



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>

2021年6月·第3期

专题：边缘计算与算力网络



《中兴通讯技术》第8届编辑委员会成员名单

顾问 侯为贵(中兴通讯股份有限公司创始人) 钟义信(北京邮电大学教授) 陈锡生(南京邮电大学教授)

主任 陆建华(中国科学院院士)

副主任 李自学(中兴通讯股份有限公司董事长) 糜正琨(南京邮电大学教授)

编委(按姓名拼音排序)

陈建平 上海交通大学教授

陈前斌 重庆邮电大学教授、副校长

葛建华 西安电子科技大学教授

管海兵 上海交通大学教授

郭庆 哈尔滨工业大学教授

洪波 中兴发展股份有限公司总裁

洪伟 东南大学教授

黄宇红 中国移动研究院副院长

纪越峰 北京邮电大学教授

江涛 华中科技大学教授

蒋林涛 中国信息通信研究院科技委主任

李尔平 浙江大学教授

李红滨 北京大学教授

李厚强 中国科学技术大学教授

李建东 西安电子科技大学教授

李军 清华大学教授

李乐民 中国工程院院士

李融林 华南理工大学教授

李少谦 电子科技大学教授

李自学 中兴通讯股份有限公司董事长

林晓东 中兴通讯股份有限公司副总裁

刘健 中兴通讯股份有限公司高级副总裁

刘建伟 北京航空航天大学教授

陆建华 中国科学院院士

马建国 广东工业大学教授

孟洛明 北京邮电大学教授

糜正琨 南京邮电大学教授

任品毅 西安交通大学教授

石光明 西安电子科技大学教授、副校长

孙知信 南京邮电大学教授

谈振辉 北京交通大学教授、原校长

唐雄燕 中国联通研究院副院长

陶小峰 北京邮电大学教授

王文博 北京邮电大学教授、副校长

王文东 北京邮电大学教授

王喜瑜 中兴通讯股份有限公司执行副总裁

王翔 中兴通讯股份有限公司高级副总裁

卫国 中国科学技术大学教授

吴春明 浙江大学教授

邬贺铨 中国工程院院士

肖甫 南京邮电大学教授

解冲锋 中国电信研究院教授级高工

徐安士 北京大学教授

徐子阳 中兴通讯股份有限公司总裁

续合元 中国信息通信研究院副总工

薛向阳 复旦大学教授

薛一波 清华大学教授

杨义先 北京邮电大学教授

叶茂 电子科技大学教授

易芝玲 中国移动研究院首席科学家

张宏科 北京交通大学教授

张平 中国工程院院士

张卫 复旦大学教授

张云勇 中国联通集团产品中心总经理

赵慧玲 工业和信息化部通信科技委信息通信网络专家组组长

郑纬民 中国工程院院士

钟章队 北京交通大学教授

周亮 南京邮电大学教授

朱近康 中国科学技术大学教授

祝宁华 中国科学院半导体研究所研究员

目次

中兴通讯技术 (ZHONGXING TONGXUN JISHU)

总第 158 期 第 27 卷 第 3 期 2021 年 6 月

专题：边缘计算与算力网络

专题导读 01
赵慧玲

边缘计算与算力网络综述 03
雷波, 赵倩颖, 赵慧玲

算力感知网络架构与关键技术 07
姚惠娟, 陆璐, 段晓东

算力网络实现一体化服务的探索与实践 12
雷波, 赵倩颖, 凌泽军

基于可编程网络的算力调度机制研究 18
李铭轩, 曹畅, 杨建军

基于 SRv6 的算力网络资源和服务编排调度 23
黄光平, 史伟强, 谭斌

算力网络：以网络为中心的融合资源供给 29
李少鹤, 李泰新, 周旭

多层次算力网络集中式不可分割任务调度算法 35
巩宸宇, 舒洪峰, 张昕

专家论坛

42 夯实云网融合，迈向算网一体
唐雄燕, 张帅, 曹畅

47 零触碰与零信任
李军, 胡效赫

企业视界

51 数据中心网络架构和底层协议演进
魏月华, 陈晓, 张征

技术广角

56 共建共享下的边缘云建设
黄倩, 黄蓉

62 边缘计算使能星地协同网络下的服务部署机制
卢华, 段雪飞, 李斌

2021 年第 1—6 期专题计划及策划人

1. 视频技术和用户体验评测
华中科技大学教授 江涛
中兴通讯股份有限公司副总裁 陆平

2. 6G 愿景及技术挑战
中国工程院院士 张平
北京邮电大学教授 张建华

3. 边缘计算与算力网络
工信部通信科技委信息通信网络
专家组组长 赵慧玲

4. 高铁智能通信技术与应用
北京交通大学教授 艾渤

5. 低轨卫星通信技术与应用
哈尔滨工业大学教授 郭庆

6. 触觉通信技术
南京邮电大学教授 周亮

MAIN CONTENTS

ZTE TECHNOLOGY JOURNAL Vol. 27 No. 3 Jun. 2021

Special Topic: Edge Computing and Computing Power Network

Editorial **01**
ZHAO Huiling

Overview of Edge Computing and Computing Power Network **03**
LEI Bo, ZHAO Qianying, ZHAO Huiling

Architecture and Key Technologies for Computing-Aware Networking **07**
YAO Huijuan, LU Lu, DUAN Xiaodong

Exploration and Practice to Realize Service Integration in Computing Power Network **12**
LEI Bo, ZHAO Qianying, LING Zejun

Computing Power Scheduling Mechanism Based on Programmable Network **18**
LI Mingxuan, CAO Chang, YANG Jianjun

Computing Power Network Resources Based on SRv6 and Its Service Arrangement and Scheduling **23**
HUANG Guangping, SHI Weiqiang, TAN Bin

Computing Power Network: A Network-Centric Supply Paradigm for Integrated Resources **29**
LI Shaohe, LI Taixin, ZHOU Xu

35 Centralized Unsplittable Task Scheduling Algorithm for Multi-Tier Computing Power Network
GONG Chenyu, SHU Hongfeng, ZHANG Xin

Expert Forum

42 Tamp Cloud and Network Integration, Step into Computing Power Network
TANG Xiongyan, ZHANG Shuai, CAO Chang

47 Zero Touch and Zero Trust
LI Jun, HU Xiaohe

Enterprise View

51 Data Center Infrastructure and Underlay Protocol Evolution
WEI Yuehua, CHEN Xiao, ZHANG Zheng

Technology Perspective

56 Edge Cloud Construction under Co-Construction and Sharing
HUANG Qian, HUANG Rong

62 Service Deployment Mechanism in Edge Computing Enabled Satellite Terrestrial Integrated Network
LU Hua, DUAN Xuefei, LI Bin

期刊基本参数: CN 34-1228/TN*1995*b*16*66*zh*P*¥20.00*6500*13*2021-06

敬告读者

本刊享有所发表文章的版权, 包括英文版、电子版、网络版和优先数字出版版权, 所支付的稿酬已经包含上述各版本的费用。未经本刊许可, 不得以任何形式全文转载本刊内容; 如部分引用本刊内容, 须注明该内容出自本刊。



边缘计算与算力网络专题导读

专题策划人



赵慧玲

工信部通信科技委专职常委、信息通信网络专家组组长，中国通信学会理事，中国通信学会信息通信网络技术专业委员会主任委员，中国通信标准协会网络与业务能力技术工作委员会主席，中国电信科技委常委，SDN、NFV、AI产业联盟技术委员会副主任，网络 5.0 产业联盟技术委员会副主任；长期从事电信网络领域技术和标准工作；曾获多个国家及省部级科技进步奖项；发表技术文章百余篇，出版技术专著 12 部。

5G 移动边缘计算（MEC）的快速部署，边缘计算、人工智能（AI）等技术的迅速发展，以及虚拟现实（VR）/增强现实（AR）、云游戏等新型业务的不断涌现，都需要大量的算力资源。通过无处不在的网络连接，将多级算力资源进行整合，能够实现云、边、网和端的高效协同，提高算力资源利用效率，进而实现服务灵活动态部署和用户体验的一致性。多样性算力是云网融合的必然要求，也是推动产业数字化多样性的关键一环。在中国大力推动算力资源服务化的背景下，算力网络已经成为产业界和学术界的研究热点。什么是算力网络？算力网络的研究进展如何？其关键技术和标准化进程如何？针对这些问题，本期专题的论文将从不同角度论述算力网络的研究进展及相关成果，期望能对读者有所帮助。

《边缘计算与算力网络综述》一文对目前最热门的两项技术——边缘计算与算力网络的定义和产业发展等做了综合性介绍，并特别给出了通信领域边缘计算和算力网络的全球标准制定情况及最新进展。文章指出近期边缘计算和算力网络面临的主要技术挑战，让读者对其发展及研究进展有了宏观的认识。边缘计算是 5G 时代的一项重要技术，它不仅满足了新兴业务的低时延内在需求，还给现有网络带来了新需求及挑战，并催生了算力网络技术。作为 5G/B5G 时代信息通信技术（ICT）融合的两项重要技术，边缘计算和算力网络将成为驱动各行各业变革的重要技术途径。

《算力感知网络架构与关键技术》一文提出了算力感

知网络的概念，即网络可感知应用、算力和用户需求等多维资源，并协同调度算力资源和网络资源，使应用能够按需、实时调用相关的计算资源，实现边缘计算与云计算的协同联动，以提供最优的用户体验及计算和网络资源利用率。文章指出了算力网络的关键技术，包括算力度量和建模、算力路由和算力管理等，并论述了这些技术的研究进展。此外，文章还介绍了算力感知网络的部署案例，并给出了相关技术的试验结果。

《算力网络实现一体化服务的探索与实践》一文介绍了算力网络的最新探索与实践。文章从算力资源的定义及特点出发，分析并指出一体化算力资源服务的需求。基于此需求，文章提出了算力网络交易平台的基本架构，并对资源交易视图生成模型及交易系统功能模块进行了详细阐述，最后结合基于 AI 的游戏场景，对交易平台系统进行了试验验证。算力网络交易平台是算力网络助力“新基建”、推动算力资源一体化服务的关键组成部分，它可以把融合的多维资源智能化、可视化地提供给用户，并形成统一的资源供给机制，以满足各类新兴业务的多样化算力需求。

《基于可编程网络的算力调度机制研究》一文介绍了可编程网络的概念和技术架构，提出了基于可编程网络的算力调度机制和技术方案。该技术基于云原生来实现算力网络的融合调度，可以根据网络情况进行算力调度，也可以基于算力调度需求进行网络适配和可编程。该技术为实现云网融合进行了有益的技术探索。

《基于 SRv6 的算力网络资源和服务编排调度》一文介绍了基于 SRv6 的算力传送及调度的研究思路及进展。文章指出，算网一体的编排和路由是算力网络的核心特征之一。

该文分别论述了算力网络的控制面和数据转发面技术，对算力网络新型路由协议的技术功能进行了分析，并提出了一种基于聚合原则的分级分层路由表机制，在算力网络资源调度方面进行了有价值的研究和探索。

《算力网络：以网络为中心的融合资源供给》一文梳理了网络计算模型的发展历程及网络功能的变迁，介绍了算力网络的需求背景及以网络为中心的核心特征，阐述了算力网络的服务供给模式，指出算力基础设施服务形态、算力平台服务形态和算力软件服务形态是算力网络的三层服务形态，并对算力网络的发展现状以及未来研究进行了全面展望。

《多层次算力网络集中式不可分割任务调度算法》一

文提出了一种多层次算力网络模型和计算卸载系统，定义了一个由时延、能耗组成的加权代价函数，建模了一个任务调度问题，并进行了仿真实验。相关数值仿真结果表明，算力网络可以有效解决单层网络带来的算力小或时延大的问题。

算力网络目前还处于研究阶段，还需要进行深入的技术探索和实践。本期论文汇聚了各位作者现阶段的研究思路及成果，希望能给读者带来有益的收获与参考。在此，对各位作者的积极支持和辛勤工作表示衷心的感谢！

赵慧玲

2021年5月20日

边缘计算与算力网络综述

Overview of Edge Computing and Computing Power Network



雷波 /LEI Bo¹, 赵倩颖 /ZHAO Qianying¹, 赵慧玲 /ZHAO Huiling²

(1. 中国电信股份有限公司研究院, 中国 北京 102209;

2. 工业和信息化部通信科学技术委员会, 中国 北京 100035)

(1. Research Institute of China Telecom Corporation, Beijing 102209, China;

2. Communications Science and Technology Commission of the Ministry of Industry and Information Technology of the People's Republic of China, Beijing 100035, China)

摘要: 作为 5G/B5G 时代信息通信技术 (ICT) 融合的两项重要技术, 边缘计算与算力网络是新型业务发展与落地的重要支撑。对边缘计算和算力网络的定义、发展、标准化现状进行综合性阐述, 特别介绍了通信领域边缘计算及算力网络的全球标准情况及最新进展。认为边缘计算和算力网络将成为驱动各行各业变革的重要解决方案。

关键词: 边缘计算; 算力网络; 一体化服务

Abstract: As the two important technologies for the integration of information and communication technology (ICT) in the 5G/B5G era, the edge computing and computing power network are the important support for the development and application of the new business. The definition, development, and standardization of edge computing and computing power network are comprehensively elaborated, and the global standard situation and recent progress of edge computing and computing power network in the communication field are especially given. It is believed that edge computing and computing power network will be important solutions to driving changes in all walks of life.

Keywords: edge computing; computing power network; integration service

DOI: 10.12142/ZTETJ.202103002

网络出版地址: <https://kns.cnki.net/kcms/detail/34.1228.TN.20210615.1321.006.html>

网络出版日期: 2021-06-15

收稿日期: 2021-05-10

在云计算时代, 业务通常把数据传输至集中的大型或超大型云计算中心来处理。在很长一段时间里, 云计算强大的存储和计算能力满足了传统业务的各项需求。但随着 5G 与人工智能时代的发展, 各类新型应用不断涌现, 数据产生量呈爆发式增长。这些对网络时延提出了极高要求, 同时对数据安全性、可控性也提出了差异化要求。为了应对这些需求, 边缘计算应运而生。

边缘计算的诞生, 一方面满足了新型业务低时延的需求, 解决了骨干

网络中大量数据所造成的拥堵问题; 另一方面导致多级计算节点遍布网络, 改变了网络的流量流向。如何实现多级资源节点的协同调度与应用的灵活部署, 从而为用户提供一致性服务体验也变得至关重要。因此, 算力网络应运而生。通过无处不在的网络, 算力网络将大量闲散的资源连接起来并进行统一管理和调度, 从而为用户提供统一的服务。

1 边缘计算的来源与定义

随着 5G 的发展, 边缘计算的热度

变得越来越高。边缘计算并非是 5G 时代的产物, 其概念的提出已有数十年, 并随着技术和业务的发展不断扩充。

边缘计算概念的由来可以追溯至 1998 年阿卡迈 (Akamai) 公司提出的内容分发网络 (CDN), 但“edge computing” (边缘计算) 这一名词的首次出现, 是在 2013 年美国太平洋西北国家实验室的一份由 Ryan LAMOTHE 撰写的内部报告中^[1]。

经过数十年的发展, 边缘计算已被 Gartner 评为 2020 年十大最热门技术趋势之一^[2], 但因其仍处于发展的

阶段,各个标准组织对其定义并不完全一致。以下为一些具有代表性的标准组织对边缘计算的定义:

- 2015年9月,欧洲电信标准化协会(ETSI)在发布的《Mobile Edge Computing: A Key Technology Towards 5G》中指出^[3]:移动边缘计算在距离用户移动终端最近的无线接入网(RAN)内提供信息技术(IT)服务环境以及云计算能力,旨在进一步减少延迟/时延,提高网络运营效率,提高业务分发/传送能力,优化/改善终端用户体验。

- 2016年11月,边缘计算产业联盟发布了《边缘计算产业联盟白皮书》,将边缘计算定义为^[4]:边缘计算是在靠近物或数据源头的网络边缘侧,融合网络、计算、存储、应用核心能力的分布式开放平台,就近提供边缘智能服务,满足行业数字化在敏捷连接、实时业务、数据优化、应用智能、安全与隐私保护等方面的关键需求。

- 2017年1月,第3代合作伙伴计划(3GPP)在技术规范(TS 23.501)中提到^[5]:为了降低端到端时延以及回传带宽,实现业务应用内容的高效分发,边缘计算需要为运营商以及第三方业务应用提供更靠近用户的部署及运营环境。

- 2020年2月,国际标准化组织(ISO)在ISO/国际电工委员会(IEC)技术报告(TR 23188)中提到^[6]:边缘计算是一种将主要处理和数据存储放在网络的边缘节点的分布式计算形式。

可以看出,目前边缘计算的概念虽未达成统一,但各方都认同边缘计算是在更靠近终端的网络边缘上提供服务的。

2 边缘计算产业发展情况

随着产业的发展,边缘计算逐步从产业共识走向应用落地。目前,业

界一般认为边缘计算可以分为3种主要的落地形态^[7-8]:

- 云边缘。云边缘形态的边缘计算是云服务在边缘侧的延伸。云边缘在逻辑上仍属于云服务,且主要的能力依赖于云服务或与云服务紧密协同。华为云智能边缘平台(IEF)解决方案、阿里云的Link Edge解决方案、AWS(亚马逊公司的云计算服务)的Greengrass解决方案等均属于云边缘的形态。

- 边缘云。边缘云形态的边缘计算是在边缘侧构建中小规模云服务能力。边缘服务能力主要由边缘云提供;边缘云管理调度能力主要由集中式数据中心(DC)侧的云服务提供。移动边缘计算(MEC)、CDN、车联网等均属于边缘云形态。

- 边缘网关。边缘网关形态的边缘计算是以云化技术与能力重构原有嵌入式网关系统的。边缘网关在边缘侧提供通信联接、协议/接口转换、边缘计算等能力,在云侧的控制器提供边缘节点的资源调度、应用管理与业务编排等能力。软件定义广域网(SD-WAN)、新一代家庭网关、新一代工业网关等均属于边缘网关形态。

不同类型的边缘计算形态,代表着产业界不同方的观点和利益。但总体上,各方都非常看好边缘计算产业发展态势。国际数据公司(IDC)发布的《中国半年度边缘计算服务器市场(2020上半年)跟踪报告》显示:2020年上半年,中国边缘计算服务器的整体市场规模为11.13亿美元(约合人民币72.78亿元),预计全年将达到27.82亿美元(约合人民币181.93亿元),同比增长20.6%;而2019—2024年,中国边缘计算服务器市场年复合增长率将达到18.8%,远高于核心数据中心的平均增速。美国通信产业研究机构(CIR)预测:到2025年,边缘计算基础设施收入将达到179亿

美元,用于支持边缘数据传输的光模块和网络投资将增加10亿美元。

目前,中国主流企业已在边缘计算领域开展了全方位的工作,并取得了不错的成绩。其中,具有代表性成果为:

- 2020年,中国电信研究院联合中国电信多家省级公司,先后完成了自研MEC系统与5G核心网(5GC)商用版本的对接与实验,成功验证了5G网络面向MEC多种商用场景的能力,对后续5G MEC系统规模商用具有重要意义。

- 2020年8月,中国信息通信研究院、中国移动、中国联通、华为、腾讯、紫金山实验室、九州云和安恒信息联合发布业界首个5G边缘计算开源平台——EdgeGallery。该平台打造了一个以“联接+计算”为特点的5G MEC公共平台,力图实现网络能力(尤其是5G网络)开放的标准化和MEC应用开发、测试、迁移和运行等生命周期流程的通用化。

- 2020年10月,腾讯云首个5G边缘计算中心对外开放。该边缘计算中心融合腾讯云在5G、边缘计算、物联网、安全等领域的多项前沿科技,成为独具创新性的一站式边缘计算产品^[9]。

3 边缘计算标准发展情况

目前,全球有多个标准组织正在进行边缘计算的标准化工作:

- 2014年12月,ETSI与24家公司成立了MEC行业规范组(ISG),率先开展了边缘计算的标准化研究工作。ETSI关于边缘计算的标准化工作主要分为两个阶段:移动边缘计算阶段和多接入边缘计算阶段。在第1阶段,ETSI以移动边缘计算为名开展研究工作;在第2阶段,ETSI则以多接入边缘计算为名开展研究工作。2017年3月,ETSI将移动边缘计算行业规

范工作组更名为多接入边缘计算工作组,将边缘计算从电信蜂窝网络进一步延伸至其他无线接入网络(如 Wi-Fi)^[10]。目前,ETSI 已经发布了关于边缘计算平台架构、边缘计算技术要求、边缘计算应用程序编程接口(API)准则、边缘计算应用程序(APP)使能、边缘云平台管理、基于网络功能虚拟化(NFV)的边缘云部署等标准 27 项,维护、更新标准共 41 版次。

- 2016 年 4 月,3GPP SA2 也正式接受 MEC,并将之列为 5G 架构的关键技术。从 R14 版本开始,3GPP 就开始定义边缘计算的网路基础能力。3GPP 关于边缘计算的研究主要针对如何将 MEC 融入 5G 架构。3GPP 在 TS 23.501 中将 MEC 纳入 5G 网络标准化中,并基于 3GPP TS 23.501(clause 5.13)定义的功能使能器,在《MEC in 5G Networks》白皮书中明确了如何部署 MEC 并将其无缝集成至 5G^[11]。

- 作为中国重要的标准化组织,中国通信标准化协会(CCSA)也将边缘计算作为重要的工作内容,分别在互联网与应用技术工作委员会(TC1)、网络与业务能力技术工作委员会(TC3)、无线通信技术委员会(TC5)、工业互联网特设任务组(ST8)设立了边缘计算相关项目,从不同角度对边缘计算技术进行标准化。其中,CCSA TC1 重点研究面向互联网的边缘云、边缘数据中心的标准化;CCSA TC3 重点研究面向边缘计算的 IP 承载网络、边缘计算网络、算力网络等;CCSA TC5 中的三大运营商分别在边缘计算领域立项,涉及边缘计算平台架构、场景需求、关键技术研究 and 总体技术要求;CCSA ST8 重点讨论面向工业互联网的边缘计算和边缘云的标准化内容^[12]。

4 算力网络及其标准化进展

作为解决多级算力资源(云计算、

边缘计算以及端计算)并存情况下资源统一供给问题的一种新型网络技术,算力网络通过网络控制面(如集中式控制器、分布式路由协议等)分发服务节点的算力、存储、算法等资源信息,并结合网络信息和用户需求,提供计算、存储、网络等资源的分发、关联、交易与调配,从而实现整网资源的最优化配置和使用^[13]。

算力网络的产生与边缘计算息息相关,它可以重点解决资源节点泛在化后的两个重要问题:用户体验一致性和服务灵活动态部署^[14]。首先,算力网络可以解决用户体验一致性的问题:用户无须关心各类基础资源(算力、存储等)的位置和部署状态,通过网络即可协同调度各类资源,保证用户的一致体验;其次,算力网络可以解决服务灵活动态部署的问题:基于用户的服务等级协议(SLA)需求,综合考虑实时的网络、算力、存储等多维资源状况,通过网络灵活匹配与动态调度,将业务流量动态调度至最优资源节点。

从 2019 年初至今,业界对算力网络的研究仅有两年的时间。算力网络巨大的潜在需求却掀起了业界的波澜。目前,三大运营商、各厂商以及学术机构纷纷开始研究算力网络。

2020 年 6 月,CCSA TC614 成立了算力网络特别工作组,依托联盟的平台和资源,联合多方力量,共推、共创算力网络产业影响力,构建算力网络生态圈。2020 年 11 月,中国联通成立了中国联通算力网络产业技术联盟,将在“联接+计算”领域和全产业链合作伙伴携手并进,共建算力网络生态,推动商业落地,共享转型成果。

中国主流运营商还先后发布了《中国联通算力网络白皮书》《算力感知网络技术白皮书》《算力网络架构与

技术体系白皮书》等。

在各方的不懈努力下,算力网络的标准化工作取得了进展:在 ITU-T、互联网工程任务组(IETF)、宽带论坛(BBF)、ETSI、CCSA 等全球标准组织中,已立项相关的国际标准 9 项、中国标准 4 项。

- 在 ITU-T SG13 组,中国电信、中国移动、中国联通、华为等单位分别从算力网络架构、算力感知网络相关技术等方面推进了 Y.CPN-arch 标准、Y.CAN 系列标准的制定。

- ITU-T SG11 组启动了 Q.CPN 标准(算力网络的信令需求)与 Q.BNG-INC 标准(算力网络边界网关的信令要求)的制定等工作。

- 在 IETF,华为等撰写了 Computing First Network 系列文稿,研究算力路由协议;

- BBF 启动了“Metro Computing Network(SD-466)”,专门研究算力网络在城域网中的应用。

- ETSI 提出了“NFV support for network function connectivity extensions(NFV-EVE020)”。该方案以内容转发网络(CFN)为基础,研究 NFV 的计算和网络集成相结合的网络功能连接扩展方案。

- CCSA TC3 目前已经完成《算力网络需求与架构》的研究报告和面向全网的算力感知网络关键技术研究。2021 年 4 月 TC3 全会形成了算力网络系列行业标准的立项,包括算力网络总体技术要求、算力网络标识解析技术要求、算力网络路由协议要求、算力网络控制器技术要求、算力网络交易平台技术要求和算力网络开放能力研究等工作。

5 边缘计算和算力网络的主要技术挑战及展望

边缘计算发展至今已取得巨大进

步,但仍面临诸多技术挑战,目前仍有三大问题亟待解决。首先是安全性的问题。边缘计算的分布式架构增加了攻击向量的维度,客户端越智能就越容易受到恶意软件感染和安全漏洞攻击。由于网络边缘设备的资源有限,现有数据安全的保护方法并不完全适用于边缘计算架构,因此需要寻求新的解决路径。其次是云边与边边协同的问题。单个节点能力是有限的,不同场景需要多资源节点能力的整合与联动。最后是网络问题。边缘计算所呈现的优势与底层的网络连接密不可分,例如,边缘计算所带来的低时延特性,如果没有网络的支持,是无法实现的。也就是说,边缘计算并不是简单地将服务器、存储设备放到边缘机房,而是需要对底层网络基础设施进行梳理,让用户能够享受到更短的接入距离所带来的优势,避免出现物理位置接近但逻辑距离绕行的尴尬场景。

从整个信息基础设施的角度来看,边缘计算的出现与部署,推动了网络、计算、存储等多类信息基础资源的融合与演进。也就是说,随着技术与业务的发展,各类网络资源需要与计算、存储能力进行深度融合,并借助数据资源和算力资源等形式对外输出,以实现多类资源的统一供给,实现信息网络基础设施能力的聚合和开放,为构架在网络基础之上的多行业、全产业创新提供便捷的条件。

为了进一步分析边缘计算对网络的影响,CCSA TC3 在 2021 年 1 月完成了《边缘计算 IP 承载网技术架构研究报告》,提出了以边缘计算为视角,将网络划分为边缘计算接入网(ECA)、边缘计算内部网络(ECN)和边缘计算互联网络(ECI),并以此重新梳理了各项新型网络技术的发展趋势。在这些新型网络技术中,有一项是被称为边缘计算原生的网络技术,即算力网络技术。

目前,算力网络的研究工作主要围绕 4 个方面展开:

(1) 算力度量。目前计算资源的衡量缺少一个统一且简单的度量单位,因此如何评估不同类型算力资源的大小成为一个亟需解决的难题。

(2) 信息分发。信息分发即如何将算力等资源信息通过网络控制面广而告之。

(3) 资源视图。如何给每个用户生成以其为中心的资源视图,让其可以智能选择最佳资源组合也是需要关注的内容。

(4) 可信交易。由于算力网络中的各类资源归属不同所有者,算力网络作为一个中间平台,需要考虑如何确保资源交易真实有效且可溯源。

6 结束语

边缘计算与算力网络的相关研究和标准化制定工作正在如火如荼地展开,并已取得初步成效。可以预期,未来边缘计算和算力网络将成为驱动各行各业变革的重要解决方案。

参考文献

- [1] 施巍松, 张星洲, 王一帆, 等. 边缘计算的发展路径[J]. 计算机研究与发展, 2019, 56(1), 69-89
- [2] Gartner. Top 10 strategic technology trends for 2020 [EB/OL]. (2019-10-12)[2021-04-22]. <https://www.gartner.com/smarterwithgartner/gartner-top-10-strategic-technology-trends-for-2020/>
- [3] ETSI. Mobile edge computing: a key technology towards 5G [R]. 2015
- [4] 边缘计算产业联盟. 边缘计算产业联盟白皮书[R]. 2016
- [5] 3GPP. System architecture for the 5G system (5GS): 3GPP TS 23.501.2017.01 [S]. 2017
- [6] ISO. 2020(en): information technology-cloud computing-edge computing landscape: ISO/IEC TR 23188 [S]. 2021
- [7] 边缘计算产业联盟(ECC). 网络 5.0 产业和技术创新联盟(N5A). 运营商边缘计算网络技术白皮书[R]. ECNI 工作组, 2019
- [8] 雷波, 宋军, 曹畅, 等. 边缘计算 2.0: 网络架构与技术体系[M]. 北京: 电子工业出版社, 2021
- [9] 边缘计算产业联盟. 边缘计算产业观察

[EB/OL]. (2021-01-10)[2021-04-22]. <http://www.econsortium.net/Uploads/file/20210224/1614140492544624.pdf>

- [10] ETSI. ETSI Multi-access edge computing starts second phase and renews leadership team [EB/OL]. (2017-03-28)[2021-04-22]. <https://www.etsi.org/newsroom/news/1180-2017-03-news-etsi-multi-access-edge-computing-starts-second-phase-and-renews-leadership-team>
- [11] KEKKI S, REANKIK A. 3GPP enables MEC over a 5G core [EB/OL]. (2018-07-04)[2021-04-22]. <https://www.3gpp.org/news-events/partners-news/1969-mec>
- [12] 吕华章, 陈丹. 边缘计算标准化进展与案例分析[J]. 计算机研究与发展, 2018, 55(3):487-511
- [13] 雷波, 陈运清. 边缘计算与算力网络——5G+AI 时代的新型算力平台与网络连接[M]. 北京: 电子工业出版社, 2020
- [14] 中国移动. 算力感知网络技术白皮书[R]. 2019

作者简介



雷波, 中国电信股份有限公司研究院高级工程师、边缘计算产业联盟 ECNI 工作组联席主席、CCSA“网络 5.0 技术标准推进委员会”管理与运营组组长; 主要研究方向为未来网络架构、新型 IP 网络技术等; 发表论文 10 余篇, 出版《边缘计算与算力网络》《边缘计算 2.0: 网络架构与技术体系》等书籍。



赵倩颖, 中国电信股份有限公司研究院工程师; 研究方向为未来网络、算力网络等; 发表论文 3 篇, 参与出版《边缘计算与算力网络》《边缘计算 2.0: 网络架构与技术体系》等书籍。



赵慧玲, 工信部通信科技委专职常委、信息通信网络专家组长, 中国通信学会理事、信息通信网络技术专业委员会主任委员, 中国通信标准协会网络与业务能力技术工作委员会主席, 中国电信科技委常委, SDN、NFV、AI 产业联盟技术委员会副主任, 网络 5.0 产业联盟技术委员会副主任; 长期从事电信网络领域技术和标准工作; 曾获多个国家和省部级科技进步奖项; 发表论文 100 余篇, 出版专著 12 部。

算力感知网络架构与关键技术

Architecture and Key Technologies for Computing-Aware Networking

姚惠娟 / YAO Huijuan, 陆璐 / LU Lu, 段晓东 / DUAN Xiaodong

(中国移动通信研究院, 中国 北京 100053)
(China Mobile Research Institute, Beijing 100053, China)



摘要: 针对运营商信息通信技术 (ICT) 基础设施面向云网融合、算网一体技术演进中的协同问题, 提出了算力感知网络 (CAN), 即网络可感知应用、网络、算力和用户需求等多维资源, 并协同调度算力资源和网络资源, 使应用能够按需、实时调用不同地方的计算资源, 实现边缘计算与云计算的协同联动, 提供最优的用户体验以及计算和网络资源利用率。

关键词: 算力感知网络; 算力路由; 算力服务信息

Abstract: Aiming at the synergy of cloud network integration and computing network integration technology of operator information and communications technology (ICT) infrastructure, computing-aware networking (CAN) is proposed. The network can perceive multi-dimensional resources such as application, network, computing power, and user demand. CAN jointly schedule computing power resources and network resources so that the application can call computing resources in different places on-demand and real-time, realize the collaborative linkage between edge computing and cloud computing, provide the optimal user experience and computing and network resource utilization rate.

Keywords: CAN; computing-aware routing; computing service information

DOI: 10.12142/ZTETJ.202103003

网络出版地址: <https://kns.cnki.net/kcms/detail/34.1228.TN.20210615.1146.004.html>

网络出版日期: 2021-06-15

收稿日期: 2021-05-12

1 算网融合发展的背景

1.1 面向算网融合演进的驱动力

目前 5G 网络发展的关键时期, 边缘计算和网络功能虚拟化 (NFV) 等技术都要求网络与计算协同发展。同时, 随着物理世界和数字世界的进一步融合, 行业数字化转型获得了全方位的提升和改变, 给运营商带来全新的市场和发展空间, 但同时也面临很多挑战。

(1) 应用需求驱动力

随着 5G 的商用规模部署, 工业互联网、车联网、增强现实 (AR) /

虚拟现实 (VR) 等垂直领域蓬勃发展。据 Machina Research 报告显示: 2025 年, 全球网联设备总数将超过 270 亿, 联网设备的指数级增长对网络传输能力及中心云处理能力带来了巨大挑战。据 Gartner 预测: 2025 年, 超过 75% 的数据需要分流到网络边缘侧。这对网络灵活调度、服务质量 (QoS) 等提出了更高的要求。另外, 产业智能化的升级会带来设备的多样性, 物联网 (IoT) 传感器、摄像头等设备的应用又会产生多样化的数据。这些异构数据的处理需要泛在的算力来支持。全行业的产业化转型对网络和计算均提

出了更高的要求: 基础设施信息技术 (IT)、通信技术 (CT) 更多融合; 基础设施不仅需要提供泛在的连接, 还需要提供算力的支持。

(2) 网络技术发展驱动力

NFV 技术的引入, 使得 5G 网络开始云化^[1], 并逐步具备向 IT 技术演进的基础。这使得算力开始服务于网络, 并随着网络进行延伸。此时的 IT 资源仅是 CT 网络的一种资源提供方式, 并不直接对外提供服务。在此阶段, IT 与 CT 的融合可称为信息通信技术 (ICT) 纵向融合。同时, 5G 网络原生支持边缘计算: 5G 用户面的下

沉为边缘计算的实现创造了网络条件,并使计算资源离用户更近,从而推动网络中的计算从集中走向边缘,并嵌入网络。计算资源逐渐成为网络基础设施的重要组成部分。ICT融合的方式由NFV时代的“IT服务于CT”向“IT与CT相互感知”演进,算网协同感知成为网络演进的核心需求^[2]。面向6G,算力与网络资源将共生,IT与CT系统需要具备相互感知的能力,以实现网络和算力的联合优化调度,并能提供端到端ICT系统的服务等级协议(SLA)体验保证。

1.2 计算网络融合发展产业现状

面向计算网络融合的演进需求,业界开展了许多研究工作。在2020年第8次网络5.0全会上,中国信息通信研究院联合三大运营商、中兴通讯等成立了网络5.0创新联盟算力网络特设组,就目前算网融合趋势下的不同技术路线展开探索,这些探索包括算力网络^[3-5]、算力感知网络^[4,6-7]等。特设组就算力网络研究方面达成共识,推动产业发展^[8]。此外,IMT-2030(6G)网络工作组也成立了算力网络研究组,研究6G网络中计算、网络融合对未来网络架构的影响。此外,互联网研究专门工作组(IRTF)成立了在网计算研究组(COINRG)^[9-12]。在网计算是指,网络设备的功能不再是简单的转发,而是“转发+计算”;计算服务也不再处于网络边缘,而是嵌入网络设备中。该工作组主要针对可编程网络设备内生功能的场景、潜在有益点展开研究。其中,内生功能包括在网计算、在网存储、在网管理和在网控制等,它是计算、网络更深层次融合的下一阶段。内生功能引起了研究人员的关注。

综上所述,在网络和计算深度融合发展的大趋势下,网络演进要求网

络和计算能够相互感知、高度协同,并可以基于无处不在的连接将泛在计算互联,以实现云、边、网高效协同,提高网络资源、计算资源利用效率,进而实现以下目标:(1)保证用户体验一致性。网络可以感知无处不在的计算和服务,用户无须关心网络中计算资源的位置和部署状态。网络和计算协同调度保证用户的一致体验。(2)服务灵活动态部署。基于用户的SLA需求,网络综合考虑实时资源状况和计算资源状况,通过灵活匹配、动态调度,将业务流量动态调度至最优节点,并支持动态服务来保证用户体验。

因此,我们提出一种面向计算网络深度融合演进的新型网络——算力感知网络(CAN),以实现用户体验最优化、资源利用率最优化等。

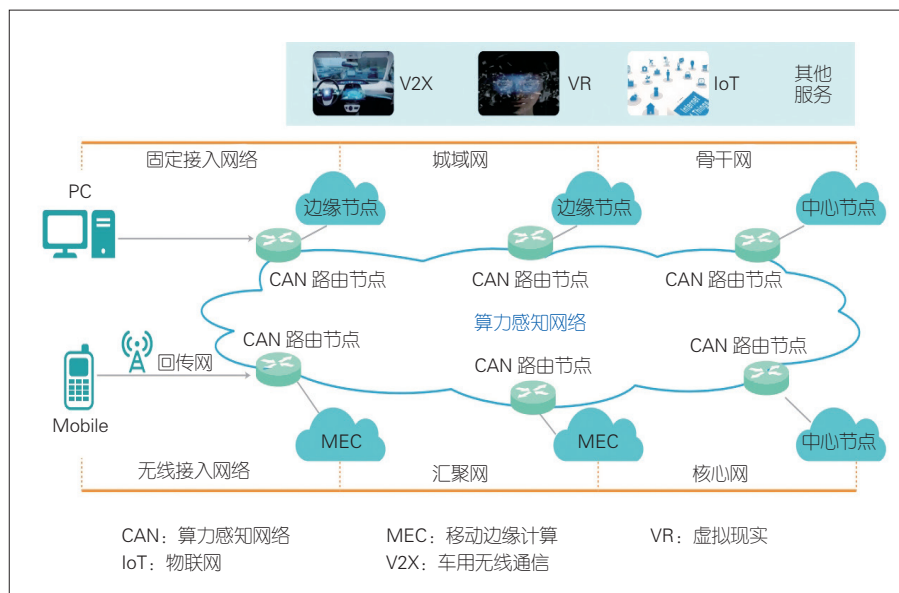
2 CAN的架构体系

2.1 CAN的概念

作为面向计算网络深度融合的新型网络架构,CAN以现有的网络技术为基础,通过无所不在的网络连接分布式的计算节点,实现服务的自动化

部署、最优路由和负载均衡,从而构建可以感知算力的全新网络基础设施,保证网络按需、实时调度不同位置的计算资源,提高网络和计算资源利用率。CAN还能进一步提升用户体验,实现网络无所不达、算力无处不在、智能无所不及的愿景。CAN的概念具体如图1所示。

基于CAN的概念,中国移动从架构、协议、度量等方面协同演进,构建面向算网一体化的新型基础网络,如图2所示。从架构层面上看,面对边缘计算、异构计算、人工智能等新业务,在基础设施即服务(IaaS)资源层编排的基础上,未来算网融合架构如何向平台即服务(PaaS)、软件即服务(SaaS)、网络即服务(NaaS)等一系列上层算法、函数、能力编排演进,需要进行研究。如何协同管理、控制数据面,以实现编排系统与网络调度系统的协作,从而实现一切即服务(XaaS)能力按需灵活部署,也需要重点考虑。从协议层面上看,传统网络优化路径仅实现信息在节点之间传输的SLA,并未考虑节点内部算力的负载。未来算网融合的网络需要感知



▲图1 算力感知网络概念图

内生算力的资源负载和 XaaS 性能,并综合考虑网络和算力两个维度的性能指标,从而进行路径和目标服务阶段的联合优化。另外,还需要考虑和数据面可编程技术的结合,如利用 SRv6 可编程性实现算网信息协同,以实现控制面和数据面的多维度创新。从度量方面看,网络体系的建模已经很成熟,但算力体系还需要综合考虑异构硬件、多样化算法以及业务算力需求,以及形成算力的度量衡和建模体系。CAN 需要依托统一的算力度量衡体系以及能力模板,为算力感知和通告、算力开放应用模型(OAM)和算力运维管理等功能提供标准度量准则。

2.2 CAN 体系架构

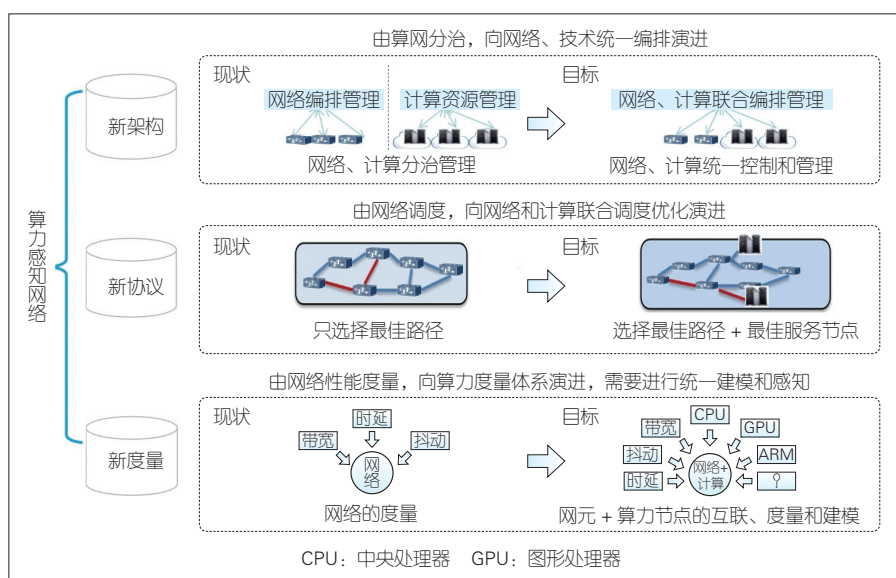
为了实现泛在计算和服务的感知、互联和协同调度,CAN 架构体系从逻辑功能上可分为算力服务层、算网管理层、算力资源层、算力路由层和网络资源层。其中,算力路由层包含控制面和转发面,如图 3 所示。

- 算力应用层:承载泛在计算的服务及应用,并将用户对业务 SLA 的请求(包括算力请求等)参数传递给算力路由层。

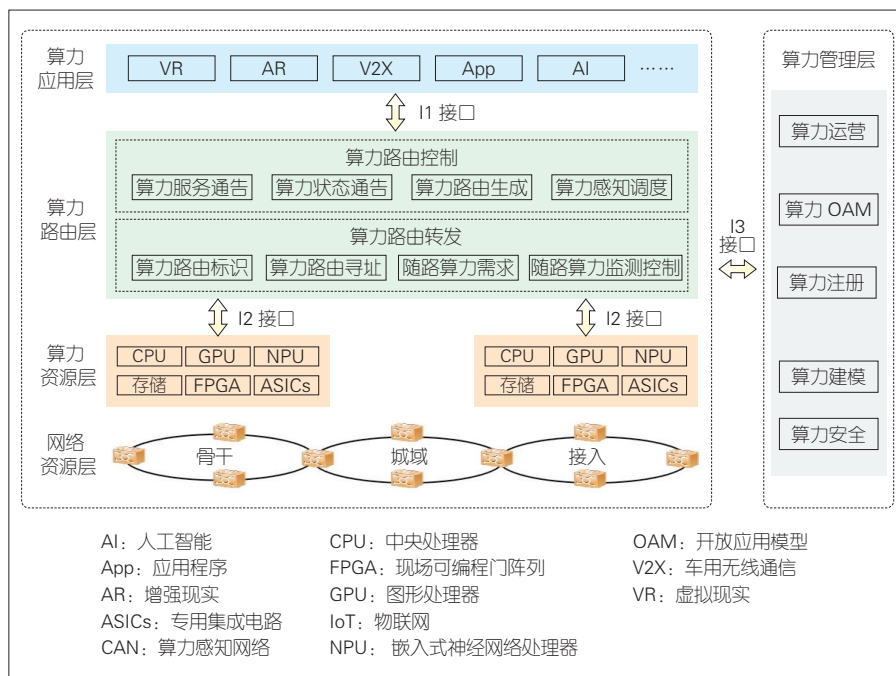
- 算力管理层:完成算力运营、算力服务编排,以及对算力资源和网络资源的管理。该层的具体工作包括对算力资源的感知、度量,以及 OAM 管理等,实现对终端用户的算网运营以及对算力路由层和网络资源层的管理。

- 算力路由层:是 CAN 的核心。基于抽象后的算网资源,并综合考虑网络状况和计算资源状况。该层可以将业务灵活按需调度到不同的计算资源节点中。

- 算力资源层:利用现有计算基础设施提供算力资源。计算基础设施包括单核中央处理器(CPU)、多核 CPU,以及 CPU+图形处理器(GPU)+现场



▲图 2 算力感知网络演进思路



▲图 3 CAN 体系架构图

可编程门阵列(FPGA)等多种计算能力的组合。为满足边缘计算领域多样性和计算需求,该层能够提供算力模型、算力应用程序编程接口(API)、算网资源标识等功能。

- 网络资源层:利用现有的网络基础设施为网络中的各个角落提供无处不在的网络连接,网络基础设施包括接入网、城域网和骨干网。

其中,算力资源层和网络资源层是 CAN 的基础设施层,算网管理层和算力路由层是实现算力感知功能体系的两大核心功能模块。基于所定义的五大数据模块,CAN 实现了对算网资源的感知、控制和调度。

总之,作为计算网络深度融合的新型网络,CAN 以无所不在的网络连接为基础,基于高度分布式的计算节点,

通过服务的自动化部署、最优路由和负载均衡,构建算力感知的全新的网络基础设施,真正实现网络的无所不达、算力无处不在、智能无所不及。海量应用、海量功能函数、海量计算资源则构成一个开放的生态。其中,海量的应用能够按需、实时调用不同地方的计算资源,提高计算资源利用效率,最终实现用户体验最优化、计算资源利用率最优化等。

3 CAN 关键技术

3.1 算力度量和建模

如何构建统一的算力模型是 CAN 的研究基础。基于算力统一的度量体系,通过对不同计算类型的异构算力资源进行统一抽象描述,形成算力能力模板,可以为算力路由、算力设备管理、算力计费提供标准的算力度量规则。首先,异构硬件设备通过统一的算力度量和建模,实现对现场可编程门阵列(FPGA)、GPU、CPU 等异构物理资源的统一资源描述,从而可以有效地提供计算服务。其次,考虑到计算过程受不同算法的影响,需要对不同的算法如人工智能(AI)、机器学习、神经网络等算法所需的算力进行度量,更有效地了解应用调用算法所需的算力,从而服务于应用。最后,由于用户的不同服务会产生不同的算力需求,需要把用户需求映射为实际所需的算力资源,从而可以使网络更充分有效地感知用户的需求,提高和用户交互的效率。

3.2 算力路由关键技术

算力路由层是 CAN 的核心功能层,支持对网络、计算、存储等多维资源、服务的感知与通告,从而实现“网络+计算”的联合调度。算力路由层包括算力路由控制技术和算力路由转发技术,这两种技术可以实现业务请

求在路由层的按需调度。

算力路由控制面可以通告算力节点的信息并生成算力拓扑,进而生成算力感知的新型路由表。算力路由控制面基于业务需求生成动态、按需的算力调度策略,实现算力感知的算网协同调度。算力路由转发需要通过 IP 协议/IPv6 扩展增强实现网络感知应用、算力需求以及随路 OAM 管理等功能。算力路由支持网络编程、灵活可扩展的新型数据面,能够实现算力服务的最优体验。

3.3 算力管理关键技术

算力管理包含算力设备的注册、OAM、运营等。统一的管理面可以对网络和算力进行管理和监测,并可生成算力服务合约以及计费策略,实现对算力的统一运营。基于统一的算力度量体系,通过对不同计算类型进行统一抽象描述,算网管理层能够形成算力能力模板,从而为算力设备的管理、合约和计费以及 OAM 提供标准的算力度量规则。算力注册需要实现对算力节点的注册、更新和注销,并对相应的路由通告策略进行管理。算力 OAM 需要实现对算力资源层的算力性能监测控制、算力计费管理等。

4 CAN 关键技术验证与测试

为了推动 CAN 的研究和标准化,中国移动搭建实验网,通过集成测试、功能测试和性能测试,多维度进行 CAN 关键技术的验证。中国移动浙江省公司完成了多个移动边缘计算(MEC)站点的 CAN 部署,具体的技术测试拓扑如图 4 所示。其中,多个节点位于杭州、金华的不同机房,平均距离约为 30 km,平均时延约 4 ms,平均通量接近 1 000 kbit/s。

集成测试验证了 CAN 组件与现有 MEC 软硬件环境及业务系统的集成能

力,实现控制面与数据面端到端的通信流程。

功能测试验证了 CAN 新增的网络能力,包括根据网络以及服务状态优选计算位置的能力、业务流粘性保持能力,以及主动触发服务质量劣化业务流重连的能力。

性能测试验证了 CAN 的 MEC 整体系统与基准 MEC 系统在数据面性能指标方面的对比情况。测试表明,CAN 调度系统单位时间完成的总任务数(QPS)有所提升,同时客户端感知的任务端到端完成时延有所降低,从而验证 CAN 调度系统可以实现系统资源利用率最优、用户体验最优。

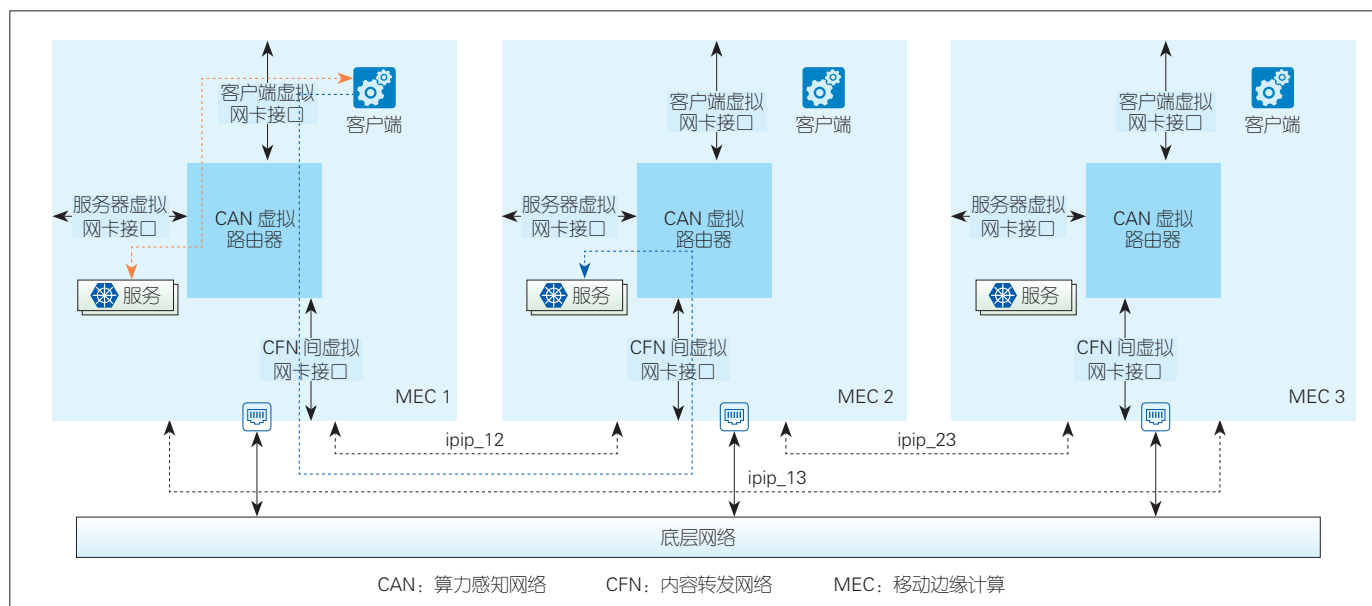
测试案例分为两大类,包含理想情况(系统算力容量和性能都比较均衡),以及系统算力容量、性能不均衡的情况。通过现网测试,我们可以对 CAN 调度系统与基准系统随机调度的 QPS 和端到端平均时延进行比较。

当系统算力容量和性能都比较均衡时,测试结果如表 1 和表 2 所示。当请求业务数处于小、中两种输入情况时,CAN 调度系统的 QPS 比基准系统随机调度的 QPS 分别提升 5.5%、33.17%,端到端时延分别降低 25.62%、16.00%。

本次测试有力地证明了 CAN 调度系统能够将业务请求分配到更优的边缘节点上,从而实现边边协同、整体系统负载均衡优化、资源利用率优化等。

5 结束语

以“新基建”为导向的一系列政策,使得新一代信息技术间的融合效应逐渐显现。“5G+云+AI”将成为推动中国数字经济持续发展的重要引擎。结合未来计算形态云-边-端泛在分布的趋势,计算与网络的融合将会更加紧密。为了提升“联接+计算”的能力,需要计算和网络两大产业进



▲图4 CAN 关键技术测试拓扑图

▼表1 系统无背景流时测试对比结果

请求业务数	系统	平均时延 /ms	QPS
5 (小)	CAN 调度系统	3.954	208.5
	随机调度	5.316	197.7
10 (中)	CAN 调度系统	4.700	402.3
	随机调度	5.595	302.1
15 (大)	CAN 调度系统	5.506	559.3
	随机调度	5.718	546.0

CAN: 算力感知网络 QPS: CAN 调度系统单位时间完成的总任务数

▼表2 系统不均衡时测试对比结果

请求业务数	系统	平均时延 /ms	QPS
5 (小)	CAN 调度系统	6.291	185.6
	随机调度	9.630	165.3
10 (中)	CAN 调度系统	6.854	360.9
	随机调度	10.592	316.3
15 (大)	CAN 调度系统	7.987	512.4
	随机调度	12.156	441.7

CAN: 算力感知网络 QPS: CAN 调度系统单位时间完成的总任务数

行有机协同，相互配合。业界也需要积极探索算力资源和算力服务的智能调度、高效分配的方式和途径，打造面向云网融合、算网一体技术演进的新型网络。

致谢

本论文由中国移动研究院算力网络团队共同完成，特向项目组成员孙滔、付月霞、刘鹏、杜宗鹏等致谢！

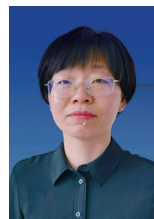
参考文献

- [1] SDN/NFV/AI 标准与产业推进委员会. 网络人工智能应用白皮书 [R]. 2019
- [2] 边缘计算网络产业联盟. 运营商边缘计算网络技术白皮书 [R]. 2019
- [3] 雷波, 刘增义, 王旭亮, 等. 基于云、网、边融合的边缘计算新方案: 算力网络 [J]. 电信科学, 2019, 35(9): 50-57
- [4] 中国联通. 中国联通算力网络白皮书 [R]. 2019
- [5] 中国移动. 算力感知网络技术白皮书 [R]. 2019
- [6] 中国通信标准化协会. 面向全网算力的算力感知网络关键技术研究 [R]. 2020
- [7] 姚惠娟, 耿亮. 面向计算网络融合下一代网络架构 [J]. 电信科学, 2019, 35(9): 38-43
- [8] 网络 5.0 技术和产业创新联盟. 网络 5.0 技术白皮书 [R]. 2019
- [9] GENG L, LEI B, FU Y, et al. IMT2020-CAN-

req: use cases and requirements of computing-aware networking for future networks including IMT-2020 [R]. ITU-T, 2020

- [10] GENG L, WILLIS P. Compute First Networking (CFN) scenarios and requirements [R]. IETF, 2019
- [11] LI Y, HE J, GENG L, et al. Framework of Compute First Networking (CFN) [R]. IETF, 2019
- [12] GU S, ZHUANG G, YAO H, et al. A report on Compute First Networking (CFN) field trial [R]. IETF, 2019

作者简介



姚惠娟, 中国移动通信研究院网络与 IT 技术研究所项目经理, 担任 CCSA TC614 架构组组长和算力特设组组长; 长期从事 IP 网络研究和标准工作, 主要涉及承载网络、边缘计算、未来网络架构等领域。



陆璐, 中国移动通信研究院网络与 IT 技术研究所副所长, 担任 CCSA 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。



段晓东, 中国移动通信研究院副院长, 担任 IMT-2030 (6G) 推进组网络技术组组长; 主要研究方向为 5G/6G 网络架构、云计算及虚拟化、IP 新技术。

算力网络实现一体化服务的探索与实践

Exploration and Practice to Realize Service Integration in Computing Power Network



雷波/LEI Bo, 赵倩颖/ZHAO Qianying, 凌泽军/LING Zejun
(中国电信股份有限公司研究院, 中国 北京 102209)
(Research Institute of China Telecom Corporation, Beijing 102209, China)

摘要:算力网络(CPN)通过网络控制面将资源信息进行分发,有机地实现多维资源信息的整合。除此之外,CPN还需要与算力交易、网络订购等业务关联起来,形成统一的体系架构,实现对多类资源的优化分配。在屏蔽底层资源的差异与异构特性的基础上,所提出的算力网络交易平台向算力需求方提供了从资源选择到使用的一体化服务,形成了统一的资源供给机制,满足各类新兴业务的多样化需求。

关键词:算力网络;算力网络交易平台;多维资源;一体化

Abstract: Computing power network (CPN) distributes computing power resources information through the network control plane, which realizes the integration of multi-dimensional resource information. To form a unified architecture and realize the optimal allocation of multiple kinds of resources, CPN also needs to be associated with computing power transactions, online orders, and other businesses. Based on the shielding differences and heterogeneous characteristics of the underlying resources, the proposed computing power network transaction platform provides integrated services from resource selection to use to the demand side of the computing power, forming a unified resource supply mechanism to meet the diversified needs of various emerging businesses.

Keywords: computing power network; computing power transaction platform; multi-dimensional resource; integration

DOI: 10.12142/ZTETJ.202103004

网络出版地址: <https://kns.cnki.net/kcms/detail/34.1228.TN.20210615.1456.008.html>

网络出版日期: 2021-06-15

收稿日期: 2021-05-13

随着人工智能、车联网、边缘计算、工业互联网等业务的兴起,算力资源作为关键生产要素,受到了业界的广泛重视,但目前仍存在度量困难、种类繁多、分布广泛、归属复杂等特点。

这些特点使得现有业务大多在

特定类型的算力资源中部署,很难在不同类型、不同级别的算力资源之间灵活调度。从长期来看,能够综合利用不同等级的资源,业务才能实现性能与成本的优化,从而才能提升竞争力。

因此,将多级算力资源整合成一体化算力服务体系已是业界共识,并成为国家产业导向^[1]。

1 算力资源的定义与特点

在分析算力资源的特点前,我们首先要明确算力的概念。算力,也称为计算力或计算能力。该词的最早来源已经不可查证,互联网上的资料大多与区块链相关。这是因为区块链技术采用的是哈希算法,即在相同时间内挖出更多的“币”,也就是说谁算得快谁就能获得更多的收益。因

基金项目:国家重点研发计划(2018YFB1800100)

此,人们就以“算力”这个词来描述计算的快慢,比如“算力大”则意味着单位时间内计算得更快。为了计算得更快、更灵活,需要将分散的算力集中起来形成资源池,这就是所谓的算力资源。

通过分析、对比不同类型的算力资源,可将算力资源的特点归纳为4个方面:度量困难、种类繁多、分布广泛、归属复杂。

(1) 度量困难

当业务在各类算力资源之间部署、迁移时,需要综合评估节点空闲算力资源余量,这就需要使用一个简单、公认的量纲来衡量算力大小。由于计算快慢不仅与算力大小相关,也与所选择的算法有关,同一个算力节点运行不同的算法会有完全不同的效果。因此,对算力的度量往往不是单一维度的,这远比电力、水力的度量困难得多。

(2) 种类繁多

算力资源从不同维度被划分成不同类型,按核心芯片类型可以分为中央处理器(CPU)、图形处理器(GPU)、专用集成电路(ASIC)等,按所在位置可以分为云、边、端等。

(3) 分布广泛

算力资源的构建具有灵活性,只要能有一定的空间,并提供电力,任何单位甚至个人都可以构建相应类型的算力资源节点。这使得各类算力资源可以分布在不同的物理空间上:越是远离人口密集区域的算力资源,规模就越大,成本就越低;越是靠近城市核心区域的算力资源,规模则越小,成本越高。

(4) 归属复杂

不同类型算力资源的建设难度相差极大。例如,对于云计算节点,算力资源的建设需要占用大量的土地、电力等资源,还需要通过国家规

定的各项审批流程,并需要规模效应来降低单位成本,技术门槛非常高,一般只有大型投资方有意愿实施;对于边缘计算节点,只要具有一定的机房空间(如室外机柜)就可构建,成本不高且不需要太复杂的技术,中小型企业能够自建;而对于端计算节点,个人就可以购买一套适合的设备对外提供服务。在整合算力资源来提供一体化算力服务时,就必须考虑到算力资源归属于多方的这一特点,因此需要尽量简化在多方之间的交易与调度过程。

2 一体化服务与算力网络

虽然算力资源存在以上4个特点,但新兴业务可以将算力资源整合起来,形成一体化的服务机制,让算力随时随地按需供给。

国家发展和改革委员会、工业和信息化部等部委在《关于加快构建全国一体化大数据中心协同创新体系的指导意见》中提出“推动算力资源服务化”,这包括两方面的要求^[1]:

(1)构建一体化算力服务体系。加快云资源接入和一体化调度机制的建立和完善,以云服务方式提供算力资源,降低算力使用成本和门槛。

(2)优化算力资源需求结构。以应用为导向,充分发挥云集约调度优势,引导各行业合理使用算力资源,以提升基础设施的利用效能。

针对以上目标,业界已出现一些解决方案。例如,云服务提供商提出了云边缘的概念,希望通过扩展云的使用范围来统一各级算力资源,提供统一服务。另外,还有以网络为平台来设计的算力网络(CPN)技术方案。CPN是一种有机整合多级算力资源、存储资源与网络资源的新型技术方案,能够提供新型的一体化算力服务。CPN技术核心在于通过网络控

制面分发多维资源信息,通过计算最佳路径的方式实现多维资源的有机结合。

目前,已有多种基于CPN的技术路线被提出,如集中式、分布式、混合式等。这些技术路线开发了CPN资源调度系统原型^[2-3]和CPN交易平台系统原型。算力资源调度系统根据资源分配策略,建立算力消费者与算力资源提供者之间的网络连接,并根据业务需求变化及时调整资源分配。在此基础上,CPN交易平台成为连接算力消费者和算力资源提供者的纽带,从商业模式上连接了算力消费者、算力资源提供者与网络运营者,实现从用户需求到资源分配、资源交易、资源使用的一体化算力资源服务。

3 CPN交易系统设计与实践

3.1 总体设计思路

为满足算力资源一体化服务的需求,CPN交易平台应具有以下功能:

(1)CPN交易平台需要将算力消费者、算力提供者以及CPN控制层结合,以实现消费者提出的资源或业务需求;交易平台制定分配策略,CPN控制层则根据分配策略,建立算力消费者与算力提供者之间连接的一体化服务。

(2)不同能力的CPN消费者的资源与业务需求的分析能力不尽相同。CPN平台还应具备对用户业务需求进行人工智能(AI)分析的能力,提供更加智能的服务,满足不同用户对CPN交易平台的使用需求。

(3)CPN交易平台还应提供可供应用开发者上传第三方应用的应用商店,实现从资源到应用的全生态服务。

根据上述需求,CPN交易系统与各方参与者之间的关系如图1所示^[4]。

在CPN基本框架中:

(1)CPN消费者是CPN交易平台的主要使用者,因此CPN交易平台需要提供消费者账户管理能力,并使CPN消费者在该平台中选择合适的资源,然后购买。

(2)作为资源供应方,算力提供者需要在CPN交易平台中进行资源注册,对资源的使用情况进行实时监测。

(3)作为底层资源和算力平台之间的枢纽,CPN控制面需要与CPN交易平台联动,将所有采集到的资源信息上报给CPN交易平台,并根据交易平台形成的调度策略,对底层资源进行调度,构建网络连接。

(4)为满足算力消费者的智能分析需求以及使用诉求,CPN交易平台还应连接AI赋能平台,对用户的需求进行智能分析,并根据用户的意图为其匹配最佳资源。

3.2 资源交易视图生成模型

多类型、多归属方的泛在资源池位于网络的各个位置。如何获得资源池的各项信息成为利用资源池的前提。在CPN中,资源信息的发现由CPN控制面实现,资源池的各项信息由集中式的管理控制系统或分布式路由算法来获得,包括但不限于资源类型、大小、功能、路由。信息由CPN控制面发送至CPN交易平台,结合用户信息后生成资源交易视图。本节中,我们将对资源交易视图生成模型^[5]进行介绍。

网络控制层所获得的资源信息模型为 $\Phi=\{C,T,X,\mathcal{A}\}$,其中计算能力为 C ,包括计算资源类型、现有资源数量;存储能力为 T ,包括存储资源类

型、资源数量;算法能力为 X ,包括算法种类、算法复杂度;路由为 \mathcal{A} 。

用户信息模型包括用户位置信息 Γ 和用户资源需求信息。用户资源需求信息包括性能指标和能力需求。性能指标为 $P_i=\{\gamma_i, \zeta_i, \eta_i, \alpha_i, \dots\}$,其中 γ_i 为计算性能指标,具体包括核数、主频等; ζ_i 为存储性能指标,具体包括内存容量、外存容量等; η_i 为网络性能指标,具体包括时延、带宽、抖动、误码率等各项指标; α_i 为算法性能指标,如时间复杂度、空间复杂度等。

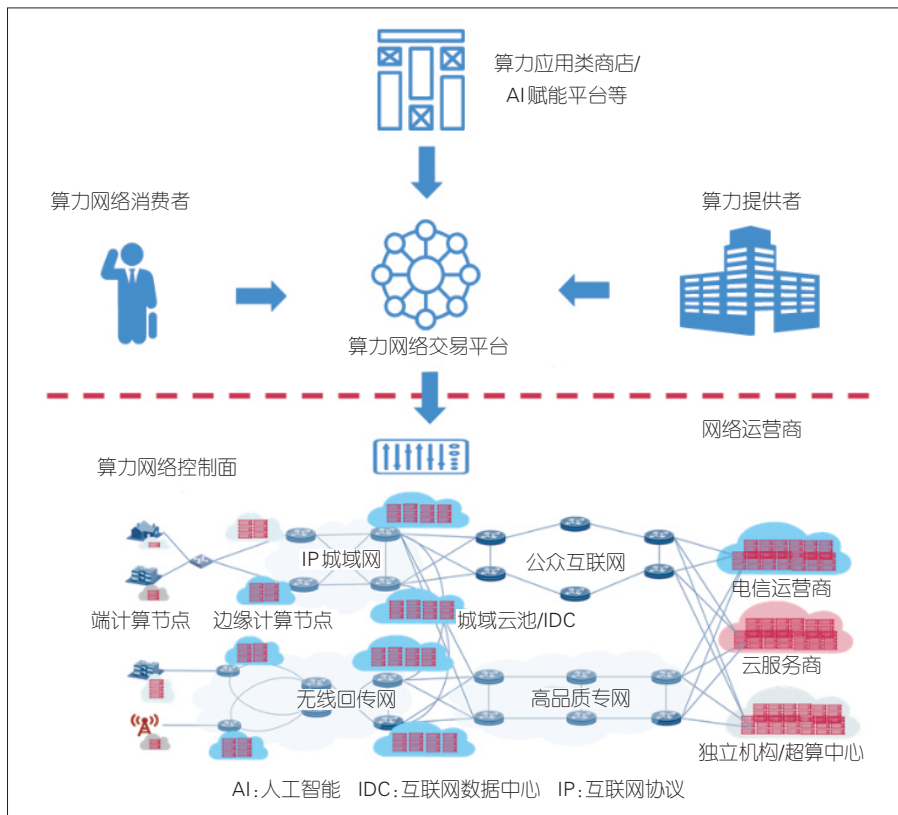
功能需求为 $F_i=\{f_i^\gamma, f_i^\zeta, f_i^\eta, f_i^\alpha, \dots\}$,其中 f_i^γ 表示计算功能,如CPU能力、GPU能力; f_i^ζ 表示存储能力,如块存储、对象存储; f_i^η 表示网络功能; f_i^α 表示算法功能如图形算法、语音算法。

综上所述,用户信息模型表示为:

$$S_i = \left\{ \Gamma, \left\{ P_i \right\} \right\} = \left\{ \Gamma, \left\{ \begin{matrix} \gamma_i & \zeta_i & \eta_i & \alpha_i & \dots \\ f_i^\gamma & f_i^\zeta & f_i^\eta & f_i^\alpha & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{matrix} \right\} \right\} \quad (1)$$

通过对资源信息以及用户信息的计算,CPN交易平台能够查找到满足用户需求的资源池,并生成资源池列表 L 。资源列表中每一资源项 l_i 的参数包括计算能力 C_i 、存储能力 T_i 、算法能力 X_i 、用户到资源池时延 D_i ,以及资源报价 E_i 。计算查找函数可以通过函数 \mathcal{P} 实现:

$$L = \left\{ \begin{matrix} \vdots \\ l_i \\ \vdots \end{matrix} \right\} = \left\{ \begin{matrix} \vdots \\ \{C_i, T_i, X_i, D_i, E_i\} \\ \vdots \end{matrix} \right\} \triangleq \mathcal{P}(\Phi, S_i) = \mathcal{P} \left(\{C, T, X, \mathcal{A}\}, \left\{ \Gamma, \left\{ \begin{matrix} \gamma_i & \zeta_i & \eta_i & \alpha_i & \dots \\ f_i^\gamma & f_i^\zeta & f_i^\eta & f_i^\alpha & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots \end{matrix} \right\} \right\} \right) \quad (2)$$



▲图1 算力网络基本框架

3.3 系统功能模块

根据总体设计思路,CPN 交易平台需要提供账户管理、交易监测控制、交易流程、日志管理、用户体验反馈、采集和监测控制、对象存储服务(OSS)接口、应用市场、增强编排调度等模块。CPN 交易平台系统的功能架构如图2所示。

- 账户管理模块:对算力消费者账户、算力提供者账户,以及权限账户进行管理,包括账户申请注册、查询、登录、退出等功能。

- 交易监测控制模块:对交易过程(如交易合约的执行过程)、交易资源、交易记录进行管理,确保交易过程的安全性,及时掌握资源的占用情况及输出交易记录。

- 交易流程模块:支持用户的交易申请、可交易资源的展示、交易套餐的选择和提交、交易的验证和生效,以及交易结束后的资源释放,处理用户从选择到购买的整个流程。

- 日志管理模块:对报警日志、故

障日志进行管理,以便更好地对交易平台信息进行跟踪、管理,对报警、故障进行诊断和解决。

- 用户体验反馈模块:对用户意见进行反馈和汇总,更好地提升交易平台的使用体验。

- 采集和监测控制模块:对可交易资源进行采集、汇总及监测控制,对资源信息及时进行更新。

- 应用市场模块:支持应用市场展示、应用上线申请和提交、应用的审核验证和批准、应用的撤销和删除以及应用版本的更新。对CPN交易平台中准备上线的应用进行安全管理和交易。

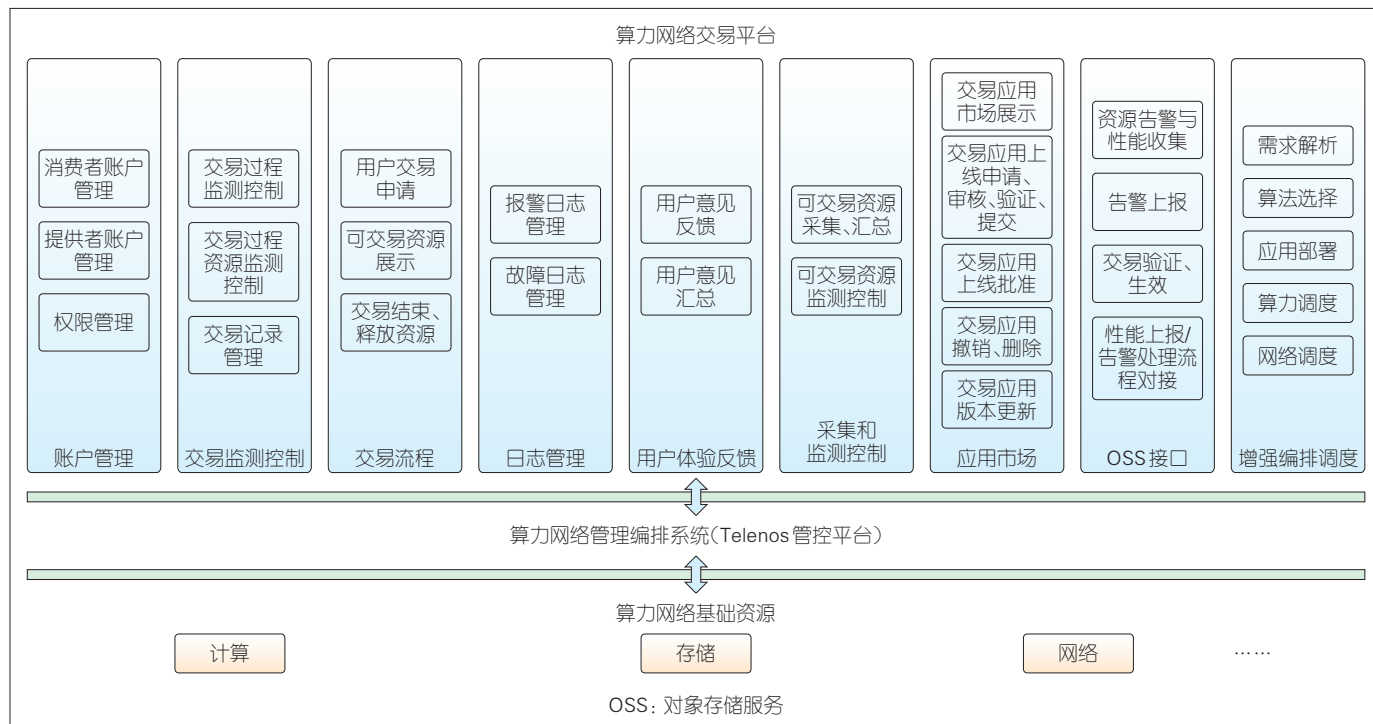
- OSS接口模块:与CPN控制面中的资源告警与性能收集、告警上报、性能上报、告警处理流程进行对接。

- 增强编排调度模块:支持需求分析、算法选择、应用部署、算力调度和网络调度,为CPN交易平台提供资源管控服务。

3.4 基于AI游戏场景下的试验验证

CPN 交易平台可以为众多新兴业务场景提供服务,如基于AI的人机互动游戏。由于应用开发者往往选择虚拟资源进行应用部署,因此,在众多资源池中选择与业务场景匹配的资源便成为关键问题。在AI交互类游戏中,时延对用户体验起到决定性作用。当端到端时延超过50 ms时,体验感开始下降;超过100 ms时,将出现明显卡顿^[6]。因此,在不考虑其他处理过程所需时间的情况下,AI交互类游戏网络时延要尽量控制在50 ms以下甚至更低。综合游戏以及AI类应用的各项指标^[7],在本文测试例中需要为AI类交互游戏匹配一个算力不小于4 TFLOPS、存储容量不小于1 TB、网络带宽不小于1 Gbit/s、网络时延不大于50 ms的算力资源。

当一名AI交互类游戏开发者(以下统称CPN消费者)想要购买合适的资源为某区域(以北京市亦庄经济开发区为例)的用户提供服务时,可以



▲图2 算力网络交易平台功能架构图

注册并登录 CPN 交易平台。注册登录界面如图 3 所示。

当该 CPN 消费者具有丰富的资源使用经验时,会比较了解应用与资源的匹配情况,那么可以根据自己的经验填写相应的服务位置及资源需求。依据前文分析即填写(北京,北京,亦庄经济开发区)(4,1,0,1,50),如图 4 所示。

当 CPN 消费者并不明确所需资源情况,而只清楚资源所要应用的场景是 AI 游戏类时,可以选择服务位置以及相应的应用场景,如 AI 游戏。CPN 交易平台将通过自身的 AI 增强功能,按场景对所需资源进行分析,从而查询到满足需求的资源池。

CPN 消费者输入资源需求(如图 5 所示)或业务需求后,CPN 交易平台会生成以用户为中心的资源视图,如图 6 所示。资源视图的中心位置表示应用提供服务位置,每一圈虚线表示距离用户(使用应用服务的用户)的不同时延,虚线上的点表示时延圈内的资源池,点的面积越大则代表计算资源则越多。

资源视图生成后,CPN 交易平台还会生成与可用资源相匹配的资源池列表。列表中详细介绍各资源池的资源状况与报价,如图 7 所示。CPN 消费者可以根据自己的支付能力选择合适的资源池。

CPN 消费者选择合适的资源池

▲图3 算力消费者注册及登录界面

▲图4 资源需求输入

▲图5 场景需求输入



▲图6 算力网络资源视图

资源池名称	价格	网络时延	资源池描述
资源池-BJTY-增强计算型	¥578.8	20 ms	SG侧算力-北京亦庄
资源池-BJ2-增强计算型	¥1123.1	20 ms	算力网络资源池-北京朝阳区
资源池-BJA2-增强计算型	¥506.0	20 ms	算力网络资源池-北京朝阳区
资源池-BJC1-增强计算型	¥1728.0	20 ms	算力网络资源池-北京西城区
资源池-BJC2-GPU加速型	¥1468.8	20 ms	SG边缘云算力网络资源池-北京西城区
资源池-BJ3-通用计算型	¥578.8	20 ms	算力网络资源池-北京海淀区
资源池-BJA1-通用计算型	¥446.4	20 ms	算力网络资源池-北京海淀区
资源池-BJA3-GPU加速型	¥1310.4	20 ms	算力网络资源池-北京海淀区

▲图7 算力资源列表

后,便可在支付中心进行交易支付,如图8所示。

在整个交易流程中,CPN交易平台将持续跟踪资源占用情况。交易结束时,CPN交易平台将终止服务,释放算力资源与网络资源。

3.5 未来发展方向

CPN交易平台能够实现分布式资源与资源用户之间的交易,为用户提供算力资源一体化服务的同时,保证了交易的安全性、可靠性。安全性主要体现在算力消费者、算力提供方的身份认证及算力交易过程中有安全保证。基于分布式账本的属性,区块链技术可以为基于分布式资源的CPN提供更加合理的安全保障。区块链可有效连接分布式计算、存储能力和数据资源,实现多种异构网络资源共享和数据流转。基于区块链构建的数字身份系统,可以对算力消费者及算力提供者进行有效的身份认证。区块链技术可以支持用户按需购买算力资源,并将购买记录和资源使用情况上链存储,业务运营方就可以根据记录进行计费 and 结算。在未

来,区块链技术将是保证算力交易的一种重要技术,CPN也将借助区块链技术,为用户提供更加全面更加可靠的一体化服务^[8-9]。

4 结束语

CPN技术在标准制定、原型开发等方面已取得了重大进展。CPN交易平台为用户提供了一体化的算力资源服务,将融合的多维资源智能化、可视化地提供给用户,创新性地提供了一种融合各算力参与方的商业模式。CPN商业模式的相关研究正在开展,前景逐渐清晰,但在算力平台的安全性、如何实现AI能力增强等方面仍需进行更加深入的研究。

本研究得到北京邮电大学梅杰的帮助,谨致谢意!

参考文献

- [1] 关于加快构建全国一体化大数据中心协同创新体系的指导意见(发改高技〔2020〕1922号) [R]. 国家发展和改革委员会, 2020
- [2] 雷波, 刘增义, 王旭亮, 等. 基于云、网、边融合的边缘计算新方案: 算力网络 [J]. 电信科学, 2019, 35(9):44-51

- [3] 雷波, 赵倩颖. CPN:一种计算/网络资源联合优化方案探讨 [J]. 数据与计算发展前沿, 2020, 2(4): 55-64
- [4] 雷波, 陈运清. 边缘计算与算力网络——5G+AI时代的新型算力平台与网络连接 [M]. 北京: 电子工业出版社, 2020
- [5] 胡宇翔, 伊鹏. 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
- [6] 中国宽带发展联盟. 千兆宽带网络商业应用场景白皮书 [R]. 2019
- [7] 中国联通. 面向业务体验的算力需求量化与建模研究 [R]. 2020
- [8] 任梦璇. 区块链+边缘计算应用研究与探讨 [EB/OL]. (2021-01-25)[2021-06-07]. https://blog.cs-dn.net/weixin_41033724/article/details/113153834
- [9] 方军. 超入门区块链 [M]. 北京: 机械工业出版社, 2019

作者简介



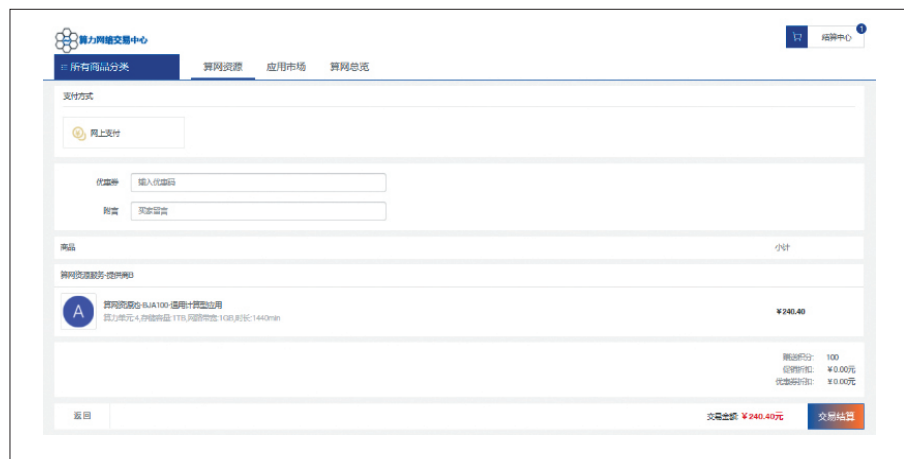
雷波, 中国电信股份有限公司研究院高级工程师,担任边缘计算产业联盟ECNI工作组联席主席、CCSA“网络5.0技术标准推进委员会”管理与运营组组长等职务;主要研究方向为未来网络架构、新型IP网络技术等;发表论文10余篇,出版图书《边缘计算与算力网络》《边缘计算2.0:网络架构与技术体系》。



赵倩颖, 中国电信股份有限公司研究院工程师;主要研究方向为未来网络、算力网络等;发表论文3篇,参与出版图书《边缘计算与算力网络》《边缘计算2.0:网络架构与技术体系》。



凌泽军, 中国电信股份有限公司研究院高级工程师;主要研究方向为未来网络、算力网络、软件开发、终端研究等;发表论文10余篇,出版图书《构建运营级的LTE网络》。



▲图8 算力网络交易平台支付界面

基于可编程网络的 算力调度机制研究

Computing Power Scheduling Mechanism Based on Programmable Network



李铭轩 / LI Mingxuan, 曹畅 / CAO Chang, 杨建军 / YANG Jianjun

(中国联合网络通信有限公司研究院, 中国 北京 100048)
(China United Network Communication Research Institute, Beijing 100048, China)

摘要: 结合最新的可编程网络技术, 提出了算力资源调度技术, 并介绍了技术架构和算力调度机制。在算力资源调度技术架构的基础上, 进一步提出了整体平台功能架构和编程架构。基于可编程网络的算力资源调度技术解决了目前算力调度过程中无法实现的网络参数问题, 从而能够更好地实现网络和算力的融合。

关键词: 可编程网络; 云原生; P4; 无服务

Abstract: Combined with the latest programmable network technology, the computing power resource scheduling technology is proposed, and the overall technical architecture and computing power scheduling mechanism are introduced. Based on the technical architecture of computing power resource scheduling, the overall platform functional architecture and programming architecture are further proposed. The computing power scheduling mechanism based on the programmable network solves the current bottleneck of network parameters that cannot be achieved in the current computing power scheduling process, which can better achieve the integration of network and computing power.

Keywords: programmable network; cloud native; P4; serverless

DOI: 10.12142/ZTETJ.202103005

网络出版地址: <https://kns.cnki.net/kcms/detail/34.1228.TN.20210615.1137.002.html>

网络出版日期: 2021-06-15

收稿日期: 2021-05-15

软件定义网络(SDN)通过网络控制逻辑(控制平面)与转发流量(数据平面)的分离, 将传统封闭的网络体系解耦为数据平面、控制平面和应用平面, 简化策略实施和网络配置^[1]。2008年, 以斯坦福大学 Nick MCKEOWN 教授为首的研究团队提出了 OpenFlow 以及 SDN 技术。自此, SDN 技术获得了业界的高度关注, 一系列相关应用被提出, 极大地促进了网络创新发展。2014年, 在 SDN 基础上,

研究者又提出了可变成数据平台技术, 将网络编程能力扩展到数据平面, 进一步开放了网络设备的可编程能力^[2]。

在从原有的虚拟化技术向云原生技术的演进过程中, 传统的算力资源调度技术往往基于网络互通, 并通过集群自身的调度策略来实现算力的动态调度和应用分配。在这一过程中, 固化的组网方式已经无法满足业务需求, 网络编排能力已成为算力调度能力的制约因素。文章中, 我们着重研究基于可编程网络的算力调度机制, 以期能够将可编程网络的最大优势应

用于传统的算力调度机制中。

1 可编程网络介绍

现有的网络技术尤其是 SDN 技术的发展, 使得传统网络转发设备能够从固化在芯片上的转发机制向基于通用芯片承载的转发机制转变, 同时也可以为 SDN 的实现带来可能性^[3]。现有的以 SDN 为代表的可编程网络实现技术, 主要是基于可编程的网络协议和转发控制协议: 应用平面的网络应用通过控制平面的控制器向底层的数据平面 SDN 数据转发设备下发路由协

基金项目: 国家重点研发计划(2019YFB1802800)

议和转发策略,从而实现网络转发机制^[4]。SDN 网络架构如图 1 所示。

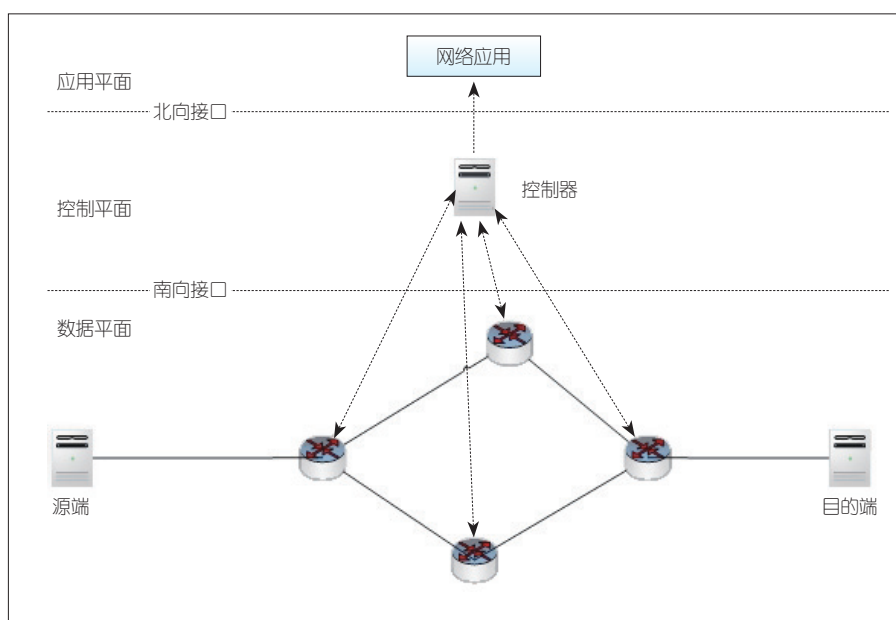
基于上述 SDN 网络架构,网络应用通过控制器对 SDN 路由器的转发机制进行控制。数据报文从源端服务器,经 SDN 路由器,发送至目的端服务器。在此过程中,网络应用程序可以通过控制器来选择合适的转发路径,进行数据转发链路的路由。在整个网络架构中,为了实现两层解耦,可以通过南向接口对底层数据平面的转发功能进行封装,实现控制器和路由器之间的对接;再通过北向接口对 SDN 控制能力进行封装,对上层应用提供统一的能力开放。网络应用则通过调用北向接口来实现对控制器的调用和控制^[5]。

1.1 可编程网络技术架构

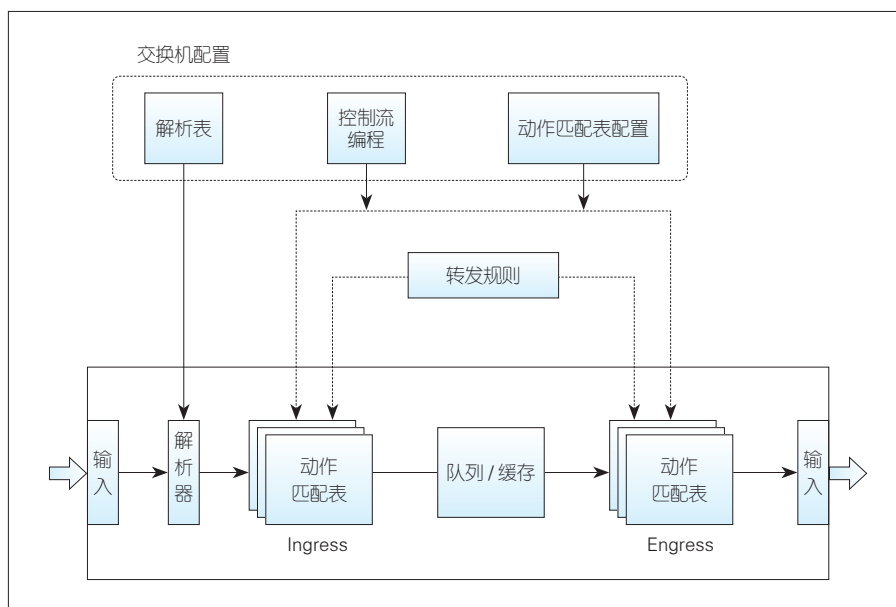
SDN 的可编程网络技术架构,可以实现 SDN 路由器的流量转发策略控制和软件定义设定。在数据转发平台,现有技术从原有的基于 OpenFlow 协议相关的数据转发,逐渐演进到与协议无关的面向高级编程的数据转发平面。可编程网络技术通过代码级的自定义网络数据平面来实现可编程能力,同时还可以实现数据转发控制和策略^[6],具体实现流程如图 2 所示。

根据交换机的配置,传统的可编程网络数据转发对解析表和控制流进行编程,并且将解析表下发至数据转发平台的解析器中,再通过控制流编程将动作匹配配置下发至数据转发平台的动作匹配表中^[7]。

数据流在输入时,首先经过解析器进行报文解析,并通过 Ingress 负载均衡器进行转发设置,以实现动作匹配配置,然后进入转发的队列和存储中等待;数据流从队列中出栈时,先通过对应的 Egress 反向负载均衡器的解析,再进行报文输出,从而实现了



▲图 1 软件定义网络架构图



▲图 2 可编程网络数据平面抽象转发模型图

完整的数据平面数据转发机制^[8]。

1.2 控制面 / 用户面分离实现机制

传统的 SDN 技术通过控制面和数据面的分离,实现网络数据流转发和控制策略制定之间的分离,也为基于可编程网络的算力调度提供了可能。

从上述的技术架构可以看出,现有的 SDN 技术实现了两层解耦:一方

面,将控制器集中于上层控制面,实现对 SDN 路由器或数据转发设备的统一管理;另一方面,使数据平面的 SDN 路由器和转发设备脱离了传统模式(固化在设备内),并使得转发功能和转发规则通过南北向接口开放至上层控制器和网络应用。这种控制面 / 用户面分离机制,为网络可编程的进一步实现提供了可能。

2 算力调度机制

随着云计算技术的不断发展,基础设施领域已有越来越多的企业采用云计算作为统一资源的管理方式。随着云资源池规模的不断扩大,算力节点的调度主要采用分布式的方式;而传统的基于云计算或云原生的算力资源调度是以虚拟化技术和进程共享技术来实现的^[9]。基于 OpenStack 或 Kubernetes 的算力资源调度将算力节点的空闲度作为算力调度策略的主要评判依据。本文中,我们以 Kubernetes 的资源调度组件 Scheduler 为例,重点阐述云原生的算力调度机制。

Scheduler 是 Kubernetes 的核心组件,负责为用户声明的 Pod 资源选择合适的算力节点,同时保证集群资源的最大化利用,其任务资源调度流程如图 3 所示。

现有的 Kubernetes 资源调度机制根据用户的请求,从资源管理器中获取资源信息,并且根据具体的调度策略将任务调度至具体的算力节点上。

在网络可达的情况下,现有 Kubernetes 算力节点运行状态监测控制主要通过算力节点代理监测的方式来实现。通过采集和上传算力节点上的中央处理器(CPU)、存储、内存等信息,并将这些信息上传至资源管理器,再经资源调度器进行策略调度^[10],从而将任务调度至指定的算力节点上。这种算力资源调度机制虽然能够解决分布式环境中、算力资源非均衡情况下的算力动态调度问题,但必须基于网络可达的情况。该机制并没有考虑到网络质量、算力节点的连接,以及传输过程中的网络情况。随着分布式计算,尤其是大数据等多集群甚至是跨数据中心协同处理的发展,网络的数据传输质量往往会成为影响上层用户体验的关键因素,同时也会成为跨数据中心算力调度和集群高效协同的

制约因素。传统的算力调度机制仅实现了计算资源在非均衡状态下的动态调度,却未考虑网络的服务质量(QoS)或体验质量(QoE)。在算网融合应用快速发展的趋势下,传统的算力调度方式已无法满足当前需求。

SDN 技术,尤其是可编程网络技术的发展,促使网络能力进一步开放、可编程化,并使传统算力调度机制能够更好地融合网络因素。本文中,我们通过算力和网络协同的方式来实现算力最优化调度,极大地发挥了网络在数据传输和转发方面的优势。

3 基于可编程网络的算力调度方案

基于网络的可编程、可控制等能力,并结合算力节点空闲度和计算能力等因素,基于可编程网络的算力调度机制能够在网络路由和路径选择方面实现算力调度。本文中,我们研究了通过编排调度方式来实现算力服务的编排和管理,以及可编程网络能力的开放。

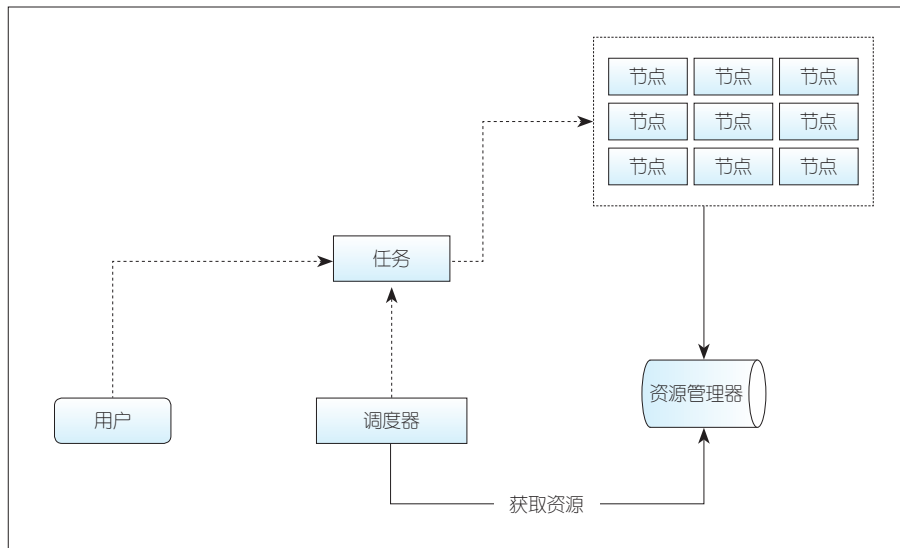
3.1 整体网络架构

基于可编程网络的管控分离能力,本文在 Kubernetes 原有调度方式基础

上,进一步研究了基于可编程网络的算力调度机制。其中,网络拓扑采用数据控制面和数据转发面分离模式,容器计算集群承载具体的算力分配和容器承载,控制集群和可编程网络的数据控制面对接以实现网络控制。基于可编程网络的技术架构如图 4 所示。

基于可编程网络的算力调度架构,面向分布式集群通过容器控制集群,向数据控制面下达网络控制指令;通过控制器,向数据转发面的转发设备发送数据转发策略和数据路由表等网络转发协议;通过数据转发面接入了计算集群,实现算力节点的调度和容器承载。基于可编程网络架构的数据转发,可以改变原有容器资源仅能在 overlay 的数据中心内调度的情况,实现基于 underlay 的跨数据中心的算力资源调度^[11]。

在每一个算力节点上,该架构采用传统的 master/agent 模式来代理、发布算力节点的计算、存储和应用输入输出(IO)等情况,并尝试将这些情况反馈至控制集群中服务器的调度器,从而实现集群内算力节点的统一管理。网络 QoS、QoE 以及转发设备等状态,通过数据转发平台上传至控制平面,



▲图 3 调度器基础资源调度机制图

然后被统一管理。在整个算力资源编排调度的过程中,容器控制集群为用户和开发者提供了统一入口,并通过统一配置脚本、开发变成方式,来实现面向服务的算力和网络的统一编排和调度,从而实现面向服务的基于可编程网络的算力调度。

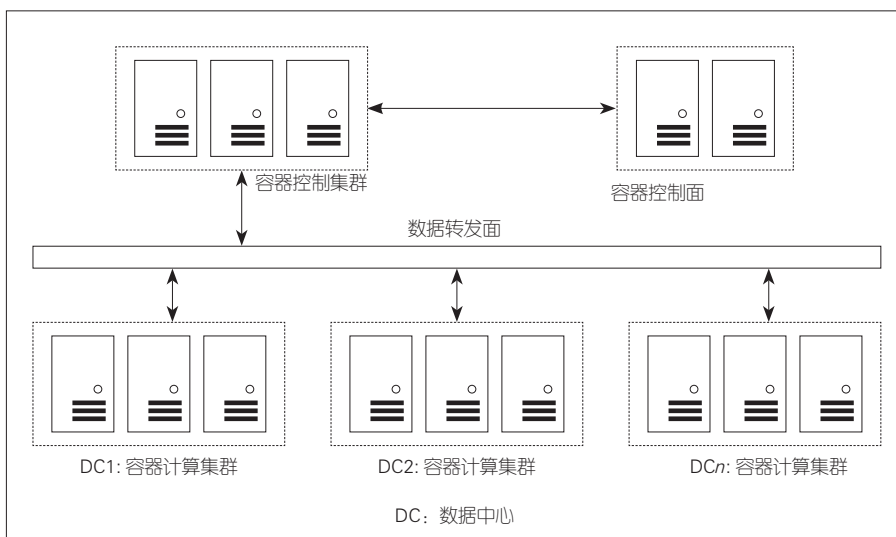
3.2 算力调度机制

在算力调度方面,Kubernetes 云原生平台提供服务编排调度能力,集成网络编排能力和计算服务编排能力,并通过 Knative 实现统一的应用能力封装和消息队列。整体算力调度层为上层门户提供应用程序编程接口(API)网关,也为上层应用提供统一的 API。这可以开放可编程网络的算力能力,屏蔽底层网络和算力的差异性,并且可以为开发者和用户提供统一门户,进一步降低了可编程网络算力调度的开发门槛。算力调度平台具体的架构如图 5 所示。

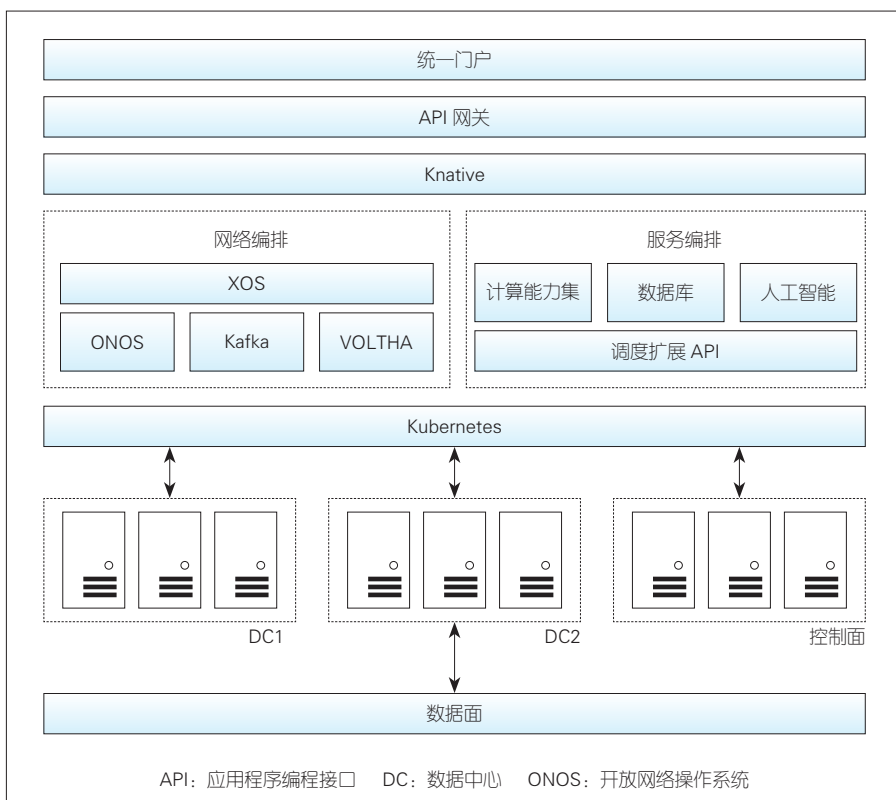
整体算力调度机制由 Kubernetes 实现统一的算力网络资源调度。其中,根据资源服务对象的不同,Kubernetes 调度能力可以分为两个方面:一方面是以基础设施平台即服务(i-PaaS)能力为主,实现对底层基础设施算力资源的调度,借助控制平面的对接来实现对网络数据面的调度和管理,通过对接不同的 Kubernetes 云原生集群实现对底层云原生集群的调度管理;另一方面是以应用层 PaaS(A-PaaS)能力为主,实现对网络编排和计算服务编排的服务能力管理。面向上层的能力调度主要包含网络编排和服务编排两个方面。

(1) 网络编排

网络编排主要是指,对底层的网络服务编排能力进行硬件资源的抽象和能力的建模,并通过服务编排来实现网络控制。本文中,我们提出基于



▲图 4 基于可编程网络的技术架构图



▲图 5 算力调度平台架构图

SDN 的宽带接入(SEBA)容器化架构,以实现 SDN 网络访问。SEBA 的核心组件主要包括开放网络操作系统(ONOS)、Kafka、VOLTHA、XOS。

- ONOS: 实现 SDN 网络操作系统,对网络服务编排实现统一的资源调度和管理。

- Kafka: 实现 REST 的消息队列管理,并通过上层的服务能力对底层硬件的访问请求消息进行统一管理。

- VOLTHA: 实现底层网络接入设备和转发设备的硬件资源抽象,从而使用和访问上层的网络功能。

- XOS: 实现网络功能虚拟化和服

务化, 并可以基于 SDN 控制器的可编程能力实现网络控制和功能软件定义能力。

(2) 服务编排

服务编排可以实现对 PaaS 和软件即服务 (SaaS) 能力的容器化调度。由于云原生具有服务化和微服务化的能力, 因此在实现算力调度的过程中, 基于不同的应用场景, 我们提出了 3 个方面的服务能力。

- 计算能力集: 集成目前云原生统一的计算型能力库, 包括 Spark、Hadoop、Hive、Flink 等。

- 数据库: 采用传统的数据库服务能力, 为上层的应用和业务场景提供一键部署式的云原生数据库, 包括 Mysql、MongoDB 等。

- 人工智能: 包括面向人工智能场景的推理和训练, 以及对硬件加速有特定需求的算力调度能力。

这些服务能力统一由 Kubernetes 来实现编排。通过 Kubernetes 的调度扩展接口和平台内部调度器对接, 从而能够实现 PaaS 和 SaaS 服务的容器化调度。

通过 Knative 来完成统一服务能力的封装和打包, 通过 Knative 的 API 网关提供统一的网络和算力调度接口, 并通过统一的门户对外开放, 开发者可以根据网络和算力调度能力进行网络编程。这样可以进一步融合底层网络和算力, 实现基于可编程网络的算力调度。同时, 用户也可以更加关注上层业务逻辑和业务流程。

3.3 可编程网络编排机制

构建在传统 SDN 架构上的可编程网络算力调度机制, 不仅实现了网络控制面和用户面的分离, 还实现了基于云原生统一 Kubernetes 平台的服务编排调度能力。在可编程网络服务编排能力方面, 随着网络转发设备的普

及, 基于 P4 的网络可编程能力实现了网络的可编程。另外, 网络组件本身也可以进行容器化, 并可以调度到具备 P4 功能的白盒交换机上。基于容器化的可编程网络编排架构如图 6 所示。

依据上述可编程网络算力编排技术架构, P4 交换机集成了专门针对 P4 的运行 (Runtime)。在通用计算节点上, 运行时集成了运行应用程序的镜像。在整个技术架构中, 面向上层的网络功能和应用程序提供了统一的容器封装能力, 用于打包和封装容器镜像。其中, 在网络功能容器化封装的过程中, P4 编译器专门用于服务网络功能程序, 即将网络功能程序编译成可在 P4 交换机上运行的可执行程序后, 再进行容器化封装。Kubernetes 平台实现了 P4 交换机和通用计算节点的算力资源调度和服务编排。根据网络功能和应用程序的不同, Kubernetes 平台分别将网络功能调度到 P4 交换机上运行, 将应用程序调度到通用计算节点上运行^[12]。面向上层开发者则提供统一的开发平台、API、网络可编程能力和应用程序开发能力, 从而实现可编程网络和应用程序。这样一来, 基于可编程网络的算力调度技术在代码开发阶段就能够进行融合开发, 满

足目前越来越多的算网融合场景下的应用程序开发需求。

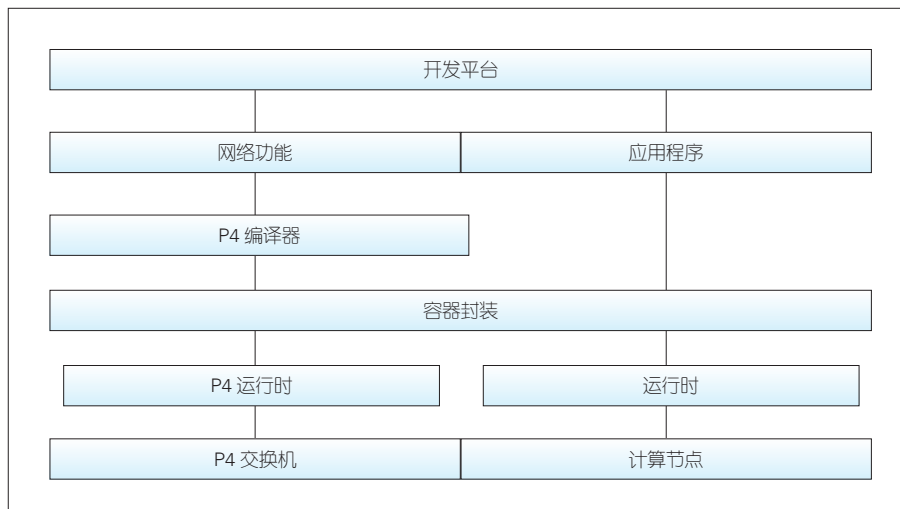
4 结束语

异构算力网络是下一代云网融合 2.0 的发展趋势。基于可编程网络的算力调度机制, 能够在网络可编程的基础上, 实现传统算力调度无法实现的基于网络的算力调度方式。该机制可以根据网络情况进行算力调度, 也可以基于算力调度需求进行网络适配和可编程, 从而真正实现云网融合^[13-15]。本文中, 我们所提出的基于可编程网络的算力调度技术, 是基于云原生技术来实现算力网络的融合调度。传统的云原生调度仅能基于 overlay 网络实现算力调度, 而该技术则可以实现基于 underlay 网络的算力调度和服务编排能力。该技术还可以提供业务感知的网络编排能力, 从而能为后续的网络感知业务提供新的研究思路和发展方向。

致谢

本研究得到中国联通研究院李建飞高级工程师的帮助, 同时也得到了中国联通研究院首席科学家唐雄燕的指导, 谨致谢意!

下转第 61 页 →



▲图 6 可编程网络算力编排架构图

基于 SRv6 的算力网络资源和服务编排调度



Computing Power Network Resources Based on SRv6 and Its Service Arrangement and Scheduling

黄光平 /HUANG Guangping, 史伟强 /SHI Weiqiang, 谭斌 /TAN Bin

(中兴通讯股份有限公司, 中国 深圳 518057)
(ZTE Corporation, Shenzhen 518057, China)

摘要: 提出一种以 IP 网络为中心的算力网络架构, 即在网络域创建云池算力资源和服务的状态, 从而实现网络层的算力编排和调度。算网一体编排和路由, 是该算力网络架构的核心特征。针对算力网络中的服务多实例应用场景, 所提架构方案对 SRv6 或基于 SRv6 的业务功能链 (SFC) 做功能增强和扩展, 以满足单服务对应动态多实例的算力路由需求。控制面架构方案采取一种分级分层状态表的维护机制, 将不同颗粒度的算力资源和服务状态在不同的网络域做同步通告, 并创建对应的分级路由表, 从而压缩节点的状态表和边界网关协议 (BGP) 的通告频率。转发面则执行算力服务标识语义封装, 承载网骨干节点仍然保持无状态转发。

关键词: 算力网络; SRv6; 算力状态; 分级路由

Abstract: An IP network-based architecture of computing power network is proposed, which creates the state of cloud pool computing power resources and services in the network domain to realize the computing power arrangement and scheduling of the network layer. Integrated computing network arrangement and routing are the core features of the computing power network architecture. For the service multi-instance application scenario in the computing power network, the proposed architecture scheme enhances and extends SRv6 or SRv6-based service function chaining (SFC) to support the single service routing requirements for dynamic multi-instances. The control surface architecture scheme adopts a maintenance mechanism of hierarchical state tables, which synchronously notifies the computing power resources and service states of different granularity in different network domains, and creates the corresponding hierarchical routing table, to compress the state table of the node and the notification frequency of the border gateway protocol (BGP). Accordingly, a dual-semantic encapsulation with IP topology and computing service identification in the forwarding plane would also be proposed, while the backbone network nodes would remain unaware of computing power metrics.

Keywords: computing power network; SRv6; computing status; classified routing

DOI: 10.12142/ZTETJ.202103006

网络出版地址: <https://kns.cnki.net/kcms/detail/34.1228.TN.20210623.1809.002.html>

网络出版日期: 2021-06-24

收稿日期: 2021-05-10

在互联网协议 (IP) 承载网络域, 通过精细化动态感知, 网络控制器或网络节点可以创建基于多云池内算力资源及服务状态的算力路由表, 并据此进行算力资源和服务的编排调度。这是以网络为基础平台的算力网

络架构的核心要素。也就是说, 在 IP 拓扑路由的基础上, 新增算力资源和服务路由, 使路由策略约束机制由当前的 IP 拓扑单约束演变为 IP 拓扑和算力双约束。这给网元控制面、转发面和管理面均带来新的挑战, 也是算

力网络为 IP 网络引入的全新议题。

当前主流的云侧应用级跨云池计算资源调度系统, 如内容分发网络 (CDN)、AWS (亚马逊公司的云计算服务) 等, 均与特定应用或应用集群硬绑定。除此之外的其他应用无法

接入该系统纳管的计算资源。此外,这种云侧算力调度系统纳管的云池资源是一种典型的封闭调度平台,仅限于在服务商自营的资源中,且从技术和运营模式上均不兼容多元云池计算资源。更重要的是,这类云侧调度系统与网络资源无关,即它的网络连接服务要么适用于公共网络的“尽力而为”服务,要么适用于专线租用或业务虚拟专用网络(VPN)的开通。网络与计算业务独立配置、独立编排、独立调度。以网络为基础平台的算力网络,构建的是一个开放平台,即与具体的应用和业务完全解耦,且兼容多元云池算力资源和服务。与云侧算力调度显著不同的是,在算力网络架构下,算力和网络的状态和路由表均由网络维护,因此这种算力网络架构内生支持算网一体编排和调度。

然而,一个开放的算力网络平台,可以创建多元云池算力资源、服务状态、路由表,其前提是算力资源和服务的标准化度量和标识。SRv6(基于IPv6的源路由技术)中间转发节点无状态的优良特征,非常适合算网一体路由策略和路由转发,但是需要在转发面和控制面进行功能增强和扩展,以满足算力网络场景下的全新需求。同时,根据应用的算网服务级别协议(SLA)需求,网络需要进行精准灵活的资源匹配和编排,并需要对应用的算力SLA进行更细颗粒度的感知。

1 算力资源和服务的颗粒化度量

当前,云池算力资源和服务的运行模式是与业务强相关,并且高度本地化的,不存在互通和交易,因此尚无系统的度量和标识方案。但是,云池内的算力资源和服务在网络域进行应用流颗粒度的编排和调度,涉及算力资源和服务的跨池跨域调度,以及平台层面的多方资源交易。因此,对

算力资源和服务进行层次化颗粒度的度量和标识,是算力网络架构的关键因素。如图1所示,从交付和执行模式来看,算力资源可以分为3个层次,或称为3种颗粒度。

1.1 算力资源和服务的层次化颗粒度

(1) 基础设施即服务(IaaS)类型算力资源

该类型算力资源属于裸资源,包括中央处理器(CPU)、图形处理器(GPU)、现场可编程门阵列(FPGA)、专用集成电路(ASIC)等。当前这些资源的度量颗粒度,比如核数,无法满足算力网络精细颗粒度的资源调度。因此,需要针对各类异构的计算裸资源进行系统的标准度量。可服务计算资源的标准量化数据,是网络对算力资源感知并创建状态的数量依据。

(2) 函数即服务(FaaS)类型算力服务

虚拟机、容器、微内核等更细颗粒度计算单元的出现,让一些基础计算功能或服务的驻留和运行模式发生根本性的变化。例如,分布式的微服务架构,将传统单一应用系统解耦成独立的微服务群组,应用层根据特定的业务逻辑调用不同的微服务,完成特定的业务功能。

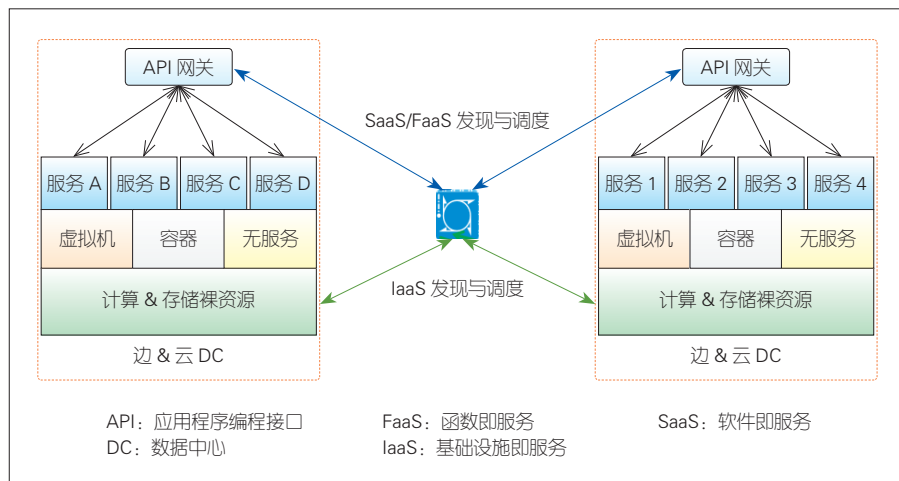
在这种架构下,一些与业务无关的基础计算功能或算法可以实现分布式灵活部署,更加快速地满足新型业务需求,缩短新业务上线周期,大幅降低部署成本。基础计算功能是算力裸资源的一种可服务形态,而算力网络需要创建基于其状态的路由表,并在网络域完成对这种计算功能服务的编排和调度。

(3) 软件即服务(SaaS)类型算力服务

相对于当前增值业务的单站点资源部署和服务模式,在算力网络目标架构下,增值算力服务的驻留和服务将由单点变为全网虚拟SaaS池的模式。同一类增值算力服务资源,在上层交易系统的支撑下,可以在算力网络域完成跨池编排和调度。

1.2 算力资源和服务的度量和标识

如1.1所述,算力资源的标准度量,需要针对上述3种颗粒度的资源和服务进行业务无关的通用度量,以及CPU、GPU等异构裸资源的度量。目前,学术界和信息技术(IT)界已经开始了一些有益的尝试。资源和服务标准化标识的实现,首先需要建立一个结构化的标识体系,对各种颗粒度的资源和服务进行收敛和标定。考虑到网络单元的存



▲图1 层次化算力资源和服务颗粒度

储和处理容量限制,网络域可感知、可编排、可调度的资源和服务标识需要优选数字化标识机制^[1]。

2 基于 SRv6 的算力网络增强控制面技术

在网络域创建、维护云池算力资源和服务的状态,也就是完成对多资源和服务颗粒度的精细化和动态感知,是控制面在算力网络架构下的首要功能。控制面有集中式和分布式两种通用架构技术。

2.1 集中式控制面架构增强

目前的控制器主要有3类。第1类是管理与编排(MANO)控制器,负责纳管移动边缘计算(MEC)内的计算和存储资源、侧重占用率之类的宏观数据,其颗粒度无法满足算力网络的精细化编排和调度需求。因此,可以基于上述算力资源的标准化度量,对MANO纳管的算力资源颗粒度进行扩展和增强。第2类是数据中心和边缘计算中心控制器,负责纳管云内网络拓扑资源。其颗粒度可达服务器对应的端口号,但无法纳管层次化的算力资源和服务。同样,它也可以进行扩展和增强,以涵盖对算力资源的精细化纳管。第3类是IP承载网控制器,负责纳管承载网络域的拓扑资源。

另一种可选方案则是新增算力资源编排器,可与上述3类控制器并列;但也可以居于更上一层,在纳管层次化算力资源的同时,统一纳管数据中心或边缘计算中心、IP承载网的网络拓扑资源,可以实现单点算网全局资源视图。

2.2 分布式控制面架构增强

跨云池的算力资源和服务分布式路由协议,目前主要是基于边界网关协议(BGP)增强和扩展。BGP在现

网通告的对象主要是节点端口、链路等拓扑资源的状态。这些资源的变化周期通常为小时、天,甚至月的数量级,网络的高并发拓扑变更会造成路由震荡等严重后果。在算力资源和服务状态(尤其是FaaS级算力服务的状态)被通告的情景下,其资源标识种类和通告频率均远大于网络拓扑资源及其通告频率。例如,在一些通用计算功能实例中,一次服务执行的生命周期最短可达毫秒级。大规模的通告量和高通告频率,对算力路由表的稳定将造成严重的后果。因此,简单地扩展BGP通告的资源种类,无法解决路由表高度不稳定的问题。本文中,我们提出一种分级通告分级路由的机制,极大地压缩BGP通告的资源数据量和通告频率;还提出一种独立于BGP的全新算力路由协议雏形。

2.2.1 基于 BGP 的分级路由机制

分级分域路由通告的算力网络路由解决方案,旨在解决两个算力网络路由的问题:多种云内算力资源及服务在路由节点上引起的超大路由表项问题、算网端到端路由问题^[2]。

我们将算力资源和服务划分为两种颗粒度:

(1) 边缘计算节点或数据中心的粗颗粒度(颗粒度记为1)算力资源,包括但不限于:

- 计算及存储资源的种类,如CPU、GPU、嵌入式神经网络处理器(NPU)、ASIC等;
- 上述资源种类的可用状态,包括但不限于量化空闲资源值,如使用率、可用核数目等;
- 提供的算力服务种类,包括SaaS/FaaS服务种类及标识,以及服务对应的忙闲状态属性,并且服务的忙闲状态阈值可配置,如90%及以上为忙的状态;

(2) 边缘计算节点或数据中心的细颗粒度(颗粒度记为2)算力服务,包括但不限于:

- 算力服务种类以及其所对应的可服务实例数;
- 每实例的处理容量;
- 算力服务与其实例之间的标识映射关系,如一个任播地址Anycast标识一个算力服务,关联的群组成员地址为实例地址。

粗颗粒的算力资源状态仅在边缘计算节点或数据中心节点之间通告,并维护对应的路由表项。首次上线的节点,通告上述粗颗粒度全集数据,此后根据可配置的变更门限值来触发变量更新通告和同步。通告可有两种方案:BGP扩展方案,即将上述粗颗粒度算力资源信息,通过扩展BGP协议载荷,通告至邻居网络边缘节点;集中式控制器方案,包括但不限于通过路径计算单元通信协议(PCEP)、边界网关协议-链路状态(BGP-LS)等通告同步上述粗颗粒度算力资源相关信息。

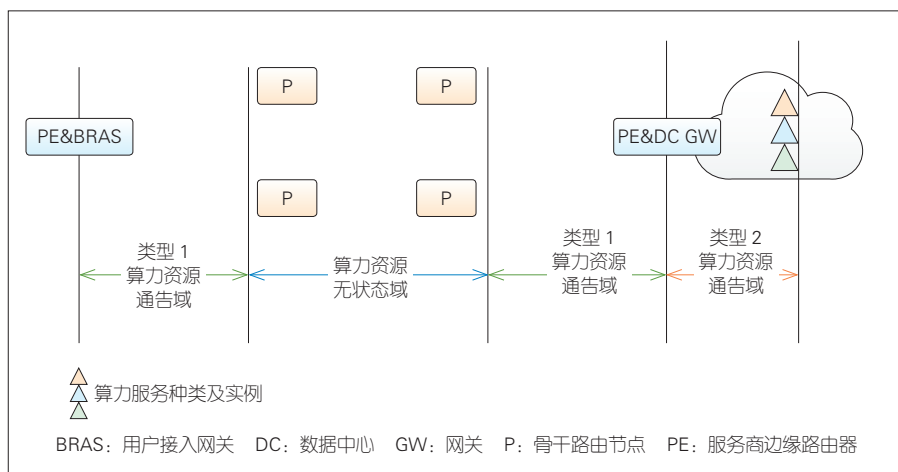
细颗粒度算力服务状态,仅在边缘计算或数据中心节点所归属的域内网络边缘节点进行维护,无须通告邻居网络边缘节点。首次上线的节点,通告或发布上述全集信息,此后根据可配置的变更门限值,触发变量更新通告和同步。细颗粒度的算力服务通过如下可选方案通告网络边缘路由节点:发布订阅的应用消息,并向网络边缘节点通告状态数据;通过内部网关协议(IGP)扩展通告,将上述细颗粒度算力服务信息通过扩展IGP协议载荷,向网络边缘节点通告。

2.2.2 基于 BGP 的地址路由和算力服务路由的两级路由表机制

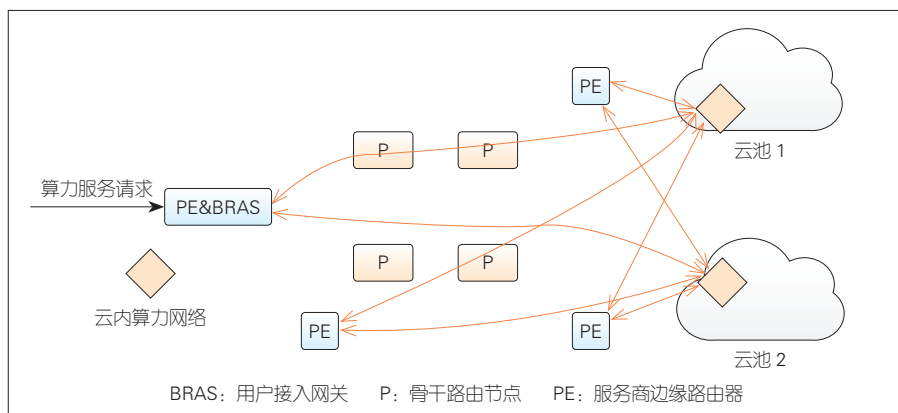
用户接入网络边缘节点维护类型1路由表,即路由节点仅感知边缘计

算或数据中心节点的粗颗粒度算力资源信息，并以此创建、维护对应的算力路由表。类型 1 的算力资源颗粒度较粗，变更频率较低，因此网络边缘节点维护的类型 1 路由表的大小与联动的边缘路由和数据中心节点数目成正比，路由表规模可以得到数量级的压缩。

边缘计算或数据中心节点归属的域内网关或网络边缘节点维护类型 2 算力服务路由表，即上述域内网关或网络边缘节点可以感知本边缘计算或数据中心节点内的算力服务状态，并以此创建、维护对应的算力服务路由表或映射表。类型 2 路由表的大小，与该网络边缘节点、网关归属的边缘计算或数据中心提供的算力服务规模成正比。由于仅做本地的或有限归属边缘计算的或数据中心节点的算力服务信息状态维护，类型 2 路由表规模得到极大的压缩。两级算力颗粒度类型路由及通告机制如图 2 所示。



▲图 2 两级算力颗粒度类型路由及通告机制



▲图 3 基于网络 L4 的新型算力路由协议通告

2.2.3 新型算力路由协议

云内算力资源和服务的种类以及状态变更频率均与现网 IP 拓扑通告有着显著区别。为了适应新型算网一体路由架构，我们提出一种全新的算力路由协议。该协议内生支持算力资源和服务的跨域通告，并将与 BGP 解耦，从而规避算力资源的动态对现网路由收敛的负面影响。网络和算力资源的融合路由策略通过算法优化解决。我们还提出了一种基于网络 L4 的新算力路由协议架构，其主要特征是算力资源和服务在云内直接发布，并由服务商边缘路由器（PE）为其创建算力路由表，如图 3 所示。

两种可能的协议模式为：发布订阅机制和定向通告机制。

（1）发布订阅机制：作为发布主体，云池内算力网关对云内层次化算

力资源进行发布，并对云池内算力资源状态信息进行结构化设计；支持增量发布，支持高频率动态更新；发布对象为网络边缘节点以及用户的接入网关。

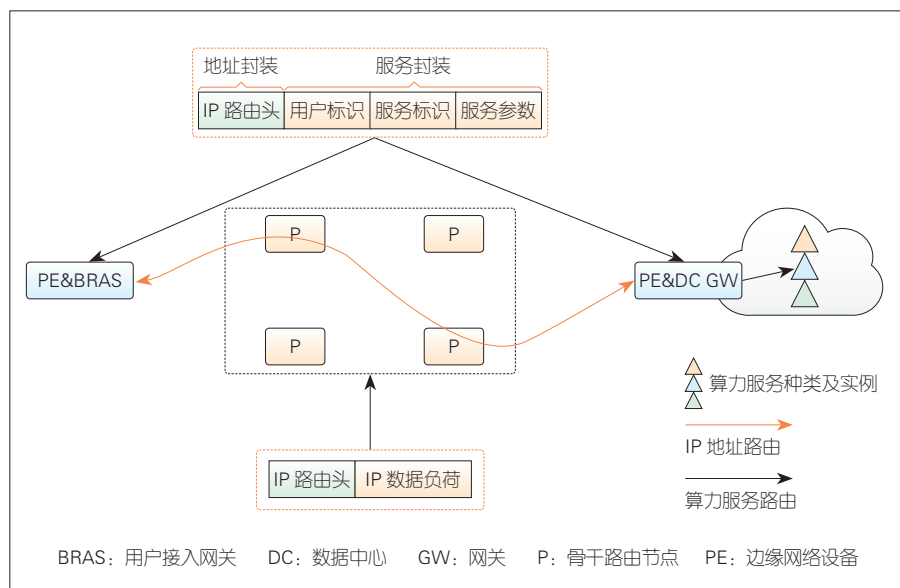
（2）定向通告机制：云内算力网关向网络边缘节点以及用户接入网关主动发起面向连接的状态通告，网络边缘节点以及用户接入网关仅接收通告并据此创建和更新路由表；支持基于隧道的高频率更新通告。

3 基于 SRv6 的算力网络增强转发面技术

算力网络路由是一种集网、云、算为一体的综合路由。在网络入口节点，算力网络路由根据用户业务的算

力和网络双 SLA 约束，制定算网路由策略。和当前 IP 拓扑路由显著不同的是，IP/多协议标签交换（MPLS）拓扑路由本质上解决的是“去哪里”，即明确路由的网络目的节点，在参数上体现为 IP 地址或标签。在算力网络架构下，网、云、算综合路由本质上解决的是“去哪里”+“干什么（执行何种计算服务）”，即在 IP 路由的基础上，叠加了算力服务路由。因此，转发面的报文头需要执行 IP 路由 + 算力服务路由由双重封装。算力网络的 IP 和算力服务双重路由机制网络流程图，如图 4 所示。

如 2.2.2 节所述，在分级路由表的机制下，网络在入口和出口节点，维护有两种不同颗粒度的算力路由表，



▲图4 算力网络 IP 和算力服务双重报文封装和路由机制

这对应转发面的 IP 拓扑和算力服务双重路由封装。在用户接入网关（如 BRAS）处，网络执行上述两级封装，并由用户接入网关根据 2.2.2 节所述本地维护的类型 1 路由表，计算生成到选定的边缘计算或数据中心节点的路由，并执行 IP 拓扑地址封装。我们有两种封装方案：（1）目的地址封装方案，即将选定的边缘计算或数据中心节点归属的网络边缘节点或网关地址，作为目的地址，封装在报文头对应的字段中，包括但不限于互联网协议第 4 版（IPv4）、互联网协议第 6 版（IPv6）、MPLS 等网络数据平面；（2）源路由地址方案，即以选定的边缘计算或数据中心节点归属的网络边缘节点或网关作为出节点，编排源路由路径，并封装在对应的报文头中，包括但不限于 SR-MPLS、SRv6 等网络数据平面^[3]。

用户接入网关（如 BRAS）根据用户算力服务请求执行算力服务标识封装，这包括：单一算力服务标识封装、基于 SRv6 的业务功能链（SFC）、多算力服务标识链封装。算力服务标识的封装包括两种方案：（1）增强 SRv6 算力服务标识编程扩展方案，

即在片段识别（SID）的 Locator + Function（定位器 + 功能）结构中，算力服务标识作为 Function 封装在 SID 中，并可选择扩展 Argument 来作为算力服务的必要输入参数；（2）算力服务标识封装在 IP 与 L4 传输层之间的 overlay 层中，如 SFC 架构下的网络业务报文头（NSH）、三层网络虚拟化 overlay（NVO3）的 Geneve 等，还可以在 IPv6 之上引入一个全新标识层，用于封装算力服务标识，从而实现与 IP 层完全解耦。在这种 IP 拓扑和算力服务双路由封装、点到点路由的机制支持下，网络中间转发节点无须识别算力服务标识，仅做普通路由转发，即平滑继承当前网络中间节点无状态的特征。

类型 1 路由的出节点执行算力服务标识解封装，并查找节点维护的所属边缘计算或数据中心算力服务的路由表或映射表，从而将用户数据路由至对应的服务实例，并终结全部端到端算力路由。

特别地，为了保持流粘性，即确保同一应用的数据流被路由至同一个算力服务实例，出节点维护应用数据

流标识与算力服务实例的映射关系，并将后续应用数据流路由至同一算力服务实例。这种映射关系的维护方法包括但不限于 5 元组方案（源 IP 地址、目的 IP 地址、源端口、目的端口、传输层协议类型）。在 IP 拓扑和算力服务双重封装的机制下，算力服务标识仅仅体现了服务类型的抽象语义，而实际服务实例节点的映射关系被维护在 2.2.2 节所述的类型 2 路由表中。由于路由表具有与业务无关的中性特征，算力业务流粘性的维护保证，需要在出入口节点维护业务相关的状态。在两级路由、两级封装的全流程下，流粘性也需要维护对应的两个颗粒度的状态，即在入口节点维护业务标识和算力服务标识的状态，业务标识可通过类似前述 5 元组的模式实现。在出口节点维护业务标识、算力服务标识和服务标识实例的状态，服务标识实例可以是虚拟局域网（VLAN）/ 虚拟扩展局域网（VxLAN）号、端口号、IP 地址等。

4 网络对算力应用的感知

在当前数据网络的转发和路由机制中，网络资源和策略对应的最小颗粒度是流甚至报文。也就是说，从本质上看，网络路由策略是与业务无关的。在算力网络架构下，网络感知云池算力资源和服务，并根据应用的算力 SLA，在网络层对算力资源和服务进行编排和调度。与当前网络策略和路由机制不同的是，算力资源和服务对应的最小颗粒度是算力应用，且必须与业务相关。当前网络路由策略的聚合服务质量（QoS）机制，无法直接对标算力 QoS 的颗粒度。算力 QoS 更加灵活，不便于聚合，因此算力网络的另一个全新技术挑战是网络层（L3）对应用的算力 SLA 的感知。

由于 ISO 层级解耦的内生架构原

则,当前网络层没有感知接口,对应用无感知。算力网络架构下,应用的算力 SLA 的感知主要有两种方案:一种是控制面方案,即所谓的带外方案,通过类似接入控制信令扩展向网络入口网关通告特定算力应用的 SLA,网络入口网关据此创建算力应用颗粒度的会话。控制面方案的优势是安全、可信、与设备硬件无关。另一种方案是转发面方案,即所谓的带内方案,通过在 IPv6 或 SRv6 的扩展头中增强封装应用标识及其 SLA,网络节点解封装即可执行对应的路由策略。转发面应用感知方案的优势是网络每个节点均可做精细化策略和资源匹配,但这也引入了额外的安全问题,以及大量的冗余硬件设备处理负荷。

5 结束语

算力资源和服务的标准化度量和标识是算力网络中一个重要的支撑要素。层次化资源和服务颗粒度下的度量和标识,带来了精细化的可编排、可调度算力资源和服务体系。在网络

域创建云池算力资源和服务的状态,给控制面尤其是路由协议如 BGP 等带来了挑战。本文中,我们提出了一种基于聚合原则的分级分层路由表机制,即将算力资源和服务分为粗和细两种颗粒度,极大地压缩了路由协议的通告频率和路由表尺寸。同样,在转发面引入基于 SRv6 可编程的增强功能,或扩展 overlay 层的 IP 拓扑和算力服务标识双重语义封装,都能较好地适应 IP 拓扑和算力服务双重路由的全新需求和场景。同样,当前网络 L3 不能感知应用的层级解耦模式,无法应对算力网络的资源匹配和调度需求。这需要通过带外模式,即控制面增强扩展方案来实现网络层对算力应用感知,对现网架构以及设备的影响最小。

参考文献

- [1] 朱海东. 云网一体使能网络即服务 [J]. 中兴通讯技术, 2019, 25(2): 9-14. DOI: 10.12142/ZTETJ.201902002
- [2] 刘铎, 杨涓, 谭玉娟. 边缘存储的发展现状与挑战 [J]. 中兴通讯技术, 2019(3): 15-22. DOI: 10.12142/ZTETJ.201903003
- [3] 马洪源. 面向 5G 的边缘计算及部署思考 [J]. 中兴通讯技术, 2019(3): 77-81. DOI: 10.12142/ZTETJ.201903011

作者简介



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; 发表论文 3 篇, 申请专利 20 余件。



史伟强, 中兴通讯股份有限公司有线架构总经理; 主要研究方向为 IP 网络、光网络和 SDN 系统架构与技术, 先后从事网管、接入网和 SDN 控制器等产品的架构设计和研发管理工作; 获 2012 年国家科技进步奖二等奖等奖项; 发表论文多篇,

申请专利 3 项。



谭斌, 中兴通讯股份有限公司未来网络技术研究项目经理; 主要研究方向为 IP 网络、SDN 系统架构与技术, 先后从事有线路由器、接入产品开发、产品规划和市场等工作; 申请专利 2 项。

算力网络：以网络为中心的融合资源供给



Computing Power Network: A Network-Centric Supply Paradigm for Integrated Resources

李少鹤 /LI Shaohe^{1,2}, 李泰新 /LI Taixin¹, 周旭 /ZHOU Xu¹

(1. 中国科学院计算机网络信息中心, 中国 北京 100190;

2. 中国科学院大学, 中国 北京 100049)

(1.Computer Network Information Center, Chinese Academy of Sciences, Beijing 100190, China;
2.University of Chinese Academy of Sciences, Beijing 100049, China)

摘要：算力网络能够改善边缘和云中心、边缘和边缘的资源互通调度问题，实现算力、存储、网络等多种资源动态调度，并提供极致的服务质量。基于网络计算模型的发展历程和算力网络需求背景，提出算力网络的供给模式和3层服务模式，指出算力网络是一种以网络为中心的多融合资源供给网络计算模型。

关键词：算力网络；以网络为中心；网络计算模型；供给模式

Abstract: Computing power network can improve resource interoperability and scheduling in edge-to-cloud and edge-to-edge scenarios, realize the dynamic scheduling of multiple resources such as computing power, storage, and network, and provide ultimate service quality. Based on the analysis of development process of network computing model and the background of computing power network demand, the supply paradigm and three-layer service mode of computing power network are proposed. It is pointed out that computing power network is a network-centric new network computing model with integrated supply of multiple resources.

Keywords: computing power network; network-centric; network computing model; supply paradigm

DOI: 10.12142/ZTETJ.202103007

网络出版地址: <https://kns.cnki.net/kcms/detail/34.1228.TN.20210617.1103.008.html>

网络出版日期: 2021-06-17

收稿日期: 2021-05-15

随着5G网络时代的到来，以及人工智能、大数据技术的兴起，作为互联网基础设施的计算机网络体系面临巨大的挑战。国际数据公司（IDC）预测，2020—2025年将有超过50%的数据会在网络侧进行存储、计算和处理^[1]。在中国“新基建”战略的指引下，“新联接”和“新计算”成为建设数字基础设施的重要抓手。当前，计算能力供需关系不平衡成为产业创新升级演进的瓶颈。构建弹性开放、高效协同的计算基础设施，成为信息技术（IT）产业与通信技术（CT）产业融合发展的重要共识。

算力网络采用以网络为中心的多融合资源供给网络计算模型，依靠“云数据中心+边缘服务器+用户终端”三级协同（简称“云+边+端”协同），使计算资源从终端、云向边缘扩散，以便提供泛在网络连接和算力服务，实现算力资源的灵活调度。算力网络有望满足智能社会中新型业务对网络的需求，实现“算力无处不在、随取随用”的未来网络场景。

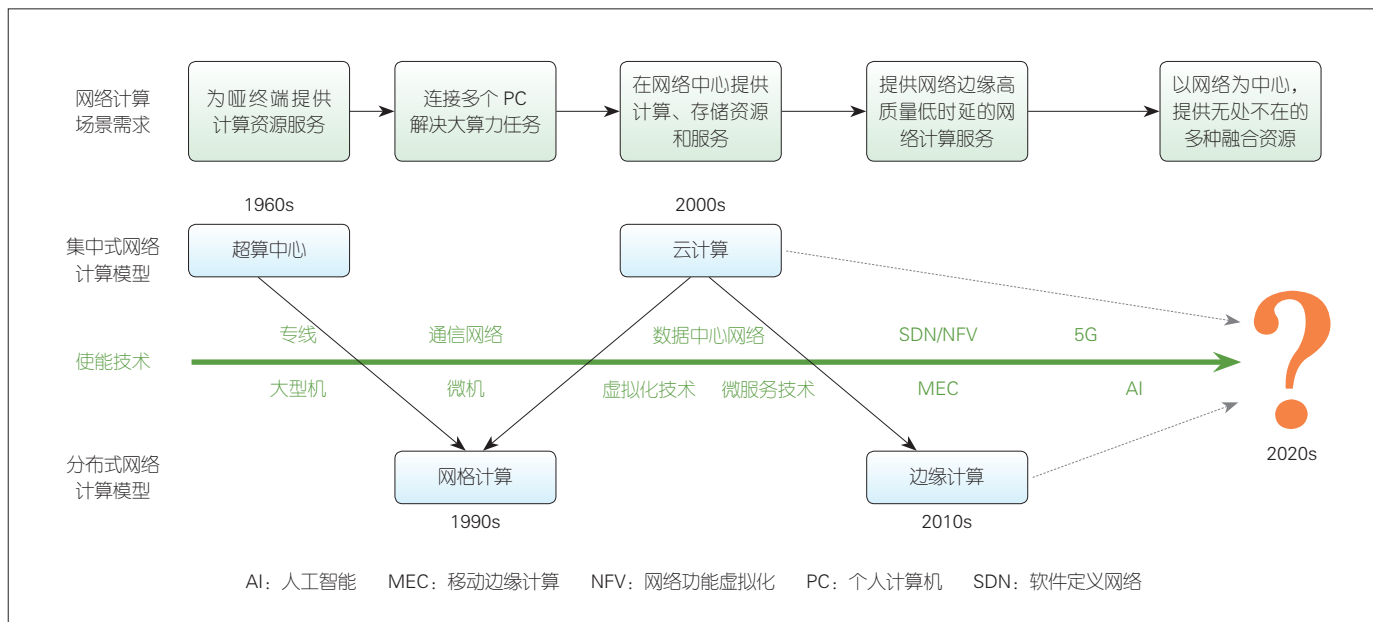
1 网络计算模型发展历程

网络和计算一直以来是计算机领域发展的主线。在两者互补融合的历史

进程中，网络计算模型的变化经历了多个阶段。图1展示了计算网络模型及其使能技术演化和网络计算场景需求变迁的脉络。

在早期的大型计算机时代，用户终端仅仅作为显示器，并通过通信线路连接到大型主机，使用集中点的算力资源。那时的计算资源完全集中，网络只起到终端登录连接的作用，功能单一。

随着个人计算机（PC）的发展，主机间的通信需求促进计算机网络的蓬勃发展。同时，随着PC的普及，计算资源也变得越来越分散。在20世纪



▲图1 网络计算模型演化

90年代，随着计算资源的分散化和计算机网络的发展，研究者提出一种分布式的网络计算模型——网格计算，即利用大量异构的计算机闲置中央处理器（CPU）计算资源和磁盘存储资源，通过网络通信技术，将其作为一个分布式的大规模计算机集群，以完成需要大量算力的计算任务^[2]。网格计算技术的愿景十分宏大。然而，由于这一技术思想过于超前，当时的PC和网络性能均不能支撑大规模分布式网络计算的场景。此外，在商业模式上，闲置资源的利用效率不高。这使得网格计算并没有获得大规模商用，多被用于志愿科学计算。网格计算的提出，为后来网络计算模型的发展提供了思路和技术基础。

随着互联网业务的飞速发展，人们对计算存储能力的需求不断攀升。在2006年提出的云计算便是脱源于网格计算思想的下一代网络计算模型^[3]。通过分布式技术，云计算将计算、存储资源存放在云网络中心，使用户仅通过网络就可以获得庞大的计算资源和存储服务。相对于网格计算完全分

散的计算资源，云计算的计算、存储资源仍是集中部署的。这使得云计算可以完成高可靠且高弹性的资源供给。这与以WEB为代表的互联网服务对资源的需求高度吻合，因此云计算一提出便掀起热潮。云计算衍生出多种云服务模式，是一个里程碑式的网络计算模型。云计算诞生于固定互联网业务需求之下，为移动互联网的发展奠定坚实的基础，并与智能手机这一新型终端一起，掀起移动互联网发展的浪潮。目前，几乎所有通用互联网、移动互联网应用都被部署在云服务器中。可以说，云计算贯穿了整个通用互联网和移动互联网的发展，成为互联网经济的核心推动力。

近年来，互联网技术开始从消费类应用场景向产业应用场景拓展。物联终端、工业设备、智能汽车等更多类型的终端开始联入网络。产业应用的新业务对实时性、可靠性、吞吐能力、能耗等的要求远远高于消费类的应用，网络环境也变得更加复杂。面向产业应用的特殊需求，传统云计算数据中心部署位置距离用户较远，无

法为时延敏感业务提供低时延服务。把海量物联终端数据传输到云计算中心进行处理，将给网络带来巨大的带宽压力。单单依靠集中式的云计算，无法有效支撑产业互联网的发展。在这种需求的推动下，边缘计算应运而生。边缘计算将数据存储、处理和计算下沉到网络边缘，并接近用户终端，可以满足低时延、大带宽、低能耗的网络需求。边缘计算这一概念最早由2009年的Cloudlets演化而来^[4]，并作为5G网络的使能技术，伴随着5G技术进入快速发展期。2014年欧洲电信标准协会（ETSI）提出移动边缘计算。随后，移动边缘计算演化为多接入边缘计算^[5]。边缘计算模型使得算力资源在网络中得到进一步丰富，地理布局更广，提供方式更为灵活，弥补了云计算集中部署带来的时延、带宽方面的弱点，是新一代网络计算模型从“集中”回归“分布”的又一次轮回。

按照时间顺序，表1给出了网络计算模型发展历程中的关键技术和思想对比。从前文描述可以看出，一项技术的成功，除了技术本身的先进性

▼表1 网络计算模型关键技术对比

名称	提出时间/年	模型架构	网络计算思想	关键技术	场景/应用	存在问题
超算中心	1964	三大网络： 管理/计算/存储网络	通过网络将待处理任务 传输到超算中心	多核处理技术、高速大容量 数据缓存技术	中科院超算中心	能耗高，场景受限，时延高
网格计算	1995	5层沙漏模型	将闲置计算、存储资源 组成网格网络	网格计算量动态选择技术	BOINC等志愿计算平台	难以商用，网络计算效率不高
云计算	2006	3层架构： IaaS/PaaS/SaaS	通过网络将待处理任务 传输到云数据中心	虚拟化技术、容器技术	Azure、阿里云、华为云、 天翼云	网络能力成为瓶颈， 用户隐私安全难以保证
边缘计算	2009	3层架构： 网络/主机/系统层	将计算资源下沉至网络 边缘，提供便捷服务	计算卸载技术	面向工业园区、物联网的 边缘计算平台	边缘算力不足， 边缘之间难调度

IaaS：基础设施即服务 PaaS：平台即服务 SaaS：软件即服务

以外，真实存在的产业需求、可行的商业模式都是决定性的因素。

2 算力网络需求背景

2.1 算力网络发展背景

在5G人工智能（AI）时代，新型网络业务持续涌现，对算力的需求飞速增长。高算力和低时延的应用场景愈加多样化，如物联网、智慧出行、虚拟现实、泛在计算等。这些场景对算力的需求亦呈现多样化爆发式增长。随着万物互联愿景的进一步推进，联网终端和设备数量将呈现指数级增长。据Statista预测，2025年全球物联网设备将超过750亿台^[6]。用户对于时延、带宽的变化更为敏感，对服务质量的要求进一步提高。

与此同时，算力资源的供给也将进入快速发展期。2020年4月20日，中国国家发展和改革委员会首次明确新型基础设施建设（简称“新基建”）的范围。其中，信息基础设施包括以数据中心、智能计算中心为代表的算力基础设施等。这是“算力基础设施”这一概念首次在国家层面的被提出。目前，多种算力供给设施正在大力建设中，如超算中心、云计算数据中心、智能计算中心、边缘计算站点。有报告指出，与已投运机柜数相比，2020年北京、上海、广州、深圳周边的在建和规划数据中心机柜增长超过

了300%，这说明算力资源供给进入快速增长期^[7]。相比于2020年，2025年以边缘计算为代表的分布式算力资源将增长790%，超过集中式算力资源。IDC预计，未来中心化算力占比将不超过12%，分布式算力将超过88%^[7]。

随着算力需求和算力供给的飞速增长，算力供需之间的不平衡问题愈加凸显。IDC数据表明，计算资源的综合利用率普遍小于15%。特别是边缘计算节点，由于均是面向特定场景建设的，计算负载的潮汐效应往往更加明显，单靠目标应用场景，难以消耗边缘计算节点的所有算力资源。目前，云计算中心和边缘计算节点之间、边缘节点和边缘节点之间的计算资源调度不够灵活，集中式算力资源和分布式算力资源发展不一致，导致算力供给与需求无法有效匹配，使局部过载而其他资源闲置的情况出现。这大大降低了算力资源作为信息社会底层基础设施能力的效率。由于现有网络系统存在局限性，业务大多属于静态部署，资源复用率低。网络配置也多为静态，路由寻址方式效率低，难以针对目前轻量级的微服务需求进行优化。算力资源需要与网络结合，更大范围、更细粒度地有效匹配和调度，才能充分发挥海量算力的真正效用。

2.2 算力资源和网络能力适配

算力在“集中-分布”模型间呈

现钟摆式变化。随着边缘计算的兴起和智能社会的算力需求发展，算力即将进入“集中+分布”的全新发展阶段。

集中式算力资源可以高效处理需要大算力的计算任务。分布式算力资源可以为终端用户提供高质量、低时延、随用随取的算力服务。面对“云+边+端”网络协同和“集中+分布”算力协同的场景需求，以及为解决算力资源供给失衡的问题，网络在新型网络计算模型中将会占据更重要的位置。网络的功能将从“连接算力”（为数据中心、算力节点和用户终端提供连接功能）转向“调度算力”（通过网络对算力节点间的算力资源分配和调度），甚至转向“组织算力”（对整个网络中的异构算力资源进行编排和组织管理）。新需求对网络能力的要求进一步提高，即要求网络可以容纳、调度、编排多种地理布局、多种物理异构，并提供海量的计算、存储、连接资源。新型网络计算模型将会以网络为中心，实现算力资源和网络能力的有效适配，最大限度地提供高效的网络算力调度和编排。

3 算力网络供给模式与核心特征

3.1 网络计算资源的组织与供给模式

在目前的网络生产关系下，产业链各方不同程度地掌握了应用需求、计算资源、网络资源。与之对应，网

络计算资源的组织与供给可能有以下 3 种模式，如表 2 所示。

(1) 以应用为中心。在这种模式下，算力资源的组织与调度以自身的业务生态为中心。具有代表性的算力服务商有百度、阿里、腾讯等大型 OTT（指互联网公司越过运营商）互联网企业。它们在进行算力资源的部署时，以自建为主、整合第三方算力资源为辅，并基于 OTT 模式，租用运营商网络资源实现传输与调度。以应用为中心的模式本质上是云计算模式的扩展。对于互联网公司来说，这种模式具有业务延续性好、技术成熟、成本相对较低的优点。然而，由于不同互联网业务生态系统之间存在互斥，这种模式的算力网络服务较难保证第三方的公立性。同时，由于采用 OTT 模式，不直接掌控网络资源，该模式难以支持高可靠、低时延业务。

(2) 以计算为中心。在这种模式下，算力资源主要来自原分散的第三方算力，通过服务网络加以组织，来执行计算任务。此模式以新兴的区块链算力网络公司为代表，例如 BHP、EXODUS、Computing Planet 等^[8]。它们的算力资源多为整合的第三方算力资源，自建算力资源比例较低。基于 OTT 模式，它们租用运营商网络资源来构建 overlay 的算力网络，在应用层实现传输与调度。然而，以计算为中心的模式也存在一些问题：难以保证服务质量，支持的业务类型较为有限，管理成本相对较高。

(3) 以网络为中心。这种模式强调直接使用底层网络对算力进行整合、组织和调度，通过优化后的算力标识、路由协议与传输协议，来实现算力资源与网络资源的高度集成与协同调度。具有代表性的算力服务商为网络运营商，其算力资源部署以自建为主、第三方为辅，且拥有大量的边缘计算节

点，同时算力资源分布广泛、类型丰富。这种模式具有完整的底层网络资源和调度管理能力，可以按照应用需求来按需调度合适的算力资源，并保证网络传输质量。以网络为中心的模式具有明显优点：算力服务的中立性最高，服务质量可保证，管理成本较低。

从以上分析可以看出，以网络为中心的算力组织与供给模式最符合未来多元化业务发展需求，也最符合算力作为智能社会底层基础设施的定位。

3.2 算力网络核心特征：以网络为中心的融合资源供给

算力网络是以网络为中心的新型网络计算模型。基于最新网络技术，如运用网络功能虚拟化技术（NFV），算力网络可以有效地将异构算力资源虚拟化。此外，通过云网融合技术和软件定义网络技术（SDN），算力网络还可以将网络中的计算、存储、连接资源进行智能化的有效编排，将计

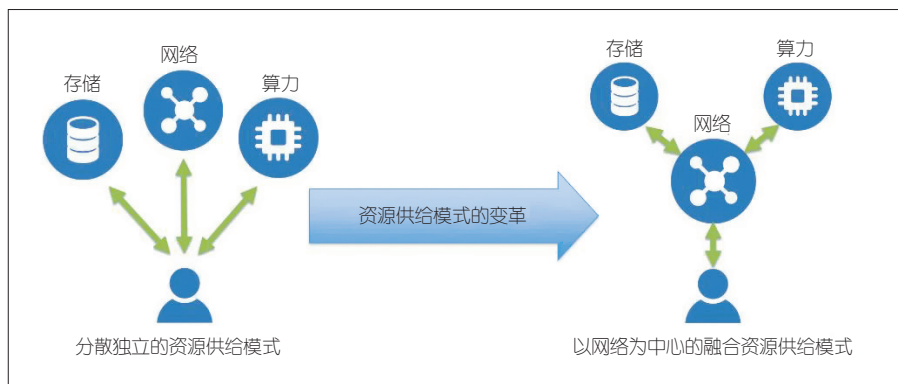
算资源、存储资源等多维异构资源完全融入到网络系统中，使各类资源节点在网络中可以通信和实时交互，并进行多维异构资源的动态调度。算力网络的出现是资源供给模式的变革，它通过网络来连通分散且碎片化的计算、存储等资源，构建一体化的信息通信技术（ICT）基础设施，并向各相关产业提供网络、计算及存储等服务，如图 2 所示。

在算力网络中，与算力任务相关的算力网络能力，比如算力资源、服务标识，在网络边缘进行收集，并向用户提供算力网络能力视图。用户通过应用，发出初始报文携带算力任务的需求，比如需要的算力种类、算力总量、服务名字、时延上限、任务拆分等。基于收集到的算力网络能力详情和相应需求到能力的映射算法，算力网络边缘网关将用户的算力任务需求组合，映射为相应的算力网络能力组合。然后基于预设语法，将算力

▼表 2 网络计算资源的组织与供给模式对比

项目	以应用为中心	以计算为中心	以网络为中心
代表	大型互联网企业	区块链算力服务企业	基础网络运营商
计算资源	自身为主（云）	第三方为主	自身为主（边缘+云）
网络调度能力	OTT 模式	OTT 模式	运营商底层调度
服务中立性	低	高	高
服务质量	可保证	难保证	可保证
成本	低	高	低
服务类型	局限	局限	丰富

OTT：互联网公司越过运营商



▲图 2 资源供给模式变革

网络能力组合解析为相应灵活的报文格式。这些报文携带调用相应算力网络能力的指令和元数据，以便完成算力任务。

在算力网络服务模式下，用户无须在海量、分散的算力供应商中选择合适的算力点，也不必担心网络能力如何与计算模型相匹配。算力网络将帮助用户发现性能最佳的算力资源，按照业务需求来规划可靠的网络路径与传输服务，并帮助用户高效率地完成业务，使成本降到最低。

4 算力网络服务模式

算力网络吸纳和调度各类分布式的算力，以统一服务的方式，并结合确定性网络输送高可靠、可度量、通用化的算力资源，来使能人工智能应用，体现网络价值。运营商或者第三方公司建设供给侧的算力供给资源池，并通过算力网络完成基础设施供给、网络连接供给和平台及业务能力供给，以满足需求侧的虚拟现实、云渲染、自动驾驶以及 AI 等应用的算力需求。

算力网络服务形态决定了算力调度方式、度量方式、盈利模式和商业模式。类比于云计算的基础设施即服务（IaaS）、平台即服务（PaaS）、软件即服务（SaaS）的 3 层服务模式，算力网络也可以自底向上分为 3 种服务形态，如图 3 所示。

（1）算力基础设施服务形态。算力服务商提供基础算力设施，算力资源以算力站点结合虚拟网络的形态存在。需求侧租赁算力网络服务商提供的算力资源，并由算力网络调度用户请求并使之到达合适的算力部署站点，同时由用户决定算力设施的使用方式。这种服务形态主要面向那些拥有较强 IT 开发能力、产品对算力依赖强、有能力对算力进行精细化管理的客户，如大型互联网公司、独角兽企业、云

计算中心。

（2）算力平台服务形态。算力服务商提供算力平台和开发环境，算力资源以算力资源池的形态存在。用户无须关注算力资源部署细节，即可基于算力平台来开发和利用高质量、低时延的算力。这种服务形态主要面向那些具备一定 IT 开发能力、产品对算力较依赖的客户，如大型工业制造企业、中小互联网企业。

（3）算力软件服务形态。算力服务商提供上层算力应用服务，算力资源以应用程序接口（API）的形态存在。用户无须关注系统和软件细节，仅提交业务需求，并通过 API 完成算力任务。这种服务形态主要面向那些 IT 开发能力弱、产品对算力较依赖、无法对算力进行精细化管理的客户，如互联网小微企业、物联网公司、制造业企业等。

算力网络服务形态亦是算力网络研究者的重点研究方向。算力网络将打破传统运营商仅贩卖网络连接和流量的盈利模式，有助于扩展算力需求的商业客户。3 层算力网络服务形态模型具有自下而上垂直拓展的特点，针对不同的算力业务需求，可以提供全方位和高自由度的实现方式，能够给予算力网络供应商更多可能的服务

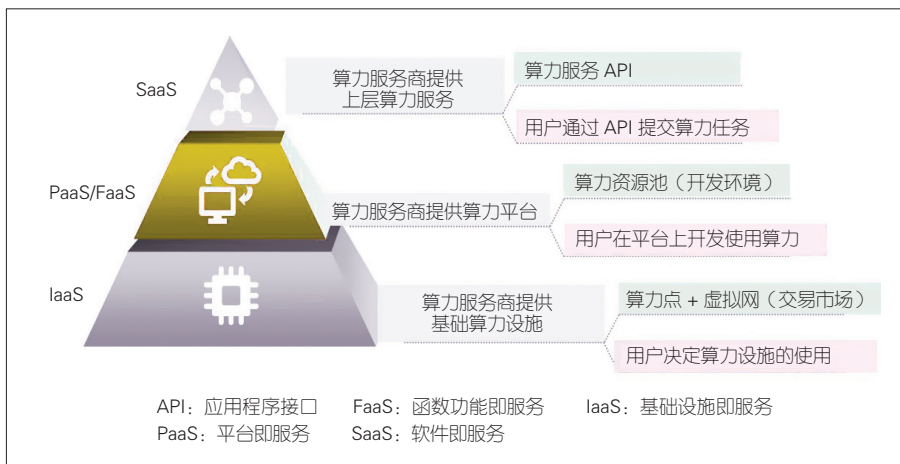
方式和业务模式。

5 算力网络发展现状

算力网络架构一提出便备受业界关注，业界对于算力网络的研究正在如火如荼地开展中。华为提出算力网络基础框架计算优先网络（CFN）。三大运营商也在着力建设算力网络架构设施，并发布对算力网络的研究成果^[9-12]。为满足未来科学研究范式向基于大数据发展的需求，中科院计算机网络信息中心提出面向科研大数据的算力网络架构。

算力网络的标准化和产业化也在持续发展中。2019 年末，华为和移动基于 CFN 技术提出 3 项国际互联网工程任务组（IETF）标准草案。目前，这 3 项标准仍在持续更新中^[9, 13-14]。2019 年 10 月，在国际电信联盟（ITU）全会上，中国移动提出算力感知网络（CAN）相关草案，华为、中国联通和中国电信也提出算力网络（CPN）的相关草案^[15-16]。同时，华为在宽带论坛（BBF）上也进行城域算力网络（MCN）的立项^[17]。目前，华为和运营商在 ETSI 和中国通信标准化协会（CCSA）也在积极推进算力网络相关标准的立项工作。

在产业动态方面，网络 5.0 产业



▲图 3 算力网络服务形态

和技术创新联盟也在积极参与算力网络的生态建设。联盟成员包括华为、三大运营商、中科院计算机网络信息中心、中国信息通信研究院、清华大学等。该联盟成立“算力网络特设工作组”，将5G智能云化网络架构推进为算力网络^[18]。目前，华为、中国移动、中国电信均已完成算力网络的初步试验部署，并展示了试验验证成果。此外，中国联通也在开展算力网络服务平台的试点工作。

6 结束语

如同电力的普及奠定了工业社会发展的基础一样，泛在算力将成为智能社会发展的基石。在政策上，算力网络是中国新基建概念中算力基础设施建设的核心之一，它响应了建设数字化、智能化社会的政策号召；在技术上，算力网络符合5G+AI技术建设乃至6G网络建设的要求，并可以起到关键作用；在商业应用上，算力网络将为各行各业提供高质量、低时延、大带宽的网络、计算、存储服务，符合未来网络业态的良性发展趋势。

参考文献

- [1] REINSEL D, GANTZ J, RYDNING J. Data age 2025: the evolution of data to life-critical [R].

- 2017
- [2] LAN F, CARL K. The grid: blueprint for a new computing infrastructure [M]. San Francisco: Morgan Kaufmann publisher, 1999
- [3] AZODOLMOLKY S, WIEDER P, YAHYAPOUR R. Cloud computing networking: challenges and opportunities for innovations [J]. IEEE communications magazine, 2013, 51(7): 54–62. DOI:10.1109/MCOM.2013.6553678
- [4] SATYANARAYANAN M, BAHL P, CACERES R, et al. The case for VM-based cloudlets in mobile computing [J]. IEEE pervasive computing, 2009, 8(4): 14–23. DOI:10.1109/MPRV.2009.82
- [5] ETSI. Multi-access edge computing (MEC) framework and reference architecture: ETSI GS MEC 003–2019 [S]. 2019
- [6] Statista Research Department. Internet of Things – number of connected devices world-wide 2015–2025 [R]. 2019
- [7] 前瞻产业研究院. 2021–2026年中国IDC（互联网数据中心）行业市场前瞻与投资战略规划分析报告 [R]. 2020
- [8] BHP工作组. BHP全球智能算力网络项目周报 [R/OL]. [2021–03–18]. <https://m.huoxing24.com/userCenter/4c1079cbe18d4e57a7722a2f-d676763a>
- [9] LI Y, HE J, GENG L, et al. Framework of compute first networking (CFN) [Z]. 2019
- [10] 中国移动研究院，华为技术有限公司. 算力感知网络白皮书 [R]. 2019
- [11] 中国联合网络通信有限公司. 中国联通算力网络白皮书 [R]. 2020
- [12] 雷波，刘增义，王旭亮，等. 基于云、网、边融合的边缘计算新方案：算力网络 [J]. 电信科学，2019, 35(9): 44–51
- [13] GU S, ZHUANG G, YAO H. A report on compute first networking (CFN) field trial [Z]. 2019
- [14] GENG L, WILLIS P. Compute first networking (CFN) scenarios and requirements [Z]. 2019
- [15] GENG L, LEI B, FU Y, et al. Use cases and requirements of computing-aware networking for future networks including IMT–2020: ITU–T 480–WP1 [S]. 2019
- [16] HUAWEI Technologies Co. Ltd, China Telecom. Framework and architecture of computing power network: ITU–T 563–WP3 [S]. 2019
- [17] HUAWEI Technologies Co. Ltd. Metro Computing Network (MCN) [EB/OL]. [2021–03–

- 18]. <https://wiki.broadband-forum.org/display/BBF/SDN+and+Nfv>

- [18] 网络5.0产业联盟. 网络5.0产业联盟CFN特设组倡议与筹备汇报 [Z]. 2019

作者简介



李少鹤，中国科学院计算机网络信息中心在读硕士研究生；主要研究方向为新型网络技术、网络流量感知和预测等。



李泰新，中国科学院计算机网络信息中心助理研究员；主要研究方向为网络协议、机器学习、天地一体化网络等；发表论文20余篇，申请专利10余项。



周旭，中国科学院计算机网络信息中心研究员；主要研究方向为未来网络架构、5G/B5G、天地一体化网络、人工智能、边缘计算等；发表论文70余篇，申请专利40余项。



多层次算力网络集中式不可分割任务调度算法

Centralized Unsplittable Task Scheduling Algorithm for Multi-Tier Computing Power Network

巩宸宇/GONG Chenyu¹, 舒洪峰/SHU Hongfeng², 张昕/ZHANG Xin²

(1. 上海科技大学, 中国 上海 200120;
2. 深圳市智慧城市科技发展集团有限公司, 中国 深圳 518046)
(1. ShanghaiTech University, Shanghai 200120, China;
2. Smart Cities Group, Shenzhen 518046, China)

摘要:根据算力网络不同层次的特性和各种应用的不同需求,提出一种多层次算力网络模型和计算卸载系统,并定义一个由时延、能耗组成的加权代价函数以建模一个任务调度问题。为解决这一问题,提出一个基于交叉熵的集中式不可分割任务调度(CUTS)算法。数值仿真结果表明,与其他基线算法相比,该算法在系统平均代价方面拥有较好的性能。

关键词:多层次算力网络;交叉熵;集中式;任务调度;不可分割

Abstract: According to the characteristics of different layers of computing power network and different requirements of various applications, a multi-tier computing power network model and computation offloading system are proposed. Specifically, a cost function consisting of latency and energy consumption to model a task scheduling problem is defined. To solve the problem, a centralized unsplittable task scheduling (CUTS) algorithm based on cross-entropy is introduced. Simulation results show that the algorithm provides superior performance in terms of the average system cost compared with other baseline solutions.

Keywords: multi-tier computing power network; cross-entropy; centralized; task scheduling; unsplittable

DOI: 10.12142/ZTETJ.202103008

网络出版地址: <https://kns.cnki.net/kcms/detail/34.1228.TN.20210617.1144.010.html>

网络出版日期: 2021-06-17

收稿日期: 2021-05-13

近年来,随着深度学习的不断发展,人工智能服务和应用大量涌现,比如人脸识别、自然语言处理、虚拟现实、增强现实等。这些应用通常都是计算密集型任务,将消耗大量的终端资源(如算力和能耗)。然而,由于计算能力和能量供应有限,终端设备(例如手机)可能无法提供良好的服务质量。为此,研究者们提出云计算的概念。

云计算^[1-2]是由分布式计算、并行处理、网格计算发展而来的新型计算模型。通过虚拟化技术建立强大的

资源池,云计算使各种应用和服务能够按需获取算力、存储资源及各种软件资源。云计算为海量数据的处理提供了可能,同时也为计算密集型的人工智能应用提供了强大的算力。然而,端与云之间的传输时延使得云计算无法满足时延敏感型应用的需求。因此,雾计算和边缘计算^[3-4]的概念被提出,以解决云计算传播时延大的问题。

边缘计算是指,在靠近物或者数据源头的一侧部署设备,提供计算、存储等软件服务,并通过算力和通信

资源的联合分配,满足应用的时延需求。经典的边缘计算网络由雾节点和本地用户共同组成。其中,本地用户通过任务拆分和任务卸载决策,来达到全局时延和能耗最小的最优效果。此前,学者们的研究主要集中在单用户多节点^[5-6]和多用户单节点^[7-8]。文献[9]研究了多用户多节点这一应用场景。研究表明,边缘计算可以降低传输时延。但是对于一些对算力和时延都有较高要求的应用来说,边缘计算网络将不再适用,比如自动驾驶、虚拟现实等。因此,算力网络^[10]

的概念被提出。

算力网络涉及云计算、雾计算、边缘计算等。算力网络是由云边端等设备构成的多层次资源网络,它能够将在云边端进行统一调配,但是如何实现系统的最优性能仍是一个难题。原因主要有两点:(1)云边端各有其特性。云距离端较远但算力强,多用于处理全局任务;边距离端较近但算力弱,多处理本地实时任务。(2)用户任务的需求不同。计算密集型任务可能更多地需要云的参与,时延敏感型任务可能更多地需要边的参与,对算力和时延同时有较高要求的任务则需要联合进行调度。基于以上原因,文献[11]研究了多层次算力网络,并提出一种分布式调度算法。但是该模型是边端混合的两层算力网络,并未考虑云的作用。

试想存在如下场景:一座办公大楼内有多层,每层都有多间办公室,且每间办公室都有多个用户和不同性质的任务。由于职能划分不同,不同部门通常所需要的算力不尽相同。这就容易造成算力资源的不合理利用,甚至造成任务中断。如果我们按照办公室和楼层的位置,将其构造成一个多层次算力网络,进行任务的调度和算力分配,那么就能够更好地满足计算密集型和时延敏感型应用的需求。

1 计算卸载系统建模

1.1 系统概述

本节将详细介绍一个多层次算力网络和计算卸载系统,定义一个由时延、能耗组成的加权代价函数,并建模一个任务调度问题。

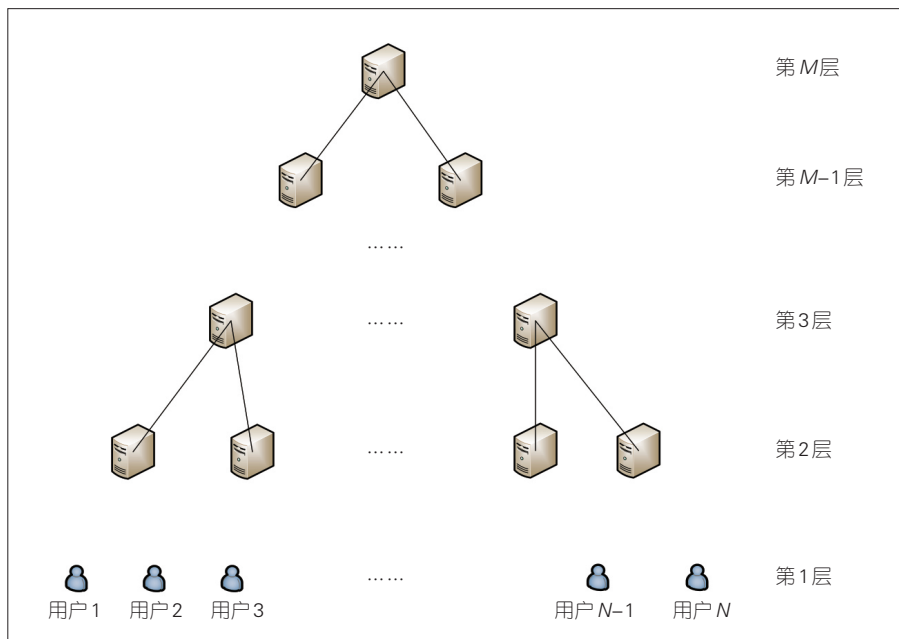
算力网络一共有多层。第1层为用户节点,其他层为雾节点。雾节点的算力随层数的增加而上升。通常,

距离用户较远的高层雾节点算力比较强大,但是往返时延较长;距离用户较近的低层雾节点往返时延较短,但是算力有限。在考虑时延和能耗的基础上,用户可以将不可拆分的任务卸载到某层的某个雾节点,也可以选择将任务在本地执行。因此,如何根据时延和能耗帮助用户做出卸载决策,是解决任务调度问题并获取全局最优解的核心。

如图1所示,算力网络一共有 M 层,第1层包含 N 个用户,每个用户都有一个任务。用户集合 $\mathcal{N} = \{1, 2, \dots, N\}$ 也可以看作任务集合。此处,我们假设每个任务都不可拆分,即如果用户选择把任务卸载到雾节点,那么只能卸载到某一个雾节点上。我们将用户 $n \in \mathcal{N}$ 的计算任务表示为 (z_n, γ_n) ,其中, z_n 是任务的大小(位), γ_n 是任务的处理密度(中央处理器转数/位, cycle/b),即处理单位比特数据所需要的(CPU)中央处理器转数。第2层至第 M 层是雾节点。其中,低层雾节点和高层雾节点均为汇

聚式连接方式,即第 $m-1$ 层的某几个雾节点向上汇聚并连接第 m 层的某一个雾节点。如果用户选择将任务卸载到第 m 层($m > 2$),因为存在汇聚式的连接方式,当第2层中转雾节点确定后,用户任务卸载到第 m 层的节点是确定的。因此,用户在3层以上的决策空间为层数。如果第2层的雾节点有 K_2 个,那么第2层的雾节点可以表示为 $\mathcal{K}_2 = \{1, 2, \dots, K_2\}$,第 m 层的雾节点表示为 \mathcal{L}_m ($m > 2$),所有雾节点的集合表示为 $\mathcal{K} = \mathcal{K}_2 \cup \mathcal{L}$,且 $\mathcal{L} = \{\mathcal{L}_3, \mathcal{L}_4, \dots, \mathcal{L}_M\}$ 。用户 n 的卸载策略 $a_n \in \{0\} \cup \mathcal{K}$,可以表示为:若 $a_n = 0$,则用户选择在本地运行任务;若 $a_n \in \mathcal{K}$,则用户选择将任务卸载到雾节点上。对于所有用户的算法调度策略,我们用 $A = \{a_1, a_2, \dots, a_n\}$ 来表示。所有选择雾节点 $k \in \mathcal{K}$ 的用户数 $n_k(A) = \sum_{n=1}^N I_{\{a_n=k\}}$,其中 $I_{\{x\}}$ 是指示函数。如果 x 为真,则 $I_{\{x\}} = 1$,否则 $I_{\{x\}} = 0$ 。

用户与第2层雾节点之间为全连接式结构。多层次算力网络不考虑



▲图1 多层次算力网络模型

隔层直接通信。

1.2 通信模型

每个用户都有两种卸载决策。当用户决定将任务在本地执行时,此时并没有通信产生的时延;而当用户决定将任务卸载到雾节点时,就会存在通信时延。

假设雾节点 k 只有一个用户 n 接入,则从用户 n 到雾节点 k 的最大传输速率表示为 $R_{n,k}$ ^[12]。但是一个雾节点往往会有多个用户接入,此时各个用户需要竞争来获得通信带宽。根据共享模型,我们将带宽进行均分。当有多个用户接入节点 k 时,用户 n 的传输速率表示为公式(1):

$$R_{n,k}(A) = \frac{R_{n,k}}{n_k(A)} \quad (1)$$

与大多数工作一样,与卸载到雾节点的任务大小相比,由于从雾节点返回用户的结果太小,并且下行速率要比上行速率大得多,因此,结果返回产生的通信时延忽略不计。用户 n 将任务卸载到第3层雾节点 k 上的通信时延如公式(2)所示:

$$T_{\text{trans},n,(2,k)} = \frac{z_n}{R_{n,k}(A)} = \frac{z_n n_k(A)}{R_{n,k}} \quad (2)$$

如果用户将任务卸载到第 m 层 ($m > 2$) 雾节点,首先用户需要将任务卸载到第2层雾节点,然后通过有线网将任务向上传输。假设任务在上传时经过的第2层雾节点为 q^* ,则

$$q^* = \arg \min_q \left\{ T_{\text{trans},n,(2,q)}; q = 1, 2, \dots, K_2 \right\} \quad (3)$$

我们假设相邻层雾节点之间的往返传输时间为常数,并用 t_c 表示。那么,用户将任务卸载到第 m 层的通信时延可用公式(4)表示:

$$T_{\text{trans},n,(m,k)} = T_{\text{trans},n,(2,q^*)} + (m-2)t_c \quad (4)$$

综上所述,用户 n 卸载到雾节点的通信时延如公式(5)所示:

$$T_{\text{trans},n,(m,k)} = T_{\text{trans},n,(2,k)} I_{\{a_n \in K_2\}} + \sum_{m=3}^M T_{\text{trans},n,(m,k)} I_{\{a_n = l_m\}} \quad (5)$$

因为用户的通信能耗只存在与第2层雾节点的通信过程中,所以用户 n 将任务卸载到雾节点 k 上的能耗可用公式(6)表示:

$$E_{\text{trans},n,(m,k)} = P_{\log,n} T_{\text{trans},n,(2,q^*)} = P_{\log,n} \frac{z_n n_k(A)}{R_{n,q^*}} \quad (6)$$

这里, $P_{\log,n}$ 表示用户 n 向雾节点 k 发送任务时的发送功率。我们假设这一功率是恒定的。

1.3 计算模型

1.3.1 本地计算模型

当用户决定在本地执行任务时,则存在计算时延和计算能耗。我们将用户 n 的计算能力表示为 f_n ,即CPU的时钟频率(CPU转数/s),那么本地计算时延可以用公式(7)来表示:

$$T_{\text{comp},n} = \frac{z_n \gamma_n}{f_n} \quad (7)$$

相应地,本地计算能耗可以用公式(8)表示:

$$E_{\text{comp},n} = \kappa_n z_n \gamma_n f_n^2 \quad (8)$$

其中, κ_n 是与硬件结构相关的常数。

1.3.2 雾节点计算模型

假设同一层雾节点的算力相等,并且算力随着层数的增加而增加,第 $m+1$ 层的算力是第 m 层算力的两

倍。如果第2层雾节点 k 所能提供的算力用 f_2 表示,那么第 m 层雾节点 k 的算力为 $f_m = 2^{m-2} f_2$ 。当用户决定把任务卸载到第 m 层雾节点时,如果雾节点 k 只有一个用户 n 接入,那么雾节点 k 的算力将完全由用户 n 使用。当有多个用户接入雾节点 k 时,多个用户需要竞争雾节点有限的算力资源。我们简单地将雾节点的算力均分给每个用户,则用户 n 此时的计算时延如公式(9)所示:

$$T_{\text{comp},n,(m,k)} = \frac{z_n \gamma_n}{f_m} = \frac{z_n \gamma_n n_k(A)}{2^{m-2} f_2} \quad (9)$$

因为任务是在雾节点上运行的,所以对于用户 n 来说并没有计算能耗。

1.4 问题建模

我们将用户的任务卸载代价建模成时延和能耗的线性组合,即定义一个加权代价函数。我们用 $\alpha, (1-\alpha)$ 分别表示时延和能耗的权重,并且规定 $0 \leq \alpha \leq 1$ 。对于不同用户,权重取值不同。时延较敏感用户的 α 取值应该更大,而能耗敏感用户的 α 取值则应该更小。

如果用户选择在本地处理任务,那么代价可以用公式(10)表示:

$$C_n = \alpha T_{\text{comp},n} + (1-\alpha) E_{\text{comp},n} = \alpha \frac{z_n \gamma_n}{f_n} + (1-\alpha) \kappa_n z_n \gamma_n f_n^2 \quad (10)$$

如果用户选择将任务卸载到雾节点 k 上,代价可以用公式(11)来表示:

$$C_{n,(m,k)} = \alpha (T_{\text{trans},n,(m,k)} + T_{\text{comp},n,(m,k)}) + (1-\alpha) E_{\text{trans},n,(m,k)} \quad (11)$$

给定策略组合 $A = \{a_1, a_2, \dots, a_n\}$, 则用户 n 在该策略组合下的代价函数

如公式(12)所示:

$$C_n(A) = C_n I_{\{a_n=0\}} + \sum_{m=2}^M C_{n,(m,k)} I_{\{a_n \in K\}} \quad (12)$$

每个用户需要做出决策来最小化自己的代价函数。这里,我们将问题建模为最小化所有用户的平均代价函数,如公式(13)所示:

$$\min_A \frac{1}{N} \sum_{n \in \mathcal{N}} C_n(A) \quad (13)$$

我们将公式(13)称为任务调度问题,并提出集中式不可分割任务调度算法来解决该问题。

2 基于交叉熵的集中式不可分割任务调度算法

交叉熵方法最初是用来解决稀有事件估计问题的,其基本思想是:通过不断迭代来修正稀有事件的发生概率,使得修正后的概率不断增大,直到此概率达到收敛。收敛概率即为最优概率。根据最优概率就一定会获得最优解。交叉熵方法后来逐渐发展成为解决优化问题(尤其是组合优化问题)的一种方法。这里,我们将使用交叉熵方法来分析任务调度问题。

我们定义所有用户的平均代价函数,如公式(14)所示:

$$Q(A) = \frac{1}{N} \sum_{n \in \mathcal{N}} C_n(A) \quad (14)$$

假设 \mathcal{A} 是 A 的取值空间, γ^* 是 $Q(A)$ 的最小值,此时,公式(14)所示的问题就可以转化为对公式(15)的求解。

$$\gamma^* = \min_{A \in \mathcal{A}} Q(A) \quad (15)$$

假设所有层的雾节点一共有 T 个,我们将所有用户的概率矩阵表示

为 P , 维度表示为 $n \times (T+1)$ 。元素 p_{ij} 表示用户 i 选择将任务卸载到设备 j 的概率, $j \in \{0\} \cup \{1, 2, \dots, T\}$ 。其中, $j \in \{0\}$ 和 $j \in \{1, 2, \dots, T\}$ 分别对应将任务在本地执行和卸载到各层的雾节点 $\{K_2, L_3, \dots, L_m\}$ 上执行。我们可以看出, $\sum_j p_{ij} = 1$ 。概率矩阵 P 的取值空间为 \mathcal{P} 。我们定义一个关于卸载策略 A 的概率密度函数 $\{f(\cdot; P \in \mathcal{P})\}$ 。由此我们构建一个估计问题,如公式(16)所示:

$$l(\gamma) = \mathbb{P}_V(Q(A) \leq \gamma) = \mathbb{E}_V I_{\{Q(A) \leq \gamma\}}, \quad (16)$$

其中, A 是符合概率密度函数 $\{f(\cdot; V \in \mathcal{P})\}$ 的一个随机决策向量, γ 是一个常数。

对于这个估计问题, $\{Q(A) \leq \gamma\}$ 是个小概率事件。我们要根据 l 的取值来使 γ 逐渐接近最优解 γ^* 。我们通过交叉熵方法来不断迭代概率密度函数,从而使概率矩阵也会发生变化,以此来获得最优解。理论上,迭代过程中产生的 $f(\cdot; V), f(\cdot; P^1), f(\cdot; P^2), \dots, f(\cdot; P^T)$ 都向着最优概率密度方向更新,从而使最优解可以获得。

2.1 更新 γ'

对于固定的 P^{t-1} , 我们使 γ' 为在概率矩阵 P^{t-1} 下 $Q(A)$ 的 $(1-\rho)$ 分位数,如公式(17)所示:

$$\begin{aligned} \mathbb{P}_{P^{t-1}}(Q(A) \leq \gamma') &\geq \rho \\ \mathbb{P}_{P^{t-1}}(Q(A) \geq \gamma') &\geq 1 - \rho, \end{aligned} \quad (17)$$

其中 $A \sim f(\cdot; P^{t-1})$ 。

根据概率密度函数 $f(\cdot; P^{t-1})$, 我们随机抽取 N_s 个样本 $\{a^1, a^2, a^3, \dots, a^{N_s}\}$, 然后计算出 $\{Q(a^1), Q(a^2), Q(a^3), \dots, Q(a^{N_s})\}$ 的值,

并将其按降序排序,那么 γ' 的估计值 $\hat{\gamma}'$ 就可以用公式(18)表示。

$$\hat{\gamma}' = S_{\lceil (1-\rho)N_s \rceil} \quad (18)$$

2.2 更新 P'

对于固定的 γ' 和 P^{t-1} , 我们通过最小化交叉熵来得到 P' , 相应的求解如公式(19)所示:

$$\begin{aligned} \max_P D(P) = \\ \max_P \mathbb{E}_{P^{t-1}} I_{\{Q(A) \leq \gamma'\}} \ln f(A; P) \end{aligned} \quad (19)$$

$f(\cdot; P)$ 的对数表示如公式(20)所示:

$$\ln f(A; P) = \sum_{i=1}^N \sum_{j=0}^T I_{\{a_i=j\}} \ln p_{ij} \quad (20)$$

利用公式(20), 我们可以得到公式(19)的拉格朗日函数,如公式(21)所示:

$$\begin{aligned} \mathcal{L}(P, \lambda) = \\ \mathbb{E}_{P^{t-1}} I_{\{Q(A) \leq \gamma'\}} \sum_{i=1}^N \sum_{j=0}^T I_{\{a_i=j\}} \ln p_{ij} + \\ \sum_{i=1}^N \lambda_i \left(\sum_{j=0}^T p_{ij} - 1 \right), \end{aligned} \quad (21)$$

其中 λ_i 为拉格朗日乘子, $i = 1, 2, \dots, N$ 。应用 KKT (Karush - Kuhn - Tucker) 条件, 并通过 $\partial \mathcal{L}(P, \lambda) / \partial p_{ij} = 0$, 我们可以得到公式(19)的最优解, 如公式(22)所示:

$$p_{ij} = \frac{\mathbb{E}_{P^{t-1}} I_{\{Q(A) \leq \gamma'\}} I_{\{a_i=j\}}}{\mathbb{E}_{P^{t-1}} I_{\{Q(A) \leq \gamma'\}}} \quad (22)$$

公式(23)可以被用来估计 \hat{p}_{ij} :

$$\hat{p}_{ij} = \frac{\sum_{k=1}^{N_s} I_{\{Q(a^k) \leq \gamma'\}} I_{\{a_i^k=j\}}}{\sum_{k=1}^{N_s} I_{\{Q(a^k) \leq \gamma'\}}} \quad (23)$$

但是通常来说, 我们并不通过公

式(23)来直接优化 P^t ,而是通过公式(24)来进行优化:

$$\hat{p}_{ij}^t = \beta \hat{p}_{ij}^{t-1} + (1 - \beta) \hat{p}_{ij}^{t-1}, \quad (24)$$

其中, \hat{p}_{ij}^t 可从公式(23)中获得, β 是平滑参数,且 $\beta \in (0,1]$ 。

2.3 算法

下面我们将给出具体算法。

Algorithm 1 基于交叉熵方法的集中式不可分割任务调度(CUTS)算法

- 1: 初始化:
- 2: $p_{ij}^0 = 1/(T+1)$,
 $\forall i \in \mathcal{N}, j \in \{0\} \cup \{1, 2, \dots, T\}$
- 3: $t = 1, N_s, p, d$
- 4: 初始化结束
- 5: 重复执行以下步骤
- 6: 根据 $f(\cdot; P^{t-1})$ 随机抽取 N_s 个样本 $\{a^1, a^2, a^3, \dots, a^{N_s}\}$,
然后计算出 $\{Q(a^1), Q(a^2), Q(a^3), \dots, Q(a^{N_s})\}$ 的值并按降序排序
- 7: 根据公式(18)更新 γ^t
- 8: 根据公式(24)更新 P^t
- 9: $t = t + 1$

10: 如果 $\gamma^t = \gamma^{t-1} = \dots = \gamma^{t-d}$ 就结束循环

3 实验与结果

3.1 仿真设置

我们假设存在这样一个多层次算力网络(参数设置如表1所示)。该网络为3层算力网络:第1层有多个用户,第2层有10个雾节点,第3层有1个雾节点。用户的任务不可拆分。用户先将任务卸载到第2层雾节点,其他层的雾节点之间通过有线进行连接。假设雾节点的初始状态都为无其他任务在运行。

与CUTS算法相对比的几种基准方法为:

(1)本地计算:每个用户都在本地运行任务;

(2)云计算:每个用户都将任务卸载到云端;

(3)随机卸载:每个用户做出的卸载决策是随机的。

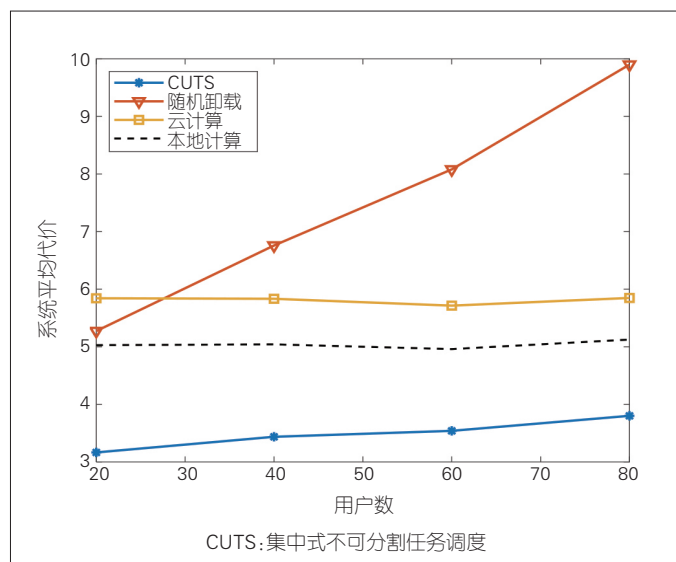
本文以下仿真结果均为400次仿真结果的平均值。

3.2 系统平均代价

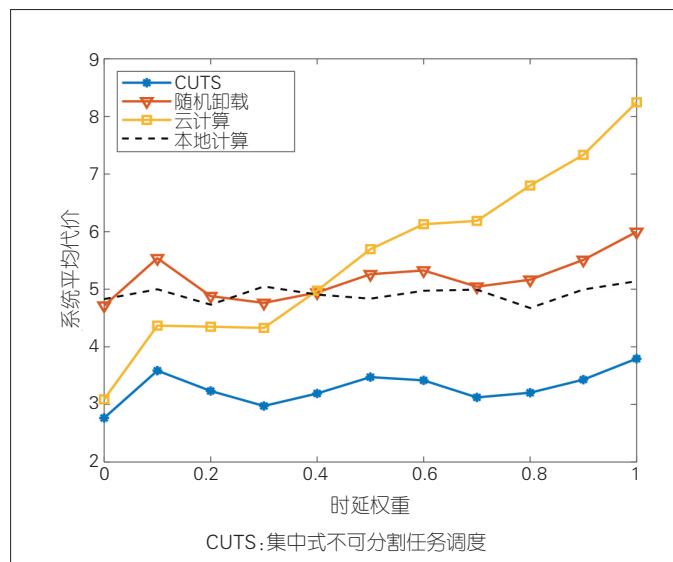
如图2所示,随着用户数的变化,CUTS算法总是能够取得最优的系统平均代价。本地计算通信时延较小,然而总代价却高于多层算力网络,这说明引入算力网络有效解决了本地计算算力较小的问题。图3展示了当

▼表1 仿真参数设置

参数	取值
网络各层节点数	$\{N, 10, 1\}$
任务大小 z_n /kB	[500, 5 000]
处理密度 γ_n /(cycle \cdot b $^{-1}$)	[500, 3 000]
本地设备的计算能力 f_n /GHz	{0.8, 0.9, 1.0, 1.1, 1.2}
能耗常数 κ_n	10^{-27}
二层雾节点的计算能力 f_2 /GHz	20
用户 n 到雾节点 k 的传输速率 $R_{n,k}$ /(Mbit \cdot s $^{-1}$)	[2.01, 4.01]
发送功率 $P_{\text{fog},n}$ /(mJ \cdot s $^{-1}$)	1 224.78
时延权重 α	(0, 1)



▲图2 系统平均代价与用户数的关系



▲图3 系统平均代价与时延权重的关系

用户的任务性质不同时,不同算法的效果。当 α 值较大时,任务性质偏向时延敏感型。因为云端距离用户较远,通常具有比较大的时延,从图3中我们可以看出,引入多层算力网络可以有效解决云计算网络存在的延迟大的问题。

3.3 受益用户数

图4展示了在不同算法下的受益用户数。受益用户是指,在当前卸载策略下降低自身处理任务代价的用户。除随即卸载算法外,其他算法的受益用户数都与总用户数呈正相关。由图4可知,CUTS算法依然表现出最优性能。

3.4 时延及能耗成本分布

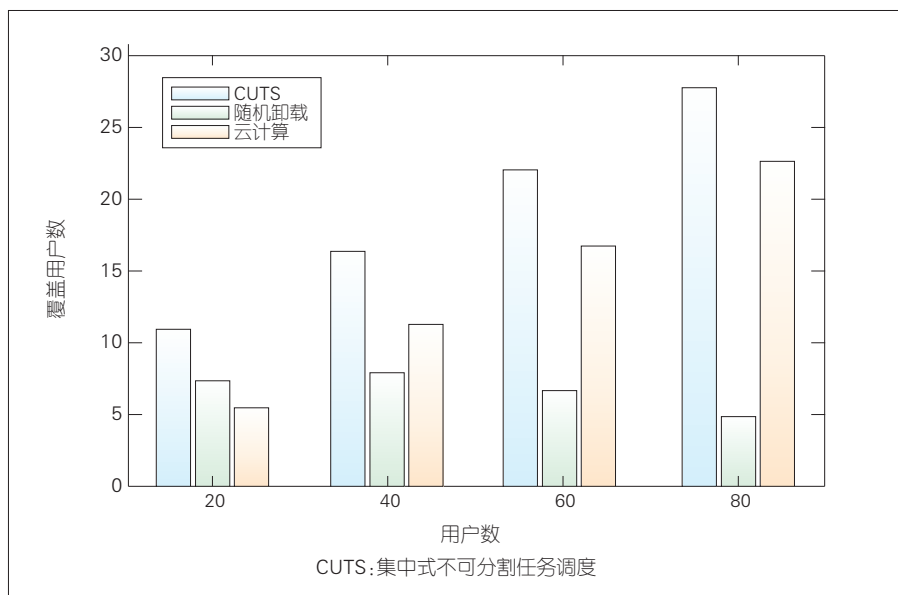
图5展示了随着用户数目的增加,时延和能耗的对比情况。可以看出,随着总用户数的增加,总的代价也在增加,但是增加幅度在减缓。此外,时延产生的代价要略高于能耗产生的代价。

3.5 本地计算、雾计算和云计算用户分布

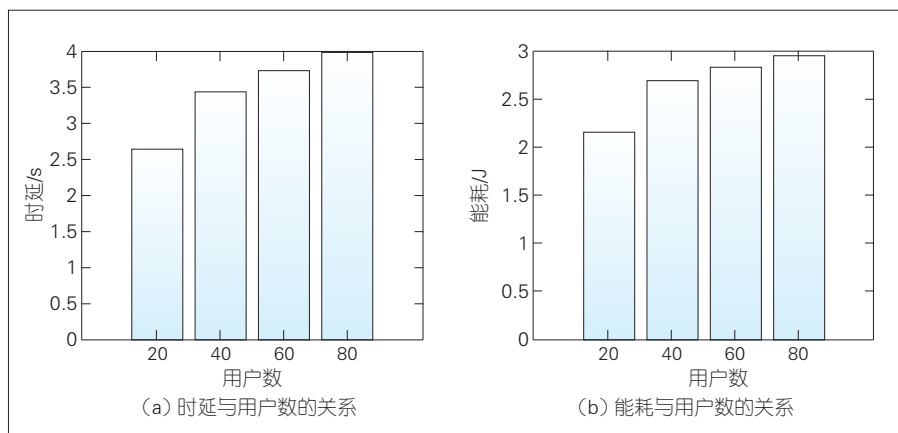
图6展示了随着用户数增加,各用户的卸载决策分布。可以看出,选择本地用户和云计算的用户数目逐渐增多,而选择雾节点的用户数却几乎不变。这是因为雾节点的算力资源接近饱和。

4 结束语

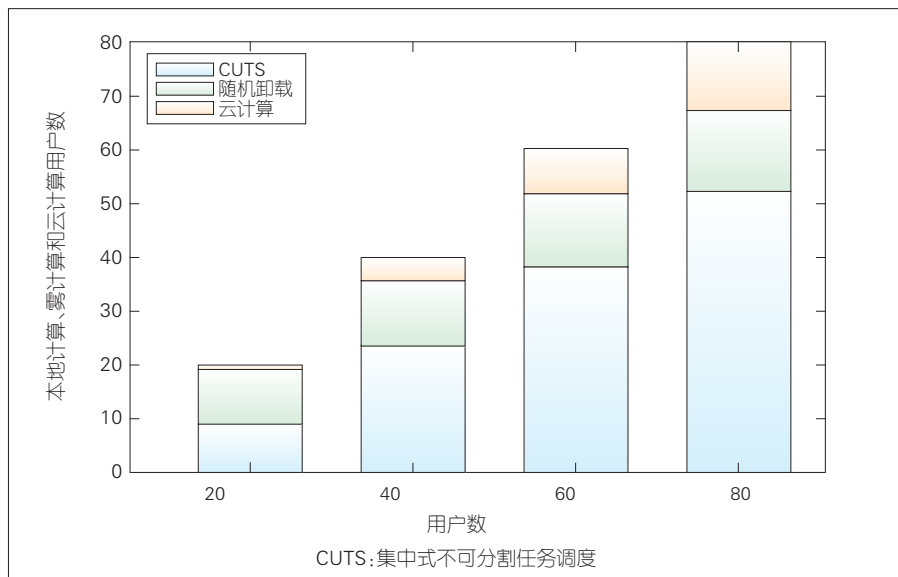
本文中,我们提出一种多层次算力网络模型和计算卸载系统,定义一个由时延、能耗组成的加权代价函数,并建模一个任务调度问题。为解决这一问题,我们提出CUTS算法,即将一个确定性问题转化成了一个估计问题,通过重要性采样和交叉熵的方法来求解问题的最优解。数值仿



▲图4 受益用户数与用户数的关系



▲图5 时延和能耗与用户数的关系



▲图6 选择本地计算、雾计算和云计算的用户数与总用户数的关系

真结果表明,CUTS算法能够在系统平均代价和受益用户数方面提供最优性能。算力网络可以有效解决单层网络带来的算力小或时延大的问题。

致谢

本研究得到上海科技大学杨旸老师、吴连涛老师的帮助,谨致谢意!

参考文献

- [1] REN J K, HE Y H, YU G D, et al. Joint communication and computation resource allocation for cloud-edge collaborative system [C]// 2019 IEEE Wireless Communications and Networking Conference (WCNC). Marrakesh, Morocco: IEEE, 2019: 1-6. DOI: 10.1109/WCNC.2019.8885877
- [2] MOURADIAN C, NABOULSI D, YANGUI S M, et al. A comprehensive survey on fog computing: state-of-the-art and research challenges [J]. IEEE communications surveys & tutorials, 2018, 20(1): 416-464. DOI: 10.1109/COMST.2017.2771153
- [3] LIU Z N, YANG Y, CHEN Y, et al. A multi-tier cost model for effective user scheduling in fog computing networks [C]//IEEE INFOCOM 2019 - IEEE Conference on Computer Communications Workshops (INFOCOM WK-SHPS). Paris, France: IEEE, 2019: 1-6. DOI: 10.1109/INFOCOMW.2019.8845252
- [4] MAO Y Y, YOU C S, ZHANG J, et al. A survey on mobile edge computing: the communication perspective [J]. IEEE communications surveys & tutorials, 2017, 19(4): 2322-2358. DOI: 10.1109/COMST.2017.2745201
- [5] YANG Y, WANG K L, ZHANG G W, et al. MEETS: maximal energy efficient task scheduling in homogeneous fog networks [J]. IEEE Internet of Things journal, 2018, 5(5): 4076-4087. DOI: 10.1109/JIOT.2018.2846644
- [6] DINH T Q, TANG J H, LA Q D, et al. Off-loading in mobile edge computing: task allocation and computational frequency scaling [J]. IEEE transactions on communications, 2017, 65(8): 3571-3584. DOI: 10.1109/TCOMM.2017.2699660
- [7] CHEN X, JIAO L, LI W Z, et al. Efficient multi-user computation offloading for mobile-edge cloud computing [J]. IEEE/ACM transactions on networking, 2016, 24(5): 2795-2808. DOI: 10.1109/TNET.2015.2487344
- [8] NOWAK D, MAHN T, AL-SHATRI H, et al. A generalized Nash game for mobile edge computation offloading [C]//2018 6th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (Mobile-Cloud). Bamberg, Germany: IEEE, 2018: 95-102. DOI: 10.1109/MobileCloud.2018.00022
- [9] YANG Y, LIU Z N, YANG X M, et al. POMT: paired offloading of multiple tasks in heterogeneous fog networks [J]. IEEE Internet of Things journal, 2019, 6(5): 8658-8669. DOI: 10.1109/JIOT.2019.2922324
- [10] YANG Y. Multi-tier computing networks for intelligent IoT [J]. Nature electronics, 2019, 2(1): 4-5. DOI: 10.1038/s41928-018-0195-9
- [11] LIU Z N, YANG Y, ZHOU M T, et al. A unified cross-entropy based task scheduling algorithm for heterogeneous fog networks [C]//Proceedings of the 1st ACM International Workshop on Smart Cities and Fog Computing. New York, NY, USA: ACM, 2018: 1-6. DOI: 10.1145/3277893.3277896
- [12] SHAH-MANSOURI H, WONG V W S. Hierarchical fog-cloud computing for IoT systems: a computation offloading game [J]. IEEE Internet of Things journal, 2018, 5(4): 3246-3257. DOI: 10.1109/JIOT.2018.2838022

作者简介



巩宸宇, 上海科技大学信息与技术学院在读硕士研究生;研究领域主要包括物联网与无线通信、雾计算等。



舒洪峰, 深圳市智慧城市科技发展集团有限公司副总经理, 曾担任深圳市盐田港集团有限公司办公室副主任, 深圳市特区建设发展集团有限公司办公室主任、董事会秘书;主要研究领域包括大数据与云计算、5G及城域物联网、数字经济等。



张昕, 教授级高级工程师, 深圳市智慧城市科技发展集团有限公司解决方案部部长, 深圳市智能交通标准化技术委员会委员、深圳市政府采购中心资深专家、深圳市后备级领军人才;从事智慧城市、智能交通等政府信息化工作, 主持并完成综合交通运行指挥中心、智慧宝安总体规划、路边停车系统、侨香路智慧道路、智慧国资管理展示中心及智慧国资大数据中心等项目;获华夏建设科学技术奖一等奖、中国智能交通协会一等奖、深圳市科技创新奖等省部级奖励(7项);发表论文与专著20余篇, 申请发明专利3项, 参与编制深圳市地方标准9项。



夯实云网融合，迈向算网一体

Tamp Cloud and Network Integration, Step into Computing Power Network

唐雄燕 /TANG Xiongyan, 张帅 /ZHANG Shuai,
曹畅 /CAO Chang
(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.202103009
网络出版地址: <https://kns.cnki.net/kcms/detail/34.1228.TN.20210622.1436.002.html>
网络出版日期: 2021-06-22
收稿日期: 2021-05-15

摘要: 新一代信息网络正在从以信息传递为核心的网络基础设施, 向融合计算、存储、传送资源的智能化云网基础设施演进。结合云网融合发展趋势, 倡导积极发展算力网络, 并通过算网一体实现深层次的云网融合。同时, 指出算力网络重点创新领域及核心能力, 提出促进算力网络发展的若干建议。

关键词: 云网融合; 算力网络; 算网一体; 云化

Abstract: The new generation of information network is evolving from the network infrastructure focusing on information transmission to the intelligent cloud network infrastructure integrating computing power, storage, and transmission resources. In the trend of cloud network integration development, it needs to develop computing power network and realize deep cloud network integration through computing power network integration. Key innovation areas and core capabilities of computing power network are pointed out, and some suggestions to promote the development of computing power network are put forward.

Keywords: cloud network integration; computing power network; integration of computing power and networking; cloudification

1 云网融合的背景与趋势

2021年是“十四五”开局之年, 信息通信网络肩负着赋能经济社会数字化转型的新使命, 也迎来创新发展的新机遇。作为新一轮科技革命和产业变革的主导力量, 信息通信产业深刻影响和改变了经济社会发展模式和人们生产生活方式, 成为了科技创新和经济增长的重要引擎。新一代信息通信网络正在从以信息传递为核心的网络基础设施, 向融合计算、存储、传送资源的智能化云网基础设施发生转变。

(1) 数据成为重要生产要素, 数字经济潜力正逐步释放

中国“十四五”规划高度重视数字经济的发展, 把“网络强国”“数

字中国”作为新发展阶段的重要战略进行部署。2019年, 全球47个经济体数字经济规模达到31.8万亿美元, 同比增长5.4%, 高于同期全球国内生产总值(GDP)增速的3.1%, 数字经济成为全球经济增长的主要引擎^[1]。数字产业化、产业数字化纵深推进, 不断促进虚拟世界与物理世界紧密结合, 而工业互联网、车联网、智慧医疗、智慧城市等也将成为推动经济发展的新动力, 同时网络需要承载更大的带宽流量、更多的类型业务, 以及响应更加迅速的各类需求。现实虚拟化、虚拟真实化成为新兴数字经济的重要助推器, 两者的交互融合将给业务创新带来巨大的发展空间, 这就要求新型网络基础设施能够充分适应这些快速变化的新业态。

(2) 企业服务云化趋势加速, 驱动通信网络信息技术(IT)化转型

近年来, 以云计算为代表的计算产业对通信网络变革产生了巨大影响: 企业上云节奏不断加快, 云流量持续增长。这给运营商带来机遇的同时, 也给通信产业带来了巨大的运营与竞争压力, 同时还驱动着通信网络IT化转型。经济社会数字化将带动信息通信技术(ICT)产业步入增长新轨道: ICT技术创新进入“新领域”和“无人区”。局域业务、本地业务、低时延业务全面兴起, 以5G为代表的新一代通信技术赋予通信产业新活力。网络作为云、网、边、端中承上启下的关键环节, 将从纯粹的管道角色, 转变成成为承载更多价值可能性的数字经济中枢。此外, 2020年突发的新冠肺炎

炎疫情,作为“黑天鹅”事件,使无接触服务成为主流,并驱使企业加速数字化转型。在线教育、家庭办公、远程医疗、城市治理等数字化生活与工作方式将会得到长久保留,从而为网络服务和信息通信产业注入了更强劲的发展动力。

(3) 泛在“联接+计算”紧密结合,构建 ICT 智能融合新格局

2020 年 4 月,国家发展和改革委员会首次对新基建的具体含义进行了阐述,在信息基础设施部分,提出构建以数据中心、智能计算中心为代表的算力基础设施,提升各行业的“联接+计算”能力,引领重大科技创新、重塑产业升级模式,为社会发展注入更强动力。随着 5G、移动边缘计算(MEC)和人工智能(AI)的发展,算力和智能将无处不在,网络需要为云、边、端算力的高效协同提供更加智能的服务,计算与网络将深度融合。为满足现场级业务的计算需求,计算能力进一步下沉,出现了以移动设备和物联网(IoT)设备为主的端侧计算。在未来计算需求持续增加的情况下,虽然“网络化”的计算有效补充了单设备无法满足的大部分算力需求,但是仍然有部分计算任务受到网络带宽及时延限制,因此未来形成云、边、端多级计算协同部署是必然趋势,即云侧负责大体量复杂的计算、边缘侧负责简单的计算和执行、终端侧负责感知交互的泛在计算模式^[2]。新基建政策给以算力网络技术为基础的转-算-存主体分离、联合服务的新商业模式提供了重要的发展机遇。

2 从云网融合到算网一体,网络成为价值中心

5 年前,全球主要电信运营商纷纷开启了面向 2020 年的下一代网络转型规划,以云计算为中心、实现云网

融合是这一阶段网络转型的主旋律。5 年后的今天,运营商已实现了移动接入从 4G 到 5G、固定接入从以太网无源光网络(EPON)/无源光纤网络(GPON)到千兆光网(F5G)的转变,构建了面向个人、家庭和企业的泛在千兆接入网,部分领先运营商和海外主流互联网交换中心(IXP)还构建了超宽的云互联(DCI)骨干网。得益于这一阶段的架构转型,在新冠肺炎疫情期间,中国的通信网络成功支撑了数亿用户居家办公,实现了互联网服务能力从支撑消费视频娱乐到满足居家视频办公的显著提升。

过去 5 年的云网融合,网络位于云与端之间,解决了云与端的连通性。云上丰富的内容可以自上而下,顺畅地呈现在各种智能终端上;网络支持了下行流量为主的云端互联,为终端提供了内容服务。未来 5 年,随着大量实时性业务的出现,如云虚拟现实(VR)、机器视觉、自动驾驶等,终端产生的大量数据需要上传到边、云的计算节点进行处理,并将结果实时送回终端;网络需要支持上行流量爆发的云、边、端互联,并为终端提供确定性的智能服务。

边缘计算的出现改变了传统云和网的相互独立性,使计算进入网络内部。边缘计算的效率、可信度与网络的带宽、时延、安全性、隔离度等都将发生深度耦合,算网一体才能实现高效服务。

从云网融合到算网一体,网络的作用和价值将发生变化。对于云网融合,网络是以云为中心的。从云的视角看,一云多网对网络的主要需求是连通性、开放性,对服务质量的要求是尽力而为,网络起到支撑作用。对于算网一体,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、安全可信通信,对服务

的质量要求是确定性,网络成为价值中心。

这两个阶段是相辅相成的,云网融合为算网一体提供必要的云网基础能力,算网一体是云网融合的升级^[3]。

2.1 夯实云网融合,持续打造网络服务核心竞争力

(1) 一网联多云,构建新生态,助力产业发展

在云网融合时代,网络提供包括带宽、时延、抖动、安全、算力、可视可控等多样化能力,从而可以深度参与行业应用和智能终端的业务逻辑,与端、云产业生态实现合作共赢。新服务呼唤新生态,需要满足客户对应用、云、网、端的一体化需求,产业链的开放协作是必然要求。网络基础服务能力的夯实,对外需要与云服务商、应用服务商、终端提供商等广泛合作,依托外部的内容服务和智能应用生态。面向个人和家庭消费者,运营商提供极致的信息生活体验;面向垂直行业和政府,提供丰富的智能融合应用。

(2) 协同一体智能安全,多层次、多维度安全防护

利用大数据技术持续对业务流量和各类网络、安全设备日志进行关联分析,并结合 AI 智能推理,能够及时发现潜在威胁,提供全网安全态势可视化能力。网络内生安全将多种安全基因注入网络,使得安全边界更加细化,安全能力与云网边资源能够更好地协同配合,安全策略与业务需求进一步贴合,原生和外挂的安全能力得到进一步整合,从而能提供端到端一体化防护服务。同时,网络内生安全能够结合客户的诉求,灵活调度安全资源,提供不同级别的安全防护等级以及安全服务组合。

(3) 网络切片和行业专网,助力

行业用户服务升级

根据自身需求，行业用户可以选择一种或多种标准化的网络服务，如云服务、网络切片、确定性等。云网融合可为用户提供电商化的网络切片即时服务新体验。其中，切片带宽灵活多样、动态可调整，服务平台支持自助式即开即享的租户级切片服务，云业务驱动切片创建，分钟级业务开通，实时感知切片状态。同时，基于切片理念的多种隔离技术及其组合将为客户提供端到端灵活多样的业务隔离服务，保证客户的带宽和延迟等网络性能不受其他切片的影响。

面向 VIP 大客户需求，云网融合除了提供标准网络能力和服务外，还能够为客户提供量身定制的专网建设与运营服务，包括高带宽保障、确定性时延、有界时延抖动、高精度定位以及网络无损传输等，并可以基于确定和稳定的基础网络能力，针对行业客户需求提供可管可控和安全可靠的确定性服务。

(4) 加强和延伸云网端协同能力，支撑业务融合创新

端云协同就是根据业务特点、网络、终端能力及运行环境等，通过终端与云端协商，智能化地将原来由终端执行的非实时复杂计算和存储转移至云端或边缘计算节点，并将运算结果返回终端执行。端云协同可以进一步充分、有效利用网络与云端资源，提升用户的使用体验，减少对终端软硬件能力的需求，降低功耗和成本。端云协同需要提供端云协同的智能调度机制和策略算法，实现系统最优。另外，面向行业业务需确保终端本地、边缘及云端数据的隐私安全。

跨云网络互通，需保障多云服务商和云资源池的多种接入和互联能力，保障不同云之间的网络互通，以实现云网无缝对接。跨云连接需要保障网

络连接的高度确定性，基于云业务要求提供确定性网络连接关键绩效指标（KPI）指标。多云协同支撑业务融合创新，有效地控制负载和成本，多云共管提高运维效率，提升数据的可移植性和互操作性。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^[4]。

2.2 迈向算网一体，加快构建算力服务新能力

数字经济促进数字产业化，而算力将是数字经济的重要引擎。随着算力下沉到边缘，城域数据中心（DC）需要互联，这对业务属性的感知和计算资源的感知提出了更高的要求。运营商网络云化的加速和以算力基础设施为代表的新基建，给 DC 算力资源社会化共享提供了商业机遇。算力经营将成为运营商新的重要业务抓手，使运营商不再是纯粹的管道服务商。

基于云网融合的发展，算网一体不能一蹴而就，需要分步进行技术攻关，渐进打造核心能力。

(1) 强化算力建模与管理底层技术研究

算力网络中的算力资源包括通用服务器架构下的中央处理器（CPU）计算单元，专门适用于处理类似图形、图像等数据类型统一的图形处理器（GPU）并行计算芯片，专业加速处理神经网络的神经网络处理器（NPU）和张量处理器（TPU），广泛应用于边缘侧嵌入式设备的 CPU，以及半定制化处理器现场可编程门阵列（FPGA）等。根据处理器运行算法及涉及的数据计算类型的不同，从业务角度出发，算力可以分为可提供逻辑运算的算力、可提供并行计算的算力与神经网络加速的算力等^[5]。

泛在算力资源的统一建模度量是算力调度的基础。针对泛在的算力资源，

通过模型函数，可将不同类型的算力资源映射到统一的量纲维度，形成业务层可理解、可阅读的零散算力资源池，为算力网络的资源匹配调度提供基础保障。将业务运行所需的算力需求按照一定的分级标准划分为多个等级，这可作为算力提供者在设计业务套餐时的参考，也可作为算力平台设计者在设计算力平台时的选型依据。

(2) 基于泛在算力需求，完善算力网承载能力

• 算力资源信息感知技术

算力网络整合计算资源，并以服务的形式为用户提供算力。与基于链路度量值进行路径计算的网路路由协议类似，在算力网络中，网络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。

算力网络的实现不可能一蹴而就，面向算力承载的网络应遵循“目标一致、分期建设”的原则，通过 DC 网关设备联网可以搭建 MEC 节点之间的算力“薄层”。可以首先在 overlay 层面引入 SRv6 与内容转发网络（CFN）等协议，进而逐步扩大到承载网全网 underlay 层面的算力感知和算网联合优化。

• 增强确定性网络技术

确定性网络协议（DIP）是在互联网协议（IP）网络上，通过增强的周期排队和转发技术实现的一种新型网络转发技术。确定性 IP 网络能够保证网络报文传输时延上限、时延抖动上限、丢包率上限。它既适用于中小规模网络，又适用于解决大规模、长距离 IP 网络端到端确定性传送问题。DIP 技术通过在原生报文转发机制中，加入周期排队和转发技术，通过资源预留、周期映射、路径绑定、聚合调度等手段实现大网的确定性转发能力。通过确定性技术和算力的结合，可以

提供精确保障的业务体验, 满足算力抖动敏感型业务的需求。

- 应用感知网络 (APN) 技术

基于 APN 技术, 利用第 6 版互联网协议 (IPv6) 扩展头将应用信息及其需求传递给网络, 网络根据这些信息, 通过业务的部署和资源调整来保证应用的服务等级协议 (SLA) 要求。特别是当站点部署在网络边缘 (即边缘计算) 时, APN 技术有效衔接网络与应用, 以适应边缘服务的需求, 将流量引向可以满足其要求的网络路径, 从而充分释放边缘计算的优势。

- 业务链技术

业务链使得不同算力服务链接成为现实, 从而可以快速提供新型业务。业务链是一种业务功能的有序集合, 可以使业务流按照指定的顺序依次经过指定的增值业务设备, 以便业务流量获取一种或者多种增值服务。

业务链在算力网络中的本质是驱动算力服务, 即依据客户的意图, 实现不同算力服务的连接, 并结合 SRv6 安全标识符 (SID) 即服务, 构建算力交易平台。各种生态算力将自己的服务以 SRv6 SID 的形式注册到网络中, 购买者通过购买服务使用算力, 而网络则通过业务链将算力服务链接, 从而无感知地将服务提供给购买者。

(3) 构建算网服务编排能力, 实现算网资源的能力开放

算力网络是融合计算、存储、传送资源的智能化新型网络, 通过全面引入云原生技术, 实现业务逻辑和底层资源的完全解耦。通过打造面向服务的容器编排调度能力, 可以实现服务编排向算网资源的能力开放。同时, 可结合底层基础设施的资源调度管理能力, 对数据中心内的异构计算资源、存储资源和网络资源进行有效管理, 实现对泛在计算能力的统一纳管和去中心化的算力交易, 构建统一的服务

平台。

(4) 打造算力服务和交易平台, 促进算力安全有效流通

算力网络中的算力服务与交易依托于区块链去中心化、低成本、保护隐私的可信算力交易平台。该平台由算力卖家、算力买家、算力交易平台 3 种角色组成, 在以往的交易模式中, 买家和卖家彼此之间信息并不透明。在未来泛在计算场景中, 网络可以将算力作为透明和公开的服务能力提供给用户。在算力交易过程中, 算力的贡献者 (算力卖家) 与算力的使用者 (算力买家) 分离。这样可以通过可拓展的区块链技术和容器化编排技术, 整合算力贡献者的零散算力, 为算力使用者和算力服务的其他参与方提供经济、高效、去中心化、实时、便捷的算力服务。

3 算力网络研究概况

据国际数据公司 (IDC) 预测, 到 2023 年, 数字经济产值将占到中国 GDP 的 67%, 超过全球平均水平, 发展潜力巨大。以 5G、云和 AI 为代表的数字基础设施发展将带动全网的算力密集分布、快速下沉, 从而逐步实现联网服务。目前, 算力网络的愿景已在业界得到广泛认可, 算力网络在标准制定、生态建设、试验验证等领域均取得了一定进展。算力网络作为中国的原创技术成果, 开始走向国际舞台。在标准制定方面, 中国移动、中国电信与中国联通分别在国际电信联盟 (ITU-T) SG11 与 SG13 工作组立项了 Y. CPN^[6]、Y. CAN 和 Q. CPN^[7] 等系列标准, 并在互联网工程任务组 (IETF) 开展了《Computing First Network Framework》^[8] 等系列标准的研究; 华为联合中国运营商在欧洲电信标准化协会 (ETSI) 和世界宽带论坛 (BBF) 启动了包括 NWI、城域算网

在内的多个项目; 中国通信标准化协会 (CCSA) 正有序开展算力网络总体架构和技术要求、标识解析技术要求、集中控制系统技术要求等 6 项系列标准工作。面向未来的 6G 时代, 中国的 IMT-2030 6G 网络工作组已将算力网络列为研究课题之一, 开展算力网络与 6G 通信技术的融合研究。在生态建设方面, 中国未来数据通信研究的主要组织——网络 5.0 产业联盟, 专门成立了算力网络特设工作组。2019 年, 中国联通、中国移动和边缘计算网络产业联盟 (ECNI) 均发布了算力网络领域相关白皮书, 进一步阐述了算网融合的重要观点。2021 年初, 三大运营商与华为、边缘计算网络产业联盟联合出版《边缘计算 2.0: 网络架构与技术体系》; 2020 年, 中国联通率先成立中国联通算力网络产业技术联盟, 作为首个运营商牵头的算力网络研究组织, 该联盟结合自身业务发展, 对相关先进网络协议的制定提出了明确需求。在试验验证方面, 2019 年中国电信与中国移动均已完成算力网络领域的实验室原型验证, 并在全球移动通信系统协会 (GSMA) 巴塞罗那展、ITU-T 和全球网络技术大会 (GNTC) 相关展会上发布成果。2020 年年底, 中国联通在江苏南京开通了中国首个集成开放网络设备、算力服务平台和 AI 应用的一体化试验局^[9]。

4 算力网络发展建议

4.1 进一步推动算力网络的标准化工作

算力网络的标准化工作虽已有序开展, 但目前仍处于前期。中国的研究成果目前处于国际领先, 建议运营商和设备商结合自身标准研究与应用实践, 将标准推向国际, 加快算力网络技术的标准化进程。为了更好地解决泛在计算和服务感知、互联以及

资源控制和调度中存在的问题,并满足未来新应用场景需求^[10],需要重点推进一些标准化工作。

(1) 应用及算力感知研究: 研究算力、网络和存储等多维资源感知,实现多维资源感知、调度的协同机制。

(2) 算力需求量化与建模研究: 针对泛在的算力资源,通过模型函数,将不同类型的算力资源映射到统一的量纲维度,形成业务层可理解、可阅读的零散算力资源池。

(3) 算网资源可信与协同: 解决资源可信与协同问题,为需求方提供更多选择,促使算力流动起来,促进应用发展。

4.2 注重算力网络产业的自主可控

在当前国际竞争背景下,网络领域的自主可控是一项突出问题,算力网络也不例外。为此,需要把算力网络技术的自主可控作为重要研究内容。通过“以算联网,以网促算”的方式进行计算和网络的联合布局优化,并通过计算成网,弥补中国计算芯片单体的自主可控这一短板,解决“卡脖子”问题。具体来说,需要加强计算处理单元和网络控制系统双方的开放性,以便更加快速、便捷地响应对方的需求。这个过程不仅需要国家相关部门牵头组织、政策性扶持,同时更需要产学研用各个参与方的积极推动。但是,大力发展自主可控并不意味着故步自封、闭门造车。自主可控的策略应该是在中国企业掌握核心竞争力的基础上,以积极开放的态度拥抱开源,在全球范围内共建共享算力网络技术

产业生态。

5 结束语

随着国家新发展格局的构建和新发展理念贯彻,数字经济蓬勃发展,以“联接+计算”为根基的数字基础设施的重要性进一步凸显。夯实云网融合,向算网一体演进,实现 CT、IT 和数据技术(DT)能力的融合服务,是顺应千行百业数字化转型的必然要求。发展算力网络,实现“算力即服务,网络即平台”的目标涉及到 IT 产业、CT 产业和 DT 产业,有赖技术融合、产业协同和生态重构。算力网络的技术理念已逐步在行业形成了共识,未来需要通过市场牵引、技术驱动和开放创新推进算力网络大发展,实现网络与计算的超级融合,赋能数字经济。

参考文献

- [1] 中国联通 CUBE-Net 3.0 网络创新体系白皮书[R]. 北京: 中国联合网络通信有限公司研究院, 2021
- [2] 算力网络架构与技术体系白皮书[R]. 北京: 中国联合网络通信有限公司研究院, 2020
- [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]. 中国通信, 2021, 18(2):175-185
- [4] 云网融合向算网一体技术演进白皮书[R]. 北京: 中国联合网络通信有限公司研究院, 2021
- [5] 异构算力统一标识与服务白皮书[R]. 北京: 中国联合网络通信有限公司研究院, 2021
- [6] ITU-T. Draft recommendation ITU-T Y. CPN-arch: framework and architecture of computing power network[R]. 2020
- [7] ITU-T. Draft recommendation ITU-T Q. CPN: signaling requirement of computing power network[R]. 2019
- [8] IETF. Framework of Compute First Networking (CFN) draft-li-rtgwg-cfn-framework-00[R]. 2019

[9] 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

[10] 算力网络前沿报告[R]. 北京: 中国通信学会, 2020

作者简介



唐雄燕, 中国联合网络通信有限公司研究院副院长、首席科学家, “新世纪百万人才工程”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组长, 北京通信学会副理事长, 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, 国际开放网络基金会 ONF 董事; 拥有 20 余年电信新技术、新业务研发与技术管理经验, 主要专业领域为宽带通信、光纤传输、互联网/物联网、SDN/NFV 与新一代网络等。



张帅, 中国联合网络通信有限公司研究院未来网络研究部工程师、中国通信标准化协会 CCSA “网络 5.0 技术标准推进委员会”需求组副组长; 主要专业领域为 IP 网络宽带通信、SDN/NFV、新一代网络编排技术等。



曹畅, 中国联合网络通信有限公司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 CCSA “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; 主要专业领域为 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 20 余项。



零触碰与零信任

Zero Touch and Zero Trust

李军 /LI Jun, 胡效赫 /HU Xiaohu

(清华大学, 中国 北京 100084)
(Tsinghua University, Beijing 100084, China)

DOI: 10.12142/ZTETJ.202103010

网络出版地址: <https://kns.cnki.net/kcms/detail/34.1228.TN.20210617.0922.004.html>

网络出版日期: 2021-06-17

收稿日期: 2021-05-10

摘要: 随着网络规模持续增加、应用日益复杂以及动态性不断增强, 网络自动化的需求愈发迫切。网络转发呈现零触碰的趋势, 以实现策略编排的自动化为目标。网络安全呈现零信任的趋势, 以实现身份访问的自动化为目标。从基本理念、核心组成以及工业实践的角度对零触碰和零信任进行分析, 阐述网络自动化的必要性与发展情形。

关键词: 网络自动化; 网络转发; 零触碰; 网络安全; 零信任

Abstract: With the increasing scale of networks, complexity of applications, and dynamics of scenarios, there has been an urgent demand of network automation. Network forwarding is becoming zero touch, automating the policy orchestration. Network security is becoming zero trust, automating the identity and access management. Zero touch and zero trust networks are analyzed in three aspects, i.e., basic concept, core components, and industrial practice, and the necessity and development of network automation are described.

Keywords: network automation; network forwarding; zero touch; network security; zero trust

在学术研究中, 研究者有时会把网络的转发与安全“正交化”, 即把网络转发与网络安全“解耦”开来, 以便简化问题、“分而治之”。按照这个原则设计出来的系统架构和解决方案, 符合实际管理体系中的人员分工, 自然也就比较容易落地应用。然而, 网络的转发与安全事实上是高度相关、密不可分的。不发生网络转发的行为, 就不存在网络安全的问题。而没有网络安全的保障, 网络转发就容易受到攻击。网络转发协议、拓扑和流量的变化, 必定会引发网络安全的机制设计、技术构成和实现方式的改变。而网络安全的完备性和有效性, 又是与网络转发的私密性、完整性和可靠性交织在一起的。

通常所说的网络, 主要是指交换和路由机制对网包的操作。伴随着网络规模和流速的指数增长, 协议和应

用的日趋纷繁, 以及人类生活对网络依赖性的不断加剧, 通过人工配置来管理网络的方法经常捉襟见肘、状况百出。网络管理人员难以应对“复杂大系统”。网络自动化成为日益迫切的需求。在这样的背景下, 零触碰网络应运而生。

同时, 随着网络虚拟化和动态化的不断增强, 不但以工作环境为网络物理边界的安全防护早已无法满足移动办公和居家办公的普及要求, 大量企业信息系统“云化”对逻辑边界的安全需求, 以及企业核心资产面临的巨大数据安全挑战, 也使得原有的“一次认证、一路畅通”的信任模式不再可靠。因此, 我们需要建立零信任网络。或者说, 对用户或终端的信任只能建立在持续的认证、鉴权、“健康体检”和行为监测控制的基础之上。

无论是零触碰还是零信任, 本质

上都是对网络自动化迫切需求的反映。

1 网络转发与零触碰

网络转发主要依靠路由器和交换机来完成, 而这些物理资源并不是随时随地更换或增减的(至少变化频率不会很高)。此外, 它们在运行软件时所遵从的协议, 相对而言也是“静态”的, 不会频繁切换或升级、卸载。网络转发设备或虚拟设备相互连接构成的网络拓扑及其承载的具体功能, 主要是由策略(涵盖配置和规则等说法)编排决定的。

严格来说, 这些策略也是程序的一部分, 但不是在硬编码和预编译(机器语言, 即二进制代码)之后才被绑定到物理设备或逻辑设备中的, 而是在运行期间甚至运行时, 依据网络设计目标和资源情形“动态”注入的。通常这些策略也会需要预编译, 以紧

凑的数据结构来满足性能要求。

传统上,策略编排是由网络管理员来完成的。他们运用特定的网络连线和设备配置来完成网络设计目标。Google 研究报告显示,超过7成的网络故障发生在网络管理操作过程中^[1]。在网络的规模、复杂性和动态性远超过人工能力范畴的情况下,策略编排需要新的范式来保障网络设计的效率和稳定性。

零触碰的宗旨是最小化网络管理生命周期中的人工介入,并最大化程序与工具在网络管理中的占比。其中的关键在于实现自动化的策略编排:管理员只需要关注网络设计的目标“是什么”,无须考虑连线与配置“怎么做”。零触碰网络的实现可以参照计算机程序编译和芯片设计的电子设计自动化(EDA)工具。网络自动化和程序设计都是把高级语义映射为机器代码的过程,而芯片设计面对的布局布线约束与网元资源及其连接的约束也有相似之处。

零触碰网络的提出不仅源自云计算和网络代际升级带来的管理挑战,还得益于软件定义网络(SDN)为网络开放创新打下的基础。在SDN的理念下,零触碰进一步融入闭环控制、软件验证、编程语言等多领域机制。这也充分说明计算机网络是信息科学中典型的交叉融合应用场景。

零触碰网络依托于一整套网络自动化系统和一系列策略编排核心技术。这主要体现在:(1)在系统设计方面,零触碰网络遵循抽象化原则,通过管理闭环来保障网络稳定性;(2)在管理员接口设计方面,零触碰网络遵循声明式编程范式,以提高系统易用性;(3)在模块实现方面,零触碰网络通过算法优化来应对规模增长带来的效率问题。

从网络自动化架构来看(如图1所示),零触碰分别体现在3个层面上。

最上层是意图驱动网络(IBN)

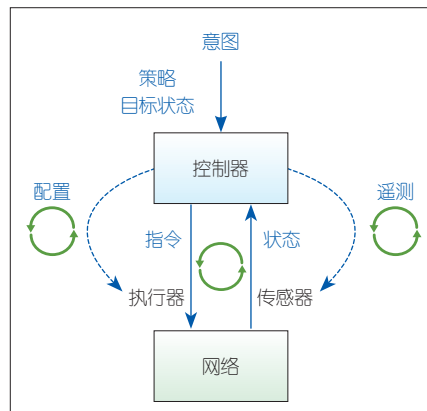
重点关注的,也是零触碰的关键。该层将管理平面用户或应用的意图,映射为逻辑集中的控制平面可以理解的策略。这使网络管理员从庞杂细碎的手工配置中抽出身来,以便他们把主要精力聚焦在网络设计目标的确定和达成上。这里涉及的主要是策略语言,它包括策略描述的定义和编译(也称综合)。网络策略通常被分为转发策略和其他网络服务策略,而服务策略主要包括安全策略。

中间层是SDN关注的核心,以控制器的能力来支撑零触碰。基于网络拓扑与协议,同时根据网络策略和状态,该层可求解策略部署方案,并验证策略的一致性。其中,策略和状态组成闭环控制的整体。零触碰根据特定的网络状态部署相应的策略以实现管理员意图。细化来看,策略部署的正确性也需要一个下发和校验的闭环来保障,状态获取的针对性也同样需要一个配置和监测的闭环来支撑。

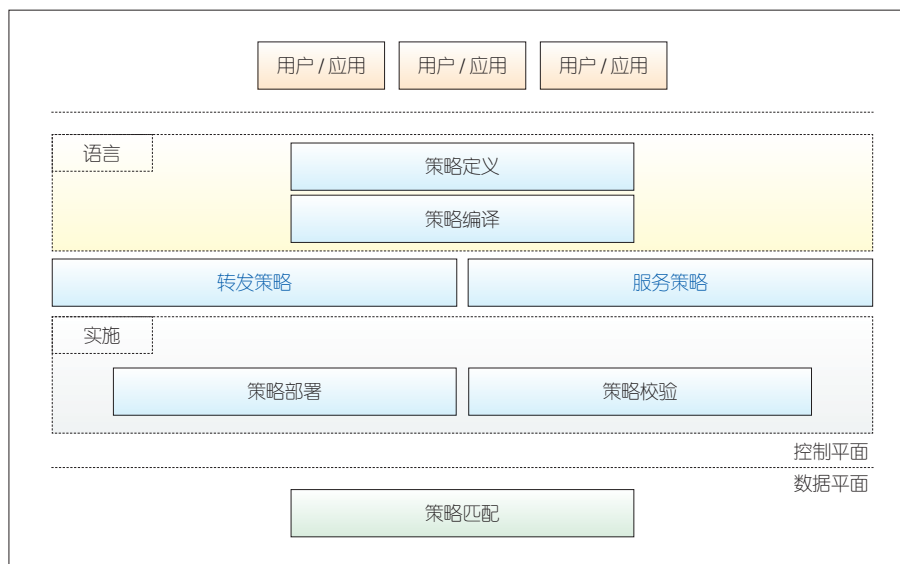
最下层是分布的数据平面。该层完成网络流量的策略匹配与状态监测,并执行相应的网络功能。

图2展示了策略编排在网络自动化系统架构中对应的核心技术构成。

策略语言中定义和编译环节的关键在于意图基元的设计。这种设计背后反映的是对网络流量、协议和拓扑的描述,以及对生成、修复等不同管理操作的构建。在策略实施中,策略部署的关键在于求解给定网络约束下策略分配和部署网元(或节点)选取的优化问题,策略校验的关键在于对拓扑和网元的建模以及适应性算法的设计。策略匹配的关键在于软件算法优化和硬件平台加速。整体来看,当前的策略编排技术能够针对具体的网元和拓扑需求抽象出特定模型,并能应用特定算法。但从系统性的角度看,实现零触碰的统一化和插件化还有很多工作要做。



▲图1 网络自动化系统的逻辑架构



▲图2 策略编排的核心技术构成

零触碰网络的理念来源于 Google 在 2016 年发表的学术论文^[1]。欧洲电信标准化协会 (ETSI) 后续成立了面向 5G 的零触碰网络与服务管理工作组。与零触碰相类似的理念包括 2016 年 Juniper 公司提出的自动驾驶网络 (SelfDN) 和 2017 年 Gartner 公司提出的 IBN。这些理念的共性都是要实现自动化的策略编排, 减少网络基础设施的交付时间, 并降低网络故障的发生频次。

在零触碰方面上, 走在业界前沿的是 Google、Microsoft、阿里云等大型云计算厂商。零触碰网络的业界实践可参考 Google 的设计方案 (如图 3 所示)。该方案在系统架构上与图 1 所描述的网络自动化架构相似。阿里云在 2019 年发表了针对骨干网接入控制的策略编排, 定义并实现了特定领域的意图。拥有面向云数据中心和混合云场景的网络自动化方案的代表性厂商有 Apstra、Intentionet 等。在零触碰产品的市场认可和推广方面, 向下需要扩充对底层网元功能和协议类型的

适配广度, 向上需要丰富典型业务场景的操作意图参考设计, 以满足网络连通到网络安全等多样需求, 实现系统的“能观”与“能控”。

2 网络安全与零信任

信任是人际交往和交易中影响效率的重要因素, 因此网络访问或网络交互也必然依赖于信任的建立。关于零信任, 业界流传着这样一句话: “永远不要信任, 始终进行验证, 实施最小权限。”其实, 不是“永远不要信任”, 而是“不要永远信任”。也就是说, 应将传统的一次性认证改为经常性查验, 而且不是严格意义上的不间断地“始终进行验证”。通常, 人们还是采用定期采样或事件触发的验证, 并给予一定期限和条件的授权。

零信任的实现, 不仅需要引入新技术或新设备, 还需要借助微隔离 (MSG) 以及细粒度的边界策略。在虚拟网络特别是云场景中, 相对于传统模式来说, 规模和复杂性增加很多,

因而工程实现的难度也将增加。在解决新问题、引进新能力的同时, 零信任的实施也牵涉更多的人力和物力资源。适度的信任和自动化, 是在特定安全等级下降低成本、提升效率的良好途径。

当然, 面对自带设备 (BYOD) 和 5G 带来的接入多样化, 应用与系统的日趋云化, 以及社会组织、网络安全的态势改变, 信任和风险总是相互关联的。

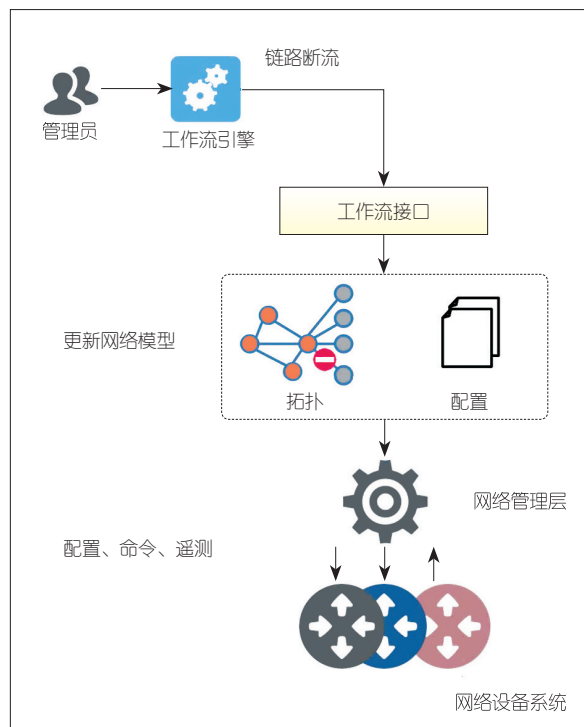
零信任的前提是实施最小特权, 而零信任的基础是实践动态认证。不同应用场景下的零信任系统, 会因地制宜地选择和融合

多种网络安全和数据安全技术。但身份认证和边界控制是所有零信任实践的核心。

最小特权也称最小授权, 是信息安全的基本模型之一。最小特权的概念最早源于容错理论。它涉及特权分离和完全仲裁等一系列原则。在网络安全实战中, 它可以减少应用的公共可见性, 从而显著减小攻击面。一般情况下, 传统网络物理边界的失效, 并不意味着“无边界”。网络不同安全区域之间的逻辑边界是实施安全分级 (等级保护) 治理的前提条件。普通用户对网络边界无感并不意味着完全没有边界。零信任恰恰是在软件定义边界 (SDP) 和云访问安全代理 (CASB) 等技术体系的基础上整合出来的解决方案。当然, 边界的极致就是终端。零信任也可以在终端侧部署, 以解决传统边界防护在端到端远程访问场景下的诸多问题。

动态认证则是身份权限管理 (IAM) 中身份验证和鉴权的延伸。从最简单的“五元组”接入控制表 (ACL), 到基于角色的接入控制 (RBAC), 再到基于属性的接入控制 (ABAC), 都是网络安全中身份认证不断精细化的体现。然而, 除了简单的定期复检 (超时) 外, 它们大多不会“与时俱进”, 反映即时的变化。而对用户、设备、网络“身份”的识别和认证, 以及对网络安全相关状态、流量、行为甚至“全息”数据的掌控, 则是安全身份判定逐步动态化的体现。实际上, 这和大数技术中常用的“用户画像”概念十分类似, 它们都通过精准定位来服务对象, 以达到提供精准服务的目的。零信任正是通过精细化、动态化的风险评估, 才实现了相应的接入或准入控制, 在空间和时间上为网络提供最大程度的防御。

目前, 业界参考的零信任理念大



▲图 3 Google 的零触碰网络架构^[1]

多基于美国国家标准研究院 (NIST) 于 2020 年发布的《零信任架构 (ZTA)》(2019 年以建议为名发布第一版草案)。其中, 零信任的核心技术被归纳为“SIM”。这里, “S”是 SDP, “I”是 IAM, “M”是 MSG。市场上已经推出的零信任产品则是在厂家原有技术基础上扩充而来的。这些产品或是基于云安全联盟 (CSA) 的 SDP 规范 (如图 4 所示), 或是参考 Google 的 BeyondCorp (如图 5 所示)。

其实, 还有一个技术框架可以被整合、升级, 并作为零信任技术的基础, 那就是 TNC (可信网络连接)。TNC 经历过 Cisco 和 Microsoft 两大阵

营的多年竞合, 有 IETF 系列协议层面的标准 RFC 支撑, 在身份和安全状态、情报等数据交换格式上具备扎实的基础, 利于推进开放兼容, 避免形成厂商锁定。

3 结束语

近年来, 网络领域最主要的变革都源于网络虚拟化和 SDN。无论是意图驱动网络, 还是自动驾驶网络, 它们都指向网络自动化, 即零触碰。而这一切在网络安全方面的反映, 就是零信任。

无论是零触碰还是零信任, 它们的关键都是实现闭环。从业务的视

角看, 它们实质上是基于反馈控制的网络编排和信任管控自动化。有趣的是, SDN 和零信任都是 Google 率先验证和推出的。SDN 将控制平面与数据平面分离, 并将分布、自治的网络升级为集中与分布式结合的控制系统。作为经典实践, Google 的 B4 显著提升了网络带宽的有效利用率。较为完整的零信任概念由市场分析和咨询公司 Forrester 于 2010 年提出。之后, Google 经过 7 年时间, 成功地将零信任全面上线部署。

总之, 以零触碰和零信任为目标的网络自动化, 已经成为大势所趋。它正在不断推进技术创新、产品研发和服务落地。

参考文献

- [1] KOLEY B. The zero touch network [EB/OL]. [2021-03-18]. <http://www.cnsn-conf.org/2016/presentations/CNSM2016-Keynote1-Koley.pdf>
- [2] 绿盟科技. 零信任安全解决方案 [EB/OL]. [2021-03-18]. https://www.nsfocus.com.cn/html/2020/210_0608/129.html
- [3] BeyondCorp. Run zero trust security like Google [EB/OL]. [2021-03-18]. <https://beyondcorp.com>

作者简介

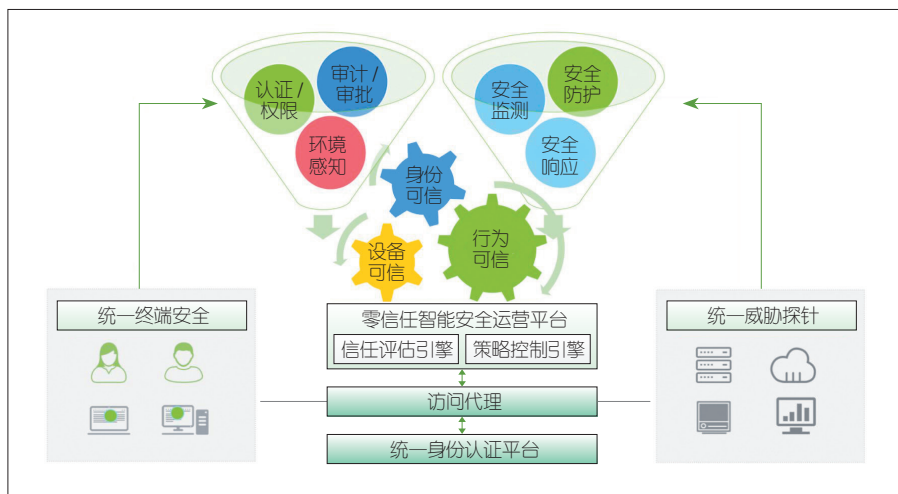


李军, 清华大学自动化系研究员、博士生导师, 中国电子学会计算机工程与应用分会副主任委员; 主要从事网络与网络安全等领域的教学和研究工作; 主持了多个“863”、国家重点研发计划和自然科学基金等项目; 作为第一完成人

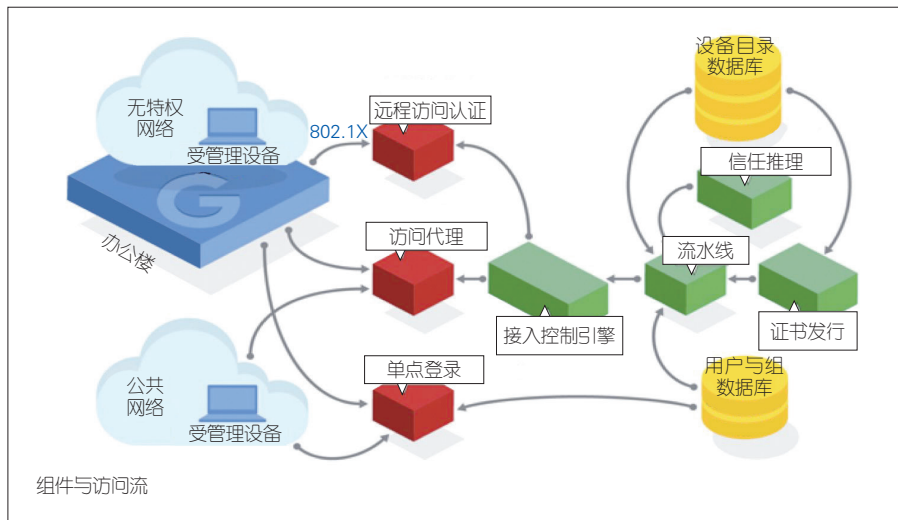
荣获 2014 年中国电子学会科学技术奖二等奖; 著译中外教材 3 部, 发表学术论文 100 余篇, 获得美国专利 2 项、中国发明专利 20 余项, 且多数成果已商用。



胡效赫, 清华大学计算机系博士后; 主要从事软件定义网络、高性能网络处理、安全隐私等方向的研究工作; 先后参与国家重点研发计划和自然科学基金等项目; 发表论文 10 余篇, 获美国专利 1 项、中国发明专利 2 项。



▲图 4 绿盟科技的零信任网络安全架构^[2]



▲图 5 Google 的零信任企业安全^[3]



数据中心网络架构和底层协议演进

Data Center Infrastructure and Underlay Protocol Evolution

魏月华 /WEI Yuehua
陈晓 /CHEN Xiao
张征 /ZHANG Zheng

(中兴通讯股份有限公司, 中国 深圳 518057)
(ZTE Corporation, Shenzhen 518057, China)

DOI: 10.12142/ZTETJ.202103011

网络出版地址: <https://kns.cnki.net/kcms/detail/34.1228.tn.20210617.0922.002.html>

网络出版日期: 2021-06-17

收稿日期: 2021-05-10

摘要: 受计算规模的驱动, 数据中心物理拓扑从接入-汇聚-核心三级网络架构演进到基于 Clos 的 Spine-and-Leaf 架构。计算资源的基本单位经历了物理服务器、虚拟机、容器化 3 个阶段。数据中心底层 (underlay) 连接协议逐步从以二层协议为主演进到以 IP 路由协议为主。但传统路由协议存在可扩展性、拓扑可见性、自动化部署能力等诸多问题。结合链路状态和距离矢量的胖树路由协议, 解决了超大规模数据中心部署的痛点问题, 有望逐渐成为超大规模数据中心底层网络的主流技术。

关键词: Spine-and-Leaf; 路由; 数据中心

Abstract: Driven by the scale of computing, the physical topology of the data center has evolved from an access-aggregation-core three-level network architecture to a Clos-based Spine-and-Leaf architecture. The basic unit of computing resources has gone through three stages: physical server, virtual machine, and containerization. The underlay connection protocol of the data center has gradually evolved from layer 2 protocol to IP routing protocol. However, traditional routing protocols have many problems, such as scalability, topology visibility, and automated provision capabilities. The fat-tree routing protocol, which combines link state and distance vector, solves the pain points of ultra-large-scale data center deployment, and is expected to gradually become the mainstream technology for ultra-large-scale data center underlay networks.

Keywords: Spine-and-Leaf; routing; data center

1 接入-汇聚-核心三级网络架构协议方案演进

受计算规模的驱动, 数据中心的网络架构和解决方案, 在过去 20 年里发生了很大变化。总的来说, 数据中心物理拓扑从接入-汇聚-核心三级网络架构演进到基于 Clos 的 Spine-and-Leaf 架构。计算资源的基本单位经历了从物理服务器到虚拟机再到容器化 3 个阶段。

在物理服务器阶段, 应用直接在物理服务器上运行, 数据中心物理拓扑为经典的接入-汇聚-核心三级网络架构, 整张网络采用二层协议互联, 应用访问模式为客户端-服务器模式,

并且南北向流量远大于东西向流量。其中, 南北向流量在核心交换机处理, 数据中心内跨网段需要经过核心交换机, 内部子网的网关一般也配置在核心。在这种模型中, 由于节点之间的通信都可能经过核心, 因此核心交换机需要记录所有节点的互联网协议 (IP) 和介质访问控制 (MAC) 地址信息。在这种网络方案中, 与计算节点规模相关的瓶颈最可能出现在核心交换机中。

2008 年, 传统的数据中心逐步演进到云计算时代的数据中心。云计算时代计算资源的基本单位从物理机变成了虚拟机。计算资源的数量和密度都有数量级的提高。应用广泛采用微

服务访问模式。这种模式带来的网络变化是: 东西向流量超过南北向流量, 成为数据中心的主要流量。

随后, 网络虚拟化应运而生。数据中心网络中的每个宿主机都运行一个虚拟交换机 (vSwitch)。虚拟交换机向上连接物理交换机, 向下连接多个虚拟机。网络的边界从原来的接入交换机 (置顶交换机) 层, 下沉到宿主机内部。这使得整张网络变成一个大的二层网络。在这个大二层网络内, 虚拟机生命周期内的 IP 地址和 MAC 地址均保持不变。对于同网段的虚拟机, 不管它们是否在同一台宿主机上, 彼此都能够通过二层 (MAC 地址) 访问对方。此时, 核心交换机不仅需要

记录宿主机的 IP/MAC 信息，还需要记录所有虚拟机的 IP/MAC 信息，以便支持虚拟机全网可迁移。

2016 年以后，数据中心进入大规模容器时代。容器也被称为轻量级虚拟机，可进一步提高部署密度。虚拟机与容器的最大区别在于：虚拟机平台交付的是虚拟机实例，抽象的是计算资源，而容器平台交付的是服务，访问入口为服务的 IP 地址，同时服务屏蔽了计算资源的细节（如虚拟机实例的 IP 地址或 MAC 地址）。

当把虚拟机换成容器后，考虑到容器的部署密度，如果继续采用大二层模型，交换机转发表容量将会成为网络瓶颈。为此，在每个服务器节点内可用虚拟路由器（vRouter）替换虚拟交换机。一个虚拟路由器管理一个网段。服务器域内是一个二层网络。服务器节点运行边界网关协议（BGP）代理，并负责节点之间或者节点和数据中心网络之间的路由同步。核心交换机只需要记录服务器节点本身的 IP 和它所管理的网段。表项与服务器的数量保持同一量级，但与容器的数量没有关系。

因此，数据中心网络拥有一个在三层网络下有无数个小二层网络的架构，如图 1 所示。这种以三层路由为主的数据中心协议架构，可以满足现代数据中心规模不断扩大和服务器数量不断增加的需求。

2 带宽与流量模型的变化

传统数据中心的流量主要是进出数据中心的流量，通常被称为南北向流量。即使在网络层之间存在很高的收敛比，传统的“树”拓扑也足以容纳这样的流量。如果需要更多的带宽，则可以通过“扩展”网络元素来增加带宽。例如，升级设备的线路板，或者采用端口密度更高的设备。

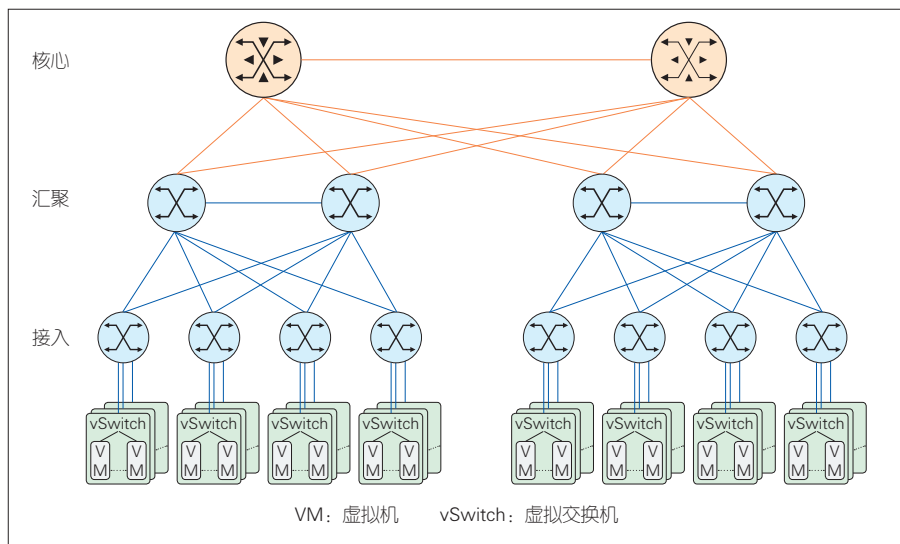
如今，许多大型数据中心承载着大量服务器到服务器的流量。这些流量并不会离开数据中心，通常被称为东西向流量。例如，某些应用程序需要集群之间的海量数据进行复制，或者需要虚拟机进行迁移。由于受到物理限制（例如交换机的端口密度低），采用扩展传统的树形拓扑来满足带宽需求的方式，不仅成本很高，而且难以实现。

3 基于 Clos 的 Spine-and-Leaf 结构演进

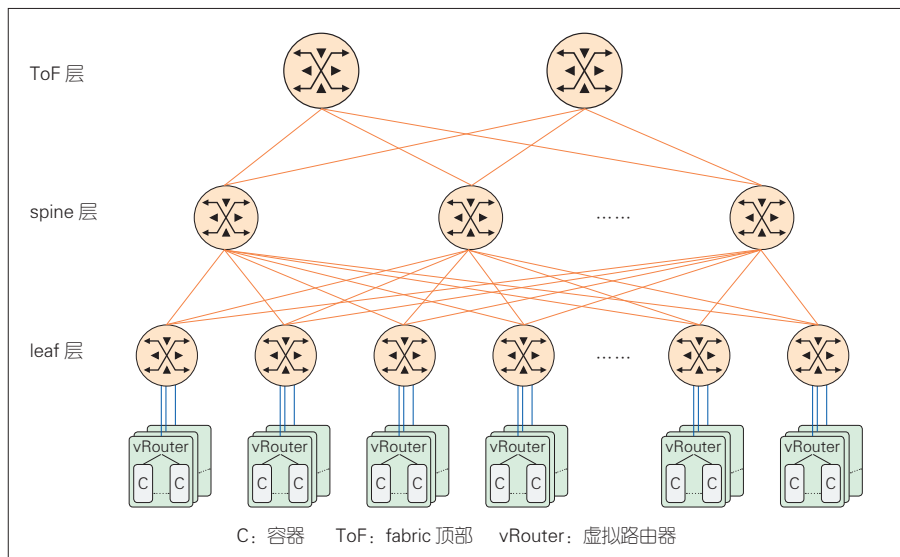
东西向流量的增加使三层数据中

心架构中的带宽成为瓶颈。此外，服务器到服务器的延迟会随着流量路径的不同而不同。为了解决这两个问题，基于 Clos 网络的 Spine-and-Leaf 架构被提出。

在如图 2 所示的三级 Clos 架构中，每个低层级的 leaf 交换机都与所有高层级的 spine 交换机相连，并形成全网状连接拓扑。leaf 交换机用于连接服务器等设备，spine 层则负责将所有的 leaf 连接起来。当 leaf 层的接入端口和上行链路都没有瓶颈时，这个架构就实现了无阻塞连接。



▲图 1 网络虚拟化与接入 - 汇聚 - 核心网络结构



▲图 2 典型的 Spine-and-Leaf 拓扑

在 Spine-and-Leaf 架构中,任意一个服务器到另一个服务器的连接,都需要相同数量的设备(除非这两个服务器都在同一个 leaf 下)。这使得延迟可以被预测。由于东西向带宽更高,因此它更适合现代微服务的场景。

当 Spine-and-Leaf 中任意一层存在带宽瓶颈时,只需要添加一台新设备,并将其和另外一层的所有设备相连即可。这种横向扩展的方法比较容易实施。

4 数据中心协议的选择与设计

4.1 选择三层路由的 Spine-and-Leaf 架构

Spine-and-Leaf 结构相当于传统网络架构中的“接入层-汇聚层”。如果采用二层交换技术,则生成树协议(STP)生成的无环树形结构会大大减少活跃可用的链路。

如果采用三层路由,Spine-and-Leaf 则可以充分利用 spine 和 leaf 之间的全网状连接,并选择最短路径。如果为了获得更高的整体利用率,该架构也可以选择特定的路径。

4.2 BGP 路由协议部署技术与特征^[1-2]

BGP 在应用于数据中心之前,主要用于运营商网络。BGP 数据中心与运营商网络最大的区别在于连接的密度:超大型数据中心的连接密度远大于运营商网络的连接密度。因此,BGP 协议在应用于数据中心之前需要经过适当的“改造”。

BGP 协议具有一些突出优势,主要包括:

(1) 作为距离矢量协议,BGP 采用传输控制协议(TCP),互操作性好,总体上很成熟,目前已经获得广泛应用。设备商和各种开源平台都实现了 BGP 部署,并获得了良好的测试结果。

(2) 由于 BGP 本身在广域通信网络上是一个广泛部署的路由协议,因此,从技术和运维的角度上看,将 BGP 应用于超大规模数据中心网络具有很高的接受度;

(3) 相比于其他内部网关路由协议,BGP 具有较高的可扩展性;

(4) BGP 协议有诸多前缀过滤、路由标记和流量工程的能力选项,在过滤、修改路由参数和控制流量方面具有优势;

(5) BGP 可以同时用于底层(underlay)网络和叠加(overlay)网络。通常在这种情况下,底层网络使用外部 BGP(eBGP)对等体,叠加网络使用内部 BGP(iBGP)对等体。这使得网络的整体配置变得更简单。

BGP 协议作为数据中心的底层也面临一些挑战,具体包括:

(1) 由于 BGP 协议具有易于扩展的特性,BGP 上逐步增加的多地址族、以太网虚拟专用网(EVPN)、虚拟专用局域网业务(VPLS)、BGP 链路状态(BGP-LS)等能力,使得 BGP 协议变得非常复杂。虽然可以通过一些开关来关闭这些功能,但是实际上仍无法避免实现 BGP 功能的软件代码漏洞和错误配置等问题;

(2) BGP 协议在自动化能力方面不足以满足大规模数据中心的需求;

(3) 在数据中心 fabric 中的高密度拓扑中,需要大量专业的手动配置来使 BGP 快速收敛。例如,当流量从 fabric 上的一个位置移动到另一位置,或者当由 anycast 地址代表的一个服务实例从 fabric 上被删除时,BGP 收敛时间会很长。这将影响在 fabric 上正常运行的应用。

4.3 链路状态路由协议的演进^[3]

自 RFC 7938(在大规模数据中心路由中使用 BGP 的标准)发布起,

BGP 几乎成了大规模数据中心的缺省选择。考虑到标准和部署的多种因素(如收敛速度、数据遥测等),业界提出在数据中心 fabric 中采用链路状态路由协议来代替 BGP 协议。

在超大规模数据中心采用链路状态路由协议的最大的挑战是,存在用于可达性计算和拓扑计算的路由信息洪泛问题。目前,国际互联网工程任务组(IETF)正在针对中间系统到中间系统(IS-IS)开展洪泛优化和集中计算优化泛洪树的工作。

在数据中心 fabric 中,与 BGP 协议相比,链路状态协议具有收敛速度快的优点。当一个可达目的地在 fabric 中从一个地方移动到另一个地方,或者完全从 fabric 上断开时,链路状态协议的收敛速度将远快于 BGP 的收敛速度。从 IS-IS 的角度来看,任何可达目标的更改都只是叶子连接的更改。这意味着系统无须运行最短路径优先(SPF)算法。这种方法被称为部分 SPF。它的速度非常快,并且每个交换矩阵设备只需要进行最少量的处理。

与数据中心结构中的 BGP 相比,链路状态协议的第二个优势是拓扑可见性。链路状态协议要求每个设备都拥有维护拓扑的完整视图。该拓扑(称为链接状态数据库)必须与网络洪泛域中的每个路由器同步。在使用控制器时,为了获得链路状态数据库的副本,链路状态协议仅需要连接光纤网络中的一个路由器。链接状态数据库对于流量工程和流量导流很有用,也有利于做数据遥测。

数据中心结构中链路状态协议面临的第一个挑战是扩展问题,这主要与消息洪泛有关。由于消息量大,链路状态协议会在大型结构中造成严重的洪泛。

此外,链路状态协议还面临另外

两个挑战：存在可达目的地数量的扩展性问题和计算无环路径集 SPF 算法所需的时间较长的问题。通过更快的处理器和 SPF 优化，虽然不能使链路状态协议的扩展性达到 BGP 的级别，但是足以支持运营商构建大部分的数据中心结构。

4.4 胖树路由协议特征分析^[4-6]

业界对数据中心 fabric 中路由技术的探索从未停止。针对基于 Clos 网络的 Spine-and-Leaf 结构，IETF 启动了结合距离矢量路由与链路状态路由的胖树路由协议的标准化工作。

胖树路由协议可将链路状态协议和距离矢量协议的优点结合起来，以最大程度地实现网络路由配置自动化和故障管理自动化，并用于 Spine-and-Leaf 结构的大规模数据中心中。胖树路由协议支持多线程，可匹配多核 CPU 的处理能力。因此，胖树路由协议可以极大地节省操作和运维成本，并减少人为错误。

4.4.1 拓扑适用性分析

如前所述，在数据中心进入云计算时代以后，东西向流量就超过了南北向流量，成为数据中心的主要流量。东西向流量在虚拟服务器与虚拟服务器之间，以及容器与容器之间的转发，本质上还是在胖树的北向与南向运动。只不过东西向流量的转发是最大程度的就近转发。

流量从 Spine-and-Leaf 结构底部的 leaf 节点向北到达结构的顶部，然后向南回到 leaf 节点。从所需的可达性信息角度来看，这种服务器到服务器的流量模式，所需的可达信息很少。例如，在三级 Clos 中，leaf 节点流量仅需要默认路由即可到达 spine 节点。同时 spine 节点流量不需要整个路由表即可到达 leaf 节点，只需要向南一级

的节点可达信息。因此，胖树路由协议具有方向特性，具体表现为：向北为链路状态协议，向南则为距离矢量协议。

胖树结构（Spine-and-Leaf 结构）天然分层：结构顶部的节点保持在最高级别，而底部节点（leaf 节点）保持在最低级别。胖树路由协议用方向性来描述拓扑中不同级别之间的关系，并利用拓扑的这种特性，通过零接触部署（ZTP）功能进行错误布线检测。另外，这种协议在设计时也考虑了容错性，因此能够应对胖树结构的变异，比如同一层节点之间的水平链路或跨层的垂直直连链路。

4.4.2 拓扑发现

胖树路由协议通过交换链路信元（LIE）自动发现邻居，协商 ZTP，并检测错误布线。LIE 交换采用用户数据报协议（UDP），并且将互联网协议第 4 版（IPv4）报文中的生存时间值（TTL）（或互联网协议第 6 版报文中的 Hoplimit）设置为 1。LIE 包含的关键信息有本地链路 ID、SystemID、最大传输单元（MTU）、本地节点的交付点（PoD）值、所属层值等。

胖树路由协议通过交换拓扑信元来携带一个节点连接的邻居、前缀和能力等信息。由于胖树路由协议具有方向特性，拓扑信元可分为北拓扑信元和南拓扑信元。

无论是南拓扑信元还是北拓扑信元，拓扑信元都包括 6 种类别：节点拓扑信元、前缀拓扑信元、积极解聚合拓扑信元、消极解聚合拓扑信元、外部前缀拓扑信元和键值拓扑信元。

拓扑信元交换（洪泛）采用 UDP 协议，具有方向性。所有的北拓扑信元都是向北洪泛的，目的在于为更高层提供以南网络的完整拓扑视图。这可以保证从特定层节点（或低于特定

层节点）收到的流量始终采用最具体的路由来到达目的节点。

所有南节点拓扑信元都被往南泛洪，而其他类型的南拓扑信元仅往南泛洪本节点为发起者的拓扑信元。这样，低一级的节点就会拥有去往上层节点所需要的路由信息。这些信息也可以到达 fabric 的其他地方。

胖树路由协议采用类似 IS-IS 协议的方式来保持链路状态数据库的同步。在计算最短路径时，胖树路由协议也是基于南向或北向的。两个方向的最短路径算法都不会产生环路：往北向的最短路径算法只利用北向（和东西向）邻居来计算“北拓扑信元”，往南向的最短路径算法只利用南向邻居来计算“南拓扑信元”

4.4.3 负载均衡

IP 网络中的负载均衡一直是个难题。BGP 负载均衡实施困难，而内部网关协议（IGP）仅能做到等价路径负载均衡。在胖树路由协议中，负载均衡只需要在北向的缺省路由上来实现（也可以在解聚合前缀和南向路由上来实现）。胖树路由协议自动计算并继续使用所有可用最短路径上的可用带宽，使流量不会在 fabric 中迂回打转。

在正常情况下，每个前缀都带有一个关联的距离值（相当于典型的度量值）。当链路发生故障时，SPF 计算必须考虑当前不可用的带宽，并计算带宽调整后的距离（BAD），然后使用 BAD 值来代替初始距离值，以评估可用链接上的流量。

4.4.4 南向反射与路由解聚合

这种反射机制是指，只有节点的南向拓扑信元会被往北反射到上一层。因此，同一层的所有节点都能够相互感知对方。

反射机制可以触发积极解聚合。

为了解决流量黑洞问题，路由解聚合在发布缺省路由的基础上，会再发布一个更详细的路由。

解聚合包括两种类型：积极的解聚合和消极的解聚合。节点发布积极路由表示它可以到达某个前缀。而当节点不能到达某个前缀时，则通告消极路由。不管是哪种情况，解聚合的路由总是被通告为前缀或外部南拓扑信元，并且永远不会被重发。同时，其他节点不需要知道哪个节点正在发布解聚合的路由。

积极解聚合很简单。它是一种额外的路由通告。这样，南方的节点可以根据典型的最长匹配原则来进行路由布置，即胖树路由在默认路由中为部分连接的前缀打一个洞。

积极解聚合是非传递性的，以免给节点增加无用的路由信息。对于未解聚合的前缀，默认路由将为其提供可达性。

消极解聚合相对比较复杂。当 fabric 包含多个平面时，消极解聚合就是必需的。当某个节点失去某前缀的可达性时，该平面中所有上一层的节点都会触发消极解聚合。与积极路由不同，消极路由是可传递的。消极路由可以一直向南广播，直到解除流量黑洞。

4.4.5 零接触部署

胖树路由协议内置了零接触部署模式。除了 ToF 节点之外（ToF 节点需要预先设定一个层值），其他节点无需任何初始化配置就可以自动接入 fabric 中。每个节点都以竞争在 fabric 中的最高点为原则。层决策算法利用

相邻节点之间的位置信息进行运算，以确保所有节点找到在 fabric 中的稳定位置，从而自动完成一个稳定的胖树拓扑构建，并自动实现南向和北向路由策略。零接触部署能力能够有效消除可能的由错误布线对 fabric 构建产生的干扰。

零接触部署是胖树路由协议最突出的特性之一，对于提升超大规模数据中心网络构建的效率意义重大。

5 结束语

在未来，BGP 将继续成为数据中心架构底层的重要选择。它最终会具备一些链路状态协议功能，例如更快的收敛和更接近自动化的部署。然而，BGP 很难复制链路状态协议的某些功能，例如从一个位置获取整个拓扑的完整视图。同时，BGP 的收敛速度很可能总是落后于链路状态协议。对此，IETF 已经启动改进链路状态协议的标准化工作。但由于改动较大，同时协议复杂度较高，因此协议应用前景不明。胖树路由协议可将链路状态和距离矢量相结合：当数据报文沿 fabric 向上传递到 ToF 时，可采用类似链路状态的操作；当数据报文向 fabric 的边缘传递可达性和拓扑信息时，可采用类似距离矢量的操作。胖树路由协议解决了现有路由协议在 Spine-and-Leaf IP 结构中面临的诸多问题，具有扩展性好、运维简单的优点，可有效节省部署开销。

中兴通讯在 IETF 深入参与了胖树路由协议的标准化工作。我们认为，胖树路由协议有望成为超大规模数据中心底层网络的主流技术。

参考文献

- [1] IETF. Use of BGP for routing in large-scale data centers: RFC 7938 [S]. 2016
- [2] Dinesh G D. BGP in the data center [M]. California: O' Reilly Media, Inc. 2017
- [3] IETF. Dynamic flooding on dense graphs: draft-ietf-lsr-dynamic-flooding-08 [S]. 2020
- [4] IETF. RIFT: routing in fat trees: draft-ietf-rift-rift-12 [S]. 2021
- [5] IETF. RIFT applicability: draft-ietf-rift-applicability-06 [S]. 2021
- [6] IETF. A YANG data model for Routing in Fat Trees (RIFT): draft-ietf-rtgwg-policy-model-27 [S]. 2021

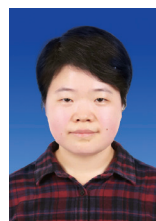
作者简介



魏月华，中兴通讯股份有限公司承载网标准预研总工；拥有 15 年以上数据网络产品研发、设计及新技术预研经验；从事以太网、IP 路由、云计算数据中心网络、SDN 等技术和标准研究；发表论文 3 篇，获授权专利 40 余项。



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。



张征，中兴通讯股份有限公司标准专家；拥有 20 年的数据网络产品研发与设计经验；从事 IP 单播/组播路由、数据中心网络、SDN 等技术与标准研究；主持多个 IETF 工作组标准的制定和 RFC 的发布；申请发明专利 40 余项。



共建共享下的边缘云建设

Edge Cloud Construction under Co-Construction and Sharing

摘要: 分析 5G 共建共享后边缘云共享对边缘云建设的可能影响。基于接入网共享和异网漫游两种 5G 共建共享主流策略,介绍 5G 非独立组网 (NSA) / 独立组网 (SA) 下边缘云的组网架构,以及承建方与共享方运营商用户在不同情况下访问共享边缘云的方式,并从用户身份鉴权、计费方式、服务质量 (QoS) 策略等方面剖析相应的网络能力要求。

关键词: 5G; SA; NSA; 共建共享; 边缘云共享

Abstract: The possible impact of edge cloud sharing on edge cloud construction after 5G co-construction and sharing is analyzed. Based on the two mainstream 5G co-construction and sharing strategies of access network sharing and off-network roaming, the network architecture of edge cloud under 5G non-standalone (NSA) / standalone (SA), as well as the ways for users of the contractor and the sharing operator to access the shared edge cloud under different circumstances is introduced. Besides, the corresponding network capability requirements from the aspects of user authentication, billing method, and quality of service (QoS) strategy are also analyzed.

Keywords: 5G; SA; NSA; co-construction and sharing; edge cloud sharing

黄倩 /HUANG Qian
黄蓉 /HUANG Rong

(中国联合网络通信有限公司研究院, 中国 北京 100044)
(China Unicom Research Institute, Beijing 100044, China)

DOI: 10.12142/ZTETJ.202103012
网络出版地址: <https://kns.cnki.net/kcms/detail/34.1228.TN.20210617.0942.006.html>

网络出版日期: 2020-06-17
收稿日期: 2021-05-11

在 5G 牌照发放以后,运营商紧锣密鼓地推进 5G 网络建设。为降低网络建设和运维成本,提升网络效益和资产运营效率,运营商将共同承建 5G 网络。运营商在某个信号覆盖区域内,可以允许其他用户进行 5G 接入。5G 共建共享的方式可快速形成 5G 服务能力。其中,作为 5G 网络的重要组成部分,边缘云通常被部署在运营商基站接入侧、汇聚机房或更高层级的区域数据中心。在 5G 共建共享后,某一区域内可能只存在一家运营商基站接入。这种部署仅能满足 5G 一般业务的需求,当面临访问边缘云平台或本地分流等场景时,还存在一些不足。

这里,我们提出两种假设。(1) 假设 5G 共建共享,边缘云不共享。首先,承建方拥有 5G 基站和机房资源,可同步部署边缘云扩展业务。然而,共享方只能共享使用承建方的 5G 基

站。是否在基站侧部署边缘云与基站互通,目前仍无法确定。其次,如果边缘云不共享,那么业务流将依赖承载网进行互通。在哪一层级实现互通与边缘云实际部署的位置有关。不同层级承载网互通的难度和成本均不相同。最后,边缘云的相关平台能力的实现,如业务分流、无线信息开放等,是否会受到 5G 共享的影响,尚未明确。(2) 假设 5G 共建共享,边缘云也共享。这种情况势必会对边缘云相关技术和策略产生影响。

本文中,我们重点讨论在第 2 种假设情形下,5G 共建共享网络策略、边缘云组网架构、边缘云共享对共享方和承建方用户访问移动边缘计算 (MEC) 方式的影响和存在的问题,并从用户身份鉴权、计费方式、服务质量 (QoS) 策略等方面剖析相应的网络能力要求。

1 5G 共建共享网络策略

5G 非独立组网 (NSA) / 独立组网 (SA) 下的网络共建共享存在两种方式:接入网共享和异网漫游。这两种方式也是主流共建共享方案,具有较强的实际指导意义。

1.1 接入网共享

5G NSA 下的接入网共享方案是指,运营商 A 和运营商 B 共享接入网,双方的运营商用户均可接入共享基站,并且各自接入核心网,如图 1 所示。为使不同运营商用户接入各自的核心网,承载网需要被共享,即在承载网的某层实现东西向互通。从用户体验来讲,这基本等同于自建网络。在接入网共享方案中,由于 NSA 架构仍然需要 4G 作为锚点站,以实现控制面 and 用户面的信令传输,因此,4G 和 5G 基站均共享。

1.2 异网漫游

5G NSA 异网漫游方案是指, 5G 基站仅接入承建方核心网, 双方核心网对接互通, 如图 2 所示。在异网漫游方案中, 运营商 A 和运营商 B 的用户均接入共享接入网。然而, 非承建方用户需经过建设方核心网, 并通过漫游方式访问核心网。这就像用户通过国际漫游的方式来享受 5G 服务一样。在异网漫游方案中, 由于 NSA 架构仍然需要 4G 作为锚点站, 以实现控制面和用户面的信令传输, 因此, 4G 和 5G 基站也均共享。

在 5G SA 架构下, 共享方案也分为基站共享和异网漫游方案。相比于 5G NSA 下的两种共建共享方案, 它们的总体架构相同, 唯一的区别在于: 由于 5G SA 共享方案不需要将 4G 作为锚点站, 因此, 仅需要所有的 5G 基站共享建设, 运营商各自的 4G 基站无须共享。

实际上, 基于建设成本和业务开展的综合考虑, 目前中国运营商采用接入网共享方案。

1.3 边缘云组网架构

1.3.1 5G NSA MEC 组网架构

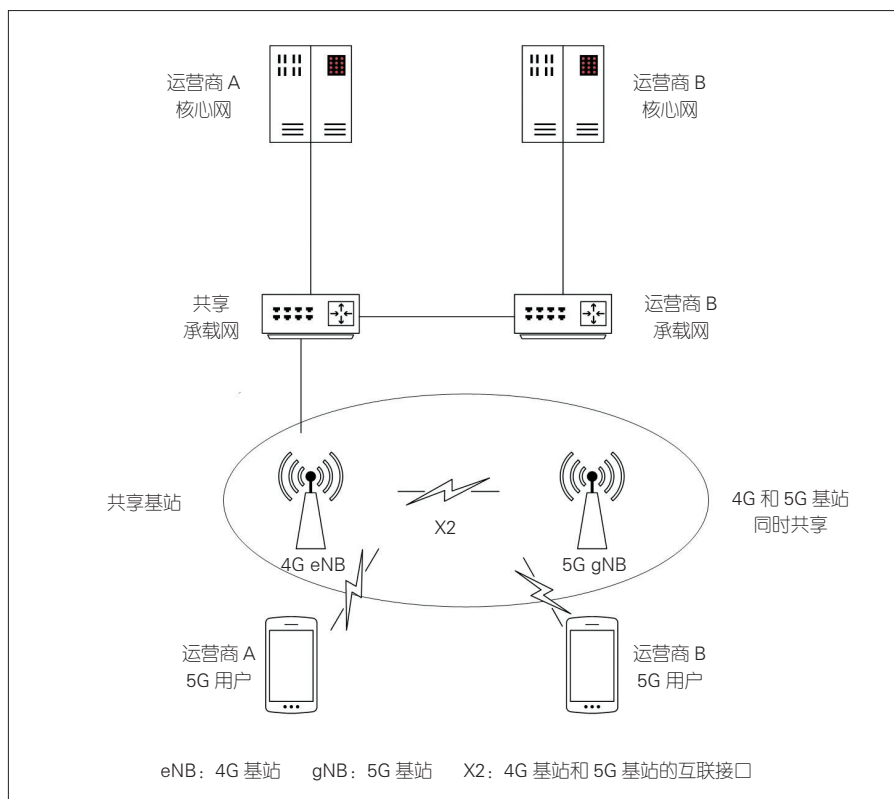
如图 3 所示, NSA 下 MEC 边缘云的部署位置与 4G 相同, 即依然串接在 S1-U 接口上, 并在核心分组网 (EPC) 和新空口 (NR) 之间。MEC 边缘云部署可以是分流+业务服务器分开部署, 也可以是一体化部署, 以实现计费等功能。

1.3.2 5G SA MEC 组网架构

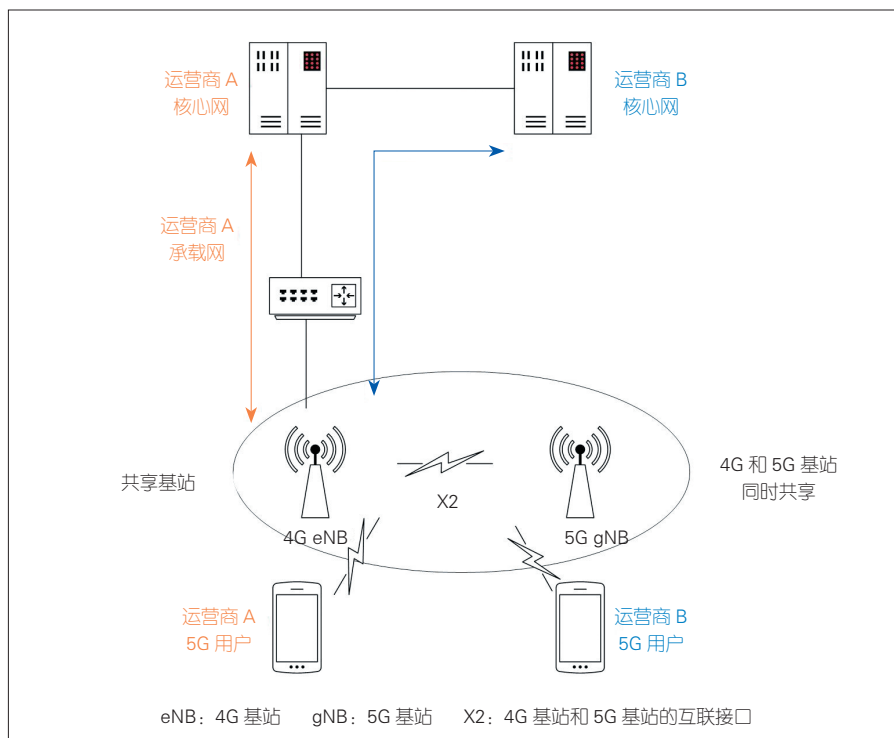
5G 网络架构下, MEC 边缘云平台一般以虚拟化的形式部署, 如图 4 所示。MEC 与网络功能虚拟化 (NFV) 技术相融合, 可以实现按需调用和灵活部署。MEC 边缘云部署位置在用

户面功能 (UPF) 后面。因此我们可以根据 UPF 位置和业务要求来部署

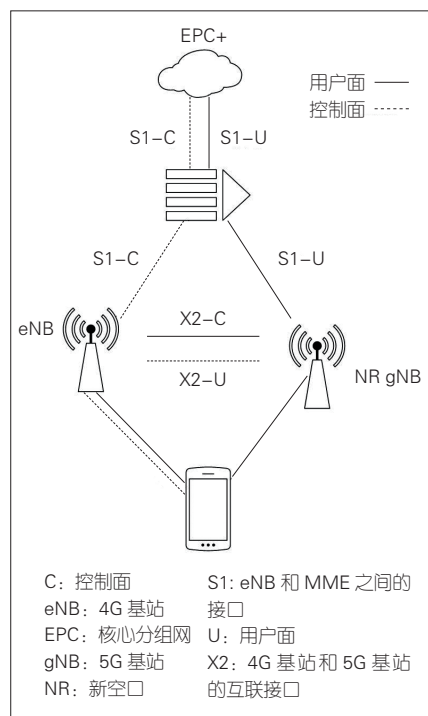
MEC 边缘云。在 5G SA 中, UPF 和 MEC 平台 (MEP) 将作为两个部分各



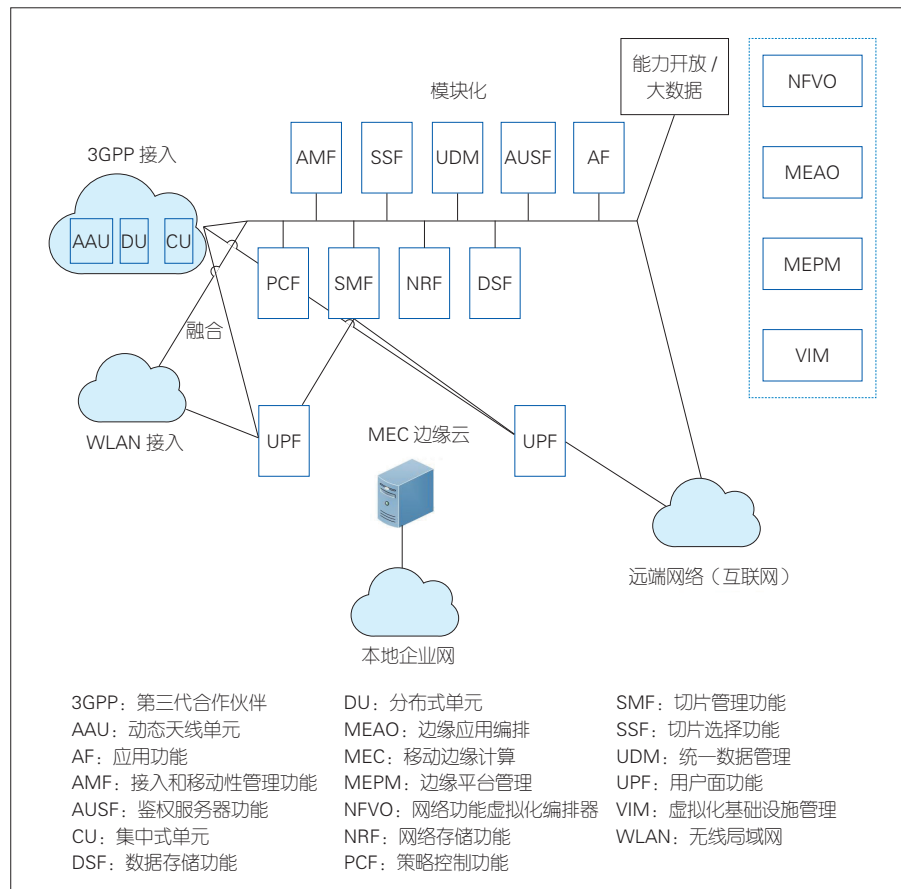
▲图1 5G 非独立组网接入网共享^[1-2]



▲图2 5G 非独立组网异网漫游^[1-2]



▲图3 5G非独立组网下的移动边缘计算组网方案^[1]



▲图4 5G独立组网下MEC组网方案^[1,3]

自部署对接。

2 对边缘云建设影响

2.1 访问边缘云方式

在5G共建共享后，某一区域内可能只存在一家运营商的基站接入。此时，边缘云的访问方式是一个亟待解决的问题。假设在共建共享后，边缘云也共享，其中MEC边缘云由运营商A建设，并且存在3类用户访问该边缘云的方式。

以下内容仅为预研使用，实际内容以建设为准。

2.1.1 在A运营商基站接入的A用户

这种情况下，用户可直接通过光纤直连链路访问MEC边缘云。这是目

前运营商建设和访问MEC边缘云最基本的方式。

2.1.2 在B运营商基站接入的A用户

此时需要考虑建立不同的互通双跨机制。5G NSA可能存在如下几种方式^[4]：

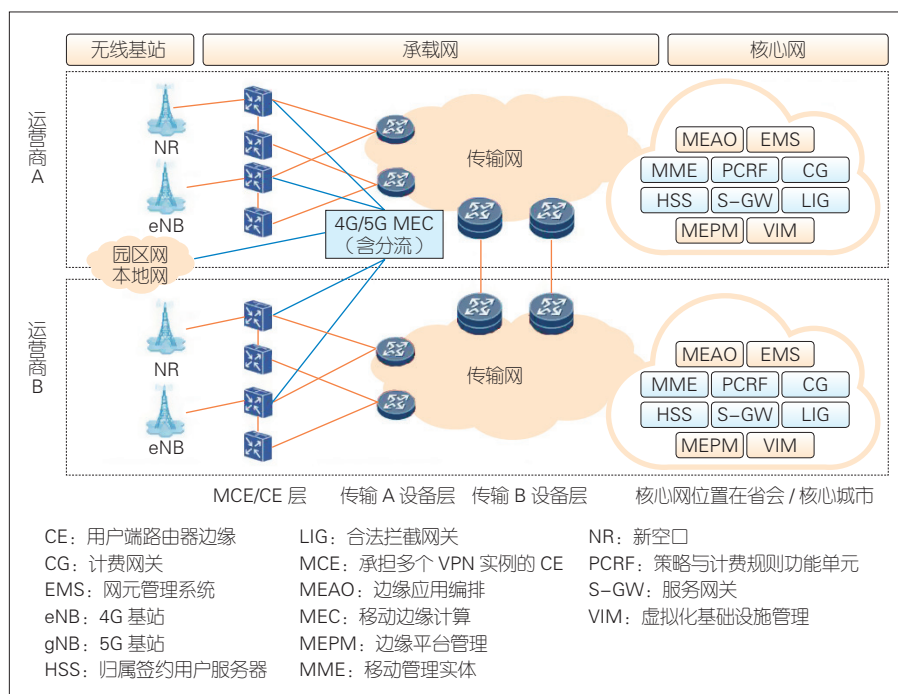
（1）MEC双跨。如图5所示，如果运营商A用户访问MEC边缘云，则需要通过运营商B的基站接入。当流量来到运营商B的MCE（承担多个VPN实例的CE）/CE（用户端路由器边缘）层时，由于运营商B的MCE/CE层同运营商A的MEC边缘云实现了双跨，因此，在运营商B的5G覆盖下的运营商A用户，可以直接访问运营商A的边缘云，也可以被分流至本地网。

（2）基站双跨。如图6所示，运营商B的基站，通过光纤直连的方式，与运营商A的CE设备相连，然后再复用运营商A的CE设备与MCE或A设备的连接链路，实现对MEC边缘云的访问和本地分流。

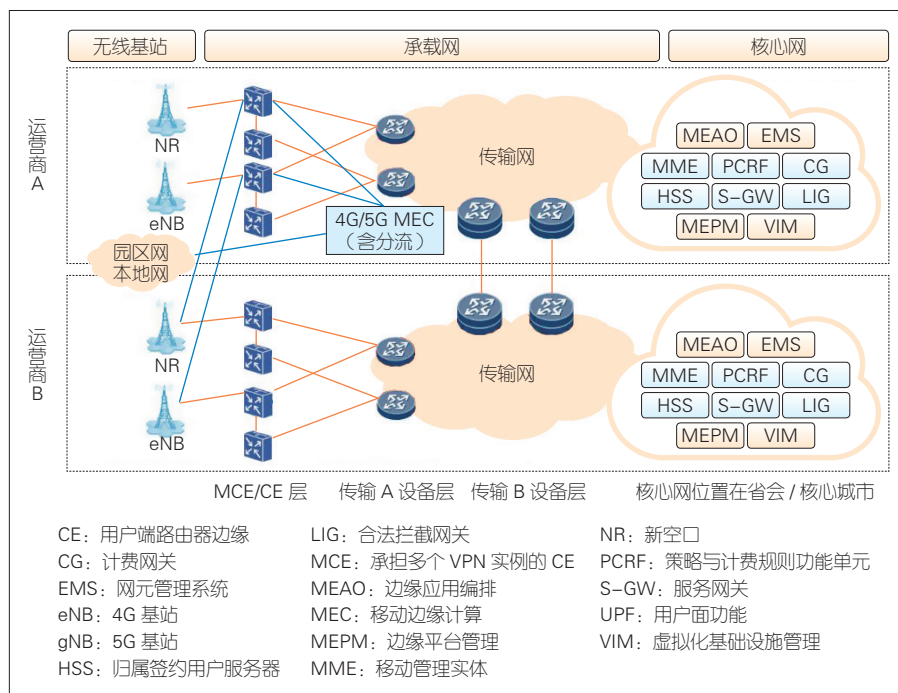
（3）传输互通。如图7所示，运营商A和运营商B的MCE/CE传输设备进行互通。所有访问运营商A的MEC边缘云的流量均通过运营商A的传输设备。这种方式需要将运营商A和运营商B的综合接入机房或汇聚机房的传输设备打通。

5G SA可能存在UPF双跨、基站双跨、传输互通等方式^[4]。其中，基站双跨、传输互通方式与5G NSA类似。这里我们将重点讨论UPF双跨。

运营商A用户如果访问MEC边缘云，则需要通过运营商B基站完成接入。当流量来到运营商B的MCE/CE层时，由于运营商B的MCE/CE层与运营商A的UPF实现了双跨，流量会先通过光纤直连的方式同UPF相连，再复用UPF到边缘云的连接，如图8所示。



▲图 5 MEC 双跨



▲图 6 基站双跨

2.1.3 在 A 运营商基站接入的 B 用户

运营商 B 的用户直接访问运营商 A 的 MEC 边缘云, 可实现本地分流和使用边缘云平台等功能。对于这种情况, 有 3 点需要明确: (1) 运营商 B 接入的用户是否能够寻址到运营商 A

的 MEC 边缘云设备; (2) 运营商 A 的边缘云能够对运营商 B 接入的用户进行分流, 并能进行无差别的 IP 五元组或域名系统 (DNS) 解析分流; (3) 运营商 A 的边缘云平台应该能够对运营商 B 用户进行鉴权和注册, 以确保

该用户为合法用户^[5-7]。

总之, 对于运营商 A 而言, 双跨的方案可以实现承建区域和共享区域内自身用户的接入。对于运营商 B 而言, 共享方区域的用户可直接访问承建方 MEC 边缘云, 以尽可能避免流量绕经的问题。但双方运营商之间需要讨论合作分成的问题。

2.2 存在的问题

(1) MEC 双跨存在交换机路由策略控制困难的问题。一方面, 哪些用户需要接入 MEC, 哪些用户不需要接入, 都要进行路由策略的控制; 另一方面, 由于 MEC 双跨涉及的互通位置在运营商承载的接入层, 场景位置太低, 虽然在技术上实现没有问题, 但是实际施工较为困难。

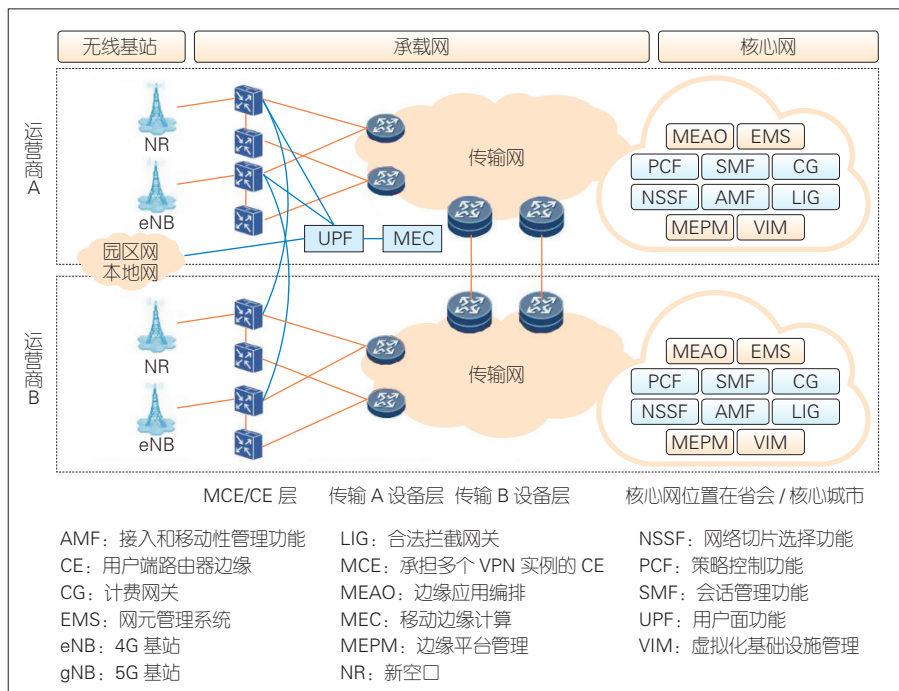
(2) 对于基站双跨来说, 基站部署的位置多样化, 数量较多, 分布也比较广, 需要打通的链路和环节也较多。

(3) 在传输互通方面, 互通层级的提高可以减少互通接口数量。虽然传输以上互通可以减少跟基站的连接, 使接口数大大减少, 但是互通传输位置的变高会导致流量绕经。

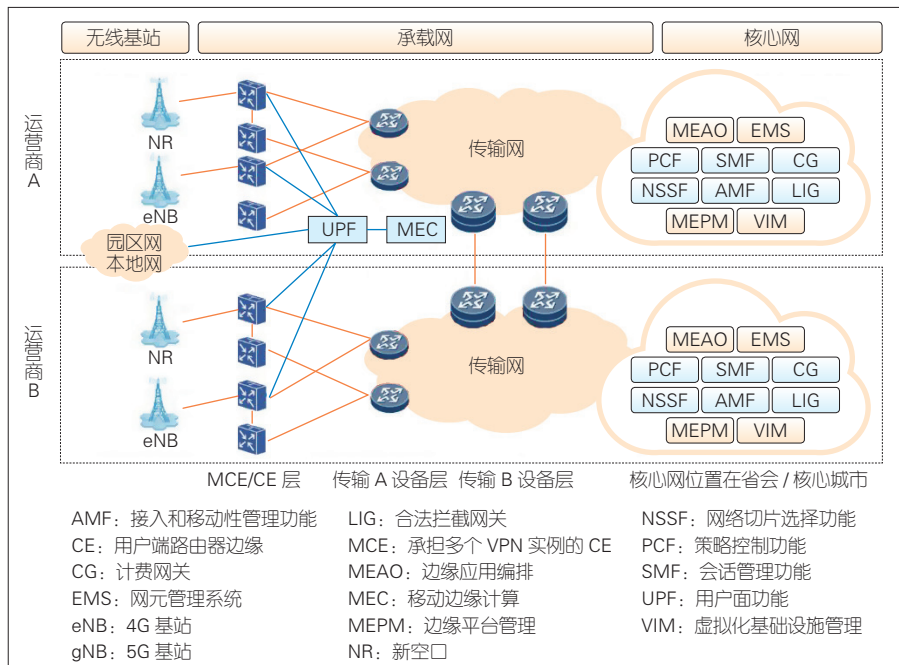
3 对网络能力的要求

3.1 用户身份鉴权

由于 MEC 边缘云会对第三方应用的身份进行认证, 因此只有经过授权的第三方应用发出的应用程序接口 (API) 请求, 才会被 MEC 平台接纳, 并被转发到 MEC 内部的服务中去。经过授权的 MEC 应用实例, 可根据用户身份激活或者去激活与之关联的配置规则。利用身份识别服务, 第三方应用可将外部应用标识映射为用户在移动网络内部的标识, 并实现面向特定用户的数据操作。因此, 在 5G MEC 边缘云共享中, 运营商的 MEC 边缘云



▲图 7 MCE/CE 层传输互通



▲图 8 UPF 双跨

平台需要对双方运营商的用户都进行身份注册和鉴权,以保证双方用户可接入共享平台^[8]。

3.2 计费方式

5G 共建共享下 MEC 的计费主要

包括两个方面:

(1) 边缘侧消耗的网络流量

5G NSA 下,由于承建方的 MEC 边缘云平台可能同时接受承建方和共享方两类用户,因此,MEC 平台需要对不同运营商的用户进行区分。对此,

可以在承建方 MEC 边缘云生成话单,即采取承建方计费方式。为解决流量区分困难的问题,可考虑采用包月方式。此外,在共享方接入承建方 MEC 边缘云平台之前,设置流量网关,将有助于对整体进入承建方的流量进行统计^[8]。

5G SA 下,流量统计和计费均由 UPF、切片选择功能(NSSF)来完成,并形成话单,无须打通计费网关(CG)。另外,UPF 还需要对不同运营商用户进行区分,并进行流量统计,以生成话单和其他计费详情。

(2) 用户向边缘云请求的云资源

MEC 边缘云资源的计费相对简单。例如,可以依据一定的物理资源分配、虚拟机或容器数量、API 调用次数、使能平台能力次数等,并按照用户级进行计费。如果双方运营商用户具有相同的等级计费,则无须修改。如果计费方式有差异,则 MEC 平台应首先识别来自运营商的用户,然后再使用计费的模板进行计费^[9-10]。

3.3 QoS 策略方式

在 5G 共建共享中,NSA 架构以承载的形式进行 QoS 保障。对于 SA 架构,由于存在共享的可能,因此,UPF 在双方运营商网络架构中需要同时进行质量保证。对于 OTT (指互联网公司越过运营商) 的切片组业务,我们建议根据本身签约信息来选择对应的 AMF,并由各自运营商来完成 QoS 的差异化保障。共建共享运营商之间应事先商定好一致的 QoS 保障策略,以获得更好的协同保障效果。

4 结束语

本文中,我们以 5G NSA 架构为例进行说明。对于部署在综合接入机房以上位置的 MEC 分流网关和 MEP 平台,我们建议各家独立部署。MEP

平台位于 MEC 分流网关之后, 不受共建共享影响。承载网互通位置决定用户访问 MEP 业务流量的走向。互通位置越高, 流量迂回就越大, 对 MEC 业务、本地化业务、低时延业务的影响也就越大。然而, 对于综合接入机房来说, 在基站侧部署的 MEC 业务, 可共用 MEC 分流网关。将共享方 MEP 平台部署至承建方基站机房内的商务模式需要做进一步讨论。此外, 在共建共享的背景下, 为解决用户访问共享边缘云方式变化带来的影响, 除了要考虑组网方式和网络能力保障外, 还需要考虑交换机、UPF 等网络设备。

致谢

本研究得到中国联合网络通信有限公司研究院王友祥博士、陈杲博士的帮助, 谨致谢意!

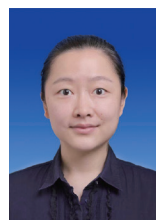
参考文献

- [1] 3GPP. System architecture for the 5G system: 3GPP TR 23.501 [S]. 2019
- [2] 3GPP. Network sharing: architecture and functional description: 3GPP TR 23.251 [S]. 2015
- [3] ETSI. Mobile edge computing (MEC): framework and reference architecture: ETSI GS MEC003 [S]. 2019
- [4] 黄倩. 5G 共享边缘云技术研究: GB/T B04-2021 [S]. 北京: 中华人民共和国工业和信息化部, 2021
- [5] ETSI. Mobile edge computing (MEC): general principles for mobile edge service APIs: ETSI GS MEC009 [S]. 2017
- [6] ETSI. Mobile edge computing (MEC): mobile edge management; part 2: application lifecycle, rules and requirements management: ETSI GS MEC010-2 [S]. 2017
- [7] ETSI. Mobile edge computing (MEC): mobile edge platform application enablement: ETSI GS MEC010-1 [S]. 2017
- [8] ETSI. Mobile edge computing (MEC): bandwidth management API: ETSI GS MEC015 [S]. 2017
- [9] 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. 信息技术 云计算 参考架构: GB/T 32399—2015 [S]. 北京: 中国标准出版社, 2017
- [10] 中华人民共和国国家质量监督检验检疫总局, 中国国家标准化管理委员会. 信息技术 云计算 概览与词汇: GB/T 32400—2015 [S]. 北京: 中国标准出版社, 2017

作者简介



黄倩, 中国联合网络通信有限公司研究院工程师; 主要从事边缘计算、开源技术、5G 标准化研究、5G 垂直行业咨询等工作。



黄莹, 中国联合网络通信有限公司研究院高级工程师; 主要从事白盒基站、边缘计算的研究。

← 上接第 22 页

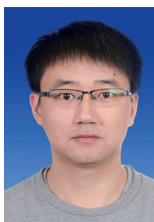
参考文献

- [1] 耿俊杰, 颜金尧. 基于可编程数据平面的网络体系架构综述 [J]. 中国传媒大学学报 (自然科学版), 2019, 26(5): 38-43
- [2] 左志斌, 常朝稳, 祝现威. 一种基于数据平面可编程的软件定义网络报文转发验证机制 [J]. 电子与信息学报, 2020, 42(5): 1110-1117
- [3] 李铭轩, 邢鑫, 王本忠. 面向电信运营商的容器云 SDN 云网一体化方案研究 [J]. 信息通信技术, 2019, 13(2): 7-12+25
- [4] 房秉毅, 张云勇, 陈清金, 等. 云计算网络虚拟化技术 [J]. 信息通信技术, 2011, 5(1): 50-53
- [5] 程莹, 张云勇. SDN 应用及北向接口技术研究 [J]. 信息通信技术, 2014, 8(1): 36-39
- [6] 林耘森, 毕军, 周禹, 等. 基于 P4 的可编程数据平面研究及其应用 [J]. 计算机学报, 2019, 42(11): 2539-2560
- [7] 衣晓玉, 秦斌, 吴文斐, 等. 基于 P4 交换机的 MAP 卸载技术设计与实现 [J]. 深圳大学学报 (理工版), 2020, 37(S1): 112-117
- [8] WANG S, WU J, YANG W, et al. Novel architectures and security solutions of program-mable software-defined networking: a comprehensive survey [J]. Frontiers of information technology & electronic engineering, 2018, 19(12): 1500-1521
- [9] 李铭轩, 曹畅, 唐雄燕, 等. 面向算力网络的边缘资源调度解决方案研究 [J]. 数据与计算发展前沿, 2020, 2(4): 80-91
- [10] 刘畅, 毋涛, 徐雷. 基于无服务器架构的边缘 AI 计算平台 [J]. 信息通信技术, 2018, 12(5): 45-49
- [11] 何涛, 曹畅, 唐雄燕, 等. 面向 6G 需求的算力

网络技术 [J]. 移动通信, 2020, 44(6): 131-135

- [12] 李铭轩, 魏进武, 张云勇. 面向电信运营商的 IT 资源微服务化方案 [J]. 信息通信技术, 2017, 11(2): 48-55
- [13] 彭新玉. 基于软件定义的可虚拟化未来网络关键技术研究及产业化 [J]. 电子世界, 2020(9): 5-6
- [14] SIVARAMAN A, MASON T, PANDA A, et al. Network architecture in the age of programmability [J]. ACM SIGCOMM computer communication review, 2020, 50(1): 38-44
- [15] VULETIC P, BOSAK B, DIMOLIANIS M, et al. Localization of network service performance degradation in multi-tenant networks [J]. Computer networks, 2020, 168: 107050

作者简介



李铭轩, 中国联合网络通信有限公司研究院高级工程师、IEEE 高级会员、中国电子学会高级会员; 主要研究方向为大数据技术、云计算技术、业务平台技术和 IT 支撑系统技术; 已发表论文 20 余篇, 授权专利 10 余篇。



曹畅, 中国联合网络通信有限公司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; 主要研究方向为 IP 网宽带通信、SDN/NFV、新一代网络编排技术等。



杨建军, 中国联合网络通信有限公司研究院未来网络研究部高级工程师, 主要研究方向为开放网络、未来网络、SDN/NFV、开放硬件、网络开源软件等; 获得授权专利 20 余篇。



边缘计算使能星地协同网络下的服务部署机制

Service Deployment Mechanism in Edge Computing Enabled Satellite Terrestrial Integrated Network

摘要:在移动边缘计算(MEC)与星地协同网络(STIN)融合的网络架构中,针对卫星网络和边缘计算对时延与资源敏感的特点,以最大化用户服务质量(QoS)为目标,提出基于强化学习的深度Q网络(DQN)算法部署机制。将部署问题描述为一个马尔可夫决策过程(MDP),并把卫星节点的状态和部署行为分别建模为DQN中的状态和动作。通过卫星的计算资源与卫星和用户的通信时延给出奖励值,在神经网络中训练以优化部署行为,进而实现最优部署策略,并对提出的算法做仿真。与其他算法对比的结果表明,在相同的优化目标条件下,DQN算法有较好的性能。

关键词:边缘计算;服务部署;强化学习

Abstract: In the network architecture of mobile edge computing (MEC) and satellite terrestrial integrated network (STIN), the satellite network and edge computing are sensitive to delay and resources. To maximize user's quality of service (QoS), a deployment mechanism based on the reinforcement learning deep Q network (DQN) algorithm is proposed. The deployment problem is described as a Markov Decision Process (MDP). The state and deployment behavior of the satellite nodes are modelled as the state and action in the DQN. The reward value is given by the satellite computing resources and the communication delay between the satellite and the user. Training in the neural network to optimize the deployment behavior achieves the optimal deployment strategy. The proposed algorithm is simulated and compared with other algorithms. The result shows that under the same optimization target conditions, the DQN algorithm has better performance.

Keywords: edge computing; service deployment; reinforcement learning

卢华/LU Hua¹,
段雪飞/DUAN Xuefei¹,
李斌/LI Bin²

(1. 广东省新一代通信与网络创新研究院, 中国 广州 510663;
2. 中兴通讯股份有限公司, 中国 深圳 518057)
(1. Guangdong Communications & Networks
Institute, Guangzhou 510663, China;
2. ZTE Corporation, Shenzhen 518057,
China)

DOI: 10.12142/ZTETJ.202103013
网络出版地址: <https://kns.cnki.net/kcms/detail/34.1228.TN.20210621.1738.002.html>
网络出版日期: 2021-06-22
收稿日期: 2021-05-12

近年来,互联网与通信技术都取得了长足进步。大数据、云计算等新兴技术已经得到广泛运用并成为当前的基础性技术^[1]。受益于5G的大规模使用,物联网(IoT)、虚拟现实(VR)/增强现实(AR)/混合现实(MR)、高分辨率(4K/8K)视频传输得到了进一步推广。然而,以车联网(IoV)、远程医疗、高帧率游戏等为代表的要求响应速度快、时延超低、占用带宽较大的应用,对现有网络体系

架构提出很大的挑战。虽然5G的应用可以缓解部分需求,但是用户与云计算中心通信产生的时延,以及海量数据传输对带宽的占用,与云计算技术本身都是矛盾的。为了解决这些问题,我们需要在数据中心之外,让计算、存储、网络延展到互联网的边缘,甚至到每个家庭的互联网网关上,使服务更加靠近用户。这种技术就是边缘计算^[2-3]。星地协同网络虽然有着很好的发展前景,但也面临着

和上述云计算类似的高数据速率、低通信时延等挑战。移动边缘计算(MEC)技术的引入可以更好地保障用户服务质量(QoS)。

关于边缘计算中服务部署问题的研究有很多。文献[4]将边缘计算系统中的服务部署建模为一个多阶段随机规划问题,设计了一个样本平均近似(SAA)方法以估计多阶段模型中资源函数的期望值,并提出贪心算法来解决基于SAA的并行算法中

每个阶段都需要解决的整数优化问题。针对把服务完全部署到本地的情况,文献[5]将问题建模为非线性整数规划问题,并采用元启发式算法求出近似解。文献[6]将服务部署问题建模为马尔可夫决策过程(MDP),并设计了一种在线算法,同时证明该算法是成本最优的。文献[7]同样将服务部署问题建模为MDP,但采用强化学习中的Dueling-DQN算法(一种改进的DQN算法)进行求解。

不同部署问题的解决方案虽然有很大不同,但基本可以归纳为传统算法和基于学习的方法。传统算法一般将问题描述为规划问题或优化问题,但通常由于问题的复杂性以及多目标约束的存在而变为非确定性多项式(NP)问题,使求解变得困难。而部署问题能够容易被建模为MDP过程,可采用强化学习中的Q-Learning或DQN等算法进行求解。

1 服务部署模型与算法设计

1.1 服务部署模型设计

这里,我们首先对研究问题做一些说明和假设:

- (1)对于每个卫星,除运行轨迹不同外,其他完全相同;
- (2)用户请求的服务相同;
- (3)用户与卫星的距离用时延来描述;
- (4)卫星的可用计算能力与中央处理器(CPU)、内存占用率成反比;
- (5)卫星的CPU和内存消耗是线性的;
- (6)服务在节点上并行计算;
- (7)卫星计算能力存在上限和下限。

为了使用强化学习算法解决服务部署问题,我们需要将其建模为MDP,具体过程如下:

我们需要先明确优化指标和具体的影响因素。本文中,我们选择最小化处理特定数量服务并交付给用户所用总时间为优化目标,如公式(1)所示:

$$\min \sum_{e \in E} \sum_{u \in U_e} (proc_e + delay_{u,e}). \quad (1)$$

公式(1)中, E 表示边缘节点集合, U_e 表示服务部署在节点 e 上的用户集合, $proc_e$ 表示在节点 e 上处理服务需要的时间(根据假设,相同节点上的 $proc$ 相同), $delay_{u,e}$ 表示用户 u 与节点 e 的通信时延。需要说明的是,这里的 $delay$ 不仅代表时延,还代表用户与卫星的物理距离。因此,我们可将时延进行适当的放大,以扩大其在问题中的影响。

MDP是一个四元组 $\langle S, A, P, R \rangle$,分别代表状态、动作、状态转移概率和奖励。本问题中的状态转移概率均为1。下面我们将讨论 S 、 A 与 R 。

边缘节点共有9个,即 $E_{size} = 9$ 。此外,我们还需要确定希望部署的服务数量 n (假设一个用户请求一个服务,即 $U_{size} = n$)。这样我们就可以将状态集定义为 $\dots S = \{s_0, s_1, \dots, s_n\}$ 。 s_i 表示当前已经部署 i 个服务,它仍是一个集合,所包含的状态数可以用简单的排列组合计算得到。其中 s_0 为初始状态, s_n 为终止状态。我们可以将这样的状态集称为简化状态集。相应地,我们可以定义具体状态集,以描述每个服务的具体部署位置(即在哪个节点上)。这个集合共有 $\sum_{i=0}^n 9^i$ 个状态,其中 9^i 是简化状态集中 s_i 扩展到具体状态集的个数。基于前述假设,我们在算法设计中使用简化状态集。可以看出,本问题与传统强化学习问题有所不同:传统问题的状态

转移步数是不确定的,而本问题的状态转移步数是确定的,并且当经过 n 步之后就一定会达到终止状态。

在本问题中,MDP中的动作是把服务部署在某个边缘节点上。我们可以规定服务的部署顺序。对于某个状态集 s_i 而言,要部署的服务就是确定的。此时,动作数量与边缘节点数量一致。本问题的MDP在状态集 s_i 中执行一个动作 a ,随后进入状态集 s_{i+1} 。

奖励是决定算法最终效果的核心。在使用简化状态集时,我们显然不能为状态集 s_i 中的所有状态设置同一个奖励值。单纯地为简化状态集中的每一个状态而定义一个奖励值也是不合理的。因此,在设置奖励值时,我们要按具体状态集来处理。

具体奖励值的设置要参考公式(1)的优化目标。我们可以把每一个状态都当作终止状态。此时 E 代表部署中有服务的节点。我们就可以利用公式(1)来计算当前状态消耗的时间,并计算处理相同服务所花费的基本时间。由于存在时延和节点计算能力下降等因素,实际时间会比基本时间长,因此我们可以通过时间差来确定奖励值。实际时间越长,奖励值就越低。相应的计算公式如公式(2)所示:

$$R = \frac{t_{basic}}{\sum_{e \in E} \sum_{u \in U_e} (proc_e + delay_{u,e})}. \quad (2)$$

1.2 基于服务部署模型的算法设计

当利用强化学习来求解MDP模型时,我们可以采用Q-Learning或DQN算法。在本问题中,即使我们采用简化状态集,随着服务数量的增加,其规模也呈指数级增长,此时不宜采用Q-Learning算法进行求解。因此,本文中我们采用DQN算法。

算法的模型如图1所示。操作环境输入选择的动作,并执行该动作,随后进入下一状态,同时反馈这一步的奖励值和是否到达终止态等信息。这些信息会形成一条记录被存入经验回放区。当经验回放区存储一定数量的记录后,神经网络会从中随机选取一些记录来进行训练,并更新相应的网络参数,选择基于当前网络参数选出的价值最大的动作来让环境执行。新的记录生成后会被继续存入经验回放区。当经验回放区的数据足够多时,新记录将逐渐代替旧记录,以便于那些之前使用价值不大的记录不会再被学习。本文中,我们使用的神经网络有两个隐藏层。神经网络通过反向传播当前Q网络与目标Q网络的差值来优化参数。

奖励值的计算方法可参照公式(2)。假设节点在最佳性能时处理一个服务消耗的时间为 t_0 ,则基本时间 t_{basic} 是所有已部署服务 t_0 的简单求和,如公式(3)所示:

$$t_{\text{basic}} = i \times t_0 \quad (3)$$

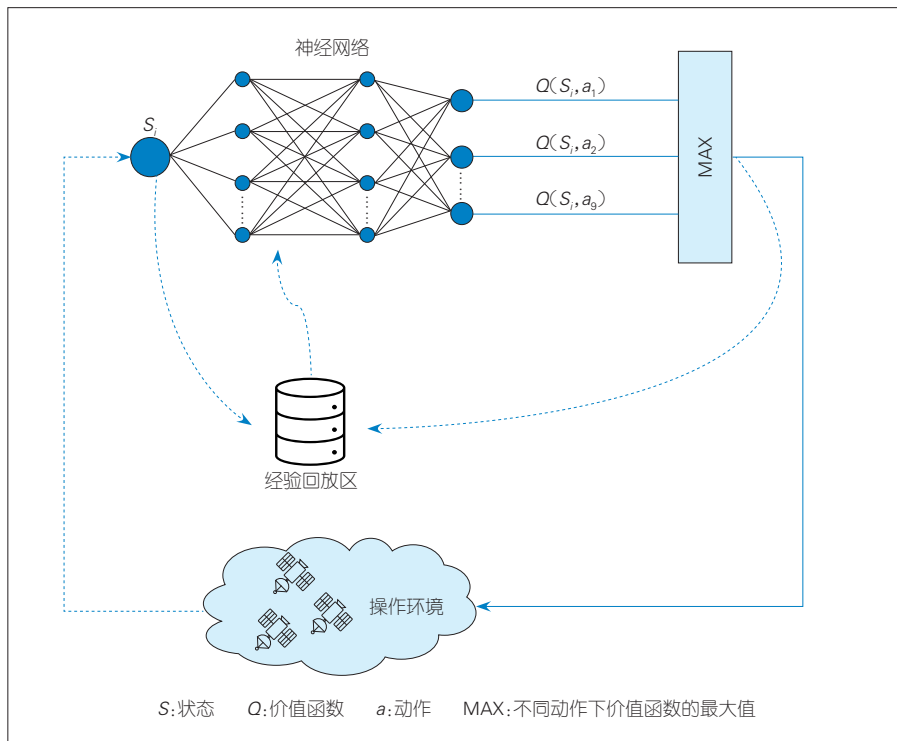
公式(3)中, i 表示已部署的服务数。对于实际时间,节点处理服务的时间 $proc$ 与1.1节作出的假设有关。假设在一个节点上部署了3个服务,且CPU空闲率为80%,则每个服务的处理时间均为 $\frac{kt_0}{80\%}$,其中 k 为比例系数。在上述假设的前提下,我们可以直接令 $k=1$ 。因此,公式(2)可以写为公式(4):

$$R = \frac{i \times t_0}{\sum_{e \in E} \sum_{u \in U_e} \left(\frac{t_0}{cpu_free} + delay_{u,e} \right)} \quad (4)$$

神经网络中更新Q值的方式如公式(5)所示:

$$y_j = \begin{cases} R_j & \text{到达终止状态} \\ R_j + \gamma \max_a Q'(\phi(S'_j), A'_j, w) & \text{未到终止状态} \end{cases} \quad (5)$$

公式(5)中, R_j 代表当前奖励值。



▲图1 服务部署的深度Q网络算法模型

γ 为衰减因子($0 \leq \gamma \leq 1$),表示后续奖励值对当前Q值的影响。 Q' 是目标Q网络, $\phi(S'_j)$ 表示下一状态的特征向量, A'_j 表示下一步动作, w 为Q网络中的状态价值函数。

2 实验仿真与结果分析

2.1 实验环境及参数

实验中,我们假定边缘节点数量为9个,用户(服务)数量 n 为20~50个,服务的最短执行时间 t_0 为60 s。为了简化问题,我们假设每个服务都会消耗节点10%的CPU。同时,节点CPU空闲率的下限为10%,即一个节点最多可以同时为9个用户提供服务。如果部署服务多于9个就需要排队等候。显然,在一个节点部署过多服务,不仅会导致每个服务的计算时间变长,还会使需要等待的节点产生更多不必要的等待时延。在上文假设的服务数量下,这显然不是最优策略。强化学习过程中的随机选择动作会导致这些策略被执行和学习,因此,我们要在算法中避免这种情况的发生,即如果采取某个动作后会进入需要排队的状态,就令这一动作无效且下一状态仍为原状态,同时给这次动作一个很低的奖励值,以避免再次作出同样的选择。

用户与卫星的时延是一个难以准确评估的参数。本文1.1节已经指出,时延可代表用户与卫星的物理距离。为了在仿真中模拟现实情况,我们需要对其进行适当放大。经过调试,我们认为,时延分布在1~20 s之间是比较合理的。

此外,本文同时设计了随机部署算法、最短时延贪心算法、均匀部署算法3个参考算法^[8]。我们分析了在不同服务数量条件下4个算法的性能。为了控制无关变量,这3个参考

算法中每一个节点部署服务的数量均不会超过9个,且满足如下条件:

(1)对于随机部署算法,每次部署随机选择节点;

(2)对于最短时延贪心算法,每次部署选择时延最小的节点;

(3)对于均匀部署算法,将服务平均部署到节点中。

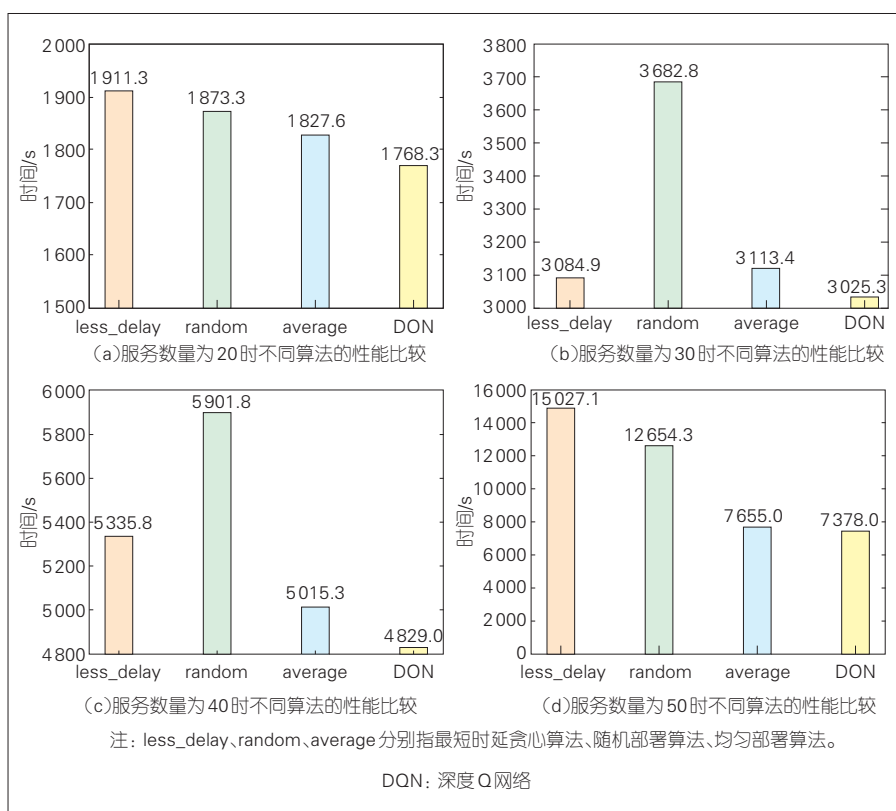
2.2 结果分析

我们选择服务数量 n 分别为20、30、40和50,并进行测试比较。得到的柱状图结果如图2所示。其中,纵坐标表示每种算法处理时延与传输时延之和。为了直观地显示不同情况的算法结果,我们对纵坐标的范围进行适当调整。图3是将柱状图绘制成折线图的结果。

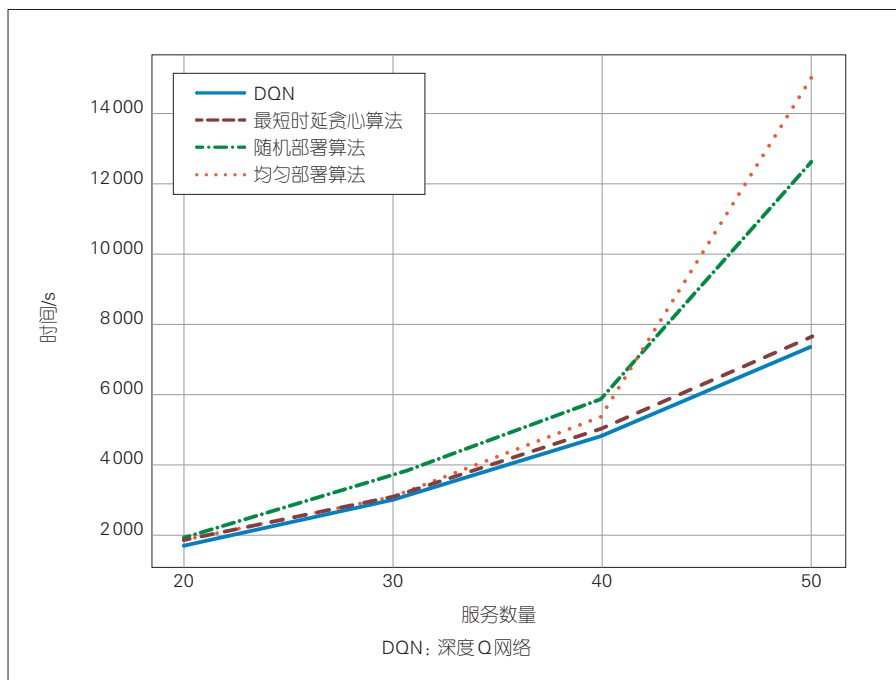
由图2和图3可知,在不同服务数量的情况下,DQN算法的性能均优于另外3种算法。由于对问题作出的一系列假设使最优部署方案接近于均匀部署,因此仿真中的平均部署算法性能与DQN较为接近。在实际问题中,服务对CPU的影响没有那么剧烈,平均部署算法与DQN的真实差距要大于仿真中的差距。此外,在算法设计中,时延对结果的影响小于节点计算能力对结果的影响。因此,基于时延的贪婪算法的性能并不出色,甚至在某些情况下要比随机算法性能更低。

3 结束语

本文中,我们围绕边缘计算使能星地协同网络中的服务部署问题展开研究,将服务部署问题建模为MDP过程,用DQN算法对模型进行求解,并提出详细的算法步骤。我们通过设定基本参数,对算法进行仿真,并将DQN算法与随机部署算法、时延优先贪婪算法、平均部署算法这3个参



▲图2 不同服务数量下的的算法性能比较



▲图3 算法性能比较折线图

考算法进行性能比较,发现DQN算法是解决边缘计算服务部署问题的一种有效算法。

参考文献

- [1] ZHAO J J, XU C Z, MENG T H. Big data-driven residents' travel mode choice: a research overview [J]. ZTE communications, 2019, 17 (3): 9-14. DOI: 10.12142/ZTECOM.201903003

- [2] 丁春涛, 曹建农, 杨磊, 等. 边缘计算综述: 应用、现状及挑战 [J]. 中兴通讯技术, 2019, 25(3): 2–7. DOI: 10.12142/ZTETJ.201903001
- [3] 秦永彬, 韩蒙, 杨清亮. 边缘计算中数据驱动的智能应用: 前景与挑战 [J]. 中兴通讯技术, 2019, 25(3): 68–76. DOI: 10.12142/ZTETJ.201903010
- [4] BADRI H, BAHREINI T, GROSU D, et al. A sample average approximation-based parallel algorithm for application placement in edge computing systems [C]//2018 IEEE International Conference on Cloud Engineering (IC2E). Orlando, USA: IEEE, 2018:198–203
- [5] CHENG Z X, LI P, WANG J B, et al. Just-in-time code offloading for wearable computing [J]. IEEE transactions on emerging topics in computing, 2015, 3(1): 74–83
- [6] WANG S Q, URGANKAR R, ZAFER M, et al. Dynamic service migration in mobile edge computing based on Markov decision process [EB/OL]. [2021–04–06]. <https://arxiv.org/abs/1506.05261>
- [7] ZHAI Y L, BAO T H, ZHU L H, et al. Toward reinforcement-learning-based service deployment of 5G mobile edge computing with request-aware scheduling [J]. IEEE wireless communications, 2020, 27(1): 84–91. DOI: 10.1109/MWC.001.1900298
- [8] 严蔚敏, 吴伟民. 数据结构: C语言版 [M]. 北京: 清华大学出版社, 1997

作者简介



卢华, 广东省新一代通信与网络创新研究院网络技术创新中心主任; 研究方向包括 5G 核心网、边缘计算、新型网络架构、软件定义网络、P4 可编程、虚拟化等。



段雪飞, 广东省新一代通信与网络创新研究院网络技术创新中心核心网部门负责人; 研究方向包括 5G 核心网架构、6G 网络架构、空天一体化通信系统等。



李斌, 中兴通讯股份有限公司系统架构师; 主要从事 IP 网络相关技术的研究; 曾获国家科学技术进步奖二等奖。

《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 158 期
2021 年 6 月 第 27 卷 第 3 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
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
外联: 卢丹
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

传真: (0551)65850139
发行方式: 自办发行
印刷: 合肥添彩包装有限公司
出版日期: 2021 年 6 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