩展和整合各种网络能力，以“网络内生”方式满足未来业务的共性需求（如确定性、安全性等）。未来网络将成为新型数字服务平台，为各行各业提供互联网的公共能力。

（3）统一平台。未来网络将打破目前云服务商的“烟囱式”模式，实现服务的统一定义和互联互通，向用户和应用提供泛在的、归属无关的服务。

4.1 泛在服务网络架构

泛在服务即服务无处不在。服务需求方无须事先知道服务提供方的身份和所处位置。网络根据服务需求方的需求选择最优服务提供方，以实现高效的服务发现和服务分发。泛在服务有3个具体含义：

（1）依托于算网一体的部署，实现服务无处不在。针对未来业务对于低时延、低抖动的需求，泛在服务从集中式的云计算向分布式的边缘计算、网内计算、端计算发展。

（2）打破服务提供商的界限，实现归属无关。目前的云服务是烟囱式的，各个云服务商和业务提供商提供的服务都是私有的，不能互通；而未来网络提供的服务是统一定义、互联互通的服务。

（3）全网一体化的无边界微服务架构。目前云内和云外的微服务架构是不同的。比如，云外采用应用程序接口（API）网关模式，对来自云外的服务请求通过7层代理的方式来提供服务；云内则采用Service Mesh架构，并将Sidecar作为分布式容器级别的服务治理和通信代理[6]。未来网络将采用统一的分布式微服务架构，使得资源和算力能够更加迅速和便捷地随需弹性部署和动态调度。

围绕泛在服务新功能范式，我们提出未来网络泛在服务感知的新架构，如图4所示。

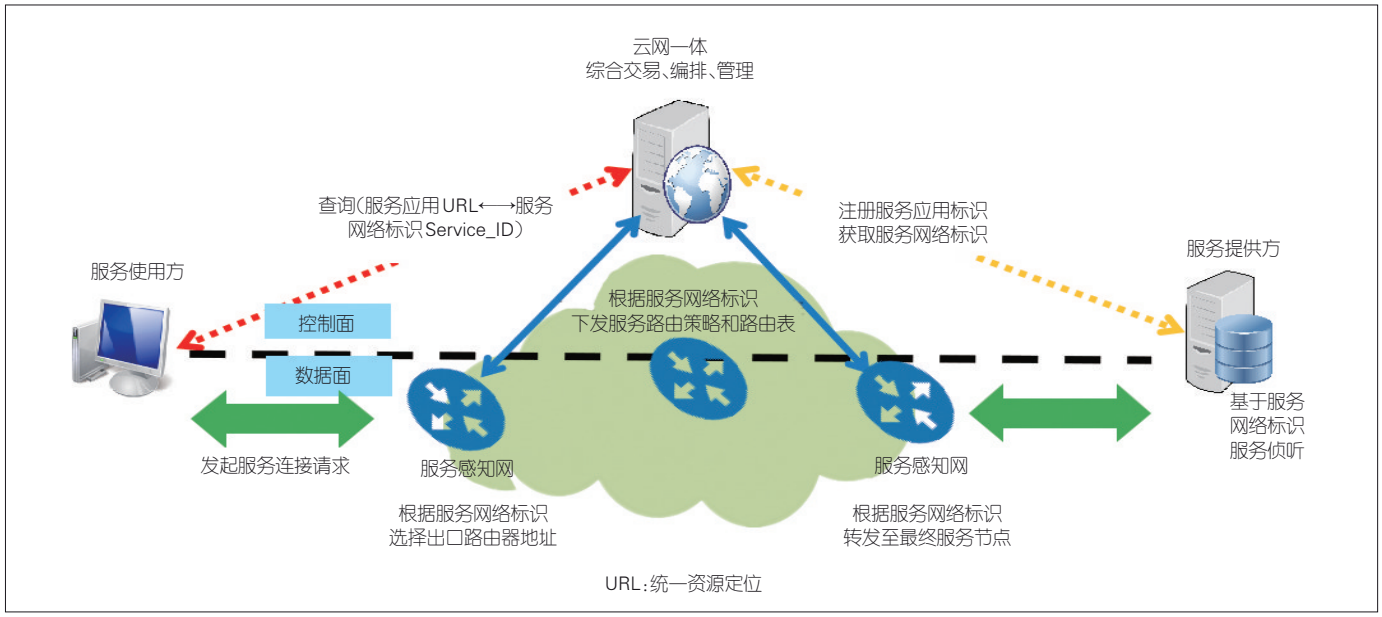
泛在服务网络架构的两个核心要点是归属无关机制和统一服务命名体系。

（1）归属无关机制保证用户可以随时、随地、随愿地请求和获取服务。在该这种制下，应用无须关心服务的提供者和位置，只须聚焦应用自身的需求和逻辑。归属无关机制负责建立服务连接，管理和维护连接状态，并提供面向服务的拥塞控制、移动性、保序、多路径/多归属、内生安全等增强传输功能。

（2）引入统一的服务命名体系有助于打通应用和网络的统一分配和调度，提供最优化的网络调度和服务质量。服务标识既可以作为开放服务网络接口，提供归属无关的泛在服务连接管理，实现对算力、存储、内容、能力等可虚拟资源的连接，又可以保证网络感知服务，并对服务进行统一注册、管理和索引。

4.2 智能化管控与算网一体编排、调度

在控制面上，新架构中的云网一体综合交易、编排、管理控制器负责服务的注册、发布和查询，并动态智能关联服务标识与网络层地址的映射。服务标识须由网络分配且全网唯一，并且在服务连接过程中保持不变，以确保移动连续性。服务路由策略和路由表通过控制通道下发到网络层设备。



▲图4 泛在服务网络新架构

在数据面上，服务使用方携带服务标识发起服务连接请求，服务提供方则基于服务标识进行服务的请求侦听。网络边缘节点根据服务标识选择最优服务目的节点和对网络资源的编排，并执行对应的SLA策略。

泛在服务网络架构实现了应用与服务的解耦，从而让应用聚焦于自身的应用逻辑创新，将共性的归属无关服务交由网络内生提供。由于应用与网络实现了机制上的解耦，从而去除了域名系统（DNS）的低效查找过程，使服务寻址更加高效。例如，在移动边缘计算（MEC）场景中，该机制至少可以提升100%的服务寻址效率。

泛在服务感知IP技术基于IP第6版（IPv6）底座，沿用了良好的IP生态链，使应用和网络均可以实现平滑演进，有利于数字经济的快速部署，很好地保护了前期的资源建设投资。通过网络和服务的一体化供给，充分形成网络即服务的转化能力，将网络的价值从管道上升为赋能平台，将加速数字经济的部署，实现IP网络的能力倍增。

5 网络能力提升

未来网络在引入新架构、实现智能化服务化平台的同时，网络自身能力需要进一步提升，以满足未来业务发展需求。

5.1 网络带宽能力

网络带宽提升是永恒的话题。随着元宇宙的提出，系统对于网络带宽的需求又将提升数个量级。然而，随着无线、有线频谱利用率都已逼近香农极限，未来的带宽提升需要更多的技术创新。

依据第3代合作伙伴计划（3GPP）R17有关5G新服务需求的研究结果[7]，结合高清、高自由度、人眼极限视频带宽与可靠性要求，6G网络预计将支持1 Tbit/s的峰值数据率、20 Gbit/s的用户体验数据率、10 Gbit·s^{-1}·m^{-2}的区域业务容量密度、100 Gbit·s^{-1}·m^{-2}的空间容量密度。

6G无线网络提升带宽和带宽利用率的技术包括：扩展频谱（高频、超高频）、自治自动网络、智能三维连接、智能大规模天线阵、按需网络拓扑、按需网络计算、超硅计算与通信等。

有线接入目前已经实现10G-PON规模商用，后续将向单波50G-PON和100G-PON发展。预计在2023年有线接入将实现50G-PON园区级的应用，在2025年左右实现50G-PON家庭级的应用。

骨干网光传输已经开始进行单波400 Gbit/s的部署，预计2023年800 Gbit/s将在城域网商用。但随着频谱效率接近香农极限，带宽提升的困难逐渐加大。尽管如此，我们有以下几条路径可以尝试：一是继续提升光电器件的波特率，并在2026年演进到170 Baud单波1.6 Tbit/s的光系统；二是扩展光纤传输的波段，从C+L波段向C++、L++扩展，并进一步向U波段、S波段扩展；三是采用空分复用技术，比如少模多芯技术。第3种技术路线可大幅度扩展带宽，但该技术仍不成熟，要实现商用还需要至少5年的时间。

5.2 网络确定性能力

在未来网络发展中，无线移动网络依然是网络发展的重点。随着移动网络技术的发展，移动网络不再仅仅提供高速的上网业务服务，还为各行各业提供网络通信服务，实现一网万用、按需服务。在垂直行业中，传统的尽力而为机制[8]已经不能满足相应需求。例如，工业控制等某些特定领域需要网络支持有界的时延和抖动、极低的丢包率和超高可靠的保障。

早期的工业网络大都采用专线（如现场总线）方式来保障特定业务流的传输，但是随着全球新一轮科技革命和产业变革的加速发展，工业互联网成为未来工业制造智能化和信息化的关键技术。在工业互联网中，IT网络与OT网络相互融合，在同一个网络中，能够同时满足互联网与信息化数据所需的大带宽以及工业控制数据的实时性与确定性要求。

在当前尽力而为的网络中，不同业务的数据在转发时，遵循先进先出（FIFO）、优先级抢占等服务质量调度机制。然而，该机制无法避免网络冲突，难以提供稳定可靠的网络传输，一旦发生报文冲突就需要等待或者重传，可能会导致较长的转发时延和不可控的抖动。这在高精度的工业控制中是不允许的，因为这可能会导致生产系统出错，甚至崩溃。为了让移动网络可以为对时延极其敏感的行业提供网络服务，我们需要引入严苛、精准的确定性保障能力。

确定性网络是指，在一个网络域内为承载的业务提供确定性业务保证的能力，包括有界的时延、抖动和丢包率等指标。在网络中为关键的业务流协调各个转发节点的调度转发资源，有助于保障该业务流在网络中的畅行无阻，可以实现超低时延和消除抖动的转发能力。此外，通过对业务流复制和多链路冗余传输还可满足超低丢包率的高可靠传输需求。

确定性网络（DetNet）技术的标准主要包括：国际电信联盟电信标准分局（ITU-T）5G回传技术标准城域传送网（MTN）、电气与电子工程师协会（IEEE）802.1 TSN、国际互联网工程任务组（IETF）DetNet以及3GPP时间敏感通信（TSC）技术。

MTN技术属于L1层的确定性技术，在以太网物理编码子层（PCS）实现时分复用（TDM），层次化时隙划分以及固定时隙位置交叉，能够满足超低时延和抖动的业务需求。MTN是对FlexE技术的扩展，能够把时隙颗粒的速率降低到10 Mbit/s，实现更加灵活的业务带宽分配。目前MTN已经进入商用部署阶段。

TSN技术是IEEE制定的基于L2 Ethernet的DetNet标准技术。目前已经有802.1AS、802.1Qbv、802.1CB、802.1Qcc等10多个比较成熟的802.1 TSN相关标准规范。此外，业界也推出了多种TSN交换机和支持TSN的芯片与工业终端等，正在逐渐开始商用。

2015年IETF成立DetNet工作组并制订相关标准，当前已发布Use Case、IP/MPLS等多种数据面架构、流模型等10多个征求意见稿（RFC）规范。与TSN仅支持L2 Ethernet网络不同，DetNet将相关技术扩展到的L3网络，实现了在IP/MPLS的确定性传输和与TSN网络叠加互通等，为实现广域的确定性传输提供技术基础。

TSC是由3GPP在2020年7月发布的R16标准中开始引入的。在R16标准中，整个5G系统被作为一个TSN逻辑网桥，以实现与TSN网络的互联互通。在目前正在制订的R17标准中，5G系统引入内生确定性，无须对接外部TSN网络，就能实现用户设备（UE）之间的确定性传输。预计在将来5G-A/6G的标准中，3GPP还将实现与DetNet的互联互通。

未来网络面临的一大难题是，如何协同利用以上多个层次的确定性技术以满足多样性的确定性需求，同时能够合理规划网络路径以高效利用各种网络资源。目前网络5.0联盟已经在研究“内生确定性路由”技术（包括资源确定性、路由确定性、时间确定性3个要素），通过高效利用跨层确定性资源能力，精准满足分类分级的确定性业务需求。

5.3 网络安全能力

当前通信网络通过补丁式、被动式、外挂式等措施来进行安全防护。网络安全风险等级只能通过静态方式来评估。通过网络价值、安全漏洞、安全事件的发生频率等因素可以粗略地对网络的风险状态进行评估，但不能对网络正在遭受的攻击进行实时的检测与防护。

网络安全防护部署的维护成本较高，难以实现动态策略调整和自动化维护，已经无法满足当前复杂的电信网络业务需求和应用场景。未来网络在安全方面须具备以下4个特征。

（1）全面自动化：基于自动化安全引擎，为网络基础设施、软件等提供自动化部署、自动化检测、自动化修复等主动防御能力，包括设备节点、基础设施、网络服务、数据、用户、管理节点、操作系统、中间件、数据库、软件服务等；增强各资产系统内部的安全防范能力，实现完整可信的防护、服务、访问、数据；动态度量系统状态，检测系统安全，构建网元级内生安全，从而提高网络级内生安全能力。

通过安全引擎实现统一的安全能力编排。对云和网进行调度和编排有助于实现安全资源自动分配、安全业务自动化发放、安全策略自动适应网络业务变化（网络安全协同）、网络高级威胁实时响应防护（安全分析联动）等能力。多网

元、多层次协同保障网络安全，可实现安全策略集中管理和编排，为用户按需提供安全服务。

（2）安全防御：根据行业防护的安全需求，提升网络及网络服务功能安全能力，并实现安全能力的弹性部署，降低安全风险，提升韧性；引入区块链技术帮助网络构建安全可信的通信环境，以实现系统的防篡改能力和恢复能力；通过可信计算技术可以实现网元的可信启动、可信度量和远程可信管理，使得网络中的硬件、软件功能运行持续符合预期，为网络基础设施提供主动防御能力；引入零信任技术，对网络进行精细化可视化管理；部署相应的安全组件，构建端到端的网络云安全体系。

（3）安全自适应：能够利用AI、联邦学习技术实现网络预测与修复；安全服务能够随时监测并感知网络云的安全动态，使资产安全风险可见，并第一时间快速自动预测、告警，对安全事件及时发现和修复或进行平衡处置，以保障网络服务的可用性；能进行网络服务升级和安全系统换代升级；在业务系统流程再改造时，安全能力能够动态提升。

当网络局部被入侵时，安全引擎会采用阻断威胁流量和启动安全加固流程的方式快速规避或消除威胁。同时，安全服务能共享威胁情报，有助于做到全网“免疫”同类威胁。

（4）安全自演进：通过端、边、网、云的智能协同，准确感知整个网络的安全态势，敏捷处置安全风险；形成自适应安全模型，构建细粒度、多角度、可持续动态安全防御体系；网络各层须嵌入AI能力、联邦学习（分布式机器学习）能力以实现网络自适应、自感知、自运维；通过快速学习和训练，AI、联邦学习技术可以更加准确地对网络流量与异常行为进行检测、回溯和根因分析。

电信网络建立端、边、网、云智能主体间的泛在交互和协同机制，有助于系统准确感知网络安全态势并预测潜在风险。通过智能共识决策机制完成自主优化演进，可实现主动纵深安全防御和安全风险自动处置，提供实用化的安全分析与告警，抵御各类高级持续性威胁（APT）攻击。

5.4 网络业务连续性能力

移动性管理是移动通信的重要研究内容之一。在未来网络中，多种网络接入方式并存。网络为用户提供多种制式接入一致的业务连续性服务。对于现有移动网络，无论是网络本身还是应用场景，在支持移动业务连续性方面都将发生巨大变化，这对网络的移动性管理技术提出了新的挑战。

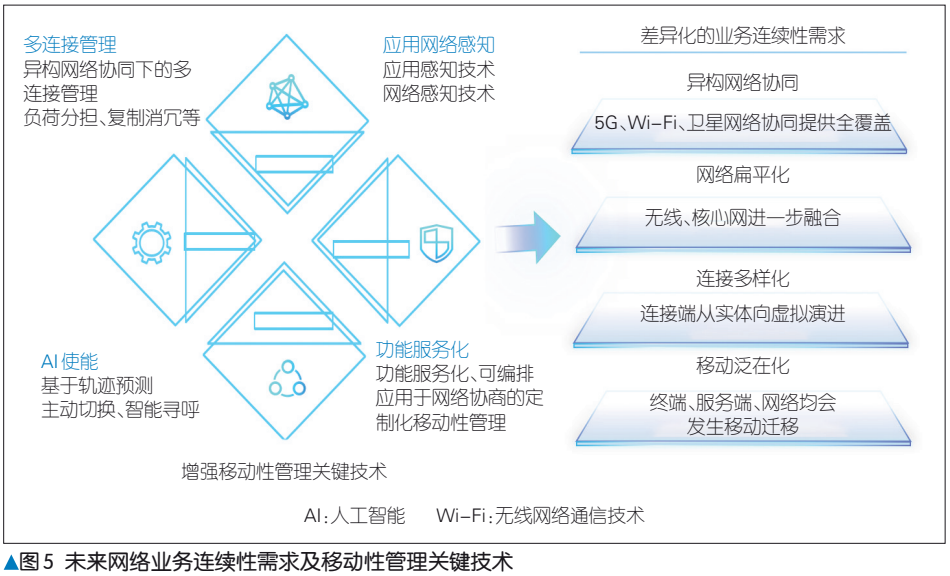
从移动和切换场景看，移动主体更加泛在化，水平切换更加频繁，垂直切换将常态化。随着云网融合、算网一体的发展，网络连接从有形的实体连接向无形的内容、服务、算力等虚拟连接发展。随之而来的是移动场景和主体更加泛在化，包括终端、服务端、网络等的泛在移动。网络的进一步扁平化，无线侧高密集组网，毫米波、太赫兹的应用，均使得水平切换更加频繁。异构网络协同提供全场景覆盖，IPv6以及多宿主终端的普及，均使得跨网络类型的垂直切换常态化。

从业务连续性需求看，不同应用对业务连续性的要求存在巨大差异。面向C端的浏览类、视频类业务对移动切换引起的连接中断并不敏感；而车联网、无人机、工业互联网等B端场景要求提供无缝的切换管理及确定性的网络连接服务。

因此，在未来网络泛在连接和移动场景下，移动性管理技术需要至少解决以下几个问题：

- 为不同场景提供差异化的业务连续性服务；
- 提供零中断、零丢包的网络连接，满足无缝切换需求；
- 保障切换前后网络性能一致性，实现确定性网络服务的快速切换。

针对差异化的业务连续性需求，移动性管理也需要同步演进。这包括异构网络协同的多连接管理、应用网络双向感知、AI使能以及服务化的架构设计，如图5所示。



▲图5 未来网络业务连续性需求及移动性管理关键技术

（1）多链接管理。集中式移动性管理数据面锚点固定，但会增加传输时延，使网络性能降低。分布式移动性数据面锚点是移动的，却难以保障切换过程连接不中断。

针对超高可靠通信需求，3GPP制定了基于双连接的端到端冗余用户面传输方案。该方案利用两条冗余的协议数据单元（PDU）来传输数据，使可靠性高于99.9999%。

将双连接方案与移动数据面锚点结合，可有效解决在移动锚点下切换过程中连接中断和切换前后网络性能不一致的问题。此时，数据面采用锚点移动分布式部署，移动终端则采用双连接方式与网络保持通信连接。在终端移动过程中，切换策略控制机制使得同一时刻仅会发生一个连接的切换，并使另一个连接仍可用。双连接切换机制保障网络始终处于可用状态，避免了连接中断。选择接近移动终端位置的数据面锚点，并结合显示路径设计等技术，可保障切换前后网络性能一致。

（2）应用网络双向感知。未来网络是云、网、边、端、业相互协同的网络。网络通过感知应用需求，选择与之匹配的移动性管理策略，进行针对性的目标网络和路径选择，切换触发并执行策略。反之，应用也可以通过实时感知网络的时延、带宽、拥塞等信息，选择最佳的切换时机，调整数据收发策略等，来获得更好的应用体验。

深度报文分析（DPI）技术已经广泛应用于现网，通过对数据包的深度检测，可实现应用及内容信息的感知。基于IPv6的应用感知网络（APN6）在IPv6扩展报文头中携带应用及应用对网络的需求信息。网络层据此进行应用颗粒度的网络资源和服务响应。应用可以通过网络能力开放或者简单双向主动测量协议（STAMP）、双向主动测量协议（TWAMP）、操作维护管理（OAM）等技术，完成端到端的网络质量检测来获得网络信息。

（3）AI使能。AI作为基本要素将与网络深度融合，使网络、业务、运维、运营实现全方位的智能化，有助于提高网络效能，降低运维成本。

AI在移动性管理中同样大有可为。3GPP R17开展了基于AI的移动性优化研究，对终端的位置、移动轨迹进行预测管理，并对接入管理功能（AMF）终端寻呼过程进行优化。

典型的AI包括数据采集、模型训练、模型推理以及决策执行4个部分。智能网络对移动终端、网络、应用信息的采集、分析和预测，有助于实现主动切换，以及最优传输路径和定制化移动性管理流程的选择。

（4）功能服务化。功能服务化主要体现在：对外能力开放，即可提供异构网络双连接、数据复制消冗传输、应用网络感知、基于AI的主动切换、接入网关间的隧道缓存转发、身份位置分离等独立功能接口；功能可编排，即针对不同场景，通过灵活编排形成功能链，并提供功能链标识供上层选择调用；平滑支持新功能的引入，即在保持整体框架不变下，功能组件随着技术和场景的演进更新换代。

5.5 网络算力一体化能力

网络能力和算力融合的网络是新型ICT融合服务与算力网络。网络基础设施具备对算力资源的感知、调度和编排，同时网络层提供网络、计算、存储。未来网络需要更迅捷高效地响应业务的请求。算力资源从集中部署模式向异构多样、分布式的部署模式演进[9-10]。网络基础设施通过其成熟发达的连接感知触角，将多级分布的算力资源进行统一的动态纳管、调度和编排，发挥全网资源的虚拟算力池化优势，在提升服务质量和资源利用率的同时，助力网络运营商为全新的业务打造网络能力和算网融合的商业模式[11]。

算力网络的核心优势是实现算力和网络的协同调度，满足未来业务对算力资源和网络连接的需求。比如，高分辨率的VR云游戏，既需要专用图形处理器（GPU）计算资源完成渲染，又需要近似确定性的网络连接来满足10 ms以内的端到端时延要求。算力网络就可以完成这种一站式的资源调度。

算力网络包括3个技术要素：兼容多种算力资源的算力度量和感知、基于现有IP路由协议的算力路由、增强基于IPv6的段路由（SRv6）的算网一体转发面。

算力资源是分层管理的。基础设施即服务（IaaS）、平台即服务（PaaS）及云原生资源与服务[12]，都属于广义的算力，均可纳入算网平台以便统一编排和调度。推进阶段上，可先聚焦IaaS基础算力。随着度量和标准进程逐步纳入PaaS、云原生基础服务等，层次化的算力资源体系逐渐形成。

算力路由有基于SDN的集中式和基于边界网关协议（BGP）的分布式两种模式。为了减少路由表规模，避免路由震荡，系统需要对算力状态按照高低频种类进行分离维护，执行两级路由机制。

转发面要基于现有IPv6平滑演进，以避免网络设备大量改动。因此，增强SRv6的业务功能编程，纳入通用算力服务功能，可实现算网业务无缝拉通。

6 关键技术

在未来网络的发展中，无论是架构创新，还是智能化服务平台的搭建，抑或是网络自身能力的增强，都需要关键技

术支撑。

6.1 带宽增强技术

（1）光电集成。该技术可提升网络传输效率，解决网络功耗和成本上升、传输距离下降的问题。

传统网络设备的内部连接，包括单板和单板之间、芯片和芯片之间的连接，都采用电串行总线连接（Serdes）方式。随着网络带宽的提升，Serdes连接的速率急剧上升，电连接在高速信号下的损耗较大，导致网络整体功耗与成本上升、传输距离下降。

解决这个问题的主要方法是采用共封装光学（CPO）技术，即把光收发器件高度集成到一个光引擎上，并将光引擎与电芯片封装在一起，形成芯片直接出光的形态，用光连接代替电连接。该技术可以大大降低功耗、延长传输距离。

预计到2023年CPO技术将在数据中心得到应用，并在2025年开始规模普及。2027左右路由器也将采用CPO技术。CPO光电集成技术对网络设备的形态也将产生影响。传统的具有背板的框式设备架构将被盒式设备互联的无背板架构取代。

CPO的关键技术包括光调制解调芯片、电驱动和放大芯片、光纤耦合技术、封装技术等。此外，CPO的大规模生产工艺也是一个很大挑战。

（2）异构算力。异构算力是提升网络传输能力的关键技术。在未来网络中，通用算力无法应对AI、视频渲染、高性能并发计算、高性能存储数据的实时处理，将造成网络拥塞，使网络带宽无法得到有效利用。现场可编程门阵列（FPGA）、图形处理单元（GPU）、专用集成电路（ASIC）等异构算力能有效满足计算节点高性能算力需求及数据高性能搬迁需求，在统一的智能编排管理下与通用算力、高带宽网络一起完成AR/VR/XR、云游戏和高性能数据的搬迁、并发计算等工作。

6.2 网络智能技术

（1）机器学习。机器学习是网络智能化的基础。传统运维方式依赖人工、静态规则，无法适应动态变化场景。机器学习技术则可以在动态变化的复杂条件下进行高效准确的决策判断。引入机器学习技术可以实现基于“专家经验”到“机器学习”的转变。在未来网络中，数据较为分散且可能有隐私要求。联邦学习能帮助多方在满足用户隐私保护的要求下，进行数据使用和机器学习建模。联邦学习的应用在很多方面还面临挑战，比如通信效率、数据非独立同分布、安全性、健壮性等。

（2）意图驱动。意图驱动是实现“零接触”运维的基础。意图规定了期望，包括对特定服务或网络管理工作流的要求、目标和约束。意图网络瞄准的是用户意图或商业目标，强调网络运维和架构人员（整个网络用户）的意图，相关应用场景包括网络规划和设计。例如，意图驱动的容量规划、覆盖优化、站址规划等，网络和业务部署，意图驱动的业务部署，意图驱动的网络和业务维护、优化及保障。然而，意图网络存在声明式的API构建、意图的分解和翻译、组件和设备的相互兼容问题。

（3）数字孪生。数字孪生提供了更好的仿真验证能力。数字孪生网络构建物理网络的实时镜像，可增强物理网络所缺少的系统性仿真、优化、验证和控制能力。将数字孪生技术应用于网络，有助于创建物理网络设施的虚拟镜像，搭建数字孪生网络平台。物理网络和孪生网络的实时交互，可以实现低成本试错、智能化决策和高效率创新。数字孪生网络系统的构建面临兼容性不佳、建模难度大、实时性挑战高等问题。

6.3 网络安全技术

（1）区块链。利用区块链数据不可篡改的特点，对交互信息、执行信息进行监督，可使媒体流的转移和流通过程更加公开透明、真实可信，进而满足IT之间、CT之间、IT与CT之间的安全、追溯、结算的互信要求。用户与网络的唯一特征可被用来建立互信关系。只有具备相互信任关系的用户和网络才被允许进行网络互通和流量交互。

（2）可信计算。可信计算可以实现网络设备的可信启动、可信度量和远程可信管理，使网络中的硬件、软件运行持续符合预期，为网络基础设施提供主动防御能力。在网络中，通用的设备指纹制作可实现初始启动、基本输入输出系统（BIOS）/统一可扩展固件接口（UEFI）加载和OS启动等过程的完整性保护，使任何恶意修改都可以被监测出来，并对该过程进行阻止。可信计算对各资产组件主动进行动态度量和静态度量，并依据度量结果进行主动裁决和控制；对相关安全活动进行可视化展现，并用于安全审计。基于可信根对隐私数据进行完整性和保密性的可信验证，可提供端到端数据可信认证。通过可信计算建立一种主动免疫可信计算的新模式，使得设备在提供服务的同时进行安全自防护。

（3）零信任。零信任的核心思想为“从来不信任，始终在校验”，即不再默认信任物理安全边界内部的任何用户、设备或者系统、应用，而是以身份认证作为核心，将认证和授权作为访问控制的基础。零信任能够对所有流量进行可视化、分析检查，并设置安全策略；对所有访问采用最小授权

策略和严格访问控制策略；可实现可视化管理，并提供按需、动态配置的安全隔离网络。建立与所有不安全网络隔离的可信网络，有助于避免网络攻击。

7 总结与展望

全球产业界和学术界广泛关注未来网络的发展及关键技术的研究，各国均从国家战略层面高度重视未来网络的发展与布局，并从未来网络的体系架构、关键技术、试验床等方面开展创新研究。

国际电信联盟（ITU）成立了网络2030焦点组，旨在探索面向2030年以后的网络技术发展。中国成立了网络5.0产业和技术创新联盟，旨在打造一个由中国主导、面向国际、开放并有影响力的下一代数据通信网络技术标准组织，探索面向未来的网络5.0创新架构。

中兴通讯在未来网络架构创新、精准网络技术、网络智能技术、算力网络技术、网络内生安全技术等领域不断深耕细作，愿携手业界伙伴共同探索、积极合作，推进中国未来网络演进升级。

致谢

中兴通讯股份有限公司无线经营部郭雪峰、毛磊、郑兴明、牛娇红、杨春建在论文撰写中给予了很多帮助，在此表示特别感谢！同时，中兴通讯股份有限公司未来网络研究项目经理谭斌在本文起草中给予指导意见，在此一并致谢！

参考文献

[1] 网络5.0技术和产业创新联盟. 网络5.0技术白皮书 [R]. 2019
[2] 黄奇帆. 数字经济时代，算力是国家与国家之间竞争的核心竞争力 [J]. 中国经济周刊, 2020, (21): 106–109
[3] 史炜. “5G+工业互联网”建设的技术经济模式 [J]. 信息通信技术与政策, 2020, (9): 1–7. DOI: 10.3969/j.issn.1008–9217.2020.09.001
[4] 黄兵, 谭斌, 罗鉴, 等. 面向业务和网络协同的未来IP网络架构演进 [J]. 电信科学, 2021, 37(10): 39–46. DOI: 10.11959/j.issn.1000–0801.2021234
[5] CARPENTER B. Architectural Principles of the Internet [R]. RFC Editor, 1996. DOI: 10.17487/rfc1958
[6] 中国信息通信研究院. 云计算发展研究 [J]. 大数据时代, 2020, (8): 28–39
[7] 方敏, 段向阳, 胡留军. 6G技术挑战、创新与展望 [J]. 中兴通讯技术, 2020, 26(3): 61–70. DOI: 10.12142/ZTETJ.202003012
[8] 黄韬, 汪硕, 黄玉栋, 等. 确定性网络研究综述 [J]. 通信学报, 2019, 40(6): 160–176. DOI: 10.11959/j.issn.1000–436x.2019119
[9] 中国联通. 算力网络架构与技术体系白皮书 [R]. 2020
[10] TANG X Y, CAO C, WANG Y X, et al. Computing power network: the architecture of convergence of computing and networking towards 6G requirement [J]. China communications, 2021, 18(2): 175–185. DOI: 10.23919/JCC.2021.02.011
[11] 雷波. 整合多方资源 算力网络有望实现计算资源利用率最优 [J]. 通信世界, 2020, (8): 39–40. DOI: 10.13571/j.cnki.cww.2020.08.018
[12] 蒋林涛. 云计算、边缘计算和算力网络 [J]. 信息通信技术, 2020, 14(4): 4–8. DOI: 10.3969/j.issn.1674–1285.2020.04.001

作 者 简 介

王卫斌，中兴通讯股份有限公司产品规划首席科学家、电信云及核心网总工、移动网络和移动多媒体技术国家重点实验室未来网络研究中心副主任、教授级高工；从事SDN/NFV、电信云研究，以及核心网产品规划；获中国通信学会科技进步奖一等奖、广东省科技进步奖一等奖，相关产品和解决方案荣获曾5G世界论坛、SDN/NFV全球论坛等多项奖项；发表核心期刊论文10余篇，合译专著2本，拥有20余项发明或实用新型专利。

周建锋，中兴通讯股份有限公司核心网产品规划总工；长期从事核心网技术研究、规划、设计等工作。

黄兵，中兴通讯股份有限公司技术规划部技术总监、高级工程师；主要从事数据通信和光通信的中长期技术规划；拥有超过20年的通信产品规划和研发的经验。

大规模网络向IPv6单栈演进的技术方案

Technical Solution of Transition to Large-Scale IPv6-Only Networks

解冲锋/XIE Chongfeng[1]， 李星/LI Xing[2]， 李震/LI Zhen[3]，余勇志/YU Yongzhi[4]

（1. 中国电信集团有限公司，中国 北京 102209；
2. 清华大学， 中国 北京 100084；
3. 下一代互联网国家工程中心，中国 北京100176；
4. 中国电信股份有限公司重庆分公司，中国 重庆 401121）
(1. China Telecom Group Co. , Ltd. , Beijing 102209, China;
2. Tsinghua University, Beijing 100084, China;
3. China Future Internet Engineering Center, Beijing 100176, China;
4. China Telecom Chongqing Branch, Chongqing 401121, China)

DOI：10.12142/ZTETJ.202201012
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20220217.1716.004.html
网络出版日期：2022-02-18
收稿日期：2021-12-15

摘要：中国互联网全面进入了互联网协议第4版/互联网协议第6版（IPv4/IPv6）双栈运行阶段，未来如何向IPv6单栈网络演进是新的技术和产业挑战。结合中国的最新政策要求，在分析已有IPv6单栈过渡技术的基础上，提出了面向大规模网络的多域纯IPv6网络的总体框架、演进路线和相关方案，并给出发展IPv6单栈网络的策略建议。

关键词：IPv6单栈；全局映射规则；多域；IPv6段路由（SRv6）；边界网关协议4.0扩展（BGP4+）

Abstract: With the Internet in China fully entering the Internet protocol version 4/Internet protocol version 6 (IPv4/IPv6) dual-stack operation stage, how to evolve into IPv6 single-stack network in the future is a new technical and industrial challenge. Combined with the latest policy requirements of China and based on analyzing the existing IPv6 single-stack transition technology, the overall framework, evolution route, and related solutions of multi-domain IPv6-only network for large-scale networks are put forward, and the strategic suggestions for the development of IPv6 single-stack networks are proposed.

Keywords: IPv6-only network; global mapping rule; multi-domain; SRv6; BGP4+

作为新一代互联网网络协议，互联网协议第6版（IPv6）是全球互联网升级演进的必然趋势和网络技术创新的重要方向，其地位和趋势在全球已基本无争议。IPv6的发展在中国一直受到高度重视，特别是自从2017年中共中央办公厅、国务院办公厅印发《推进互联网协议第六版（IPv6）规模部署行动计划》以后，中国的IPv6部署大大加速，并已实现全面部署，网络基础设施也已进入互联网协议第4版（IPv4）/IPv6双栈运行阶段。

随着IPv6的规模应用，IPv6单栈网络的部署已经成为新的阶段性重要目标。从双栈向IPv6单栈演进的趋势日益明显，并成为IPv6与经济社会深度融合的先决条件。在国际上，发达国家也在积极推进网络和业务的IPv6单栈化。例如，2020年11月美国白宫管理和预算办公室发布指南，要求“美国各机构尽快完成向IPv6的过渡，确保到2025财年末，联邦网络上超过80%的IP资源是IPv6单栈”。2021年7月，中国中央网络安全和信息化委员会办公室、国家发展和改革委员会、工业和信息化部联合发布的《关于加快推进互联网协议第六版（IPv6）规模部署和应用工作的通知》指出，增强IPv6网络互联互通能力，积极推进IPv6单栈网络部署，是我国未来推进IPv6工作的重点任务之一[1]。

网络协议栈是互联网运行的基石。单一网络协议栈是互联网的最基本要求，是互联互通的前提。IPv4/IPv6双栈网络维护成本高，安全风险大。这不仅给设备提出很高的要求，还给业务发展带来困惑甚至阻碍。从双栈向IPv6单栈演进有利于网络的运营维护，并且具有多个优点：单协议栈的运营工作量和成本比双栈显著降低；单栈网络的风险暴露面减少，使得网络更加安全；网络架构更简单，路由表数量减少，对设备的要求降低，转发能力得到提升。对于互联网

业务来说，IPv6单栈运营更为容易，减少了在支持IPv4能力方面的不必要成本投入。

本文围绕中国IPv6发展的总体规划，分析了大规模网络向IPv6单栈演进的内涵和思路，然后面向大规模、多域、多场景网络提出了向IPv6单栈演进的网络总体架构和实施策略建议。

1 IPv6单栈的概念分析

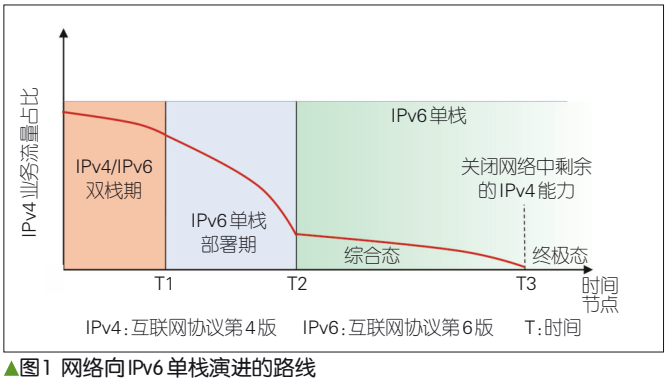
通俗地说，IPv6单栈网络就是在网络中关闭IPv4协议栈并以IPv6协议为核心进行编址、路由和转发的网络，它支持基于SRv6[2]的能力创新。IPv6单栈是网络架构和能力的全新变化。IPv6单栈化的目的是构建极简、智能、安全、绿色的新型网络，它是IPv6发展的最终方向。作为新型的网络基础设施，IPv6单栈网络主要承载各种互联网业务。按照互联网业务对于协议支持的程度不同，IPv6单栈网络呈现两种不同的形态：综合态和终极态。

（1）综合态。网络不仅要承载原生IPv6业务流量，还要承载现存的IPv4业务流量，并在边缘支持IPv4业务流量的接入和穿越。如图1所示，在T2～T3阶段，SRv6等IPv6创新技术逐渐得到广泛应用。

（2）终极态。在互联网中的业务全部IPv6化的情况下，网络本身不再接入IPv4业务流。在图1T3以后的状态中，网络将关闭边缘的IPv4属性。T3以后的状态是向IPv6网络过渡的终极态。

从双栈进入到IPv6单栈的阶段，需要经过T1～T2之间的IPv6单栈部署阶段。这一阶段所需的时间一般为5～6年（时间长短受到实际IPv6流量占比大小的影响）。网络向IPv6单栈的推进也会促进互联网业务向IPv6迁移，进而对网络中IPv6流量的占比产生影响，因此单栈部署期是网络和业务互动的过程。

目前来看，互联网业务要全部实现IPv6化还需要较长时间，因此短期内实施终极态IPv6单栈是不现实的；但如果不推进IPv6单栈化而让双栈长久持续下去，那么业务的IPv6改造动力就会更加不足，这会影响IPv6流量的提升和生态的形成。综合以上因素，当前推动综合态的IPv6单栈方案更具有现实意义和紧迫性。



▲图1 网络向IPv6单栈演进的路线

综合态IPv6单栈的关键问题是，在逐渐关闭IPv4协议栈以后，网络中设备如何有效地支持剩余IPv4业务的承载，以确保用户体验不降低。

截至目前，互联网工程任务组（IETF）针对不同的场景设计了多种纯IPv6过渡方案，如双栈（DS）-Lite[3]、轻量级4over6[4]、翻译型地址+端口映射转换机制（MAP-T）[5]、封装型地址+端口映射转换机制（MAT-E）[6]、NAT64/464XLAT[7]等。这些方案均将IPv4作为IPv6网络的一种业务，因此属于综合态的IPv6单栈方案，并已获得商用。464XLAT是移动网中最成熟的IPv6单栈方案，该方案已获得苹果IOS和安卓等操作系统的支持，并得到了广泛应用。在实践方面，中国分别在不同场景进行了IPv6单栈的试点研究。例如，CERNET是国际上首次采用IPv6单栈的网络，也是迄今为止全球规模最大的IPv6单栈主干网。此外，中国有些运营商也开始在4G和5G移动网络上进行IPv6单栈试点。

大型网络的IPv6单栈化必须具有如下几个特性。

- 开放性：支持与外部各种网络（IPv4单栈、IPv4/IPv6双栈和IPv6单栈）的互通，确保对网络内外业务的正常访问，使用户体验不降低；
- 协议单纯性：以IPv6协议为唯一基础协议进行编址、路由和转发等；
- 支持IPv6基础上的技术创新：与智能化网络架构协同，可根据业务需求和网络维护需求实现基于SRv6的流量调度和网络服务化；
- 地址真实可信：支持互联网真实源地址的精确定位和地址溯源，并可在地址中嵌入用户或者终端身份信息，增强对网络和用户的安全管控能力。

目前，中国数字化转型方兴未艾，云计算、边缘计算和工业互联网等业务都需要一个先进的网络基础设施来支撑。这个基础设施应该是以IPv6为核心协议的新一代基础网络，应避免进入IPv4误区。因此，网络的IPv6单栈化也和数字化转型密切相关，其在中国的部署具有必要性和紧迫性。另外，中国推动互联网的单栈化有利于引领国际IPv6发展潮流，可为全球互联网发展贡献力量。

2 多域网络引入IPv6单栈时的主要问题

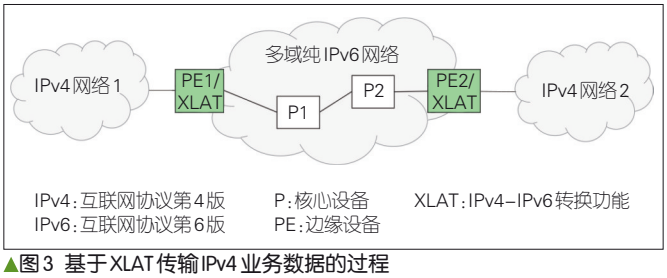
如前所述，目前业界已有多种IPv6单栈过渡技术方案。

这些方案在支持IPv4业务的承载时，需要不同类型的IPv4和IPv6转换技术。例如，464XLAT采用有状态NAT64的翻译技术，IVI采用无状态翻译NAT64技术，DS-Lite采用基于地址族转换路由器（AFTR）的4over6隧道技术，而主干网则采用GRE隧道或无状态翻译技术等。对于运营商来说，在引入IPv6单栈时面临的首要问题是采用何种技术。通常情况下，大规模IP网络是由多个自治系统组成的。各自治系统服务不同的场景，而且经常由不同的组织来管理，并采用不同的路由和安全策略。即便是在同一个运营商内，每个自治系统都由不同的机构或部门来管理和运营。在引入IPv6单栈方案时，如果自治系统之间缺乏协同，就需要在数据路径上将IPv6数据包转回到IPv4，然后在下一个域又转回IPv6。这样网络中就会出现较多功能不同的IPv4-IPv6数据包转换网关，而且转换次数会随着自制系统数量的增加而增加。如图2所示，在含3个域的网络上，各自治系统采用不同的IPv6单栈过渡技术。IPv4业务数据在跨域时需要恢复出IPv4数据包，这会导致网络中出现5次基于网关的IPv4和IPv6转换。过多的IPv4和IPv6转换使网络变得复杂，也使设备投资相应增大。因此，我们迫切需要协同域间的单栈方案，以消除不必要的转换功能，提高数据转发效率。

3 多域纯IPv6技术总体方案

针对以上问题，本文提出了面向大规模、多域和多场景的纯IPv6组网架构及相关技术方案，为不同自治系统间的IPv6单栈方案协同组网提供框架支持。

多域纯IPv6网络的目标是以IPv6为基础协议构建多域互连的网络基础设施，在进行IPv4业务的承载时遵循“IPv4 As A Service”的思路，即把IPv4作为一种业务。对于客户侧发出的IPv4数据包，为了不让IPv4数据包进入网络，网络边缘设备会将IPv4数据包适配成IPv6数据包，如图3所示。为支持在多域纯IPv6网络中传送IPv4业务数据包，在入口边缘设备（PE）上运行的XLAT需要将IPv4业务数据包转换成IPv6数据包，并通过IPv6路由系统将该数据包送到正确的出口PE，然后IPv6数据包被恢复成IPv4数据包；而在域间IPv6数据包不需要被恢复成IPv4数据包，这样系统最多需要两次IPv4和IPv6的转换。为此，各域间需要“达成转换共识”，即采用全局映射规则。具体来说，多域纯IPv6技术方案包括如下几个部分：



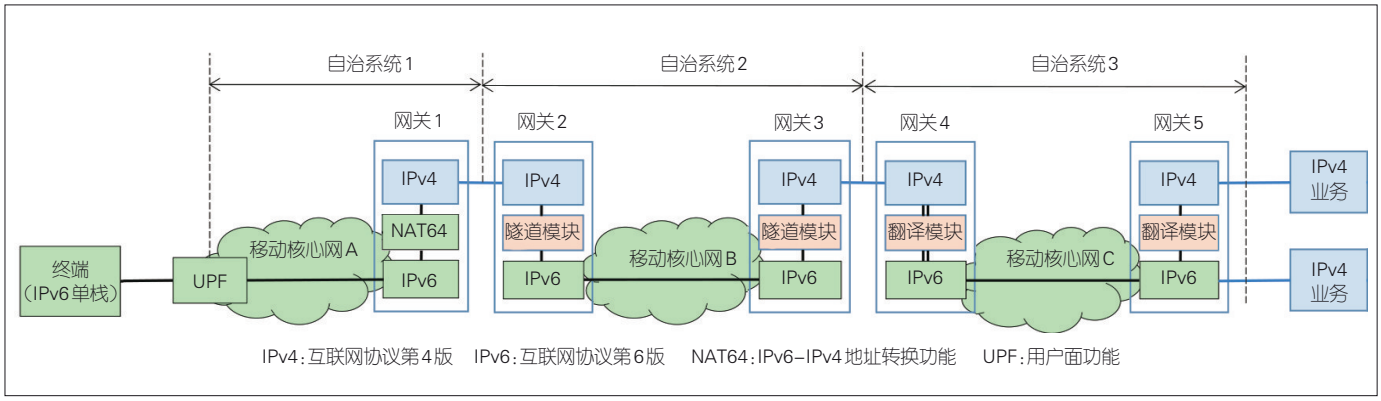
▲图3 基于XLAT传输IPv4业务数据的过程

（1） 基于全局映射规则的IPv4-IPv6地址映射

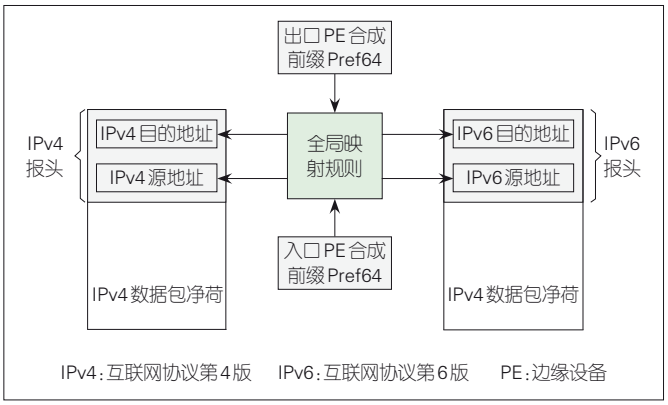
为了提升转化效率，本方案将整个IPv4地址空间“映射”到IPv6地址空间中，即把IPv4看作IPv6的一个“子集”。具体实现方式是通过添加IPv6合成前缀将IPv4地址映射到IPv6地址。全局映射规则就是所有IPv4地址块与IPv6合成前缀的映射关系：{IPv4 address block: Pref64}。对于需要通过纯IPv6网络传送的IPv4数据包，入口PE设备可基于统一定义的全局映射规则对其IPv4源和目的地址映射生成对应的IPv6源和目的地址。同时该IPv4数据包被转换成IPv6数据包，如图4所示，然后在纯IPv6的网络中进行域内和跨域传送。为了支持地址转换，全网PE都配置IPv4地址块对应的IPv6合成前缀。

（2） 基于BGP4+协议扩展跨域交换映射规则

为了将IPv4业务数据包正确地传送到网络出口，需要在PE中将IPv4地址块对应的路由映射合成IPv6路由。对于



▲图2 域间独立采用不同的IPv6单栈技术方案



▲图4 基于全局映射规则的IPv4域IPv6地址转换

特定的IPv4地址块来说，在纯IPv6网络中，可发布IPv6合成路由的PE就是该IPv4地址块的关联PE。关联PE也同样都有相应的Pref64。本方案采用PE的前缀Pref64来生成IPv4地址块所对应的IPv6合成路由，然后在全网交换IPv4地址块在纯IPv6网络中关联的边缘PE（出口PE）的位置，即IPv4地址块的IPv4–IPv6映射规则。为此，可通过BGP4+协议扩展[8]在域内和域间的设备间交换IPv6合成路由和IPv4–IPv6映射规则。

（3） 兼容隧道和翻译的灵活转发方式

当用户发出的IPv4数据包达到IPv6网络边缘时，入口PE的XLAT模块根据本地的IPv4–IPv6映射规则将其转换为IPv6数据包，然后该数据包被转发到对应的网络出口PE。要说明的是，XLAT在转发面既可以支持翻译方式，也可以支持4over6的封装方式。

本方案支持域间IPv6单栈能力协同，包括接入域和主干域的协同、主干域间的协同、运营商间的协同，通过协同降低IPv4业务承载中的转换次数，来确保数据传输的效率。如图5所示，接入域为某个移动核心网，可通过464XLAT为移动终端提供IPv6单栈接入服务。在464XLAT方案中，为了在IPv6网络中支持对IPv4业务的访问，客户端翻译器（CLAT）[9]需要将IPv4客户端（APPv4）发出的IPv4数据包转换成IPv6数据包。主干域A和B分别为提供IPv6转发服务的主干网。本方案可以融合接入域、主干域A、主干域B的IPv6单栈能力，形成从终端到网络主干域B出口的端到端IPv6单栈能力，并且只需要在出口PE路由器进行一次IPv4–IPv6转换，不需要在域间互通的位置（M和N处）将IPv6数据包恢复成IPv4数据包。

4 向IPv6单栈过渡的策略建议

为了在中国尽早推动IPv6单栈网络的产业成熟和网络部署，需要做好如下几方面的工作：

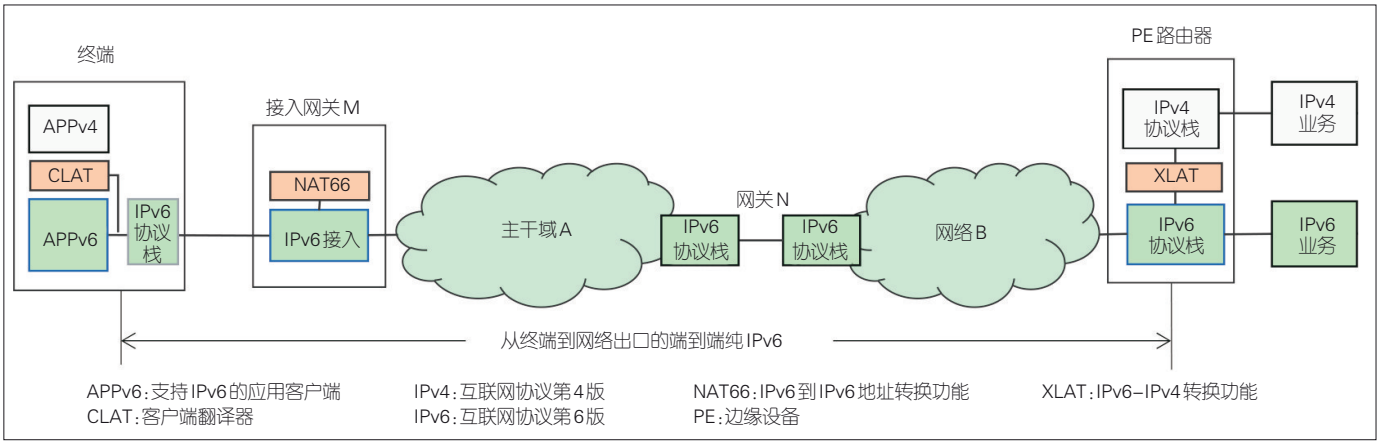
• 从政策方面对IPv6单栈给予引导支持，并在必要时给出关闭网络中IPv4协议栈的实施路线、推进策略和时间点。

• 网络边缘的设备须满足多域纯IPv6标准的能力要求。新入网操作系统和终端（4G/5G移动终端、固网CPE终端、物联网终端和其他新型终端等）都应当具有标准要求的IPv6单栈能力，并优先以纯IPv6方式接入网络。

• 推动互联网应用向IPv6迁移，提升IPv6网络的流量占比，为IPv6单栈化创造良好的流量条件。新上线的云计算、工业互联网、物联网等应按照纯IPv6要求来建设和运营，能在IPv6单栈网络环境下运行。

• 网络运营商和大型互联网公司应按照标准尽早进行IPv6单栈商用试点，并加快在IPv6单栈网络方面的相互协作；鼓励并支持在中国开展IPv6单栈示范区的建设和实践。

• 强化IPv6对网络安全的提升，确保网络安全领域内新的系统、软硬件、安全策略能够在纯IPv6的基础上进行规划设计，加速中国网络信息安全体系向IPv6单栈的同步



▲图5 基于多域纯IPv6实现的端到端IPv6数据通信

过渡。

5 结束语

随着IPv6在中国的规模应用，IPv6单栈网络已经成为新的重要目标，也是中国IPv6部署迈上新台阶的必由之路。IPv6单栈的本质是做“减法”，即消除网络中不必要的功能和协议冗余，优化网络架构。网络的IPv6单栈化和中国的数字化转型密切相关，其部署具有必要性和紧迫性。在实施层面，IPv6单栈化网络在较长时期内仍需要支持剩余IPv4业务的综合承载，因此网络的边缘还须维持IPv4特性。只有等IPv6流量增长大大超过IPv4后，IPv4协议才可彻底退网。本文提出了大规模、多域、多场景下的纯IPv6组网架构，在支持域间和运营商间协同的基础上，整合不同域间的IPv6单栈能力，提高了数据传输效率，并在最后提出了IPv6单栈部署的策略建议。

致谢

阿里巴巴集团IPv6首席架构师宋林健、中国电信股份有限公司研究院马晨昊和李聪为本文的方案研究和现网试验做了大量工作，对他们表示特别感谢！

参考文献

[1] 中央网络安全和信息化委员会办公室, 国家发展和改革委员会, 工业和信息化部. 关于加快推进互联网协议第六版(IPv6)规模部署和应用工作的通知 [EB/OL]. (2021-07-12) [2021-12-10]. http://www.cac.gov.cn/2021-07/23/c_1628629122784001.htm
[2] IETF. SRv6 network programming: IETF RFC 8996 [S]. 2019
[3] IETF. Dual-stack lite broadband deployments following IPv4 exhaustion: IETF RFC 6333 [S]. 2011
[4] IETF. Lightweight 4over6: an extension to the dual-stack lite architecture: IETF RFC 7596 [S]. 2015
[5] IETF. Mapping of address and port using translation (MAP-T): IETF RFC 7599 [S]. 2015
[6] IETF. Mapping of address and port with encapsulation (MAP-E): IETF RFC 7597 [S]. 2015
[7] IETF. 464XLAT: combination of stateful and stateless translation: IETF RFC 6877 [S]. 2013
[8] IETF. Multiprotocol extensions for BGP-4: RFC 4760 [S]. 2007
[9] IETF. IP/ICMP translation algorithm: RFC 6145 [S]. 2011

作 者 简 介

解冲锋，中国电信集团公司科技委委员、教授级高工，欧洲电信标准化协会(ETSI)IPE工组副主席，北京市IPv6重点实验室主任；主要研究方向为下一代互联网、IPv6、网络架构、SDN/NFV、物联网、云网融合等；负责中国电信的多项科研项目，以及多项国家“863”计划、CNGI、国家重大专项等科研项目；在IETF发布RFC 5项，发表论文10余篇，拥有授权专利50项。

李星，清华大学教授、博士生导师，CERNET网络中心副主任，中国互联网协会学术工作委员会主任，曾担任亚太网络工作组(APNG)主席、亚太网络信息中心(APNIC)理事会委员，中国首位在互联网架构董事会(IAB)的成员，IETF Softwire工作组的发起人和技术顾问，中国首个非中文相关标准RFC4925的第一作者，国际互联网协会“互联网名人堂”入选者；发明了无状态、基于运营商前缀的翻译技术IVI，该技术已成被列为IETF核心系列标准(RFC6052、RFC6144、RFC6145、RFC6219等)；在IETF发布RFC 11项。

李震，下一代互联网国家工程中心总工程师，IPv6 Forum Fellow；研究方向为IPv6、SDN/NFV、DNS、IOT等；主持和参与多个重大科研项目，包括GO4IT中欧合作项目、国家发改委CNGI重大专项、北京市重大科技专项等，在下一代互联网标准开发设计、物联网应用以及网络、应用、DNS等基础设施IPv6过渡方面拥有丰富的研究和开发实践经验。

余勇志，中国电信股份有限公司重庆分公司高级工程师、中国电信股份有限公司高级技术专家(IP专业)；主要研究方向为新型IP网络技术、移动互联网技术、云网融合等。

大容量、智能化光传输系统：机遇、挑战与应对策略

Towards Large-Capacity and Intelligent Optical Transmission Systems: Opportunities, Challenges, and Solutions

冯振华/FENG Zhenhua[1,2]，方瑜/FANG Yu[1,2]，施鹄/SHI Hu[1,2]
（1. 中兴通讯股份有限公司，中国 深圳 518057；
2. 移动网络和移动多媒体技术国家重点实验室，中国 深圳 518055）
(1. ZTE Corporation, Shenzhen 518057, China;
2. State Key Laboratory of Mobile Network and Mobile Multimedia Technology, Shenzhen 518055, China;)

DOI：10.12142/ZTETJ.202201013
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.TN.20220218.1930.005.html
网络出版日期：2022-02-21
收稿日期：2021-12-25

摘要：分析了长距大容量、智能化光传输系统的五大关键技术：单波超400 Gbit/s、波段扩展、空分复用（SDM）、光层操作维护管理（OAM）和备用路径性能检测技术，并从学术研究、业界标准化动态等方面介绍了这些技术的进展。针对光传输系统技术发展趋势，从硬件、软件两个层面讨论了光通信的发展机遇和面临的挑战。基于中兴通讯光网络智能化平台框架，并结合中兴通讯在大容量、高速相干光通信方面的研究与产品开发工作实际，介绍了4个典型案例：灵活调制与光域均衡相结合来有效减少滤波代价、C+L波段扩展助力单波400 Gbit/s长距传输、高频光标签实现在线光性能监测、光探针和全局功率分析算法（GPA）确保备用路径快速可靠恢复。基于这些技术的产品化，中兴通讯将持续为客户创造价值，为用户提供更好的网络服务体验。

关键词：大容量传输；扩展波段；光传输质量（QoT）；光域均衡；快速可靠恢复；光标签

Abstract: Five key technologies of long-distance, high-capacity, and intelligent optical transmission systems are analyzed including ultra-high-speed transmission beyond 400 Gbit/s per wave, waveband expansion, space division multiplexing (SDM), optical layer operation and maintenance management (OAM), and performance monitoring for idle paths. The progress of these technologies is also introduced from the aspects of academic research and industry standardization dynamics. According to the technology trends of the optical transmission systems, the development opportunities and challenges in terms of hardware and software are discussed. Based on the intelligent platform framework of ZTE Corporation and related research and development experience in the optical networks, four typical cases are presented including filtering penalty reduction enabled by flexible modulation and optical domain equalization, single wave 400 Gbit/s long-distance transmission together with C + L-band expansion, online optical performance monitoring realized by high-frequency optical label, fast and reliable optical restoration aided by the optical probe and global power analysis algorithm (GPA). With such kinds of novel techniques leading in, ZTE will continue to provide customers with improved value and better network experience.

Keywords: large-capacity transmission; waveband expansion; quality of optical transmission; optical domain equalization; fast and reliable optical restoration; optical label

5G商用能够提升网络带宽，改善用户体验，并促进新型带宽密集型业务和应用的发展。随着“6G”“元宇宙”等概念的提出，扩展现实（XR）、全息通信、智慧交互等沉浸式体验应用，将进一步提升网络对带宽、时延和可靠性的要求[1]。据预测，2030年人类将进入尧字节级别的数据量时代，网络通信需要处理2 000亿个连接，接入带宽需求高达太比特每秒，单纤容量突破100 Tbit/s[2]。毫无疑问，光通信网络基础设施将在带宽扩容和智能化运维方面面临巨大压力。

目前波分复用网络商用系统最高单波速率为800 Gbit/s。随着波特率提升到200 Gbd以上[3]，单波速率预计可以达到1.6 Tbit/s[4]。商用系统单纤最大容量为48 Tbit/s。波段扩展技术的引入可使相关容量成倍增加，如S+C+L系统最高容量可达150 Tbit/s[5]。大量研究证明，以多芯、少模光纤为代表的空分复用（SDM）技术将是实现下一代超大容量光传输的重要技术手段。目前采用38芯3模光纤最大单纤容量已高达10.66 Pbit/s[6]。波段扩展和SDM技术的扩容效率和潜力都十

基金项目：国家重点研发计划（2018YFB1800905）

分可观，但在实现商业化方面还需要应对一系列挑战，如新器件和新算法的设计与实现。

光传输系统灵活组网和智能化运维能力的提升也是业界近期关注的焦点。以数据中心为核心的云化网络将向全光化和Mesh化发展[7]。全光可重构光分插复用器（ROADM）骨干网络支持大颗粒业务波长级灵活调度，使光层一跳直达，无需电中继，有助于降低时延和成本。在工作路径和恢复路径上，业务性能和链路状态的高精度检测与实时化感知是光网智能化的基础。这对实现端到端大容量、低延时、高可靠传输而言具有的重要意义。如果要实现快速业务开通和故障定位的智能化运维，为在线业务提供低成本的单波功率和带内光信噪比（OSNR）检测功能就必不可少；而要实现低延时、高可靠的业务恢复，提前考虑链路光参预调和光损伤验证将至关重要。

本文将介绍大容量、智能化光传输系统的关键技术及其研究进展、产业现状，分析光网络升级转型时在单波速率提升、光纤扩容、恢复时延降低等方面面临的机遇与挑战，结合中兴通讯在大容量、智能化光传输相关的研发实践，展示针对挑战的应对举措及取得的成效，最后总结未来光网络技术的发展趋势。

1 大容量、智能化光传输关键技术

1.1 单波超400 Gbit/s技术

在保证传输距离几乎不变、单比特成本有所下降的前提下，提升单波速率是运营商不变的诉求。表1总结了不同单波速率商用系统的特征和传输能力。当前100、200 Gbit/s系统具备长距骨干网应用的传输能力。而现有400 Gbit/s技术由于传输性能不足，无法满足1 000 km以上长距传输的应用需求。128+Gbd四相相移键控（QPSK）被认为是骨干网升级扩容的最佳方案。受限于当前高波特率相干光DSP（oDSP）芯片、大带宽光器件的商用进展，目前尚无相关成熟产品。5 nm工艺制程的高性能oDSP芯片的采用和3D封装的高集成度光器件的成熟，将加快400 Gbit/s QPSK长距传输解决方案的商用进程（预计在2022～2023年）。

在标准进展方面[8]，光互联网论坛（OIF）已发布400ZR实施协议（IA），采用DP-16QAM+C-FEC（一种调制编码方式），实现了异厂家模块和设备的互操作测试，近期还启动了相干800 Gbit/s LR/ZR/ZR+（指10 km、80 km及80～450 km的光互连）和共封装光学（CPO）标准化研究工作。在400ZR标准框架下，电气与电子工程师协会（IEEE）立项了802.3 ct/cw，分别讨论面向80 km密集波分复用（DWDM）100 GE/400 GE标准化工作。相关标准将在未来1～2年内发布。目前来看，800 GE/1.6 TE很有可能成为下一代以太网的标准速率。国际电信联盟第15研究组（ITU-T SG15）开展了200 Gbit/s/400 Gbit/s接口的物理层规范研究，将DP-16QAM作为400 Gbit/s城域应用的标准码型，推动了开放前向纠错编码（oFEC）的标准化进程。此外，多个多源协议组织（MSA）相继发布了超100 Gbit/s的技术标准。例如，OpenROADM/OpenZR+发布的100～400 Gbit/s相干光模块规范支持CFP2-DCO和QSFP-DD/OSFP封装，在400ZR帧结构的基础上增加100/200 Gbit/s QPSK、300 Gbit/s 8QAM等调制模式，并采用oFEC替代级联FEC（cFEC）的方式来支持450 km级的400 Gbit/s传输。目前，异厂家已宣布实现模块互通测试。中国通信标准化协会（CCSA）的相关标准制订工作包括：100 Gbit/s及以下速率的光传输和模块标准制订已完成，200 Gbit/s报批稿主要选择200 Gbit/s QPSK、8QAM、16QAM码型，400 Gbit/s城域标准实质上采

▼表1 不同单波速率系统特征与能力

单波速率（Gbit·s^{-1}）/间隔（GHz）	调制格式	波段与波道数	传输距离（km）/场景	备注
100/50	DP-QPSK	C/CE/C++（80/96/120）	高于2 000/干线	已规模商用
200/50	DP-16QAM		高于600/城域	
	DP-8QAM/PS-16QAM		800～1 000/城域	
200/75	DP-QPSK	C++（80）	高于2 000/干线	
400/75	DP-16QAM	C/CE/C++（53/64/80）	约400/城域或DCI	
400/100	PS-16QAM	CE/C++（48/60）	约800/城域核心	已有小规模商用
800/112.5	PS-64QAM	CE/C++（40/53）	约200/城域或DCI	
400/150	DP-QPSK	C+L（80）	高于1 500/干线	预计2年内实现商用
800/150	DP-16QAM	C+L（80）	约300/城域	

CE：扩展的常规波段　DCI：数据中心互连　DP：双偏振　PS：概率整形　QAM：正交振幅调制　QPSK：四相相移键控

用的是单波200 Gbit/s双载波方案。《$N\times$400 Gbit/s长距离增强型光波分复用（WDM）系统技术要求研究》等面向更高速率应用的标准课题研究正在开展。

1.2 波段扩展技术

自DWDM技术商用以来，长距系统不断扩展光纤传输频带：从早期C波段（C4T）扩展到CE波段（C4.8T），再到C++波段（C6T）。80波75 GHz间隔的200 Gbit/s QPSK或120波50 GHz间隔200 Gbit/s 8QAM/PS16QAM方案的商用落地将单纤容量提升50%。实际上，单模光纤的低损耗窗口不仅包含C波段，还包括O、E、S、L、U等波段。其中，L波段在日本运营商中有少量部署。L波段的部署可避免G653色散位移光纤四波混频的影响。近年来，美国也有少量运营商和互联网厂商在DCI和海缆传输中部署了C+L系统，可将光纤容量提升一倍。随着单模光纤在容量上逼近100 Tbit/s香农极限，波段扩展技术成为学术和行业研究热点。例如，武汉邮电科学研究院在2014年基于3U超大容量、超高速率、超长距离光传输平台，实现了单模光纤C+L波段共375波的100 Tbit/s 80 km大容量传输[9]。早在2016年，Acacia公司就展示了370 nm宽带全波段（O、E、S、C、L）的可调光收发硅光器件[10]。2018年欧洲科学家系统性地提出了多波段传输的概念和相关组网架构[11]。Nokia Bell Labs和NTT分别实验了在S+C+L波段,距离为100 km、容量为115 Tbit/s以及距离为40 km、容量为150 Tbit/s的光传输系统，该波段支持的单波速率高达400 Gbit/s[5–12]。这些研究表明，波段扩展对提升单纤容量具有重要意义。

在波段扩展技术商用方面，中国运营商和设备商正在积极推动C6T向C6T&L6T方向升级，使网络能够提供单纤80波400 Gbit/s QPSK长距传输能力。目前C+L相关产业链的发展情况如表2所示[8]。可以看出，供应链的发展进度符合预期。随着单波400 Gbit/s长距光模块技术日趋成熟，扩展的C+L波段光系统有望在未来2～3年内实现商用。

1.3 SDM技术

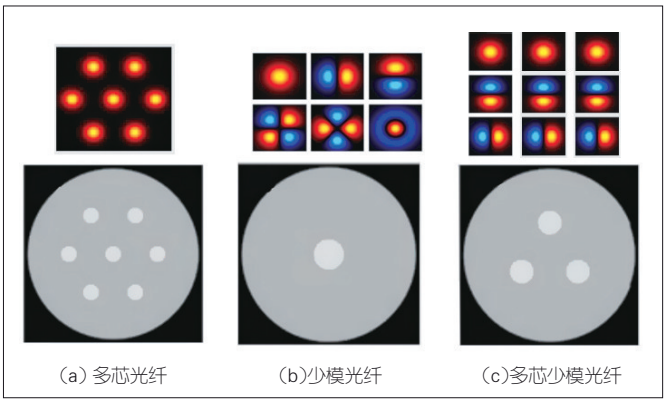
SDM需要基于新型空分复用光纤，主要包括多芯光纤、少模光纤,以及两者相结合的多芯少模光纤，相关原理如图1所示。业界报道了大量基于SDM技术的大容量传输实验，如基于19芯或22芯光纤的1+ Pbit/s传输[13]、基于15模的0.61 Pbit/s传输[14]，以及基于38芯3模的10.66 Pbit/s传输[9]。相比于普通单模光纤，SDM技术将容量提升2个数量级。中国运营商对SDM技术开展了一些研究。中国联通联合长飞光纤光缆股份有限公司、北京大学采用200 Gbit/s商用光传送网（OTN）设备在100 km弱耦合2模光纤上成功完成单纤C波段16 Tbit/s容量的实时演示[15]，充分展示了弱耦合光纤在短距传输方面的扩容优势。中国移动牵头基于弱耦合少模光纤传输技术攻关，实现了总长度为300 km的3模式×4波长×200 Gbit/s的实时模分复用传输实验验证[16]。此外，中国移动最近还联合中兴通讯验证了单波400 Gbit/s×2个模式的200 km传输可行性，为面向未来的多维复用光传输技术发展提供了重要参考。中国电信参与建设了粤港澳大湾区的“超级光网络”,开展了多芯光纤传输示范网试点验证工作，为SDM技术落地进一步奠定基础。

值得注意的是，近期关于SDM的技术研究不再一味追求超大容量，反而更加关注实用性。首先，考虑到光纤弯曲损耗、机械强度，将SDM光纤包层尺寸限制在125 um，有助于兼容现 有标准单模光纤的制备和成缆工艺。其次，考

▼表2 C6T&L6T系统关键组件产业链进展

组件	C6T	L6T	技术难点
ITLA	已商用	样品正在研发中	重新设计增益区和选频光腔
光调制接收器件	与C4T基本相同	与C6T基本相同	关注偏置点和响应度波长相关性
oDSP	与C4T相同	与C6T几乎相同	L波段色散略大,不同波段器件差异补偿
EDFA	已商用	当前铒纤可放大至L5T,优化后可到L6T	优化铒纤掺杂配比改善增益带宽;改善饱和功率和噪声系数;控制EDFA模块尺寸和功耗
DRA	已商用	增加了长波泵浦激光器	需要解决与1 502 nm OTDR的波长冲突问题
WSS	已商用	样品正在研发中，C+L10T已实现产品化	更换衍射光栅和空间光路设计
AWG	已商用	技术准备就绪	/
OPM	已商用	技术准备就绪	/
OTDR/OSC	与C4T相同	技术准备就绪,待确定波长	/

AWG：阵列波导光栅　C6T：C++波段　L5T：L+波段　oDSP：相干光DSP芯片　OSC：光监控信道
DRA：分布式拉曼放大器　EDFA：掺铒光纤光放大器　L6T：L++波段　OPM：光功率监测　OTDR：光时域反射仪
C4T：C波段　ITLA：集成可调激光器　L10T：同时支持C、L波段　WSS：波长选择开关　WSS：波长选择开关



▲图1 空分复用的主要形式

虑到长距传输系统中空间模式间的串扰，以及模式相关损耗对相干解调算法的影响，耦合芯3芯光纤或弱耦合4芯光纤结合多波段WDM传输成为近年来OFC热点。例如，NICT利用（S+C+L）光频梳在4芯光纤中实现了552波道3 001 km的传输，使单纤容量达到319 Tbit/s[17]。再者，在系统可靠性验证方面，2019年日本住友电工与拉奎拉大学合作，在意大利拉奎拉市地下隧道首次铺设了6.29 km含18根多芯光纤的光缆。现场测试表明，多芯光纤成缆及部署后仍然具有较低的损耗和模式色散。这证明了SDM传输应用初步具备从实验室理想环境走向复杂现场环境的条件[8]。最后，在标准化进展方面，ITU-T于2020年已经开始SDM光纤光缆的标准化研究工作，并重点关注光纤分类、光纤熔接和连接器等技术。目前，日本已发布SDM相关技术研究报告，着力推动SDM技术的商用。此外，中国CCSA已立项P比特超大容量光传输相关的研究课题。

1.4 光层OAM技术

大容量、低时延、绿色低碳的需求驱使以传统电交叉为主的OTN网络向以光交叉为主、电交叉为辅的光电联动全光网转型，其核心是光电深度融合和协同管控，以充分发挥光电两层技术的优势,实现网络资源和运维效率的优化。然而，当前OTN网络在光层缺乏成熟的OAM技术，导致骨干网面临升级扩容后运维难度不断增加的局面，使得光层通道性能监测、故障定位以及业务调度常需繁琐的人工性能采集和复杂的定位分析，难以适应智能化的发展趋势。基于低频调顶的随路监控光标签概念最早在1993年被提出，以用于信号识别、功率和故障管理[18]。该技术后来被朗讯公司用于实现端到端信号追踪和在线性能监测，以及故障定位、重路由和波长转换[19]。直到2006年，Tropic Networks公司提出快速傅里叶变换，用于识别不同调顶载频，为WDM网络中多载波调顶奠定基础[20]。近年来，基于调顶的光标签技术已经在动态WDM网络性能检测中得到广泛研究，如光功率、色散、偏振模色散、OSNR监测、非线性噪声监测等。与此同时，该技术也暴露出一些问题，如低载频光标签受到受激拉曼散射（SRS）串扰影响，高载频光标签受到色散衰落的影响。中国运营商将进一步关注基于光标签的OAM技术的商用落地。例如，中国移动已经着手光层OAM技术的标准化工作，并从功能、速率、开销、成帧、编码等方面进行技术规范。总体来看，基于光层OAM的业务路径追踪、连接关系识别、连接性能检测以及故障定位等功能有助于实现更高效的光电协同管控。这将在智能化光网络中发挥越来越重要的作用。

1.5 备用路径性能检测技术

专网应用对OTN网络可靠性、稳定性要求越来越高。基于波长/自动交换光网络（WASON）的网络保护恢复可以实现业务的动态重路由，可应对多次断纤故障。现有ROADM网络的动态重路由采用分布式算路策略（源节点算路）。这种机制可能存在波长冲突导致的路由回退问题，无法保证恢复时间。采用集中算路和分布式控制结合方式有望解决波长资源冲突问题。提升算路单元计算能力和算法效率、优化光转发单元（OTU）波长调谐时间和波长选择开关（WSS）切换时间，可以减少重路由业务恢复时间。另外，由工作路径倒换到恢复路径后的业务性能也是未知的。重路由后能否实现快速业务开通也是影响恢复时间的重要因素。目前通常的做法是，计算路由时考虑路径的OSNR性能和光损伤代价，然后选择OSNR满足阈值条件的路由以用于业务恢复。由于恢复路径上的器件插损、波长相关性、光交叉连接（OA）增减波增益变化的影响会导致真正业务倒换后业务不通，因此系统需要进行端到端功率优化后才能开通业务。这严重影响恢复时间，无法满足运营商普遍要求的确定性低时延需求[21]。由此可见，快速、准确地检测备用路径性能以及精细化的光参预调节对保证快速、可靠恢复至关重要。

2 大容量、智能化光传输机遇与挑战

面对更大容量、更低时延、更高可靠性和高度智能化的演进需求，光传输系统的机遇主要包含两个方面：

（1）在硬件层面上，采用更先进的芯片、器件可兼顾高集成度、高性能和绿色低碳。例如，5、3 nm超强oDSP可实现单波提速和功耗降低，高维度、多端口、多分区的WSS可实现更简洁的光交叉连接（OXC），更高自由度、更大带

宽的可编程部件能够构建灵活、超宽、极简的光传输系统，基于一体化超宽WSS可简化C+L系统光层组网。特定的应用场景需要差异化的解决方案。例如，某些中短距传输场景可采用简化的相干模块或单纤双向传输，以控制功耗和成本。新型宽带放大器、差异化光纤信道的使用拓展了大带宽、长距离、大容量传输的更多维度。宽谱半导体光放大器（SOA）、多波段拉曼放大、低噪声参量放大、大有效面积G654E、多芯少模和空芯光纤都可能是下一代光传输系统的关键技术。优化核心器件的响时间有助于缩短业务恢复时间。例如，从结构、机理上优化OTU可调谐激光器的波长切换时间，通过新材料、新设计改善硅基液晶（LCoS）芯片的响应时间，可实现WSS 10 ms级的快速切换。新材料、新工艺的使用能够持续提升系统带宽和单波速率。例如，薄膜铌酸锂、石墨烯、有机聚合物以及表面等离子材料突破了传统硅光/InP器件的限制，将器件带宽扩展到100 GHz以上[22-23]。另外，关键技术自主化与国产化也给学术研究和供应链产业带来新的机遇。例如，L6T光器件关键技术的突破与国产化，以及用于波段扩展或空分复用系统的新型器件的自主研制等。当然，通信与光层多参数感知一体化的新需求，也会促进基于通信光纤光缆资源进行环境温度应力监测、光缆风险预测、同路由识别、传感对业务影响等课题的研究，甚至会催生分布式光纤传输的广泛应用。更重要的是，产业生态的良性发展，离不开产业链的协同和标准规范的约束。各厂家应当积极迎接开放解耦趋势的机会和挑战，包括扩展波段波长标准化以实现产业聚焦，开放光线路系统以实现光电板卡解耦，实现CFP/CFP2相干模块接口标准化并与多厂家互通等。

（2）在软件层面上，光层数字化是实现光网智能化的前提，模拟光链路的精确建模在支撑光参数快速检测、光性能准确评估和业务性能在线优化方面的重要性进一步突显。长期来看，结合数字孪生技术与人工智能/机器学习算法将在性能评估与优化、软故障预测、根因分析诊断等领域发挥巨大作用。在网络管控方面，集中算路与分布式控制新架构配合更高效的选路算法、波长调度策略，可支撑更大规模的光网组建能力，如100～200 ROADM节点的区域干线网络、200～500 ROADM节点的国干全光网络以及500～1 000 ROADM节点的一二干融合大网。在底层技术上，随着波段扩展、空分复用技术的引入，光系统需要借助快速光功率调测算法去解决重路由时快速增减波功率调测问题。随着网络规模的扩大和串行链路的增加，光系统需要具备在线的全网光功率优化能力，以确保全网光功率、OSNR性能稳定。在应用上，光网络资源可视、业务性能可管、网络故障可预测等功能将会给智能化运维带来更高效的体验。当然，光层与电层、软件与硬件之间的协同，也会给整个光网络软件架构带来新的发展机会。

机遇的背后意味着挑战。当芯片和算法进入“后摩尔/香农时代”时，光模块背靠背OSNR容限潜力的挖掘举步维艰，通道间的光纤非线性效应仍缺少高效的补偿方案。受限于当前（数模转换）DA/模数转换（AD）芯片技术信号波特率难以突破200 Gbd，单通道速率向单波1 Tbit/s以上演进路线暂不明确。光电合封的共封装技术（CPO）被认为是高速光模块的终极解决方案。然而，如何提高InP基平台的集成规模并降低成本，如何有效可靠地集成SiP基平台光源，都是业界急需解决的问题。在光模块性能出现瓶颈的条件下，如果要保证超400 Gbit/s系统的传输距离就需要降低系统余量。性能和风险取决于光系统滤波、串扰及非线性等损伤代价的精准程度。频谱扩展引入更大的SRS导致更大的OSNR不平坦，这给系统性能评估和通道级光功率分配带来挑战。宽谱放大的问题也是当前急需攻克的难点。宽带SOA、低噪声相敏放大仍不能满足光线路系统的商用条件。目前，L波段铒纤的放大效率和带宽还不够理想，L6 THz的实现仍需要掺杂工艺和光路设计的优化。更大规模的光层组网要求更多级数的WSS穿通和更高维度的交叉。宽带WSS在频谱分辨率、通道谱宽、隔离度和端口数量扩展上的制约将会带来更大的串扰和滤波代价。大规模ROADM网络的路由规划、恢复路径计算对算法效率和控制时效性提出更高要求。此外，工作路径和备用路径上光层性能的有效监测技术还比较缺乏。光层OAM在进行高波特率、长距离传输时会面临SRS串扰和色散衰落等问题。基于光纤传感的同路由检测和光纤故障风险识别等应用需要考虑在线传感应用与业务信号共存的场景。传感对业务信号的影响仍需要验证，并且非通信波段的窄线宽激光器也是实现板卡式光纤传感的难点之一。

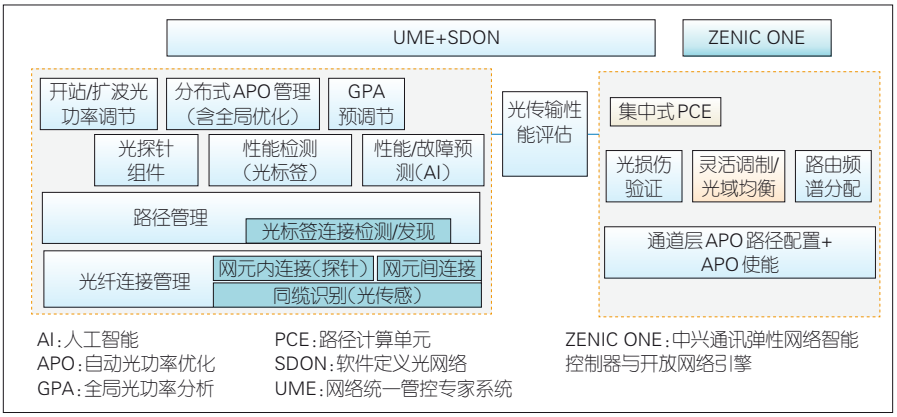
3 大容量、智能化光传输实践

针对上述挑战，我们开发了如图2所示的大容量、智能化光传输管控平台。该平台主要包括依托于网管与设备侧的全局功率和连接管理算法，以及基于WASON的业务级损伤验证、功率优化和路由频谱分配算法。下面我们将结合几个案例，来展示我们在大容量、智能化光传输方面的实践。

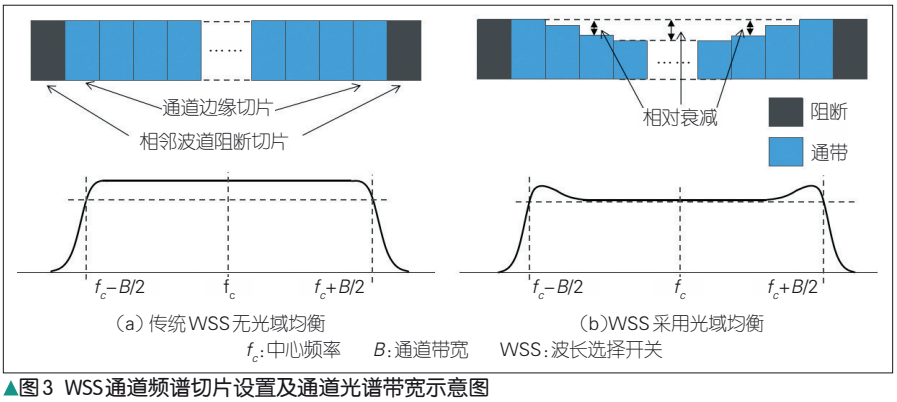
3.1 光域均衡显著提升ROADM穿通能力

如图3所示，基于商用WSS的物理特性，我们提出基于分片整形的光域均衡专利技术。该技术通过对通道内不同频

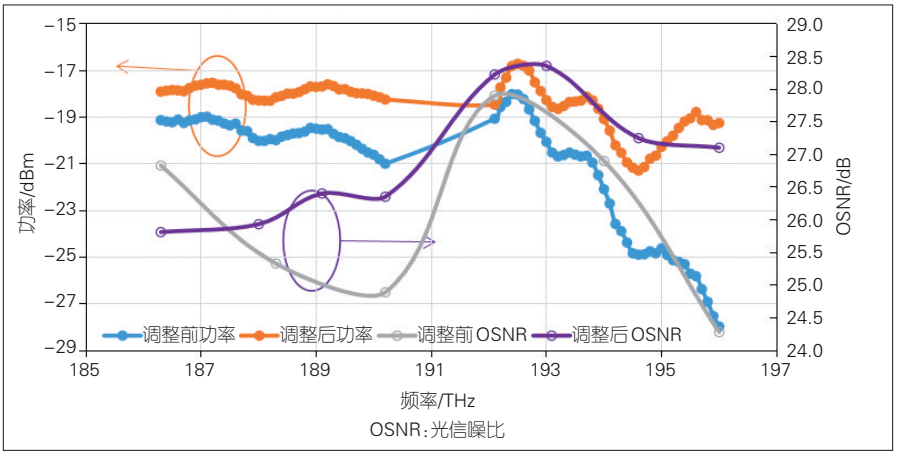
谱切片施加相对衰减使WSS的通道带宽提高了3 dB，从而减小信号的滤波代价，增加ROADM的组网能力。该方案无需硬件升级，分布式整形能力强。纯软件控制层面在线调整WSS通道内的每个频率分片的衰减并不影响业务，也不会增加额外的成本。当采用光域均衡时，200 Gbit/s QPSK信号在75 GHz通道间隔的穿通能力可提升100%[24]，200 Gbit/s PS16QAM/8QAM信号在50 GHz下的穿通能力可提升60%以上[25]。此外，基于灵活调制和光域均衡的Flex Shaping技术已经在多个运营商现网中商用。200 Gbit/s信号在37.5 GHz间隔下拥有大于10级的穿通能力和传输距离，能够有效减小不必要的电中继，节省成本，降低网络时延。



▲图2 中兴通讯智能化管控平台架构总体框架示意图



▲图3 WSS通道频谱切片设置及通道光谱带宽示意图



▲图4 5跨段G652光纤传输单波功率和OSNR在光功率优化调整前后的对比

3.2 C+L扩展波段助力单波400 Gbit/s长距离传输

受限于目前OTU的波长可调谐范围，我们在C6T+L5T光系统上传输80波400 Gbit/s信号。其中，C波段和L波段各有40波，并且波道间隔为100 GHz。由于基于128 Gbd QPSK的长距400 Gbit/s实时相干光模块目前仍处于研发阶段，本实验的长距400 Gbit/s方案为91.6 Gbd PS16QAM，并适配100 GHz波道间隔。光纤链路中有5个G652光纤跨段，每段光纤长度为75 km，损耗约22 dB。C和L波段分别采用一个EDFA来补偿跨段损耗，并在放大前后均有一个WDM合分波器。如图4所示，在没有进行功率调整前，由于C+L系统中存在强烈的SRS功率转移，5跨段末端单波功率平坦度劣化严重，无法满足系统应用需求。采用C+L功率预均衡策略对EDFA的增益和增益斜率进行调整后，系统的功率平坦度优于4.5 dB，OSNR平坦度优于2.5 dB，最小OSNR高于25 dB，满足预期的功率均衡目标。C+L波段400 Gbit/s PS16QAM的传输代价在5 dBm入纤时小于1 dB。该方案使中兴通讯成功完成业界首个单波400 Gbit/s C+L系统现网测试，G654E光纤传输距离达到1 300 km，G652光纤传输距离大于1 000 km。良好的方案结果进一步推进了单波400 Gbit/s与C+L系统的商用进程。

3.3 高频光标签实现在线光性能监测

为了避免长距离传输后SRS串扰对低载频标签信号检测性能的劣化影响，同时解决高波特率信号上加载高频标签时色散导致的功率衰落问题，我们提出将光标签的载频“搬移”到大于10 MHz的相对高频位置，并在OTU单板、OA单板、WSS单板进行逻辑电路的标签加载和检测，如图5所

示。利用在线板卡，我们开展了光标签在波长追踪、通道插损检测、单波光功率检测和单波OSNR检测的实验。在20跨段的单波100、200 Gbit/s传输系统中的实验结果表明：目前光标签接收灵敏度优于−38 dBm，功率检测精度达到1 dB；结合光放噪声系数的定标，OSNR检测精度可以做到1.5 dB，并且光标签对业务OSNR容限劣化小于0.1 dB。光性能检测的精度、速率，以及板卡集成度和成本都比传统的OPM有优势。可以预见，光标签在通道信号丢失（LOS）检测、远程功率控制等光层OAM应用方面也能发挥作用。光标签技术的商用将进一步提升光传输设备的智能化运维水平，也将为光电融合组网奠定基础。

3.4 光探针和GPA双管齐下以确保备用路径快速可靠恢复

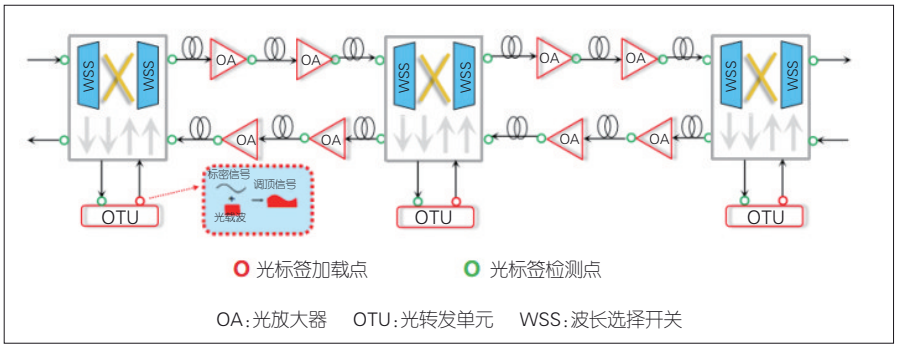
针对业务恢复的预置路径性能监测和ROADM站内插损检测等应用，我们提出光探针的概念。如图6所示，系统通过级联光放和WSS产生可编程的假波源，然后接入待测通道，利用网络中配置的OPM，来获取待测光通道的通道光功率并计算OSNR等光性能指标。目前我们已开发出基于光探针应用的完整解决方案。该技术可以有效解决空闲光通道的性能检测问题，使预置路径上的功率检测精度优于1 dB，OSNR检测精度优于1.5 dB。

对于非预置的动态重路由的恢复路径，我们提出GPA技术。该技术的总体思路是：对底层光器件建立通道级功率演进模型，以配合必要的出厂/开局定标工作；当发生业务路径倒换时，各个光层模型级联将对线路中各通道的光功率进行准确估算，以指导站点预设通道衰减值，并确保业务在倒换后的快速开通。GPA最核心的技术是光链路上的OA模型、光纤模型、站内插损的功率估算模型，如图7所示。其中，OA模型需要实现任意少波、任意输入功率、不同增益设定下的高精度功率计算；光纤模型包含波长相关损耗及SRS建模；通过在开始阶段复用段首尾OPM对模型的校准，通道功率计算模型已验证10跨段误差小于1.5 dB。

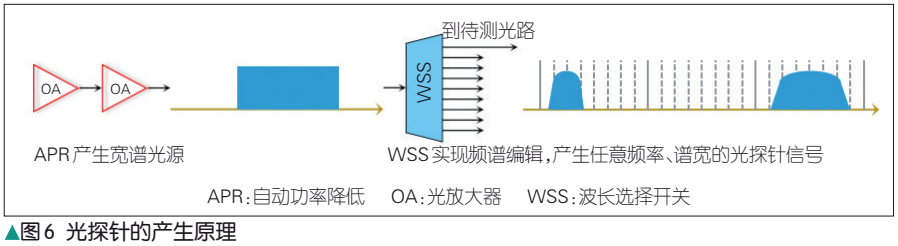
当网络发生故障时，基于光探针和GPA技术，SDON/WASON等控制模块将业务倒换到备用路径。借助已经调整好的光路参数和光传输质量（QoT）评估可验证备用路径的光性能损伤，有助于实现业务的快速、可靠恢复，提升网络的生存性和一次性恢复的成功率。

4 总结与展望

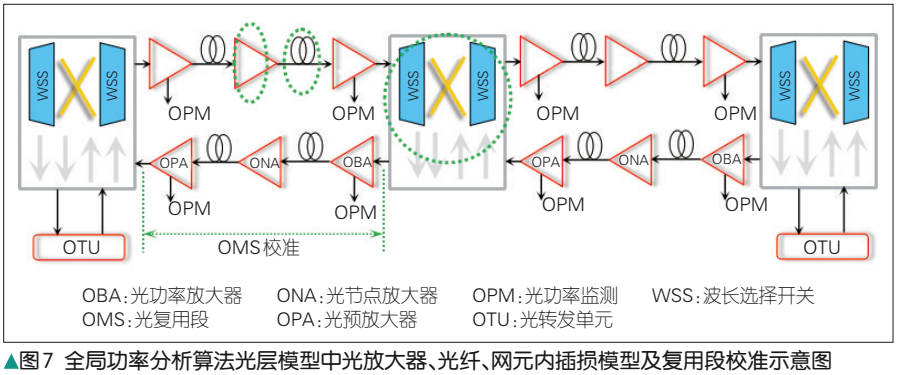
以用户体验为中心，构建无处不在与无处不及的全光连接，提供超大带宽、架构极简的超强“运力”，引入智能化“算力”，将助力全光网的高效运营，保障网络自动优化和可靠运行，有助于最终实现网络自治。大容量光网络在单波提速、波段扩展、SDM、光层OAM、备用路径性能检测和智能化管控等方面充满机遇和挑战，需要学术界和业界共同加强产学研用的协同发展，以推动技术创新和快速商用。以灵活调制和光域均衡为核心的Flex-Shaping技术可扩大200、400 Gbit/s OTN的传输距离并增强相应ROADM的组网能力，在扩展波段的基础上持续提升业务单波速率和单纤容量，加



▲图5 基于光标签的光传输系统



▲图6 光探针的产生原理



▲图7 全局功率分析算法光层模型中光放大器、光纤、网元内插损模型及复用段校准示意图

速基于光标签的光层OAM技术在连接管理、性能管理方面的应用，推动光探针、GPA、QoT等组件与人工智能（AI）、机器学习等智能化算法在现网中的快速融合应用，增强网络恢复可靠性、保障确定性低时延是我们目前重要的研究方向。

未来商用OTN将继续围绕“宽”“简”“智”发展。中兴通讯将在“大容量”“智能化”两大阵地上持续攻坚，协同推动“新速率、新波段、新站点、新算法、新运维”快速商用落地，为5G新基建甚至6G场景应用探索最合适的技术路线，持续改善网络服务体验，为用户创造价值。

致谢

本研究得到中兴通讯股份有限公司邹红兵部长、陈勇总工、贾殷秋博士、吴琼博士、高继韬博士的帮助，谨致谢意！

参考文献

[1] IMT-2030(6G)推进组. 6G总体愿景与潜在关键技术 [R]. 2021
[2] 华为. 通信网络2030 [R]. 2021
[3] PITTALÀ F, SCHAEDLER M, KHANNA G, et al. 220 GBaud signal generation enabled by a two-channel 256 GSa/s arbitrary waveform generator and advanced DSP [C]//2020 European Conference on Optical Communications (ECOC). IEEE, 2020: 1-4. DOI: 10.1109/ECOC48923.2020.9333130
[4] PITTALÀ F, BRAUN R P, BÖCHERER G, et al. 1.71 Tbit/s single-channel and 56.51 Tbit/s DWDM transmission over 96.5 km field-deployed SSMF [J]. IEEE photonics technology letters, 2022, 34(3): 157-160. DOI: 10.1109/LPT.2022.3142538
[5] HAMAOKA F, MINOGUCHI K, SASAI T, et al. 150.3-Tb/s ultra-wideband (S, C, and L bands) single-mode fibre transmission over 40-km using > 519Gb/s/A PDM-128QAM signals [C]//2018 European Conference on Optical Communication (ECOC). IEEE, 2018. DOI: 10.1109/ecoc.2018.8535140
[6] RADEMACHER G, PUTTNAM B J, LUÍS R S, et al. Highly spectral efficient C L-band transmission over a 38-core-3-mode fiber [J]. Journal of lightwave technology, 2021, 39(4): 1048-1055. DOI: 10.1109/JLT.2020.3018128
[7] 中国电信. 全光网2.0技术白皮书 [R]. 2021
[8] 冯振华, 尚文东, 陆源, 等，大容量光传输技术进展与400 G C+L系统研究 [J]. 信息通信技术与政策, 2021, 47(12): 48-61
[9] 余少华, 何炜. 光纤通信技术发展综述 [J]. 中国科学(信息科学), 2020, 50(9): 87-102
[10] DOERR C, CHEN L, NIELSEN T, et al. O, E, S, C, and L band silicon photonics coherent modulator/receiver [C]//2016 Optical Fiber Communications Conference and Exhibition (OFC). IEEE, 2016: 1-3
[11] NAPOLI A, COSTA N, FISCHER J K, et al. Towards multiband optical systems [C]//Advanced Photonics 2018 (BGPP, IPR, NP, NOMA, Sensors, Networks, SPPCom, SOF). OSA, 2018. DOI: 10.1364/networks. 2018. netu3e.1
[12] RENAUDIER J, ARNOULD A, GHAZISAEIDI A, et al. Recent advances in 100 nm ultra-wideband fiber-optic transmission systems using semiconductor optical amplifiers [J]. Journal of lightwave technology, 2020, 38(5): 1071-1079. DOI: 10.1109/JLT.2020.2966491
[13] MORIOKA T. High-capacity transmission using high-density multicore fiber [C]//Optical Fiber Communication Conference. OSA, 2017: 1-45. DOI: 10.1364/ofc.2017.th1c.3
[14] RADEMACHER G, PUTTNAM B J, LUÍS R S, et al. Peta-bit-per-second optical communications system using a standard cladding diameter 15-mode fiber [J]. Nature communications, 2021, 12: 4238. DOI: 10.1038/s41467-021-24409-w
[15] SHEN L, GE D W, SHEN S K, et al. 16-Tb/s real-time demonstration of 100-km MDM transmission using commercial 200G OTN system [C]// Optical Fiber Communication Conference (OFC) 2021. OSA, 2021. DOI: 10.1364/ofc.2021.w1i.2
[16] GE D W, ZUO M Q, ZHU J L, et al. Analysis and measurement of intra-LP-mode dispersion for weakly-coupled FMF [J]. Journal of lightwave technology, 2021, 39(22): 7238-7245. DOI: 10.1109/JLT.2021.3110821
[17] PUTTNAM B J, LUÍS R S, RADEMACHER G, et al. 319 Tb/s Transmission over 3001 km with S, C and L band signals over >120nm bandwidth in 125 μm wide 4-core fiber [C]//2021 Optical Fiber Communications Conference and Exhibition (OFC). IEEE, 2021: 1-3
[18] HILL G R, CHIDGEY P J, KAUFHOLD F, et al. A transport network layer based on optical network elements [J]. Journal of lightwave technology, 1993, 11(5/6): 667-679. DOI: 10.1109/50.233232
[19] HEISMANN F, FATEHI M T, KOROTKY S K, et al. Signal tracking and performance monitoring in multi-wavelength optical networks [C]// European Conference on Optical Communication. IEEE, 1996
[20] WAN P W , REMEDIOS D , JIN D , et al. Channel identification in communications networks: US7054556 B2 [P]. 2006
[21] NGOF/CCSA TC618工作组. 波长交换光网络(WSON)2.0技术白皮书 [R]. 2021
[22] LI H. Vision and trend analysis for transport networks in 5G era [C]//Asia Communications and Photonics Conference (ACP). OSA, 2020
[23] WANG Y, LI X, JIANG Z, et al. Ultrahigh-speed graphene-based optical coherent receiver [J]. Nature communications, 2021, 12: 5076. DOI: 10.1038/s41467-021-25374-0
[24] SHI H, SHANG W D, CHEN H, et al. Optical filtering tolerant and spectrally efficient 200 Gbps real-time transmission using flex-shaping algorithms [C]//2021 19th International Conference on Optical Communications and Networks (ICOCN). IEEE, 2021: 1-3. DOI: 10.1109/ICOCN53177.2021.9563737
[25] FENG Z H, CHEN H, SHI F, et al. ROADM traversal improvement enabled by optical domain equalization [C]//2021 IEEE 6th Optoelectronics Global Conference. IEEE, 2021: 68-72. DOI: 10.1109/OGC52961.2021.9654343

作 者 简 介

冯振华，中兴通讯股份有限公司资深预研工程师；主要从事相干光系统算法设计和验证工作。

方瑜，中兴通讯股份有限公司波分产品研发总工；主要从事光系统产品架构的设计和技术总体规划工作。

施鹄，中兴通讯股份有限公司光系统总工、预研项目经理；主要从事光系统设计、新技术预研规划相关工作。

微服务架构下的算力路由技术

Computing-Power Routing Technologies Under Architecture of Micro-Service

陈晓/CHEN Xiao[1,2]， 黄光平/HUANG Guangping[1,2]

（1. 中兴通讯股份有限公司，中国 深圳 518057；
2. 移动网络和移动多媒体技术国家重点实验室，中国 深圳 518055）
(1. ZTE Corporation, Shenzhen 518057, China;
2. State Key Laboratory of Mobile Network and Mobile Multimedia Technology, Shenzhen 518055, China;)

DOI：10.12142/ZTETJ.202201014
网络出版地址：https://kns.cnki.net/kcms/detail/34.1228.tn.20220217.1717.006.html
网络出版日期：2022-02-18
收稿日期：2021-12-26

摘要：从算力网络的目标架构和愿景出发，研究微服务集群架构下的端到端路由技术解决方案，聚焦算力路由穿透集群L4-L7代理节点并进入服务级颗粒度的场景。在确保与现网平滑兼容前提下，从协议转控面角度分析IPv6段路由（SRv6）和虚拟可扩展局域网（VxLAN）的增强算力路由解决方案。

关键词：算力路由；微服务；SRv6；VxLAN

Abstract: From the viewpoint of the designed and envisioned computing power networking architecture, an end-to-end routing solution under the architecture of micro-service is proposed, which focuses on extending the L3 routing to the computing service within the micro-service cluster. Enhanced segment routing IPv6 (SRv6) and virtual extensible local area network (VxLAN) computing-power networking solutions have been analyzed and presented in detail with the principle of smooth compatibility with the ongoing commercial network architecture.

Keywords: computing-power routing; micro-service; SRv6; VxLAN

在全行业数字化转型升级的宏观背景下，继通信网络之后，算力成为至关重要的数字化产业基础设施。在5G及后5G时代，算力向边缘乃至超边缘下沉，已经成为行业趋势。同时，随着行业算力需求的多样化和终端算力的增强，算力的泛在化将成为新的行业形态。与通信网络不同，各种算力（尤其是异构算力）之间并无统一的架构和体系，缺乏协同机制。这导致算力成为“孤岛”，泛而不强，强而不专。因此，基于公共通信网络的泛在连接，将端、边、云的泛在算力有效协同起来，使之形成统一、动态、智能的算力资源池，成为算力网络的重要目标之一[1]。

算力可分为两类：一类为基础算力，如中央处理器（CPU）、图形处理器（GPU）、数据处理器（DPU）、专用集成电路（ASIC）等，属于静态算力资源；另一类为服务算力，如算法、功能等通用服务级算力，这类算力直接面向业务数据，属于动态算力资源。算力、算法和数据构建了有机的整体。为了发挥异构算力的最大算力效能，不同的算力将采用不同的算法来处理不同类型的数据。在算力网络目标架构下，上述两类算力均被统一感知、统一调度、统一路由，从而连算成网，形成一个层次化的统一算力资源池。

在当前云网业务模式下，算力和网络独立部署、独立运营、独立服务。用户分别向云服务和网络服务提供商提交服务申请，构建服务合同，组合实现完整的应用服务。这种算网分离模式催生了互联网的繁荣，导致算力和网络服务粗放式交付模式的产生，造成巨大的资源浪费。因此，算力网络的另一个重要目标是算网深度融合，即算力和网络服务在一个平面、一个接口、一个路由策略中进行。

综上所述，网络需要将传统的感知和路由向层次化算力方向延伸，从而构建一个基于通信网络的算力和网络资源全网视图，并以此作为全新的业务交付平台，在大幅提升算力和网络资源效率的同时，为行业提供更加丰富、高效的算网融合业务能力，进而赋能信息通信技术（ICT）深度融合的全行业数字化转型升级。

1 微服务架构下的服务路由和寻址机制现状

将应用程序解耦成独立的子服务集群，并分别开发、测试、维护和交付，是微服务架构及其部署和运营模式。微服务架构是基于以更加灵活、更易扩展为主要原则的全新应用部署模式的，服务网格内部的交互和通信由应用网关在L7

层统一执行。应用仍然是最小可访问、可调度的资源颗粒度。微服务以及服务网格用户均不可见，网络也无法感知和路由。应用网关将作为应用代理终结L3层路由流量。

基于层次化算力资源感知和路由的算网一体路由机制，是算力网络架构的重要特征。而应用是算力资源的服务对象，并不是算力网络本身调度和路由的对象。因此，以应用为颗粒度的微服务部署架构，对网络屏蔽了算力服务并终结了L3层网络流量。这是端到端算力路由面临的一个行业现实问题。算力网络的资源调度，改变了传统网络以端口地址为对象的数据通信转发，它是以应用和资源为服务对象的资源匹配。应用从传统的向平台要资源转变为向网络要资源。作为应用从单体到微服务解耦的架构演进，微服务也使得网络为分布式算力协同提供网络能力支撑成为可能。因此，核心问题就是，如何打破以数据通信转发为主的网络功能与以应用为主的微服务协同功能之间的界限，让网络能够看到应用内部，从而更好地为应用服务[2]。

1.1 端到端算力路由面临的分段微服务路由挑战

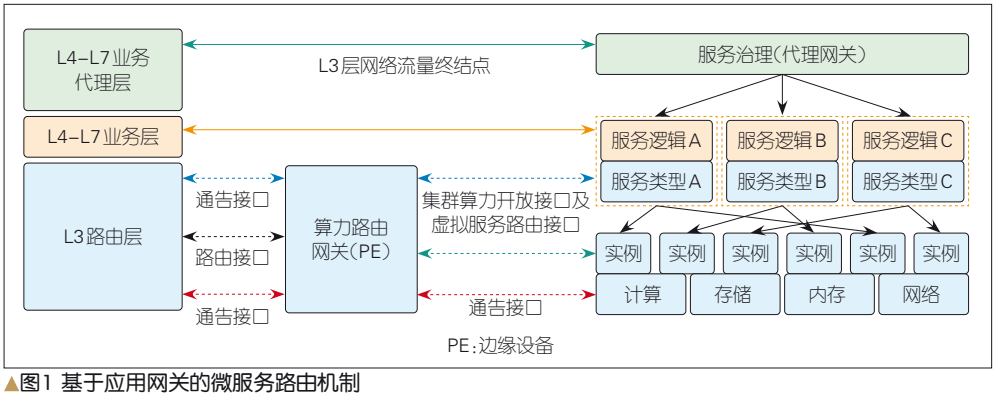
端到端微服务路由被应用网关分隔为独立的两段：一段为终端到应用网关的L3层路由，一段为应用网关到微服务实例的局部服务路由。其中，后者往往是L7层流量路由。如图1所示，应用代理网关终结L3层流量路由，即全网算力路由的最小颗粒度将被作为应用，或者微服务集群被作为应用单位。集群内的微服务仅限于局部算力调度和路由，无法执行全局算力资源协同。虽然如此，微服务集群内的基础算力资源、微服务种类及其实例状态，仍然有可能被外部网络感知。外部网络基于这类算力资源状态执行跨微服务集群路由。网络虽然可感知微服务并基于动态状态执行微服务集群路由，但是无法执行微服务的调度和路由本身。当然，这类粗颗粒度的微服务集群间路由机制也可以在集群资源调度和管理中心完成，并由后者感知和维护微服务集群内的算力服务和资源状态，从而实现基于L7层的端到端微服务路由。

1.2 算力路由面临的跨池虚机组网和寻址机制挑战

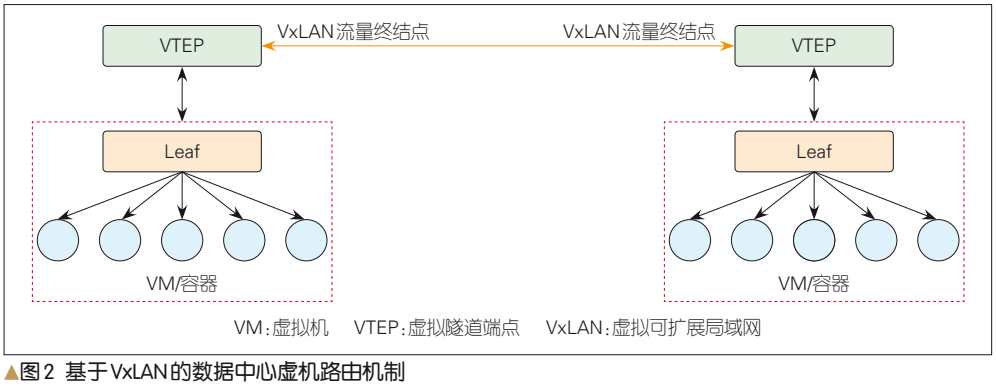
相对于虚拟局域网（VLAN），虚拟可扩展局域网（VxLAN）拥有更加庞大的寻址空间和更加灵活的组网机制，已经成为数据中心内部及数据中心之间虚机组网和寻址的主流机制。VxLAN是基于用户数据报协议（UDP）的L2层模拟网络技术。尤其是在跨数据中心虚拟网络组网场景中，虚拟隧道端点（VTEP）作为数据中心虚机集群代理，在UDP层即L4层终结了L3层网络路由流量。如1.1节所述，在跨数据中心虚机集群之间的微服务路由场景下，VTEP同样将端到端L3层微服务路由分隔成内外相互独立的两段：一段为数据中心之间的L3层外网路由，一段为数据中心内的L3层内网路由。因此，L3层端到端算力调度和路由面临着又一个现网部署的挑战。如图2所示，跨数据中心的微服务调度和路由终结于VTEP。基于L3层网络之上的跨数据中心虚机、容器及微服务集群资源感知、调度和路由虽然不失为一种可行的方案，但是缺少了与网络深度融合的算网一体调度和路由的综合优势[3]。

2 分布式微服务治理架构下的IPv6段路由（SRv6）算网端到端路由方案

在微服务架构下，各个微服务节点涉及服务治理的基础



▲图1 基于应用网关的微服务路由机制



▲图2 基于VxLAN的数据中心虚机路由机制

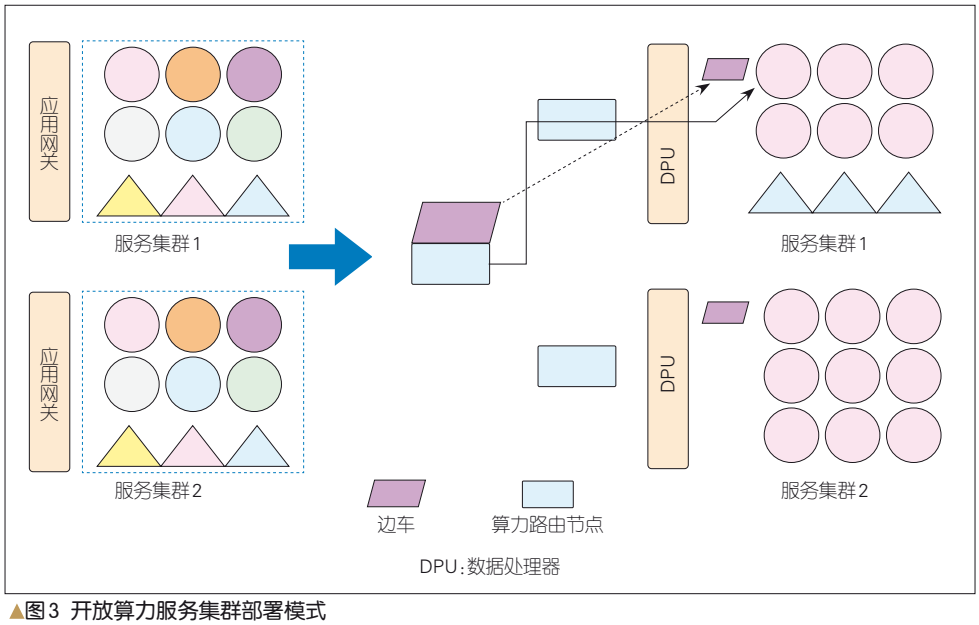
功能模块，如通信、安全、服务熔断、负载均衡等。这些模块往往被解耦成单独的模块并作为统一代理，以执行相应的功能[4]。目前行业内有两种主流的公共微服务治理模块部署模式，即应用网关模式和边车模式。其中，应用网关部署场景已在1.1节中阐述，它是一种集中式的代理入口模式；边车则是分布式模式，与微服务同节点部署。虽然如此，边车模式并不意味着微服务本身可被端到端L3层网络路由。在微服务集群的前端，系统往往通过统一的负载均衡接口对多个微服务集群执行应用代理接入。但是在分布式边车模式下，服务治理代理模块可分别在微服务的远端和近端部署。这为微服务作为一种公共算力资源对外开放和路由提供了技术支持。

2.1 基于微服务路由的服务集群部署模式变迁方案

如前所述，在当前的微服务集群治理模式下，各个集群的服务种类和服务能力是等价均衡的。这就相当于多个对等的应用服务实例以多个对等的微服务集群进行部署，从而实现资源的均衡利用。如1.1节所述，对于这种算力资源的部署模式，从全网算力资源协同和调度的视角看，应用为最小颗粒度，这跟算力网络的目标架构相比还存在较大的差距。从算力资源的部署和交付时间来看，这也是非常粗放和低效的一种模式。算力资源的部署，应根据用户发起请求的动态位置、节点的基础算力资源种类与能力，以及节点所在区域的业务需求等多因素，进行灵活编排，并据此执行最优的算网服务策略，在同等资源约束下为用户交付最优的算力服务质量。同时，算网资源的使用效率也会得到极大提升。如图3所示，对于不同的算力服务集群，服务部署的维度不再以特定应用为聚合颗粒度，而是根据不同的特征原则进行集群部署。比如，GPU算力池将部署图像处理算力服务，靠近工业控制现场的算力节点将部署工业数据采集和控制类算力服务。服务集群按照服务本身的应用场景和需求进行部署，而不以应用为颗粒度组织集群。在这种架构下，服务集群不再需要应用网关的统一入口逻辑功能，服务将可以被外部用户直接调度和路由。服务本身将成为一种公共服务算力资源颗粒度。这是与当前行业中微服务架构最根本的区别，也是全网算力网络的架构特征和关键内涵。特别地，在无需统一的对外网关接口的全新场景下，结合当前CPU算力卸载到专用加速硬件的最新发展动态，外部L3层路由流量直接对口服务集群的DPU模块，并通过DPU模块无缝路由至集群中的算力服务。

在图3服务集群部署模式中，边车可以进行远端部署。比如，在靠近算力服务请求方的网络边缘或入口处，在代理远端服务执行L7层的服务治理功能后，L3层将执行端到端算力服务路由，实现全网异构，以及跨池算力资源的灵活智能协同和调度。由于边车在当前微服务架构中主要负责执行微服务集群中的东西向微服务流量通信，在其实际的功能清单中，涉及集群以及微服务本地状态的一部分功能，并不适合直接从微服务集群中全部迁移出并在远端部署，比如鉴权、业务熔断等。因此，在这种微服务公共治理代理远端部署模式中，远端和本地模块会同时存在、同时部署，并形成互相补充和联动的关系，二者配合完成泛在微服务的端到端公共功能和治理。另外，在算力网络整体架构中，算网大脑也将执行一部分微服务治理的功能。比如，微服务提供方通过算网大脑注册本地服务种类、虚机及容器实例资源状态、微服务本身的认证等，微服务使用方则通过算网大脑完成接入认证、服务熔断等。

总的来说，在算力网络的目标架构下，算力服务的部署将呈现真正的泛在、异构、多样和层次化颗粒度的特征。通过网络统一调度和路由，算力和网络资源将成为一种深度融合、动态联动的公共基础能力，将为千行百业的业务应用提供高效、便捷、优质的算网服务。



▲图3 开放算力服务集群部署模式

2.2 微服务架构下的SRv6算网路由

基于SRv6技术拉通网络和云内业务的端到端路由，是近年来行业研究和实践的重点方向。SRv6基于网络和业务灵活可编程的功能特征[5]，同样也为算力网络提供一种优质的支撑技术。算网一体编排的核心要素是算力、网络资源和策略在统一的转发平面执行。这就意味着，SRv6的编程功能由网络向云池内的算力服务做深度延伸。如2.1节所述，算力网络架构下的云内微服务集群部署模式将使得集群内的算力服务全网路由可达。因此，算力服务将被作为SRv6端到端路由中的一个段路由，并被编排到统一的算网路由策略中。特别地，在应用需要多个集群内或集群之间的算力服务按照一定的时序组合完成服务的场景下，SRv6照样可以进行业务功能链的路由编排和策略执行。

2.3 SRv6算力路由在微服务架构下的终结模式

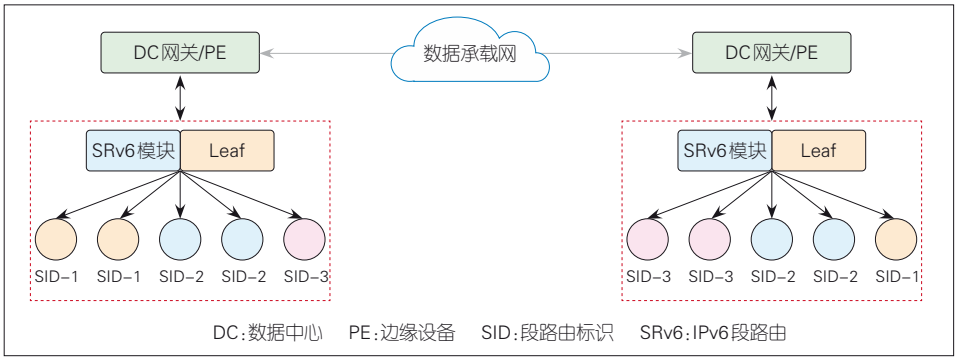
从算网端到端路由的全景视角看，云池外网络和云池内网络大多是基于两套架构、两套体系，甚至两套协议的异构网络的。近年来，两者呈现出互相渗透、互相影响的趋势。一方面，云池外网基于IPv6技术逐渐向云池内延伸；另一方面，云池内组网技术逐渐成熟，且更新迭代的周期快于外部网络，这些技术开始向外网渗透，如脊-叶（Spine-Leaf）组网架构[6]。具体到本文所述的SRv6算力路由[7]，SRv6逐渐深入到云池内网，如微服务集群内，为端到端网络+算力的综合路由提供网络基础设施条件。SRv6算力路由流量在云池内微服务架构下有两种终结模式：

• SRv6路由流量终结于数据中心（DC）网关，即SRv6轻度入云。在这种模式下，路由策略无法纳入云内业务，SRv6路由仅涉及网络侧端到端连接隧道。

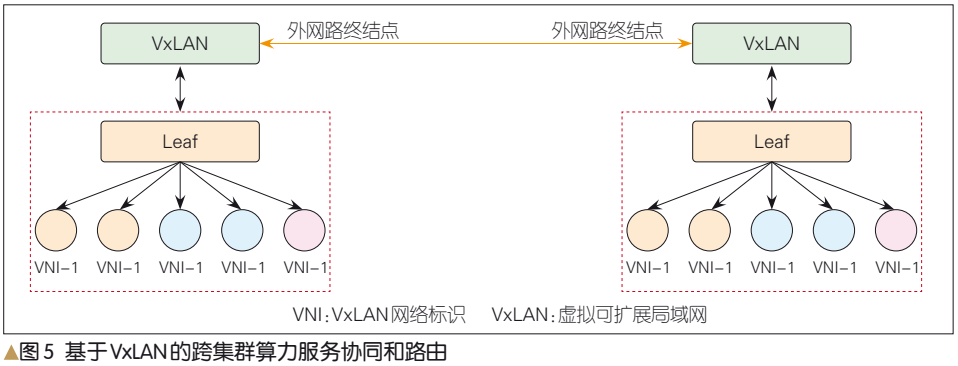
• SRv6路由流量终结于云池内Leaf节点的SRv6服务链（SFC）模块，即SRv6将深度入云，如图4所示。在这种模式下，SRv6端到端路由将执行策略编排并将自身纳入云池内的业务功能，以形成算网一体路由。但是，集群内被编排的业务仅作为一个服务节点被代理访问，并非一个独立的段路由。业务功能本身作为一个段路由被纳入算力路由策略，是最完备的SRv6算力路由场景。在该场景下，Leaf节点的SRv6模块有可能被跳过。但在实际部署中，考虑到成本和业务功能的复杂度，这种完备的SRv6算力路由方案并非最优选项。

3 VxLAN架构下跨微服务集群组网和寻址方案

VxLAN是当前云池内虚机及容器组网和寻址的主流方式。微服务集群之间的L3层寻址流量终结于VTEP，微服务集群内以及跨集群多种微服务之间的协同处理和路由面临巨大障碍。即便如此，VxLAN的网络标识空间仍非常庞大，并且部署模式也非常灵活。因此，跨集群的微服务协同和路由场景，可根据应用需求将关联微服务进行虚拟组网，即赋予同样的VxLAN网络标识（VNI），从而完成一组微服务跨集群的协同和组网路由。如图5所示，分布在两个集群资源池内的3种不同服务被赋予同样的VNI，不同服务之间在一个VNI标识的虚拟二层网络内灵活地完成数据协同处理。这种模式重用VxLAN的底层组网机制，虽然在算力服务方面实现了服务之间的灵活路由，但是在网络侧（尤其是云池外网侧），VxLAN是L3层路由之上的隧道通路，同时网络本身未被纳入端到端算力路由，应用在网络维度的SLA（服务等级协议）需求无法体现在路由策略中。这是这种方案的不足之处。当然，算力路由的主流场景应该是单算力服务的路由寻址。跨集群场景下的多服务协同组网路由，仍然可能通过SFC机制实现算网统一路由编排。



▲图4 SRv6入云模式下的算力路由



▲图5 基于VxLAN的跨集群算力服务协同和路由

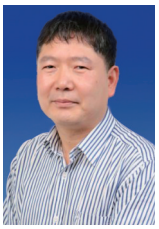
4 结束语

在算力网络架构下，端、边、云的全颗粒度算力成为全网可见、可调度、可路由的资源。网络由传统的拓扑路由进一步转变为算力路由，从而使能全新的网络架构和业务部署及交付模式，助力全行业数字化转型。其中，云内算力资源由当前的封闭模式转变为算力网络架构下的开放模式，对网络路由的颗粒度提出全新要求，在已经趋于成熟稳定的微服务架构下，增强SRv6、VxLAN等现网技术，并提供端到端算力路由解决方案，成为行业的一种优选路线。本文结合微服务架构及其部署和交付模式，对L3算网路由技术方案进行多维度的分析和探讨，为算力网络架构下端到端算力路由方案提供有益参考。

参考文献

[1] 李少鹤, 李泰新, 周旭. 算力网络: 以网络为中心的融合资源供给 [J]. 中兴通讯技术, 2021, 27(3): 29–34. DOI:10.12142/ZTETJ.202103007
[2] 兰巨龙,胡宇翔,张震, 等. 未来网络体系与核心技术 [M]. 北京: 人民邮电出版社, 2017
[3] SCHOLL B, SWANSON T, JAUSOVEC P. 云原生: 运用容器、函数计算和数据构建下一代应用 [M]. 北京: 机械工业出版社, 2020
[4] RICHARDSON C. 微服务架构设计模式 [M]. 北京: 机械工业出版社, 2019
[5] 李铭轩, 曹畅, 杨建军. 基于可编程网络的算力调度机制研究 [J]. 中兴通讯技术, 2021, 27(3): 18–22. DOI:10.12142/ZTETJ.202103005
[6] 魏月华,陈晓,张征. 数据中心网络架构和协议演进分析 [J]. 中兴通讯技术, 2021, 27(3): 51–55. DOI:10.12142/ZTETJ.202103011
[7] 黄光平, 史伟强, 谭斌. 基于SRv6的算力网络资源和服务编排调度 [J]. 中兴通讯技术, 2021, 27(3): 23–28. DOI:10.12142/ZTETJ.202103006

作 者 简 介



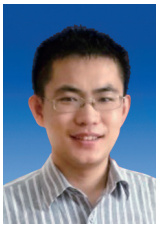
陈晓，中兴通讯股份有限公司有线架构部部长；长期从事电信产品和相关技术的研究规划。

黄光平，中兴通讯股份有限公司资深架构师；主要研究方向为下一代IP网络架构及关键技术，先后从事增值业务消息系统设计和开发、确定性网络以及远程宽带接入网关全球标准工作，近年聚焦算力网络架构、路由协议、算力标识等技术研究；发表论文6篇，申请专利20余项。

➡上接第20页

ACM, 2013: 127–132. DOI: 10.1145/2491185.2491190
[5] BOSSHART P, DALY D, GIBB G, et al. P4: programming protocol–independent packet processors [J]. ACM SIGCOMM computer communication review, 2014, 44(3): 87–95. DOI: 10.1145/2656877.2656890
[6] P4 Language Consortium. P4 [EB/OL]. (2003–07–31) [2021–12–15]. https://p4.org/
[7] WANG S H, MENG Z L, SUN C, et al. SmartChain: enabling high–performance service chain partition between SmartNIC and CPU [C]// Proceedings of ICC 2020 – 2020 IEEE International Conference on Communications. IEEE, 2020: 1–7. DOI: 10.1109/ICC40277.2020.9149136
[8] SCANO D, GIORGETTI A, SGAMBELLURI A, et al. Hierarchical control of SONiC–based packet–optical nodes encompassing coherent pluggable modules [C]//2021 European Conference on Optical Communication (ECOC). USA: IEEE, 2021: 1–3. DOI: 10.1109/ECOC52684.2021.9605850
[9] CONNOR O B, GHAFFARKHAH A, PUDELKO M, et al. Enabling the era of next generation SDN [EB/OL]. [2021–12–10]. https://opennetworking. org/stratum
[10] SANTIAGO DA SILVA J, STIMPFLING T, LUINAUD T, et al. One for all, all for one: a heterogeneous data plane for flexible P4 processing [C]// Proceedings of 2018 IEEE 26th international conference on network protocols. IEEE, 2018: 440–441. DOI: 10.1109/ICNP.2018.00063
[11] AGRAWAL A, KIM C. Intel Tofino2 – A 12.9 Tbps P4–programmable ethernet switch [C]//Proceedings of 2020 IEEE Hot Chips 32 Symposium (HCS). IEEE, 2020: 18–22. DOI: 10.1109/hcs49909.2020.9220636
[12] LUINAUD T, SANTIAGO DA SILVA J, LANGLOIS J M P, et al. Design principles for packet deparsers on FPGAs [C]//Proceedings of 2021 ACM/SIGDA International Symposium on Field–Programmable Gate Arrays. ACM, 2021: 280 – 286. DOI: 10.1145/3431920.3439303
[13] CAO Z, SU H Y, YANG Q M, et al. P4 to FPGA–A fast approach for generating efficient network processors [J]. IEEE access, 2020, 8: 23440–23456. DOI: 10.1109/ACCESS.2020.2970683
[14] LAKI S, HORPÁCSI D, VÖRÖS P, et al. High speed packet forwarding compiled from protocol independent data plane specifications [C]// Proceedings of the 2016 ACM SIGCOMM conference. ACM, 2016: 629–630. DOI: 10.1145/2934872.2959080

作 者 简 介



董永吉，解放军战略支援部队信息工程大学副研究员；长期从事路由与交换技术、网络安全和新型网络体系结构方面的研究工作；先后主持了1项国家重点研发课题，参与多项国家重点研发计划、“863”“973”项目，获得3项科研成果奖；发表论文10余篇，申请国家发明专利14项，出版专著2部。

胡宇翔，解放军战略支援部队信息工程大学教授、博士生导师；主要研究方向为新型网络体系结构、路由与交换技术。

崔鹏帅（通信作者），解放军战略支援部队信息工程大学副研究员；主要研究方向为新型网络体系结构、可编程数据平面。

《中兴通讯技术》杂志（双月刊）投稿须知

一、杂志定位

《中兴通讯技术》杂志为通信技术类学术期刊。通过介绍、探讨通信热点技术，以展现通信技术最新发展动态，并促进产学研合作，发掘和培养优秀人才，为振兴民族通信产业做贡献。

二、稿件基本要求

1. 投稿约定

（1）作者需登录《中兴通讯技术》投稿平台：tech.zte.com.cn/submission，并上传稿件。第一次投稿需完成新用户注册。

（2）编辑部将按照审稿流程聘请专家审稿，并根据审稿意见，公平、公正地录用稿件。审稿过程需要 1 个月左右。

2. 内容和格式要求

（1）稿件须具有创新性、学术性、规范性和可读性。

（2）稿件需采用 WORD 文档格式。

（3）稿件篇幅一般不超过 6 000 字（包括文、图），内容包括：中、英文题名，作者姓名及汉语拼音，作者中、英文单位，中文摘要、关键词（3 ~ 8 个），英文摘要、关键词，正文，参考文献，作者简介。

（4）中文题名一般不超过 20 个汉字，中、英文题名含义应一致。

（5）摘要尽量写成报道性摘要，包括研究的目的、方法、结果 / 结论，以 150 ~ 200 字为宜。摘要应具有独立性和自明性。中英文摘要应一致。

（6）文稿中的量和单位应符合国家标准。外文字母的正斜体、大小写等须写清楚，上下角的字母、数据和符号的位置皆应明显区别。

（7）图、表力求少而精（以 8 幅为上限），应随文出现，切忌与文字重复。图、表应保持自明性，图中缩略词和英文均要在图中加中文解释。表应采用三线表，表中缩略词和英文均要在表内加中文解释。

（8）所有文献必须在正文中引用，文献序号按其在文中出现的先后次序编排。常用参考文献的书写格式为：

- 期刊 [序号] 作者 . 题名 [J]. 刊名 , 出版年 , 卷号 (期号): 引文页码 . 数字对象唯一标识符
- 书籍 [序号] 作者 . 书名 [M]. 出版地 : 出版者 , 出版年 : 引文页码 . 数字对象唯一标识符
- 论文集中析出文献 [序号] 作者 . 题名 [C]// 论文集编者 . 论文集名 (会议名). 出版地：出版者 , 出版年 (开会年): 引文页码 . 数字对象唯一标识符
- 学位论文 [序号] 作者 . 题名 [D]. 学位授予单位所在城市名：学位授予单位，授予年份 . 数字对象唯一标识符
- 专利 [序号] 专利所有者 . 专利题名 : 专利号 [P]. 出版日期 . 数字对象唯一标识符
- 国际、国家标准 [序号] 标准名称 : 标准编号 [S]. 出版地 : 出版者 , 出版年 . 数字对象唯一标识符

（9）作者超过 3 人时，可以感谢形式在文中提及。作者简介包括：姓名、工作单位、职务或职称、学历、毕业于何校、现从事的工作、专业特长、科研成果、已发表的论文数量等。

（10）提供正面、免冠、彩色标准照片一张，最好采用 JPG 格式（文件大小超过 100 kB）。

（11）应标注出研究课题的资助基金或资助项目名称及编号。

（12）提供联系方式，如：通讯地址、电话（含手机）、Email 等。

3. 其他事项

（1）请勿一稿多投。凡在 2 个月（自来稿之日算起）以内未接到录用通知者，可致电编辑部询问。

（2）为了促进信息传播，加强学术交流，在论文发表后，本刊享有文章的转摘权（包括英文版、电子版、网络版）。作者获得的稿费包括转摘酬金。如作者不同意转摘，请在投稿时说明。

（3）编辑部地址：安徽省合肥市金寨路 329 号凯旋大厦 1201 室，邮政编码：230061。

（4）联系电话：0551-65533356，联系邮箱 : magazine@zte.com.cn。

（5）本刊只接受在线投稿，欢迎访问本刊投稿平台：tech.zte.com.cn/submission。

中兴通讯技术

(ZHONGXING TONGXUN JISHU)

办刊宗旨：

以人为本，荟萃通信技术领域精英
迎接挑战，把握世界通信技术动态
立即行动，求解通信发展疑难课题
励精图治，促进民族信息产业崛起

双月刊　1995 年创刊　总第 162 期
2022 年 2 月　第 28 卷 第 1 期

主管：安徽出版集团有限责任公司
主办：时代出版传媒股份有限公司
　　深圳航天广宇工业有限公司
出版：安徽科学技术出版社
编辑、发行：中兴通讯技术杂志社

总编辑：王喜瑜
主编：蒋贤骏
执行主编：黄新明
编辑部主任：卢丹
责任编辑：徐烨
编辑：杨广西、卢丹、朱莉、任溪溪
设计排版：徐莹
发行：王萍萍
编务：王坤

《中兴通讯技术》编辑部
地址：合肥市金寨路 329 号凯旋大厦 1201 室
邮编：230061
网址：tech.zte.com.cn
投稿平台：tech.zte.com.cn/submission
电子信箱：magazine@zte.com.cn
电话：(0551)65533356

发行方式：自办发行
印刷：合肥添彩包装有限公司
出版日期：2022 年 2 月 25 日
中国标准连续出版物号：ISSN 1009-6868
CN 34-1228/TN
定价：每册 20.00 元