



专栏：算力网络

通算一体驱动的算力内生网络技术与应用

孙杰, 马雷明, 杨爱东, 叶晓舟, 吕鹏, 王志刚, 欧阳晔
(亚信科技(中国)有限公司, 北京 100193)

摘要: 全球通信业均开展 6G 潜在技术架构与关键技术研究, 其中通过通算一体更好地赋能各项 6G 网络新技术和新业务, 从而进一步实现 6G 智能普惠愿景已成为业界共识。因此, 如何实现通信与计算的深度融合与实时协同, 以满足动态复杂网络环境下的新业务超低时延与算力实时响应等需求, 成为通算一体的挑战与关键。提出一种通算一体驱动的算力内生网络, 其通过算力时分复用与智能通算调度技术实现了基站的通信与计算算力解耦、算力池化与算力智能调度, 从而实现基站同时支持通信与计算业务。该技术已经过验证并试点应用, 实验与试点结果表明, 算力内生网络技术可有效提升无线网络的算力效率, 并实现基站同时支持通信与计算业务, 助力 6G 通算一体演进。

关键词: 通算一体; 算力内生网络; 感知单元; 智能通算调度

中图分类号: TP393

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2023153

Computing native network (CNN) technology and application driven by joint communication & computing (JC&C)

SUN Jie, MA Leiming, YANG Aidong, YE Xiaozhou, LYU Peng, WANG Zhigang, OUYANG Ye
AsiaInfo Technologies (China) Co., Ltd., Beijing 100193, China

Abstract: The global telecom industry is carrying out research and exploration of potential 6G technical architectures and key techs, of which it has become a consensus in the industry to further realize the vision of 6G smart inclusion by better enabling all new 6G network technologies and services through joint communication & computing (JC&C). Therefore, how to realize the deep integration and real-time collaboration of communication and computing to meet the needs of ultra-low latency and real-time response of new services in dynamic and complex network environments has become the challenge and key of JC&C. A computing native network (CNN) driven by JC&C was proposed, which realized the decoupling of communication and computing of base station or multiple base stations, computing pooling and intelligent JC&C scheduling through computing time division multiplexing and intelligent JC&C scheduling, thus realizing the base station to support communication and computation services simultaneously. The technology has been technically verified and piloted, and the experimental and pilot results demonstrate that CNN can effectively improve the computing utilization rate of wireless networks and realize the base station to support communication and computing services simultaneously, facilitating the evolution of 6G JC&C.

Key words: joint communication & computing, computing native network, awareness unit, intelligent JC&C scheduling



0 引言

随着行业数字化转型的不断深入,移动通信网络深刻影响着人类的生产、生活。到2030年,社会服务均衡化、社会治理科学化、社会发展绿色化,经济高质量发展、环境可持续发展等要求将驱动移动通信网络由通信基础设施向数字基础设施转变,推动5G万物互联到6G万物智联的发展。为此,全球通信行业均开展6G潜在技术架构与关键技术的研究与探索,目前产业界已发表了近百篇6G白皮书,对6G的愿景、场景、潜在技术架构与关键技术等进行了研究与探索。其中,通过设计通信与计算深度融合的6G网络架构,以更好地赋能各项6G网络新技术(如智慧内生、泛在感知)和新业务(如沉浸式扩展现实(extended reality, XR)、数字孪生、元宇宙等),从而进一步实现6G智能普惠愿景已成为业界共识^[1]。

通算一体网络的显著特征是网络内生支持各类算力和通信的相互感知与深度融合(其中的算力来源可能是终端算力、网元内生算力、移动边缘/云算力等),支持实时准确的算力发现、灵活动态的算力调度、自主适配通信动态环境的算力调整等新能力,可以提供无处不在、满足差异化服务质量(quality of service, QoS)需求的计算服务,并支撑6G网络AI即服务(AI as a service, AIaaS)的新服务能力^[2]。这其中一个重要的技术挑战是通算融合问题,即通信与计算实现更深入、更实时的协同,以在动态复杂的网络环境下,保障新业务端到端的超低时延、算力实时响应等要求。

1 通算一体技术研究现状

2014年,多接入边缘计算(multi-access edge computing, MEC)^[3]作为通算一体技术的雏形,由欧洲电信标准组织(European Telecommunications Standards Institute, ETSI)定义并提出,其作为云计算的演进,旨在将计算从集中式数据中心下

沉到通信网络接入网边缘,通过计算与网络在物理位置的融合,以分布式的方式在更靠近用户的网络边缘提供更低时延的计算服务;文献[4]通过将多个MEC节点按用户、辅助和接入点(access point, AP)节点类型进行组合,并设计了一种四时隙协议实现基于网络时延的通算融合技术,以提升MEC的计算能力;文献[5]通过将无线回传网络资源与MEC计算资源进行融合分析计算,以减少所承载业务的端到端时延。然而,MEC只是在物理位置上部署在通信网络内,在逻辑上计算与通信拥有彼此独立的编排与调度体系,动态性存在一定的缺陷,难以实现网络和计算在控制面的统一,无法及时响应实时及移动业务的需求^[6]。2019年,中国电信在国际标准组织中倡导算力网络的概念,旨在分阶段实现计算与网络的协同、融合与一体内生;然而由于运营商的网络与云计算规模较大且复杂,目前国内运营商的算力网络仍处于协同到融合的阶段,且在算力网络的标准定义方面存在一些差异,如国际电信联盟(International Telecommunications Union, ITU)定义的CPN(computing power network)^[7]、CNC(coordination of computing and network)^[8]与因特网工程任务组(Internet Engineering Task Force, IETF)定义的CFN(computing first networking)^[9]等,距离算力网络实现一体内生仍需较长的时间。2021年,宽带论坛(broadband forum, BBF)提出基于算力路由的城域算网(metro compute networking)^[10],旨在通过路由携带算力信息实现网络对算力的感知,从而实现基于计算任务的调度,这仍属于算力与网络的协同,其受限于网络时延、带宽等因素,无法实现动态、实时的通算一体调度。2022年,ETSI定义了网络功能虚拟化(network functions virtualization, NFV)中基于Dyngcast及其监控参数(如虚拟机(virtual machine, VM)或容器的CPU/内存利用率等)的虚拟化网络功能(virtualized network function, VNF)实例化与流量转发控制^[11],但网络算力仍仅用于通信业务。

文献[12]提出了一种基于开放无线电接入网络联盟 (Open Radio Access Network Alliance, O-RAN Alliance) [13]的通算一体架构,其通过对基站异构算力的统一管理,并构建虚拟机/容器管理编排平台及其与 O-RAN 服务管理编排 (service management orchestration, SMO) 的标准接口,实现将基站的空闲算力按 VM 或容器的粒度调度给非通信业务使用,但此技术不适用于非 O-RAN 架构或非云化无线网络。

为此,本文提出一种面向运营商无线网络的通用通算一体技术,其通过将大量无线基带单元 (base band unit, BBU) 的空闲算力与通信业务解耦,为计算应用提供算力支持,实现灵活、实时、低成本的算力供给。

2 算力内生网络及其关键技术

算力内生网络是基于 5G、云原生、AI 等信息与通信技术 (information and communication technology, ICT) 形成的面向全行业提供连接+算力+应用服务的一种通算一体技术体系。算力内生网络如图 1 所示,算力内生网络主要包括算力时分复用与智能通算调度两大关键技术,算力时分复用技术通过将无线网络中 BBU 的空闲算力

与通信业务所需算力解耦为计算能力与通信能力两类算力,从而为通信业务之外的应用提供算力支持。智能通算调度技术基于时间序列预测等人工智能算法,通过分析 BBU 算力消耗的“潮汐效应”,实现 BBU 资源消耗预测感知,将符合条件的 BBU 计算能力进行注册管理,并形成算力资源池,从而根据非通信业务需求实现 BBU 及 BBU 集群的最优调度组合推荐。

2.1 算力时分复用

BBU 算力时分复用如图 2 所示,其在已有第三代合作伙伴计划 (3GPP) 标准中定义的集中式单元 (centralized unit, CU) 和分布式单元 (distributed unit, DU) [14]模块架构基础上,新增了感知单元 (awareness unit, AU)。

AU 基于云计算虚拟化 (hypervisor) [15]技术,将 BBU 的算力以“CPU 核”为基本单位,采用 x86 虚拟化技术,如 Intel 的 VT-x 或 AMD 的 AMD-V (AMD virtualization) [16],将 CPU 核虚拟化为虚拟 CPU (virtual central processing unit, vCPU),每个 vCPU 能提供通信能力或计算能力;同时 AU 通过实时监控 BBU 的 CPU 算力使用情况,当监控到 BBU 的网络负载被判定闲时,可将空闲算力向智能通算调度器发起注册并被纳管,为应用提供计算

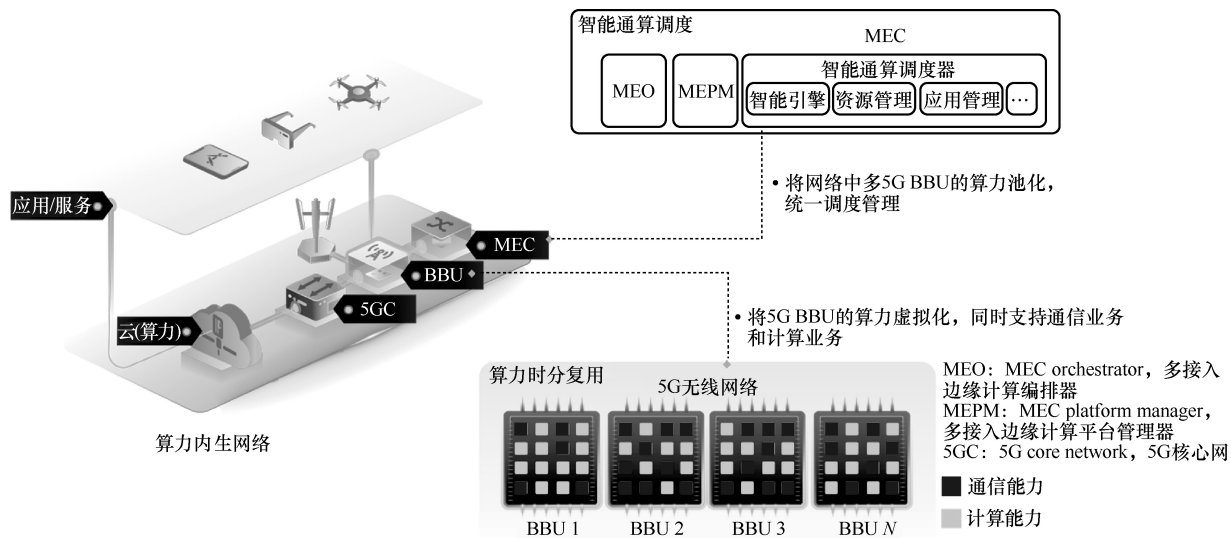


图1 算力内生网络

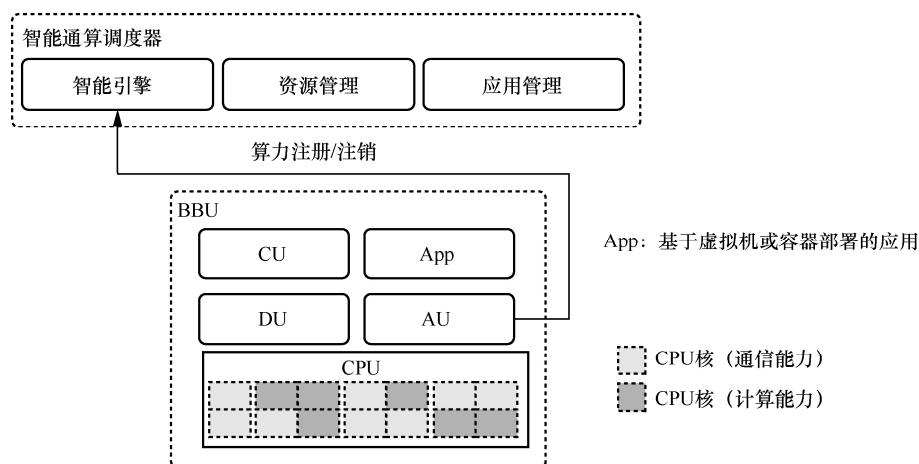


图2 BBU 算力时分复用

能力, 从而实现 BBU 通信与计算能力的智能时分复用。AU 对 BBU 算力的虚拟化技术可保证通信服务无损伤, 首先是网络架构无损伤, 该架构基于业界通用的网络部署架构, 不需要额外网络基础设施投资; 其次是网络协议无损伤, 保持 3GPP 网络控制面协议不变和用户面协议不变, 将网元空闲通信能力转化为计算能力; 最后是网络 QoS 无损伤, 从而实现通信业务 QoS 优先保障, 内生算力严格边界隔离, 安全边界智能化实时调整。

BBU 空闲算力的注册与更新/注销流程如图 3 所示, 其中注册流程主要包括以下。

步骤 1 BBU-AU 进行算力虚拟化, 并监控空闲算力。

步骤 2 BBU-AU 发现空闲算力后, 向智能通算调度器的智能引擎发起算力注册申请, 主要发送 BBU 当前空闲算力、流量与用户信息等数据。

步骤 3 智能通算调度器的智能引擎收到 BBU 算力注册申请后, 基于通算智能调度算法进行 BBU 算力态势感知与预测, 判定此 BBU 的空闲算力是否具备持续性等。如果智能通算调度器的智能引擎判定此 BBU 的空闲算力符合条件, 则通知智能通算调度器的资源管理模块完成 BBU 算力节点纳管与算力注册; 如果不符合条件, 则返回步骤 1。

算力更新/注销流程主要包括以下。

步骤 1 BBU-AU 进行算力虚拟化, 并监控空闲算力。

步骤 2 BBU-AU 定期发送算力数据。

步骤 3 智能通算调度器的智能引擎进行 BBU 算力态势感知与判定; 如果通过, 则通知智能通算调度器的资源管理模块完成 BBU 算力节点信息更新; 如未通过, 则执行下一步。

步骤 4 智能通算调度器的应用管理模块判断此节点是否部署应用, 如是, 则先进行应用迁移然后再进行 BBU 算力节点注销; 如否, 则由智能通算调度器的资源管理模块与智能引擎完成 BBU 算力节点注销。

2.2 智能通算调度

智能通算调度技术主要在 BBU 算力虚拟化的前提下, 基于 AI 技术构建通算智能调度决策算法体系, 实现 BBU 及多个 BBU 间的空闲算力池化与智能调度, 从而将 BBU 池化算力用于 AI 服务或其他非通信业务。

智能通算调度主要通过资源态势感知、节点资源动态分割、节点评估与推荐、任务编排与调度 4 个步骤完成 BBU 算力池的智能调度, 通算智能引擎算法体系如图 4 所示。

(1) 资源态势感知

在算力内生网络中, BBU 资源态势拥有不确

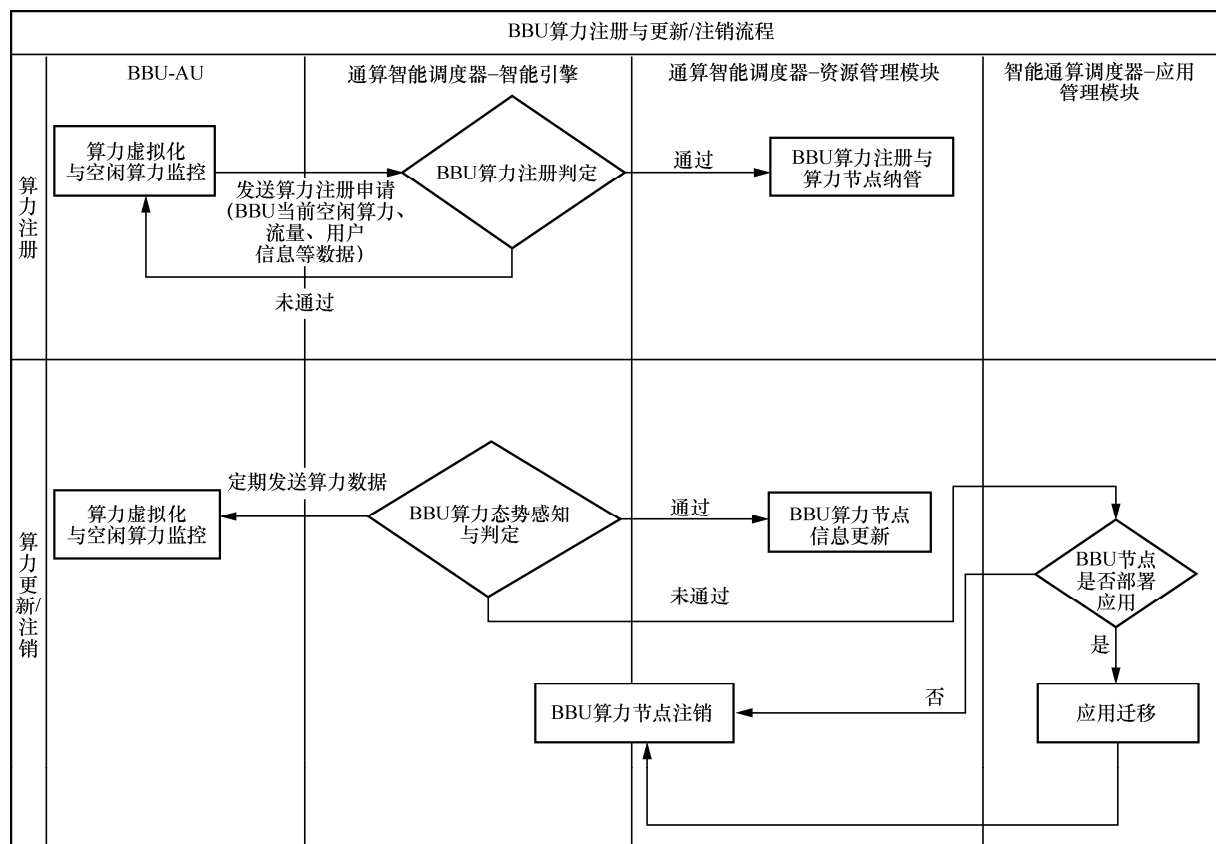


图3 BBU空闲算力的注册与更新/注销流程

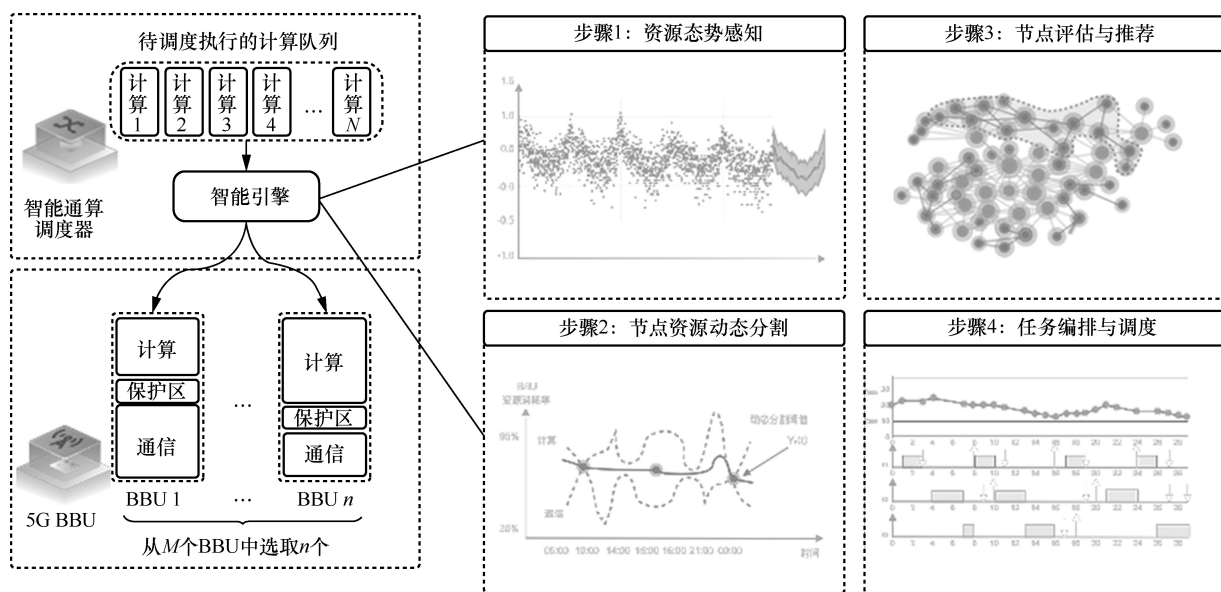


图4 通算智能引擎算法体系

定性以及复杂动态变化的问题,并且资源态势随着系统运行时间的推移不确定性逐渐增强。为了解决以上问题,提供稳定的算力服务,需要将资

源态势感知应用于算力内生网络。资源态势感知即资源消耗状态的时间预测,基于历史资源消耗的指标数据预测未来的变化趋势。常见趋势包含



长期趋势、季节变动、循环波动和不规则波动。如图4所示,收集各BBU的通信算力资源使用量,通过状态估计方法,包括最大似然和最大先验、分布式卡尔曼滤波、分布式粒子滤波以及协方差一致等,得到通信算力资源状态。然后通过时序预测算法,包括算数平均、指数平滑、差分自回归移动平均 (autoregressive integrated moving average, ARIMA) 等,对 BBU 在当前时刻的通信算力资源状态和历史通信算力资源状态进行模型计算,进而分析 BBU 资源消耗的“潮汐效应”,预测 BBU 资源消耗的未来变化趋势,以实现 BBU 集群资源的态势感知。

(2) 节点资源动态分割

在业务高峰时期, BBU 会面临算力瓶颈、存储限制等问题;业务高峰结束后, BBU 又会出现大量闲置的算力资源。节点资源动态分割可解决这个问题。动态分区作为动态资源分配策略中的关键能力,它基于资源的使用量动态调节资源分割阈值,常见算法是弹性伸缩算法,一般可分为基于某个资源指标阈值的响应式伸缩算法和基于预测型的伸缩算法。基于上述算法,实时跟踪 BBU 集群及本节点的通信与计算资源的消耗状况,为当前节点分配最优分割阈值,以实现 BBU 节点资源的最优划分。当遇到业务高峰、算力瓶颈、存储限制时,可以通过 BBU 资源动态分割快速解决;业务高峰过去后,多余的 BBU 算力资源可以快速释放。

(3) 节点评估与推荐

在算力内生网络中,不同业务的资源需求不同,如果 BBU 节点供给和业务需求不匹配,将直接导致大量业务运行效率变差,且精度也会受影响,因此需要对 BBU 节点进行评估和推荐。通过 BBU 节点进行图聚类,即对具有图结构的节点集做聚类,并研究属性图节点聚类模型(同时考虑结构特征和节点特征),可实现算力内生中的资源节点子集群推荐功能。基于图神经网络算法,分析网络拓扑结构、供给/需求侧的资源特征等数

据,计算各 BBU 之间的相似度,得到相似度矩阵。根据相似度动态匹配有效 BBU 节点集合,并实现 BBU 集群资源的最优推荐。此外,节点聚类为集群,还可以让多个 BBU 共同完成任务,可以将任务分割为若干个分片,由不同的 BBU 并行执行,以提高任务的处理效率。

(4) 任务编排与调度

在复杂动态变化的网络中,需要同时对多个不同任务进行快速高效的资源编排,且需要适配业务需求和当前的网络状态与算力状态。基于资源态势感知、节点评估与推荐、节点资源动态分割等模型提供的信息,构建综合(多因子)代价函数,从成本、性能、效率、和安全角度,对资源进行全局最优计算。此外,根据任务的资源需求,结合编排调度策略算法,包括先进先出(first in first out, FIFO)、蚁群、离线规划、强化学习等算法,对任务进行编排与调度。最终根据业务运营策略目标,如成本优先、性能优先、效率优先等优化策略,为当前任务规划最优资源编排方案,以实现 5G BBU 集群资源的最优编排与调度。

3 算力内生网络技术验证与应用

3.1 技术验证

(1) 验证环境

为了验证算力内生网络关键技术,搭建了试验环境,如图5所示。

该环境既支持 5G 网络功能,也支持验证算力内生网络通算一体的调度能力,验证环境信息见表1。

(2) 验证结果

BBU 负载情况如图6所示,描述了 BBU1 和 BBU2 的通信能力和计算能力的利用率变化曲线。通信能力代表通信业务负载情况,计算能力代表 XR 应用服务负载情况。

随着 BBU1 通信业务降低, BBU1 成功将空闲算力用于 XR 应用的部署(如图6(a)所示)。当 BBU1 的预测通信负载上升时,通知 MEC 上部

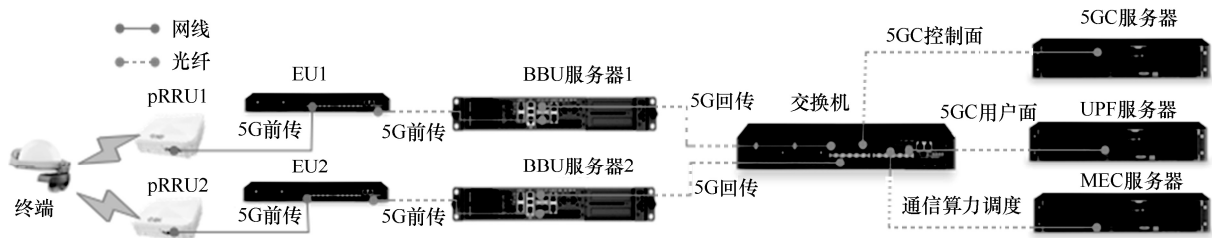


图5 试验环境

表1 验证环境信息

设备	数量/个	说明
头盔终端	1	5G 终端、XR 应用客户端
皮基站设备（pico remote radio unit, pRRU）	2	5G 网络射频单元和信号收/发天线
扩展单元（extend unit, EU）	2	承担 5G 前传接口数据汇集与分离功能
BBU 服务器	2	部署无线电接入网（radio access network, RAN）中的 AU、DU、CU 功能，同时提供内生空闲算力用于部署 XR 应用服务端
交换机	1	用于设备之间的 IP/以太网（Ethernet）数据传输与交换
5GC	1	承担 5G 核心网控制面功能，接入与移动性管理功能（access and mobility management function, AMF）、会话管理功能（session management function, SMF）、统一数据管理（unified data management, UDM）等网元
用户面功能（user plane function, UPF）服务器	1	承担 5GC 中的用户面数据功能
MEC 服务器	1	部署智能通算调度器，提供基站动态算力管理与调度、应用部署与迁移

署的智能通算调度器将空闲算力回收，将 XR 应用服务成功迁移到空闲的 BBU2 服务器中（如图 6（b）所示）。

3.2 技术应用

目前算力内生网络技术已成功在某核电专网中试点应用。在核电厂中，有些特定业务场景具有一定危险性，导致工作人员无法进入

该环境进行操作，因此需要通过远距离操控无人车进入恶劣区域执行任务。通过算力内生网络技术，可实现 5G 无线网络覆盖核电危险红区、黄区以及安全工作绿区，无人车通过 5G 网络接入。算力内生网络的 XR 远程驾驶如图 7 所示，无人车 XR 应用部署在 5G 基站内生的算力上，用户在安全区驾驶舱通过对 XR 应用进行访问，

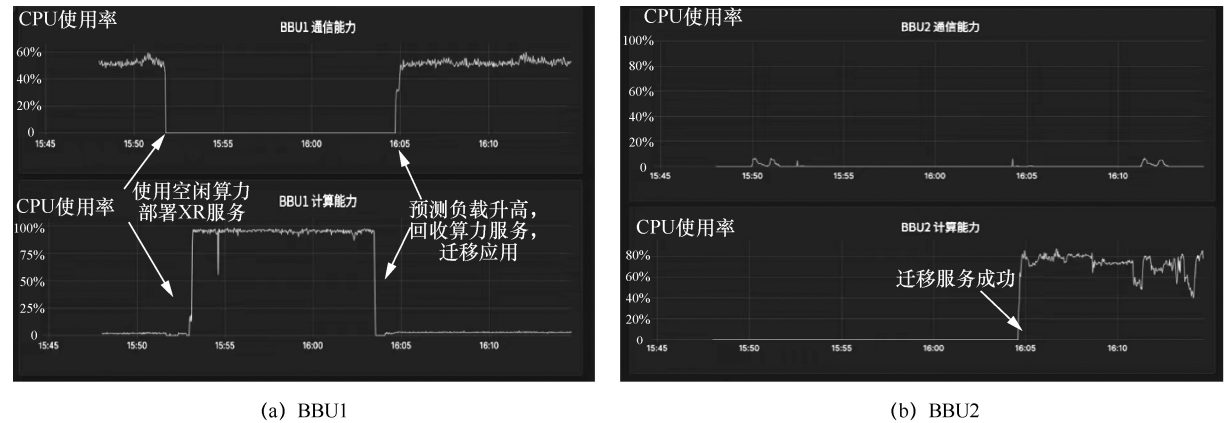
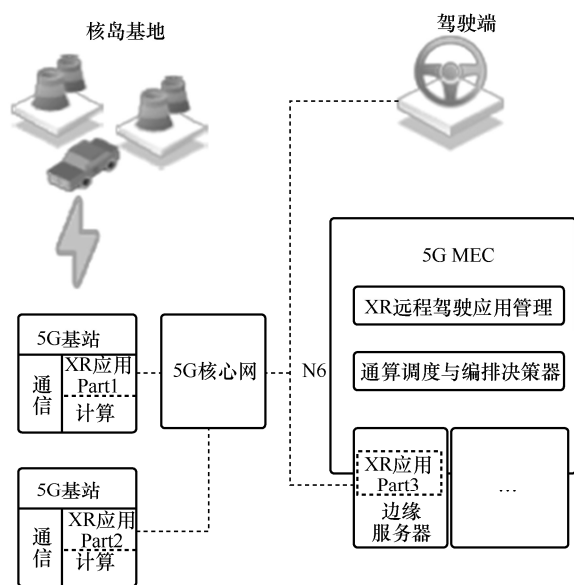


图6 BBU 负载情况



实现远程接管并操控无人车的需求。这种方式也可以实现智能装备全流程远程工作，从而代替人去特定的场景执行任务，避免人员伤亡或危险的发生。



N6: 3GPP定义的5G核心网中UPF与数据网络 (data network, DN) 的标准接口

图7 算力内生网络的XR远程驾驶

4 结束语

基于算力内生网络技术，移动网络基础设施从单纯地提供连接服务演进为同时能提供连接服务和计算服务的新型通算一体基础设施，可以满足AI所需的连接和分布式计算服务需求，从而更好地支持面向6G的通感算智一体化。

面向通算一体技术的演进，一方面，需要构建完整的算力内生网络技术标准体系，沉淀网络算力内生技术，推动其在通信领域应用的落地，如算力内生网络技术在O-RAN、SONiC-P4 (software for open networking in the cloud-P4)^[17]等网络技术架构下的部署与应用，另一方面，需要深挖算力内生网络技术在垂直行业的应用潜能，更好地赋能垂直行业，共同推进通算一体驱动的算力内生网络技术持续演进。

参考文献:

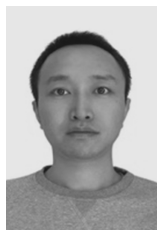
- [1] 彭程晖, 邓娟, 吴建军, 等. 6G 通算融合网络架构[J]. 无线电通信技术, 2022, 48(4): 583-591.
PENG C H, DENG J, WU J J, et al. 6G network architecture for communication and computing integration[J]. Radio Communications Technology, 2022, 48(4): 583-591.
- [2] 华为. 通算一体网络十大基础问题白皮书[R]. 2022.
HUAWEI. White paper on the ten fundamental problems of joint computing and communication networks[R]. 2022.
- [3] ETSI. Mobile-edge computing (MEC): framework and reference architecture[Z]. 2014.
- [4] CAO X W, WANG F, XU J, et al. Joint computation and communication cooperation for mobile edge computing[C]//Proceedings of 2018 16th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt). Piscataway: IEEE Press, 2018: 1-6.
- [5] REN J K, HE Y H, YU G D, et al. Joint communication and computation resource allocation for cloud-edge collaborative system[C]//Proceedings of 2019 IEEE Wireless Communications and Networking Conference (WCNC). Piscataway: IEEE Press, 2019: 1-6.
- [6] 段向阳, 杨立, 夏树强, 等. 通感算智一体化技术发展模式[J]. 电信科学, 2022, 38(3): 37-48.
DUAN X Y, YANG L, XIA S Q, et al. Technology development mode of communication/sensing/computing/intelligence integration[J]. Telecommunications Science, 2022, 38(3): 37-48.
- [7] ITU-T. Computing power network-framework and architecture: ITU-T Y.2501[S]. 2021.
- [8] ITU-T. Framework for coordination of computing and networking in IMT-2020 and beyond: ITU-T Y.IMT2020-CNC-FW[S]. 2023.
- [9] GENG L, WILLIS P. Compute first networking (CFN) scenarios and requirements[Z]. 2019.
- [10] BBF. TR-466: metro compute networking: use cases and high level requirements[R]. 2021.
- [11] ETSI. Network functions virtualisation (NFV) release 5; evolution and ecosystem; report on NFV support for network function connectivity extensions[R]. 2022.
- [12] LI N, SUN Q, LI X, et al. Towards the deep convergence of communication and computing in RAN: scenarios, architecture, key technologies, challenges and future trends[J]. China Communications, 2023, 20(3): 218-235.
- [13] O-RAN Alliance. O-RAN use cases detailed specification v04.00[Z]. 2021.

- [14] ETSI. 5G; NG-RAN; Architecture description (3GPP TS 38.401 version 17.4.0 Release 17); ETSI TS138.401 V17.4.0[S]. 2023.
- [15] GOLDBERG R P. Architectural principles for virtual computer systems[R]. 1972.
- [16] Wikipedia. x86 virtualization[Z]. 2021
- [17] SONiC. SONiC-P4 software switch[Z]. 2017.

[作者简介]



孙杰 (1983-), 男, 现就职于亚信科技(中国)有限公司研发中心云网规划部, 主要研究方向为通信与 5G 网络智能化。



马雷明 (1985-), 男, 现就职于亚信科技(中国)有限公司网络与网管产品研发中心, 主要研究方向为 5G 专网、算力网络、算力内生网络。



杨爱东 (1984-), 男, 博士, 亚信科技(中国)有限公司通信人工智能实验室首席数据科学家, 主要研究方向为 5G 无线通信、大数据挖掘、机器学习及其应用。



叶晓舟 (1980-), 男, 博士, 亚信科技(中国)有限公司通信人工智能实验室主任, 主要研究方向为通信人工智能和未来网络技术。



吕鹏 (1980-), 男, 亚信科技(中国)有限公司 5G 无线网络产品研发总监, 主要研究方向为无线网络和网络智能化技术。



王志刚 (1979-), 男, 亚信科技(中国)有限公司 5G 网络产品研发总监, 主要研究方向为核心网和网络智能化技术。



欧阳晔 (1981-), 男, 博士, IEEE Fellow, 亚信科技(中国)有限公司首席技术官、高级副总裁, 主要研究方向为移动通信、人工智能、数据科学、科技研发创新与管理。