

面向绿色计算的算网能一体化协同任务调度机制研究

罗必雄¹, 张力¹, 句赫¹, 李舒涛^{1,2}, 黄韬^{3,4}, 谢人超^{3,4}, 唐琴琴³, 彭开来⁴, 文雯³

(1. 中国电力工程顾问集团有限公司, 北京 100013; 2. 广东工业大学自动化学院, 广东广州 510006;
3. 北京邮电大学网络与交换技术全国重点实验室, 北京 100876; 4. 紫金山实验室, 江苏南京 211111)

摘要: 为满足算力网络的发展需求和低碳战略目标, 研究了算网能一体化协同任务调度机制。针对算力网络中异构节点分布分散、可再生能源动态性强以及能耗碳排优化与任务性能权衡的难题, 设计并提出了算网能一体化调度架构和综合优化调度模型。该模型在保证时延约束和负载均衡的前提下, 通过选择算力节点与网络路径, 最小化能耗与碳排放, 同时兼顾任务完成时间等用户需求。仿真结果表明, 所提模型与算法在降低能耗和碳排放的同时, 能够满足用户需求并优于基准方案。

关键词: 算力网络; 低碳; 协同创新; 任务调度; 绿色计算

中图分类号: TP393

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2025150

Computing-network-energy integrated collaborative task scheduling mechanism for green computing

LUO Bixiong¹, ZHANG Li¹, JU He¹, LI Shutao^{1,2}, HUANG Tao^{3,4}, XIE Renchao^{3,4},
TANG Qinqin³, PENG Kailai⁴, WEN Wen³

1. China Power Engineering Consulting Group Co., Ltd., Beijing 100013, China

2. School of Automation, Guangdong University of Technology, Guangzhou 510006, China

3. State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876, China

4. Purple Mountain Laboratories, Nanjing 211111, China

Abstract: To address the development needs of computing power networks and the strategic goals of low-carbon emission, the collaborative task scheduling mechanism for computing-network-energy integration was studied. In response to the challenges posed by the distributed nature of heterogeneous nodes in computing power networks, the dynamic characteristics of renewable energy, and the trade-off between energy consumption/carbon emission optimization and task performance, a scheduling architecture and a comprehensive optimization scheduling model for computing-network-energy integration were designed and proposed. This model ensured latency constraint and load balancing while minimizing energy consumption and carbon emissions through the selection of computing nodes and network paths, simultaneously considering user requirements such as task completion time. Simulation results demonstrate that the proposed model and algorithm not only reduce energy consumption and carbon emission but also meet user demands, outperforming benchmark solutions.

Keyword: computing power network, low-carbon, collaborative innovation, task scheduling, green computing

收稿日期: 2025-05-30; 修回日期: 2025-08-14

通信作者: 李舒涛, lishutao@cpecc.net

基金项目: 国家自然科学基金资助项目(No.62171046)

Foundation Item: The National Natural Science Foundation of China (No.62171046)

0 引言

在当前技术发展背景下,边缘计算、人工智能数据(AI, artificial intelligence)中心和超级计算数据中心等计算设施正大规模部署。这一趋势使普通用户更容易接入和利用计算资源,但也带来了一些挑战。例如,边缘计算节点因资源受限导致数据处理速度较慢,且计算节点分散、缺乏协同机制,任务分配调度不足,影响任务处理效率和资源均衡利用^[1-3]。为解决这些问题,产业界提出了算力网络概念^[4]。算力网络可整合网络和计算服务,高效协同处理云、网、边和端业务,提高计算资源利用率和服务质量。通过运用先进网络技术,能灵活调配计算、存储等基础资源,提升业务服务质量和用户体验,并且用户不需要关心计算资源的具体位置和部署状况^[5]。

目前,全球主要国家和经济体都在推动算力网络技术发展。2020年11月,美国发布《引领未来先进计算生态系统战略规划》,旨在建立覆盖政府、产业、学术和研究领域的国家级算力体系^[6]。2021年3月,欧盟推出“2030数字指南针”计划,目标部署1万个边缘计算节点以实现欧盟家庭千兆连接^[7]。中国同样重视算力网络技术这一算网融合研究领域,并出台一系列政策促进其发展。2021年5月,国家发展改革委等四部委联合发布《全国一体化大数据中心协同创新体系算力枢纽实施方案》,迫切需要推动数据中心合理布局、供需平衡、绿色集约和互联互通,构建数据中心、云计算、大数据一体化的新型算力网络体系,促进数据要素流通应用,实现数据中心绿色高质量发展^[8]。2022年,国务院发布《“十四五”数字经济发展规划》,提出优化升级数字基础设施,加快建设信息网络基础设施,推进云网协同和算网融合发展,有序推进基础设施智能升级^[9]。2023年2月,中共中央和国务院发布《数字中国建设整体布局规划》,提出系统优化算力基础设施布局,促进东西部算力高效互补和协同联动,引导通用数据中心、超算中心、智能计算中心、边缘数据中心等合理梯次布局^[10]。

然而,随着科技进步和算力网络发展,社会数据总量爆发式增长,数据存储、计算和应用需求大幅上升,算力网络自身能耗和碳排放也显著增加。以数据中心为例,截至2020年,我国数据中心耗电量已超900亿千瓦时,碳排放量达0.6亿吨。据中国信息通信研究院云计算与大数据研究所预测,

到2030年,我国数据中心耗电量将超3 800亿千瓦时^[11]。因此,为减轻对环境的负面影响,算力网络亟需降低能源消耗和碳排放。在此背景下,能够提高能源效率、最大限度利用可再生能源的任务调度方案将成为实现绿色算力网络的关键技术之一。

考虑到算力网络和可再生能源特性,进行算网能一体化协同任务调度需解决2个难题:1)算力网络中异构计算节点地理位置分散,能源供电构成不同,可再生能源动态性强,如何选择合适计算节点和网络路径完成计算任务?2)绿色计算核心目标是降低能源消耗和碳排放,但过于追求节能减排可能牺牲计算性能或任务完成时间,协同任务调度机制如何在能耗碳排优化和任务性能间取得平衡?

为解决上述难题,本文针对算力网络和能源协同发展,在任务调度方面设计了算网能一体化调度架构,分析描述了架构各层级主要功能和通信流程,研究了计算资源和网络资源深度集成的绿色算力网络系统中的任务调度问题,提出了算网能一体化综合优化调度模型。在保证时延约束和负载均衡前提下,通过选择任务的算力节点和网络路径,制定了一种使能耗和碳排放最小化的任务调度方案,实现绿色计算。为有效解决多目标优化问题,基于非支配排序遗传算法-II(NSGA-II, non-dominated sorting genetic algorithm II)设计了一种任务调度算法。仿真结果表明,该方案可行且优于其他基准方案。最后展望了算力网络和能源协同发展的未来研究方向并总结全文内容。

1 相关研究工作

本节将从算网能协同创新和算网能协同任务调度两方面的研究进展进行概述。

1.1 算网能协同创新

算力网络作为一种全新架构,深度融合了计算与网络,以算力为核心、网络为基础。它借助无处不在的网络连接分布式计算节点,从而达成服务的自动化部署、最优路由以及负载均衡等目标。这一创新性的架构构建起新型网络基础设施,使网络能够依据需求、实时地对不同位置的计算资源进行调度,进而提高网络和计算资源的利用率,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^[12]。算力网络概念自提出之后,就受到了产业界的广泛关注,并且发展进程正

在稳步推进之中。目前,算力网络的体系架构和关键技术依旧处于不断探索的阶段,尚未形成明确的定论。与此同时,产业和标准化相关工作也在持续开展,以满足算力网络的发展需求。

在产业层面,2024年4月,中国移动发布了《算力网络安全一体化全程可信解决方案白皮书》^[13]。该白皮书对算网安全挑战进行了分析梳理,提出了算力网络安全解决方案,并且在浙江、贵州等省份得以落地应用,围绕实现算力网络“一体化全程可信”目标开展了创新实践工作。同年7月,中国联通携手华为发布了《中国联通算力网络安全白皮书(2024)》^[14],提出了算力网络“新质安全、共链可信”的安全愿景。通过充分利用安全新技术,构建起坚实的基础设施安全底座,从而达成“开放融合+内生免疫+弹性健壮+网安智治的一体化安全”的算力网络安全目标。

在标准化方面,中国通信标准化协会牵头制定了一系列算力网络行业标准,这些标准涵盖算力路由、网络编排、设备和交易等多个领域,旨在推动算力网络的技术研究与标准化进程^[15]。此外,2021年7月,国际电信联盟电信标准部发布了首个国际标准 Y. 2501,该标准对算力网络的框架和架构进行了定义与介绍,这一举措标志着算力网络在国际学术界和产业界获得了广泛认可^[16]。2023年,国际互联网工程任务组成立了算力路由工作组(CATS, computing-aware traffic steering),致力于推动算力路由的标准化相关研究工作^[17]。

在学术领域,目前的研究主要集中在算力网络架构、调度等方面。文献[18]提出了无线算力网络的体系架构和关键技术,旨在实现无线接入网络内生算力对外开放以及原生AI支持,协同云、边、站、端多级算力,保障未来智能应用的端到端需求。文献[19]从服务的角度提出了以服务为中心的算力网络度量和建模方案,并对当前存在的问题和技术挑战进行了综述和讨论。文献[20]提出一种用于算力网络的混合式路由调度解决方案,实现按需灵活、快速路由和易于实现的融合能力。文献[21]提出了一种基于多智能体深度强化学习算法的算力路由优化方法,该算法在软件定义网络(SDN, software defined network)架构下运行,以解决算力网络中数据传输的效率问题。

随着信息技术的飞速发展和数字化转型的推

动,在未来网络中,算力需求将爆发式增长,数据中心的能耗问题日益突出。尤其是大规模AI模型的训练和应用,导致算力消耗与能耗的急剧增加。在算力网络绿色化的趋势下,算网能(算、网、能)协同创新成为一个值得深入研究的重要课题,其不仅能够提升能源利用效率、减少能源浪费,而且可为可持续发展提供重要支撑,同时亦能为应对气候变化和环境挑战做出积极贡献,进而对各行各业产生深远影响。

在此背景下,中国移动率先提出了低碳感知网络、基于SDN的“算、网、能”协同网络以及低碳算力网络总体架构,明确指出算力网络是“算网能融合”的强力助推剂,并且详细说明了算力网络结合网络能耗和能源分布情况,能够构建起低碳高效的算力网络,从而支撑国家“30/60”战略的落地实施^[22]。阿里云数据中心创新团队与华北电力大学合作的“算电协同、数能融合”系列创新成果荣获联合国工发组织 Global Call 2023 全球创新一等奖,该成果在能效优化赛道中位列全球前三名。此项技术基于业务负载的灵活调度,在时间、空间2个维度上转移数据中心电力负荷,能够根据电力系统运行信号灵活调节数据中心用能情况,进而降低数据中心用能成本和碳排放量,提升数据中心能效水平,助力新型电力系统建设以及可再生能源消纳,为绿色普惠的算力输出提供有力支撑。

1.2 算网能协同任务调度

任务调度策略在保障计算服务质量中具有较高的的重要性,目前学术界在任务调度方面已经有了一定的研究成果。常见的调度算法可以分为传统调度优化方法和基于学习的调度优化方法。

传统调度优化方法包括启发式和博弈论等方法^[23-25]。文献[23]提出了一种云边端协同下多用户细粒度任务卸载调度方法,联合考虑了时延、能耗和服务器租用成本,然后提出了多用户子任务调度方案,设计了一种改进的模拟退火粒子群算法求解最小系统总成本,从而实现最佳卸载决策。文献[24]提出了一种面向6G的节能在网计算范式,同时将在网计算的任务调度问题建模为一个多目标优化问题,提出了一种基于多目标分解的进化算法,通过实验表明该解决方案能够充分利用网络节点的计算资源,并显著降低数据中心的计算压力、网络传输开销和能耗。文献[25]考虑了用户之间的资源竞争

和移动性,在传统计算卸载模型的基础上设计了移动迁移模型,提出了一种基于博弈论的分布式算法来解决能耗最小化问题。

随着网络复杂性的提高,为适应动态的网络环境,学者提出基于学习的调度方法。文献[26]从信息与能源关联的角度出发,提出了一种面向低碳数据中心运行的在线作业调度方案,为了解决大规模异构计算任务的挑战,提出了一种基于深度强化学习的方法来解决能量感知调度问题,并实现最优在线策略。基于实际数据的案例研究结果表明,该方案可以有效地降低数据中心的碳足迹和能源成本,同时保持云产品的服务质量。文献[27]研究了考虑任务分配与资源成本不平衡的算力网络任务调度优化问题,目标是在满足任务的时延要求的情况下,使系统总利润最大化。同时提出了一种基于深度强化学习的多阶段利润感知任务调度算法来解决。文献[28]在基于信息中心网络的计算优先网络体系架构下,研究了一段时间内多用户任务的调度问题,将其建模为最小化时延和能量消耗加权和多指标优化问题,使用深度强化学习算法来得到最优任务调度解。文献[29]研究了算力网络场景下服务质量约束的能耗优化微服务调度问题,考虑了节点资源限制和网络服务质量约束,建立了能耗优化的微服务调度模型,并采用深度确定性策略梯度算法解决微服务调度问题。文献[30]在绿色算力网络中提出了一种基于广泛注意力的细粒度任务卸载方法,建立了依赖和服务约束下绿色算力网络的有向无环图子任务卸载问题,平衡总时延、能耗和匹配程度,并设计了一种元强化学习算法来解决。文献[31]针对移动边缘计算(MEC, mobile edge computing)系统中的任务调度问题,提出了一种改进的深度确定性策略梯度算法,以降低时延和能耗,为MEC系统的任务调度提供了有效的解决方案,并且该算法在复杂网络环境下具有鲁棒性和适应性。虽然以上方案在各自的场景下获得了较好的任务调度效果,但是并未考虑可再生能源的使用和不同区域计算节点的碳排放因子,无法完全应用于本文的场景中。

2 算网能一体化调度架构

2.1 设计需求

算网能一体化调度架构的设计源于数字经济发展对绿色低碳算力网络的迫切需求,主要体现在

以下3个方面。首先,能源与算力协同需求日益凸显,随着算力规模持续扩大,算力网络能耗问题愈发突出。在任务调度过程中,需要充分考虑能源供给特性与算力负载之间的动态匹配关系,以优化能源利用效率并降低碳排放。其次,多要素联合建模需求显著,算力、网络和能源资源之间的复杂交互关系需要通过构建统一的多维模型来支持全局优化决策,从而满足实时性、可靠性和经济性等综合目标。最后,低碳算网服务保障需求明确,调度架构需在确保终端用户体验(如服务响应时间、任务完成率等)的前提下,选择合适的算力节点和网络路径,以促进可再生能源消纳并降低整体碳排放。

然而,当前主流算网调度架构虽已实现算力与网络的协同编排,但在任务调度过程中对能源相关因素的考量仍显不足,尚未建立包含碳排放量、绿色能源占比等指标的多目标优化调度模型。同时,现有的算网状态信息库主要集中采集中央处理器(CPU, central processing unit)/图形处理器(GPU, graphics processing unit)利用率、网络带宽等算网指标,缺乏对发电结构数据(如火电、风电、光伏等比例)及区域电网碳排放因子等能源相关数据的实时感知能力。

针对上述问题与难点,本文提出了一种算网能一体化调度架构,旨在通过分层设计实现算力、网络和能源资源的高效协同与任务的优化调度,同时明确了相应的评价指标及其采集与量化方法。具体而言,在评价指标方面,主要包括碳排放量、绿色能源占比、任务完成时间等关键指标。针对碳排放量指标,可通过接入区域电网的碳排放因子数据库,结合实时电力消耗数据进行计算;绿色能源占比数据可通过能源管理系统获取,该系统能够实时监测不同能源类型的发电量,并将其比例作为量化依据;任务完成时间等用户体验相关指标则可通过分布式监控系统采集。

2.2 架构概述

算网能一体化调度架构如图1所示,分为3个层级:基础设施层、管控层和应用服务层。

1) 基础设施层包括算力、网络和能源基础设施,能够提供强大的计算能力,支撑各类复杂运算与数据处理任务,实现数据的高速传输与互联互通,同时保障稳定、持续的能源供给,确保整个系

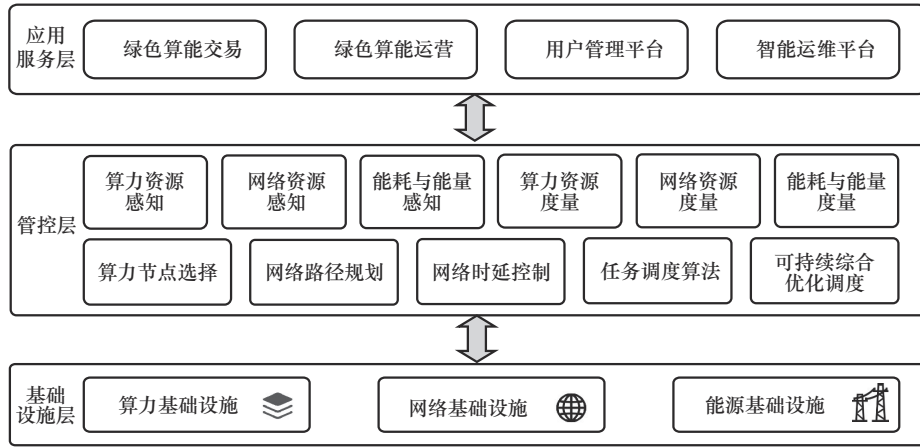


图1 算网能一体化调度架构

统架构的正常运转。

2) 管控层是调度架构的核心，包含多个功能模块。其中一些辅助调度模块主要负责感知、度量、管理与优化算力、网络、能耗及能量资源。具体职责包括监测和评估算力资源、实时监控网络状态、追踪能耗情况、能源配套比例以及电能质量和碳排放因子等。任务调度算法与可持续综合优化调度模块协同工作，根据用户的任务需求，在确保服务质量不降低的前提下，以低碳环保为目标，选择合适的算力节点、规划网络路径并控制网络时延，以支持算力任务的高效调度，促进可再生能源消纳，推动可持续发展。

3) 应用服务层包括绿色算能交易、绿色算能运营、用户管理平台、智能运维平台等服务，提供工具和技术来管理和优化业务运营，支持基于信息的决策，可提高业务效率，降低成本，并确保所需的服务质量。

2.3 工作流程

在算网能一体化调度架构中，各层之间紧密配合，以实现系统的协同工作和功能集成，如图2所示。以下是主要的工作流程。

流程1：在管控层，对基础设施层的计算资源、网络资源、能耗与能量进行感知，建立算网能状态信息库。

流程2：用户的终端设备发起计算任务请求，并为任务进行标识。

流程3：基础设施层中的算网接入侧路由器解析计算任务请求，并将其上报给管控层。

流程4：管控层分析和处理计算任务，识别终端用户、任务类型和优先级，并决策任务计算和传

输的服务级别及策略。

流程5：综合考虑算力资源、网络资源、能耗与能量等状态信息，选择最优的算力节点和网络路径来更高效和更绿色地完成用户的计算任务。

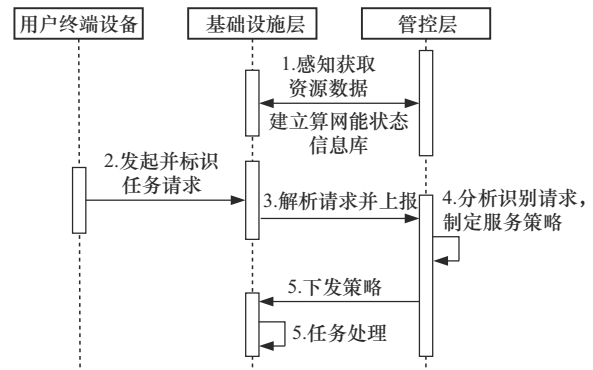


图2 调度架构工作流程

3 算网能一体化调度模型

本节将介绍算网能一体化调度系统的网络模型、任务模型、时延模型、负载均衡模型、能耗和碳排放模型以及优化问题。

3.1 系统网络模型

算网能一体化调度系统如图3所示，主要由6个主要部分组成，即用户、算力供应方、网络供应方、算网能调度平台、电力交易平台和能源供应方。令 $U = \{1, 2, \dots, u, \dots, U\}$ 表示用户的集合， $M = \{1, 2, \dots, m, \dots, M\}$ 表示算力供应方的集合， $N = \{1, 2, \dots, n, \dots, N\}$ 表示网络供应方的集合。能源供应方通过电力交易平台与算力供应方和网络供应方建立联系，提供各种电能选择（光电、风电等清洁能源以及火力发电能源）。算力供应方和网络

供应方按需购买电能, 形成各自的能源比例。算网能调度平台是系统的大脑, 通过分析用户需求和当前资源情况, 进行资源调度, 以高效完成任务。为了实现绿色计算, 解决能耗碳排优化与任务性能权衡的难题, 本文研究了用户任务的资源分配调度问题, 在满足时延和负载均衡的约束下, 最小化能耗和碳排放, 并提出了一种基于启发式的算法来制定调度策略。主要数学符号如表 1 所示。

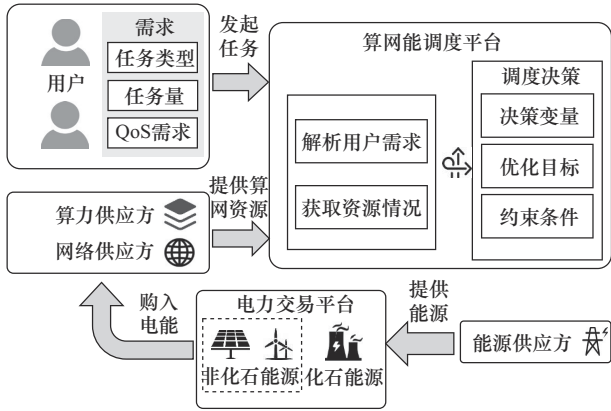


图 3 算网能一体化系统资源调度示意

表 1 主要数学符号

数学符号	描述
$\mathcal{U}, \mathcal{U}, u$	用户的集合、数量和索引
$\mathcal{M}, \mathcal{M}, m$	算力节点的集合、数量和索引
f_m, f'_m	算力节点 m 的算力资源和接收到任务所占用的算力资源总量
$\mathcal{N}, \mathcal{N}, n$	网络链路的集合、数量和索引
\mathcal{R}, R	路由器的集合和数量
\mathcal{R}^r, S, s	普通路由器的集合、数量和索引
\mathcal{R}^c, M	算力路由器的集合和数量
o_n^{\max}, ζ_n	最大网络链路容量和网络传输速率
$\mathcal{P}_{s,m}, P$	普通路由器 \mathcal{R}_s^r 和算力节点 m 之间网络路由的集合和数量
$\mathcal{P}_{s,m}^p$	普通路由器 \mathcal{R}_s^r 和算力节点 m 之间的第 p 条路由
Task	新访问的任务集合
$\mathcal{W}_s, \mathcal{W}_s $ 或 K	普通路由器 \mathcal{R}_s^r 上新到达任务的集合和数量
$w_{s,k}$	普通路由器 \mathcal{R}_s^r 上第 k 个新到达的任务
$d_{s,k}, \varepsilon_{s,k}, \tau_{s,k}$	任务 $w_{s,k}$ 的数据大小、计算负载量和最大可容忍时延
$\kappa_{s,k}$	任务 $w_{s,k}$ 到来前已有的平均任务计算负载量
$x_{w_{s,k}}^m, y_{w_{s,k}}^p$	针对任务 $w_{s,k}$, 制定的算力节点和网络路由的选择决策变量
ψ^{\max}, E^{\max}	负载均衡上限和最大可提供电能

本文将算力供应方抽象为算力网络中的不同算力节点, 为用户提供算力服务, 处理任务请求; 网络供应方抽象为算力网络中的不同网络链路, 构建起内部节点之间的网络链接。针对系统的网络拓扑, 令 $\mathcal{R} = \{\mathcal{R}^r, \mathcal{R}^c\} = \{1, \dots, R\}$ 表示路由器集合, 其中, $\mathcal{R}^r = \{\mathcal{R}_1^r, \dots, \mathcal{R}_S^r\}$ 表示负责接收和转发数据的普通路由器的集合, 类似计算机网络的网络层中的核心层路由器或汇聚层路由器, 并不直接面向用户端; $\mathcal{R}^c = \{\mathcal{R}_1^c, \dots, \mathcal{R}_M^c\}$ 表示算力路由器的集合, 直接与算力节点相连, 接收到普通路由器转发的算力任务后直接与算力节点进行数据交换。整个网络体系结构建模为无向连通图 $\mathcal{G}(\mathcal{V}, \mathcal{N})$, $\mathcal{V} = \mathcal{R} \cup \mathcal{M}$ 表示路由器和算力节点的集合, \mathcal{N} 表示网络链路的集合, $n \in \mathcal{N}$ 代表其中一条网络连接, 该网络连接具有非负的权值, 其中 o_n^{\max} 表示最大网络链路容量, ζ_n 表示网络传输速率。以源节点为普通路由器 $\mathcal{R}_s^r \in \mathcal{R}^r$ 、算力节点 $m \in \mathcal{M}$ 为目的节点的路由有多条, 表示为 $\mathcal{P}_{s,m} = \{\mathcal{P}_{s,m}^1, \dots, \mathcal{P}_{s,m}^p\}$, 其中 $\mathcal{P}_{s,m}^p \in \mathcal{P}_{s,m}$ 表示第 p 条路由。具体网络拓扑如图 4 所示。

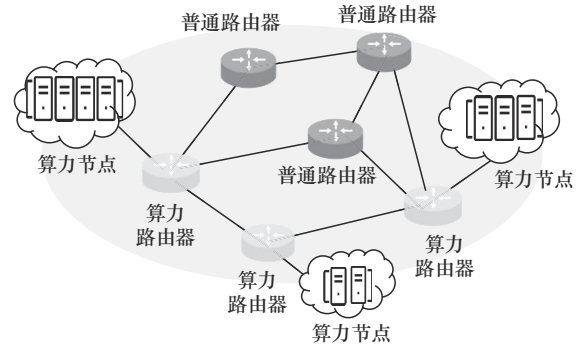


图 4 网络拓扑

3.2 任务模型

设 $\text{Task} = \{\mathcal{W}_1, \mathcal{W}_2, \dots, \mathcal{W}_s, \dots, \mathcal{W}_S\}$ 表示新访问的任务集合, 其中 $\mathcal{W}_s \in \text{Task}$ 是独立同分布的有限随机变量序列, 具有数学期望 $\mathbb{E}[\mathcal{W}_s] = \lambda$, 并满足基数约束 $|\mathcal{W}_s| \leq K^{[32]}$ 。令 $\mathcal{W}_s = \{w_{s,1}, w_{s,2}, \dots, w_{s,K}\}$ 表示新到达普通路由器 \mathcal{R}_s^r 的任务集合, 其中 $w_{s,k} \in \mathcal{W}_s$ 代表第 k 个新到达的任务 ($1 \leq k \leq K$)。每个任务 $w_{s,k}$ 定义为三维特征向量 $w_{s,k} = (d_{s,k}, \varepsilon_{s,k}, \tau_{s,k})$, 其中 $d_{s,k}$ 表示数据大小 (单位为 bit); $\varepsilon_{s,k}$ 表示计算负载量 (执行任务所需的 CPU 周期数); $\tau_{s,k}$ 表示最大可容忍

时延 (单位为s)。假设返回的数据通常小于初始数据, 因此忽略结果返回的时延和能耗等指标的计算^[33]。

3.3 时延模型

时延是衡量服务质量的关键。在本文中, 用户设备通过无线网络将任务请求发送到基站 (边缘服务器), 从而接入算力网络中, 经路由器转发到达算力节点进行处理。那么任务的处理时延包括接入时延、网络传输时延、排队时延和执行时延。

1) 接入时延

接入时延是指计算任务从终端创建, 通过无线网络到边缘服务器, 再到接入的普通路由器的时延, 主要取决于任务数据量和上行链路速率。默认边缘服务器与连接的普通路由器之间距离较短, 所造成的时延忽略不计。那么接入时延可表示为

$$D_{s,k}^{\text{access}} = \frac{d_{s,k}}{v_{s,k}} \quad (1)$$

其中, $v_{s,k}$ 为上行链路速率 (单位为 bit/s), 可利用香农公式^[34]推导如下

$$v_{s,k} = \frac{B_s}{|\mathcal{W}_s|} \text{lb} \left(1 + \frac{\delta_{s,k} \cdot \eta_{s,k}}{|\mathcal{W}_s| \cdot \sigma_{s,k}^2} \right) \quad (2)$$

其中, B_s 是信道总带宽, $\delta_{s,k}$ 是传输功率, $\eta_{s,k}$ 和 $\sigma_{s,k}^2$ 分别是信道增益和高斯白噪声功率。

2) 网络传输时延

当任务 $w_{s,k}$ 沿着网络路径 $\mathcal{P}_{s,m}^p$ 传输到算力节点 m 时, 网络传输时延可表示为

$$D_{s,k}^{\text{tran}} = d_{s,k} \cdot \sum_{n \in \mathcal{P}_{s,m}^p} \frac{1}{\zeta_n} \quad (3)$$

3) 排队时延

在任务 $w_{s,k}$ 到达算力节点 m 后, 由于算力节点无法同时处理多个任务, 因此任务 $w_{s,k}$ 需要等待。假设已知该任务到来前已有的平均任务计算负载量为 $\kappa_{s,k}$ (单位为 CPU 周期数), 那么排队时延可表示为

$$D_{s,k}^{\text{wait}} = \frac{\kappa_{s,k}}{f_m} \quad (4)$$

其中, f_m 指算力节点 m 的算力资源, 也就是每秒 CPU 周期数。

4) 执行时延

在算力节点 m 上处理任务 $w_{s,k}$ 的执行时延可表

示为

$$D_{s,k}^{\text{com}} = \frac{\varepsilon_{s,k}}{f_m} \quad (5)$$

3.4 负载均衡模型

维持长期的负载均衡对于提高资源利用率和确保系统稳定性至关重要^[35]。负载均衡模型包括算力节点和网络的均衡。

1) 算力节点的平衡

算力节点 m 的负载可表示为

$$\psi_m = \min \left\{ \frac{f'_m}{f_m}, 1 \right\} \quad (6)$$

其中, f'_m 是指算力节点 m 所接收到的任务占用的计算资源总量。对于任意 $m \in \mathcal{M}$, $\psi_m \leq 1$, 即算力节点仅能负载不超过自身算力资源的计算任务。

使用标准差来评估所有算力节点的负载均衡性能, 如式(7)所示。

$$\psi_{\text{com}} = \sqrt{\frac{\sum_{m \in \mathcal{M}} (\psi_m - \bar{\psi}_m)^2}{M}} \quad (7)$$

其中, $\bar{\psi}_m$ 表示所有算力节点的平均负载。

2) 网络的平衡

网络链路 n 的负载可表示为

$$\psi_n = \min \left\{ \frac{\Gamma_n}{\sigma_n^{\text{max}}}, 1 \right\} \quad (8)$$

其中, Γ_n 是指网络链路 n 经过的任务数据量总和。对于任意 $n \in \mathcal{N}$, $\psi_n \leq 1$, 即网络路径仅能负载不超过自身传输能力的传输数据。

那么网络的负载均衡性能可表示为

$$\psi_{\text{net}} = \sqrt{\frac{\sum_{n \in \mathcal{N}} (\psi_n - \bar{\psi}_n)^2}{N}} \quad (9)$$

其中, $\bar{\psi}_n$ 表示整个网络的平均负载。

因此, 整个系统的负载均衡性能可以表示为

$$\psi^{\text{total}} = \omega_1 \cdot \psi_{\text{com}} + \omega_2 \cdot \psi_{\text{net}} \quad (10)$$

其中, ω_1 和 ω_2 是权衡系数。

3.5 能耗模型

任务处理的能耗包括执行能耗和传输能耗。算力节点处理任务消耗的电能与任务量以及设备本身的参数有关^[36]。设 θ_m 表示算力节点 m 有关算力处理任务消耗电能的能量系数, 该系数取决于服务器的芯片架构^[37-38], p_m^{static} 表示节点 m 的静态功率^[39],

那么算力节点 m 处理任务 $w_{s,k}$ 的能耗可表示为

$$E_{s,k}^{\text{com}} = \theta_m \cdot f_m^2 \cdot \varepsilon_{s,k} + P_m^{\text{static}} \cdot D_{s,k}^{\text{com}} \quad (11)$$

为简化分析, 设 P_n^{tran} 表示网络中所有路由器的传输功率 (单位为 W), 则任务 $w_{s,k}$ 的传输能耗可表示为

$$E_{s,k}^{\text{tran}} = P_n^{\text{tran}} \cdot D_{s,k}^{\text{tran}} \quad (12)$$

3.6 碳排放模型

算力节点和网络链路的能源来源涵盖化石能源与非化石能源。化石能源发电产生大量二氧化碳, 非化石能源发电几乎无直接碳排放。不同节点和链路能源组成比例各异, 用能组合影响计算与传输任务时的碳排放。本文通过实时感知区域电网的发电结构数据及碳排放因子, 动态获取各算力节点和网络链路的能源构成。并且网络路由上的每段链路的能源结构占比可能不同, 以反映实际网络中不同区域电网的供电差异。假设能源消耗碳排放系数 (单位能源消耗所产生的二氧化碳排放量) C^B 表示化石能源的碳排放。通过下标的方式对不同区域的节点 m 和网络链路 n 进行区分, 即 C_m^B 和 C_n^B 。对于算力节点 m , 分别用 ρ_m^B 和 ρ_m^G 表示火电和绿电占比, 且 $\rho_m^B + \rho_m^G = 1$ 。同样地, ρ_n^B 和 ρ_n^G 分别表示网络链路 n 的火电和绿电占比。同时本文考虑理想情况, 任务传输和执行过程中算力节点和网络路径剩余的总能量为 0, 那么算力节点 m 在执行计算任务 $w_{s,k}$ 时造成的碳排放为

$$C_{s,k}^{\text{com}} = C_m^B \cdot \rho_m^B \cdot E_{s,k}^{\text{com}} \quad (13)$$

任务 $w_{s,k}$ 在传输过程中产生的碳排放为

$$C_{s,k}^{\text{tran}} = \sum_{n \in \mathcal{P}_{s,m}^p} C_n^B \cdot \rho_n^B \cdot \frac{E_{s,k}^{\text{tran}}}{|\mathcal{P}_{s,m}^p|} \quad (14)$$

其中, $|\mathcal{P}_{s,m}^p|$ 表示这段路由的链路总条数, 为简化分析, 本文认为路由上每条链路的能耗贡献量是近似相等的。

3.7 优化问题

调度决策主要考虑两方面: 一方面是算力节点的选择, 另一方面是网络路径的选择。使用向量 $\mathbf{x} = [x_{w_{s,k}}^m]$ 表示算力节点的选择决策, 变量取值定义为

$$x_{w_{s,k}}^m = \begin{cases} 1, & \text{任务 } w_{s,k} \text{ 调度到算力节点 } m \\ 0, & \text{其他} \end{cases} \quad (15)$$

使用向量 $\mathbf{y} = [y_{w_{s,k}}^{\mathcal{P}_{s,m}^p}]$ 表示网络路径的选择决策,

考虑任务调度执行前就需要完成调度路径的选择, 因此网络调度采取源路由方式, 变量取值定义为

$$y_{w_{s,k}}^{\mathcal{P}_{s,m}^p} = \begin{cases} 1, & \text{任务 } w_{s,k} \text{ 沿着 } \mathcal{P}_{s,m}^p \text{ 调度到算力节点 } m \\ 0, & \text{其他} \end{cases} \quad (16)$$

处理任务集合 Task 造成的总能耗可表示为

$$E^{\text{total}} = \sum_{w_{s,k} \in \text{Task}} \sum_{m=1}^M x_{w_{s,k}}^m \cdot E_{s,k}^{\text{com}} + \sum_{w_{s,k} \in \text{Task}} \sum_{m=1}^M \sum_{p=1}^P x_{w_{s,k}}^m \cdot y_{w_{s,k}}^{\mathcal{P}_{s,m}^p} \cdot E_{s,k}^{\text{tran}} \quad (17)$$

处理任务集合 Task 产生的总碳排放量可表示为

$$C^{\text{total}} = \sum_{w_{s,k} \in \text{Task}} \sum_{m=1}^M x_{w_{s,k}}^m \cdot C_{s,k}^{\text{com}} + \sum_{w_{s,k} \in \text{Task}} \sum_{m=1}^M \sum_{p=1}^P x_{w_{s,k}}^m \cdot y_{w_{s,k}}^{\mathcal{P}_{s,m}^p} \cdot C_{s,k}^{\text{tran}} \quad (18)$$

任务 $w_{s,k}$ 的处理总时延可表示为

$$D_{s,k}^{\text{total}} = \sum_{m=1}^M x_{w_{s,k}}^m \cdot (D_{s,k}^{\text{com}} + D_{s,k}^{\text{wait}}) + D_{s,k}^{\text{access}} + \sum_{m=1}^M \sum_{p=1}^P x_{w_{s,k}}^m \cdot y_{w_{s,k}}^{\mathcal{P}_{s,m}^p} \cdot D_{s,k}^{\text{tran}} \quad (19)$$

调度模型的优化目标是在保障用户需求的前提下选择合适算力节点和网络路径, 使能耗和碳排放最小, 实现绿色计算。那么总的在约束条件下的多目标优化问题可表示为

$$\min (E^{\text{total}}, C^{\text{total}}) \quad (20)$$

$$\text{s.t. C1: } D_{s,k}^{\text{total}} \leq \tau_{s,k}, \forall w_{s,k} \in \text{Task}$$

$$\text{C2: } \psi^{\text{total}} \leq \psi^{\text{max}}$$

$$\text{C3: } E^{\text{total}} \leq E^{\text{max}}$$

$$\text{C4: } \sum_{w_{s,k} \in \text{Task}} \varepsilon_{s,k} \leq \sum_{m=1}^M f_m - f'_m$$

$$\text{C5: } x_{w_{s,k}}^m \in \{0, 1\}, \sum_{m=1}^M x_{w_{s,k}}^m = 1, \forall w_{s,k} \in \text{Task}$$

$$\text{C6: } y_{w_{s,k}}^{\mathcal{P}_{s,m}^p} \in \{0, 1\}, \sum_{p=1}^P y_{w_{s,k}}^{\mathcal{P}_{s,m}^p} = x_{w_{s,k}}^m, \forall w_{s,k} \in \text{Task}, m \in \mathcal{M} \quad (21)$$

其中, ψ^{max} 是系统负载均衡的上限, E^{max} 是系统可提供的最大电能。约束条件 C4 不等式左右两边分别表示任务所需的全部计算资源和剩余的计算资源总量; C5 和 C6 保证了任务只能通过一条网络路径到达某一个算力节点上。

4 优化算法和仿真分析

本节将简单介绍所使用的优化算法，并给出上述优化问题的仿真结果和分析。

4.1 优化算法

混合求解器与启发式算法都是学术界和产业界常用的规划问题求解方案。然而考虑算力网络任务调度场景需要同时考虑多种资源的调度问题，混合求解器在大规模问题中可能遭遇“维度灾难”，导致计算时间和资源消耗急剧增加。这与算力网络的用户服务需求背道而驰。同时，求解器难以求解复杂的多目标优化问题，需要将多目标转为单目标优化问题进行处理。而启发式算法能够在较短的时间内为大规模问题提供高质量的解，对多目标优化的求解也具有较高的求解效率，这在追求调度实时性和能效优化的绿色计算中尤为关键。

为了解决多目标优化问题，采用 NSGA-II 算法，该算法通过加入快速非支配排序算法、拥挤距离和精英保持策略，可以有效地获得 Pareto 解。在原有 NSGA-II 算法的基础上，设计了一种改进的、定制的符合本文调度模型的 NSGA-II 算法。

NSGA-II 算法的时间复杂度主要由 3 个核心操作决定：快速非支配排序、拥挤距离计算和选择操作。快速非支配排序通过优化传统 NSGA 的支配关系比较过程，将时间复杂度从 $\mathcal{O}(MN^3)$ 降低至 $\mathcal{O}(MN^2)$ ，其中 M 为目标函数数量， N 为种群规模。拥挤距离计算需要对每个非支配层内的个体按目标函数排序，其复杂度为 $\mathcal{O}(MN \log N)$ 。选择操作（如二元锦标赛）的复杂度为线性 $\mathcal{O}(N)$ 。因此，NSGA-II 单次迭代的总时间复杂度为 $\mathcal{O}(MN^2)$ ，由非支配排序主导。空间复杂度方面，算法需存储 $\mathcal{O}(N^2)$ 支配关系矩阵和 $\mathcal{O}(N \cdot (D + M))$ 种群个体（ D 为决策变量数），总空间消耗为 $\mathcal{O}(N^2)$ 。这一复杂度优化使 NSGA-II 能够高效处理大规模多目标优化问题，同时通过精英策略和拥挤距离机制平衡收敛性与多样性。

用 $\mathbb{P}_j = \left\{ \{ \mathbf{x}, \mathbf{y} \}_{j,1}, \dots, \{ \mathbf{x}, \mathbf{y} \}_{j,N_{\text{pop}}} \right\}$ 表示第 j 代的种群，其中 $\{ \mathbf{x}, \mathbf{y} \}_{j,k} \in \mathbb{P}_j$ 表示第 j 代中的第 k 个个体， N_{pop} 表示总体大小。在本文中，每一个体包含算力节点和网络路径的选择决策。

算法 1 基于改进 NSGA-II 的任务调度算法

输入 系统资源、任务数据、约束限制

输出 $\left\{ \{ \mathbf{x}, \mathbf{y} \}_{N_{\text{gen}},1}, \dots, \{ \mathbf{x}, \mathbf{y} \}_{N_{\text{gen}},N_{\text{pop}}} \right\}$

- 1) 初始化种群 \mathbb{P}_0 任务调度策略
- 2) 初始化参数 N_{gen} 、 P_c 和 P_M
- 3) 计算种群 \mathbb{P}_0 中各任务调度策略下的 E^{total} 、 C^{total} 和约束条件
- 4) 计算种群 \mathbb{P}_0 中的所有非支配等级和拥挤距离
- 5) for $j = 1:N_{\text{gen}}$ do
- 6) 根据非支配等级和拥挤距离从 \mathbb{P}_j 中选择候选解
- 7) 根据每个个体的 E^{total} 和 C^{total} 更新 \mathbb{P}_j 的 P_c 和 P_M
- 8) 通过交叉突变生成子代群体 \mathbb{P}'_j
- 9) $\mathbb{P}\mathbb{P}'_j = \mathbb{P}_j \cup \mathbb{P}'_j$
- 10) for $\{ \mathbf{x}, \mathbf{y} \}_{j,k} \in \mathbb{P}\mathbb{P}'_j$ do
- 11) 计算具有约束条件的目标函数，等级和距离
- 12) end for
- 13) 根据等级和拥挤距离从 $\mathbb{P}\mathbb{P}'_j$ 中产生下一代种群 \mathbb{P}_{j+1}
- 14) 在种群中按照个体的权重与最优权重选择找到最佳个体 $\{ \mathbf{x}, \mathbf{y} \}_{j,\text{optimal}}$
- 15) 若最优个体 $\{ \mathbf{x}, \mathbf{y} \}_{j,\text{optimal}}$ 连续 N 代不变则循环终止
- 16) end for

本文提出的基于改进 NSGA-II 的任务调度算法的伪代码如算法 1 所示，具体内容如下。1) 初始化种群 \mathbb{P}_0 和相关参数，包括最大代数 N_{gen} 、交叉概率 P_c 、突变概率 P_M 。2) 计算目标函数 E^{total} 和 C^{total} 即系统在该调度策略下的能耗与碳排放总量，以及约束违反情况判断该调度策略是否可行，将所有个体分配到不同的非支配解等级。为了确保种群的多样性，根据 E^{total} 和 C^{total} 来估计每个解 $\{ \mathbf{x}, \mathbf{y} \}_{j,k}$ 的拥挤距离。然后，通过非支配排序和拥挤距离来选择候选解。3) 根据目标函数动态调整交叉概率 P_c 、突变概率 P_M ，通过交叉和突变产生新的子代群体。4) 将父代群体和子代群体整合为整个种群，并对其进行排序以生成下一代。重复步骤 2)~步骤 4)，直到算法收敛。

4.2 仿真分析

本节模拟了所提出的算力网能一体化调度系统中

任务调度资源分配方案的性能。

4.2.1 参数设置

本文在算法设计中采用了动态调整交叉概率和突变概率的创新策略。具体而言,调整机制基于双重标准。

首先,根据个体适应度值进行动态调节,当个体适应度高于种群平均值时,交叉概率从基准值0.8降至0.6以保护优良个体;低于平均值时则提升至0.9以增强种群多样性。其次,引入进化阶段自适应机制,在进化初期(前30%代数)设置较高突变概率(0.1)以促进探索,中期(30%~70%代数)线性降至0.05,后期维持在0.02以保证开发能力。该策略通过平衡探索与开发的关系,显著提升了算法收敛效率。

在仿真评估中,针对连续产生的计算任务,对任务模型^[40]进行建模。输入任务的数据总量大小 $d_{s,k}$ 随机分布为2.5~10 MB,处理每个任务所需的CPU周期数 $\varepsilon_{s,k}$ 随机分布为1~4 Gcycle;算力节点的CPU频率 f_m 设置为6 GHz,能量系数 θ_m ^[37]设置为 10^{-26} ;为全面评估系统性能,将计算任务分为2类:时延敏感型任务(占比60%)和计算密集型任务(占比40%)。任务到达过程服从泊松分布,动态生成不同权重组合的任务序列,以模拟实际场景中的混合负载情况。

每个算力节点和网络路径的火电能源占比随机分布在0.3~0.8,以更准确地反映中国区域差异:西北可再生能源富集区设置为0.3~0.5(如新疆、甘肃等光伏/风电基地),东部工业密集区维持0.6~0.8(如长三角、珠三角);火电的碳排放^[41]设置为254 g/(kW·h)。这些参数设置具有合理性:任务数据量和CPU周期需求的范围是基于视频处理任务的实际需求,反映了不同复杂度视频任务的计算和数据传输需求。6 GHz的CPU频率符合高性能计算服务器的配置,能够满足任务的计算需求。能耗和碳排放模型相关的参数则根据实际服务器能耗特性和火力发电的碳排放数据设定,反映了当前能源结构和环境影响。这些仿真条件能够有效模拟用户任务在算网能一体化调度系统中的运行情况,为后续研究提供可靠的实验基础。

对于网络拓扑而言,在算力网络中,多个边缘服务器和路由器共同构成了一个复杂的网络拓扑结构,以实现高效的用戶任务处理和数据传输。网络中包含多个边缘服务器,分别标记为Edge Server 1、

Edge Server 2、Edge Server 3等,这些边缘服务器通过无线网络与用户设备连接,作为用户接入算力网络的入口节点。网络的核心部分由多个普通路由器和算力路由器组成,这些路由器通过高速链路相互连接,形成一个高效的通信骨干网络,实现不同区域之间的数据交换和任务分配。路由器之间通过冗余链路连接,确保网络的高可靠性和容错能力。例如,当Router 1出现故障时,任务可以通过Router 2或Router 3进行转发,从而保证网络的正常运行。在算力节点方面,网络中有多个算力服务器,分别标记为Server A、Server B、Server C、Server D等,这些服务器通过高速链路连接到相应的算力路由器。

4.2.2 实验结果

图5显示了所提出的NSGA-II算法的输出结果。由于所提出的优化问题是一个多目标问题,因此每次迭代的输出都是一组解集。NSGA-II算法的收敛曲线如图6所示,其中,圆形点集为算法最终收敛到的帕累托平面,即最后一代解,不同的形状代表不同代数的解。观测解集可以发现,帕累托平面以外的点的个数较少,说明NSGA-II算法收敛速度较快,能够快速得到任务调度策略。

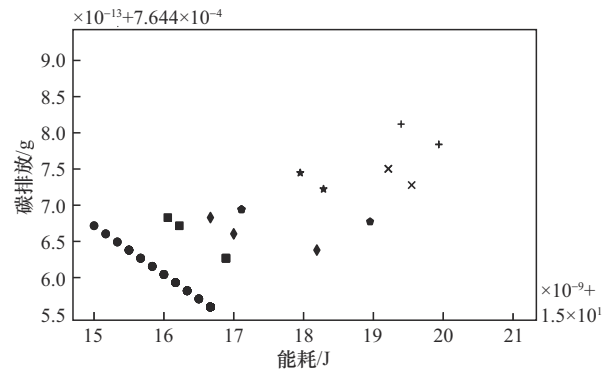


图5 NSGA-II算法的输出结果

为了表现NSGA-II算法的收敛速度,进一步将其与其他常见优化算法(如贪婪算法)进行了对比分析。通过对比实验,发现NSGA-II算法在收敛速度上显著优于贪婪算法。

具体而言,贪婪算法在迭代初期能够凭借其快速决策的特性迅速找到局部最优解。然而,这种快速收敛到局部最优的特性在后续迭代中却成为其主要局限。由于贪婪算法在每一步只考虑当前最优的选择,缺乏全局解空间的探索能力,因此它很容易

陷入局部最优解，导致收敛速度在达到一定阶段后显著减缓，且难以进一步优化以找到全局最优解。相比之下，NSGA-II 算法凭借其基于种群的优化机制和多目标优化策略，展现出显著的优势。在迭代过程中，NSGA-II 算法通过维护一个种群，可同时探索多个解，并通过遗传操作（如交叉和变异）不断生成新的解，从而更有效地探索解空间。这种基于种群的进化策略使 NSGA-II 算法能够避免陷入局部最优，从而在迭代过程中持续优化解的质量。此外，NSGA-II 算法的多目标优化策略能够同时优化多个目标函数，通过非支配排序和拥挤距离计算，确保解的多样性和分布均匀性，从而在多目标优化问题中找到一组高质量的 Pareto 最优解。

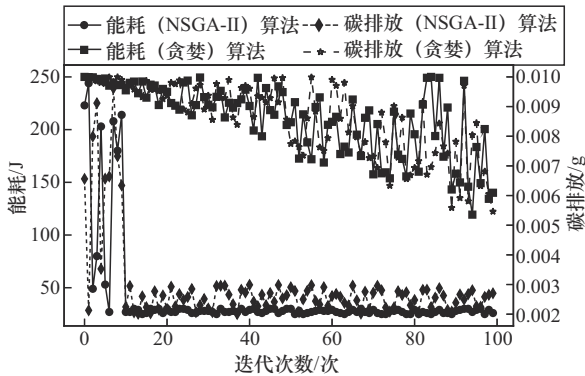


图6 NSGA-II 算法的收敛曲线

图 7 和图 8 分别展示了算力节点和网络在有无负载均衡策略下的性能对比情况。其中，纵坐标分别表示算力均衡度（对应式(7)的平方）以及网络均衡度（对应式(9)的平方）。从图 7 可以看出，随着任务所需 CPU 周期数的增加，算力节点的均衡度呈现上升趋势；从图 8 可以观察到，随着任务数据量的增加，网络的均衡度同样表现出增大的趋势。在负载均衡策略的作用下，算力节点和网络的均衡度始终显著低于无负载均衡条件下的对应值。由此，本文提出的方案通过引入负载均衡策略，能够实现任务在算力节点和网络路径上的合理分配。

为验证多目标优化特性，将本文中优化算法与单目标算法进行对照，对照方法为仅优化能耗的算法或仅优化碳排放的算法。图 9 给出了能耗碳排联合优化方案与仅能耗指标方案的性能对比，其中柱状图表示能耗指标，折线图表示碳排放指标。由图 9 结果可得，在不同计算量大小下，联合优化方案的总能耗比仅能耗指标方案高 2.66%，但总碳排放量比仅

能耗指标方案低 5.23%。这是由于该方案在实现时进一步考虑了不同算力节点的火力能源和清洁能源占比，更加偏向于将任务调度在清洁能源占比较高的算力节点上。此外，随着数据总量的增加，不同调度方案间的总能耗与碳排放差异也随之增大。

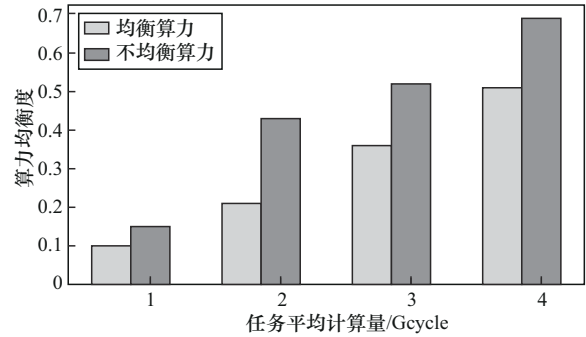


图7 算力节点负载情况对比

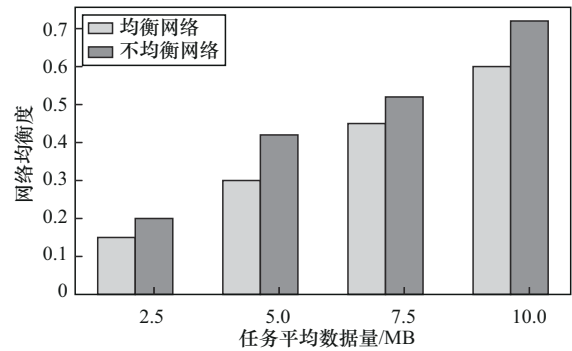


图8 网络负载情况对比

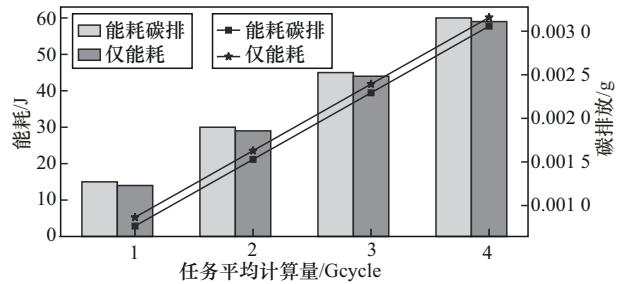


图9 联合优化方案与仅能耗指标方案的性能对比

图 10 给出了能耗碳排联合优化方案与仅碳排放指标方案的性能对比，其中，柱状图表示能耗指标，折线图表示碳排放指标。由图 10 结果可得，在不同计算量大小下，联合优化方案的总能耗比仅碳排放指标方案低 2.66%，但总碳排放量比仅碳排放指标方案高 5.23%。与图 9 相反，这是由于该方案在实现时不再着重考虑不同算力节点的火力能源和清洁能源占比，而是将能耗与碳排放进行联合优

化,使2个指标均达到较小。从优化结果的特性来看,单目标优化的结果具有明确的最优性,但可能牺牲其他未优化目标的性能。例如,仅优化能耗可能导致碳排放显著增加。多目标优化通过Pareto前沿提供多个折中解,决策者可根据实际需求选择(如优先低碳或低能耗)。

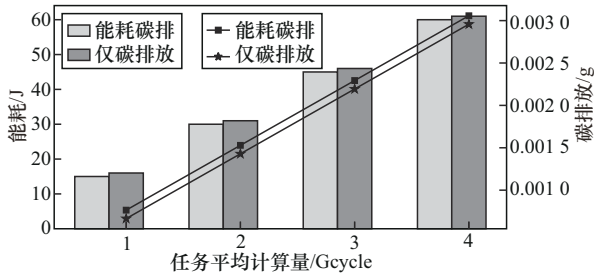


图10 联合优化方案与仅碳排放指标方案的性能对比

图11给出了能耗碳排联合优化与能耗时延联合优化方案的性能对比,其中,柱状图表示能耗指标,折线图表示碳排放指标。在新方案中,采取所有用户的总时延为一优化指标,将其与总能耗作为优化目标并利用NSGA-II算法进行优化,总数据量与上述实验相同。正因为考虑了时延的约束,总能耗比此前所提出的能耗碳排联合优化方案高3.63%,总碳排放量比此前所提出的能耗碳排联合优化方案高26.16%。由于该方案尝试寻求更低的任务执行能耗,在调度任务的过程中可能将任务分派至计算能力较强而绿电并不充足的节点上。此外,随着平均计算量的增加,不同调度方案间的能耗差异也随之略微增大而碳排放量增长较为明显。

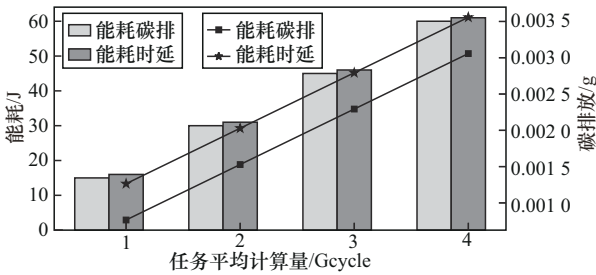


图11 能耗碳排联合优化与能耗时延联合优化方案的性能对比

图12给出了能耗碳排联合优化方案(本文方案)与其他能碳联合优化算法的性能对比,其中,柱状图表示能耗指标,折线图表示碳排放指标。在对比优化方案中,继续采取所有用户的总能耗与总

碳排放量为优化指标,并将这2种指标作为优化目标利用贪心算法与随机算法进行优化。基于贪心的调度算法(GBSA, greedy based scheduling algorithm)^[42]是一种在每一步决策中仅考虑当前局部最优解的算法策略,通过逐步构建解而不回溯或全局考量,期望通过局部最优的累积达到全局最优。其核心原理基于2个关键性质:一是贪心选择性,即每一步的局部最优选择能够导向全局最优解,如在找零问题中优先使用最大面额硬币以减少硬币总数;二是最优子结构性,即问题的最优解包含子问题的最优解,如最小生成树问题中每次选择最小边最终构成全局最优解。GBSA的高效性源于其仅关注当前状态(时间复杂度通常为 $O(n \log n)$ 或 $O(n^2)$),但局限性在于无法保证所有问题的全局最优性(如0/1背包问题)。由图12可知,本文方案相较GBSA总能耗平均降低了43.8%,总碳排放平均降低了44.1%;相较基于随机调度算法(RBSA, random based scheduling algorithm)^[43]总能耗平均降低了75.7%,总碳排放平均降低了76.1%。

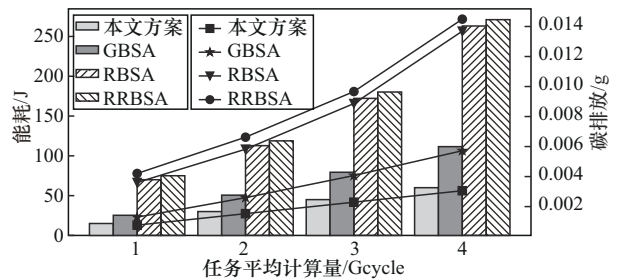


图12 本文方案与其他能碳联合优化算法的性能对比

轮询调度算法(RRBSA, round-robin based scheduling algorithm)^[44]是一种基于循环分配机制的负载均衡策略,其核心规则是通过固定顺序依次将任务或请求分配给可用资源(如服务器、处理单元等),确保资源分配的公平性与均匀性。具体实现时,系统维护一个自增指针变量(如currentIndex),每次调度时通过取模运算确定目标资源,例如,对于服务器列表[S1, S2, S3],请求按S1→S2→S3→S1→...的顺序分配,实现无状态循环遍历。该算法假设所有资源性能相同,适用于短任务、高并发的场景,但存在无法适应异构资源的问题。本文方案相较RRBSA,总能耗平均降低了76.91%,总碳排放平均降低了78.48%。

图 13 给出了本文方案与其他优化算法的时延对比。首先, RBSA 则展现出整体时延最高且波动性较高的特点。这可能是由于其随机选择任务执行节点的策略导致了系统资源的不均衡利用, 进而影响了系统的性能稳定性。另一方面, RRBSA 呈现出总时延始终保持在较高水平的趋势。这是因为在该方案下, 每个任务都是在所有算力节点上以轮循方式执行的, 且不需要记录当前所有连接的状态。这种方式导致了系统在不同计算量大小下都难以达到较低的总时延, 限制了其性能表现。GBSA 则获得了次优的性能, 尽管在某些情况下性能表现较好, 但无法保证长期优化或者在不同计算量情况下达到最优。与这些方案相比, 本文方案展现出了更为优越的性能。在所有计算量和数据量值下, 本文方案都能实现最低的总时延, 与 RBSA、RRBSA 和 GBSA 相比, 本文方案总时延分别平均降低了 20.13%、14.82% 和 3.49%。这表明本文方案在任务调度中能够有效地降低系统的总时延, 提高系统的性能表现和资源利用效率。

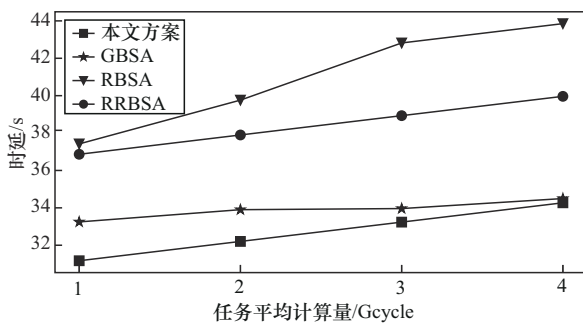


图 13 本文方案与其他优化算法的时延对比

5 结束语

面向能源领域新场景, 聚焦新业务对算力和网络的特殊要求, 本文设计了一种算网能一体化调度架构, 并基于该架构提出了一种全新的调度机制。该机制通过选择合适的算力节点与网络路径, 能够有效地降低任务调度过程的能耗与碳排放, 推动算力网络的绿色可持续发展。展望未来, 算网能一体化研究的推进需要全行业的共同努力。在具体研究方向上, 以下几方面值得关注: 首先, 算网能感知将依托大数据和机器学习技术, 实现对能源状态的精准预测与动态响应, 从而优化计算任务调度和能源利用效率; 其次, 算网能资源度量需综合评估能源消耗、任务执行效率及碳足迹等关键指标,

为系统优化提供量化依据; 再次, 算网能可信交易需通过身份认证、智能合约和隐私保护等手段, 确保交易的安全性与可靠性; 最后, 算网能数字孪生可借助实时数据更新和仿真模型, 支持系统运行优化与决策制定, 提升预测准确性与应对能力。

参考文献:

- [1] 贾庆民, 胡玉姣, 张华宇, 等. 确定性算力网络研究[J]. 通信学报, 2022, 43(10): 55-64.
JIA Q M, HU Y J, ZHANG H Y, et al. Research on deterministic computing power network[J]. Journal on Communications, 2022, 43(10): 55-64.
- [2] YUKUN S, BO L, JUNIIN L, et al. Computing power network: a survey[J]. China Communications, 2024, 21(9): 109-145.
- [3] TANG X Y, CAO C, WANG Y X, et al. Computing power network: the architecture of convergence of computing and networking towards 6G requirement[J]. China Communications, 2021, 18(2): 175-185.
- [4] LEI B, ZHAO Q Y, MEI J. Computing power network: an interworking architecture of computing and network based on IP extension[C]//Proceedings of the 2021 IEEE 22nd International Conference on High Performance Switching and Routing (HPSR). Piscataway: IEEE Press, 2021: 1-6.
- [5] 何涛, 杨振东, 曹畅, 等. 算力网络发展中的若干关键技术问题分析[J]. 电信科学, 2022, 38(6): 62-70.
HE T, YANG Z D, CAO C, et al. Analysis of some key technical problems in the development of computing power network[J]. Telecommunications Science, 2022, 38(6): 62-70.
- [6] The National Science and Technology Council (NSTC) Subcommittee on Future Advanced Computing Ecosystems. Pioneering the future advanced computing ecosystem: a strategic plan[R]. 2020.
- [7] The European Commission. 2030 digital compass: the European way for the digital decade[R]. 2021.
- [8] 国家发展改革委, 中央网信办, 工业和信息化部, 等. 全国一体化大数据中心协同创新体系算力枢纽实施方案[R]. 2021.
National Development and Reform Commission, Implementation plan of computing power hub of national integrated big data center collaborative innovation system[R]. 2021.
- [9] 国务院. “十四五”数字经济发展规划[R]. 2022.
The State Council. “14th Five Year Plan” for digital economy development[R]. 2022.
- [10] 国务院. 数字中国建设整体布局规划[R]. 2023.
The State Council. Overall layout plan for digital China construction[R]. 2023.
- [11] 李洁, 王月. 算力基础设施的现状、趋势和对策建议[J]. 信息通信技术与政策, 2022(3): 2-6.
LI J, WANG Y. The status, trends and suggestions of computing infrastructure[J]. Information and Communications Technology and Policy, 2022(3): 2-6.

- [12] 贾庆民, 丁瑞, 刘辉, 等. 算力网络研究进展综述[J]. 网络与信息安全学报, 2021, 7(5): 1-12.
JIA Q M, DING R, LIU H, et al. Survey on research progress for compute first networking[J]. Chinese Journal of Network and Information Security, 2021, 7(5): 1-12.
- [13] 中国移动. 算力网络安全一体化全程可信解决方案白皮书[R]. 2024. China Mobile. White paper on trusted solutions for computing network security integration[R]. 2024.
- [14] 中国联通研究院. 中国联通算力网络安全白皮书(2024)[R]. 2024. China Unicom Research Institute. China Unicom computing power network security white paper (2024)[R]. 2024.
- [15] 中国通信标准化协会. 算力网络总体技术要求[S]. 2021. China Communications Standards Association. General technical requirements of computing and network convergence[S]. 2021.
- [16] ITU-T. Y.2501: computing power network - framework and architecture[S]. 2021.
- [17] IETF. Computing-aware traffic steering(cats)[R]. 2023.
- [18] 郭凤仙, 闫实, 彭木根, 等. 无线算力网络: 架构与关键技术[J]. 移动通信, 2025, 49(3): 2-9.
GUO F X, YAN S, PENG M G, et al. The architecture and key technologies of radio computing power networks[J]. Mobile Communications, 2025, 49(3): 2-9.
- [19] 李一男, 唐琴琴, 彭开来, 等. 以服务为中心的算力网络度量与建模研究[J]. 信息通信技术与政策, 2023(5): 21-29.
LI Y N, TANG Q Q, PENG K L, et al. Research on measurement and modeling of service-centric computing power network[J]. Information and Communications Technology and Policy, 2023(5): 21-29.
- [20] 庞冉, 易昕昕, 辛亮, 等. 算力网络路由调度技术研究[J]. 电信科学, 2023, 39(8): 149-156.
PANG R, YI X X, XIN L, et al. Research on routing scheduling technology of computing power network[J]. Telecommunications Science, 2023, 39(8): 149-156.
- [21] CHEN Z Y, ZHANG S C, TANG Y, et al. Optimization of computing power network routing strategies based on multi-agent soft actor-critic [C]/Proceedings of the 2024 9th International Conference on Intelligent Computing and Signal Processing (ICSP). Piscataway: IEEE Press, 2024: 426-430.
- [22] 李重严, 毕成, 张晟. 面向信息能源融合的低碳算力网络架构研究[J]. 电信工程技术与标准化, 2022, 35(11): 1-6.
LI Z Y, BI C, ZHANG S. Low carbon computing network architecture for information energy collaboration[J]. Telecom Engineering Technics and Standardization, 2022, 35(11): 1-6.
- [23] 谢满德, 黄竹芳, 孙浩. 云边端协同下多用户细粒度任务卸载调度策略[J]. 电信科学, 2024, 40(4): 107-121.
XIE M D, HUANG Z F, SUN H. Multi-user fine-grained task offloading scheduling strategy under cloud-edge-end collaboration[J]. Telecommunications Science, 2024, 40(4): 107-121.
- [24] HU N, TIAN Z H, DU X J, et al. An energy-efficient in-network computing paradigm for 6G[J]. IEEE Transactions on Green Communications and Networking, 2021, 5(4): 1722-1733.
- [25] XIE Y, SUN Y Y, XU F, et al. The offloading algorithm of mobile edge computing considering mobility in the intelligent inspection scenario[J]. Transactions on Emerging Telecommunications Technologies, 2022, 33(7): e4494.
- [26] LIU W Y, YAN Y J, SUN Y M, et al. Online job scheduling scheme for low-carbon data center operation: an information and energy nexus perspective[J]. Applied Energy, 2023, 338: 120918.
- [27] HUANG X Y, LIU R R, LEI B, et al. Deep reinforcement learning based multistage profit aware task scheduling algorithm for computing power network[C]/Proceedings of the GLOBECOM 2023 - 2023 IEEE Global Communications Conference. Piscataway: IEEE Press, 2023: 3524-3529.
- [28] ZOU Z, XIE R C, REN Y Z, et al. Task scheduling for ICN-based computing first network: a deep reinforcement learning approach[C]/Proceedings of the 2022 IEEE 8th International Conference on Computer and Communications (ICCC). Piscataway: IEEE Press, 2022: 1615-1620.
- [29] 刘博文, 梁晓晨, 张桂玉, 等. 算力网络场景下SLA约束的能耗优化微服务调度策略[J]. 邮电设计技术, 2023(2): 31-36.
LIU B W, LIANG X C, ZHANG G Y, et al. Research on energy-optimized microservice scheduling strategies with SLA constraints in computing power network scenario[J]. Designing Techniques of Posts and Telecommunications, 2023(2): 31-36.
- [30] LIU Z T, QIU C, ZHAO Y F, et al. Bat-FG: a broad attention based fine-grained offloading in green computing power networks[C]/Proceedings of the ICC 2023 - IEEE International Conference on Communications. Piscataway: IEEE Press, 2023: 5117-5122.
- [31] ZHOU F, ZHAO L H, DING X M, et al. Enhanced DDPG algorithm for latency and energy-efficient task scheduling in MEC systems[J]. Discover Internet of Things, 2025, 5(1): 40.
- [32] FENG L, XIE R C, TANG Q Q, et al. Delay-prioritized task scheduling with load balancing in computing power networks[C]/Proceedings of the 2024 IEEE Wireless Communications and Networking Conference (WCNC). Piscataway: IEEE Press, 2024: 1-6.
- [33] DENG Y Q, CHEN Z G, CHEN X H, et al. Task offloading in multi-hop relay-aided multi-access edge computing[J]. IEEE Transactions on Vehicular Technology, 2023, 72(1): 1372-1376.
- [34] SUN Y P, XU J, CUI S G. User association and resource allocation for MEC-enabled IoT networks[J]. IEEE Transactions on Wireless Communications, 2022, 21(10): 8051-8062.
- [35] XIE R C, FENG L, TANG Q Q, et al. Delay-prioritized and reliable task scheduling with long-term load balancing in computing power networks[J]. IEEE Transactions on Services Computing, 2024, 17(6): 3359-3372.
- [36] MEI X X, WANG Q, CHU X W. A survey and measurement study of GPU DVFS on energy conservation[J]. Digital Communications and Networks, 2017, 3(2): 89-100.
- [37] HMIMZ Y, EL GHMARY M, CHANYOUR T, et al. Computation offloading to a mobile edge computing server with delay and energy constraints[C]/Proceedings of the 2019 International Conference on Wireless Technologies, Embedded and Intelligent Systems (WITS). Piscataway: IEEE Press, 2019: 1-6.
- [38] WEN W, XIE R C, TANG Q Q, et al. Spatiotemporal task scheduling for green computing in computing power networks[C]/Proceedings of

the GLOBECOM 2024 - 2024 IEEE Global Communications Conference. Piscataway: IEEE Press, 2024: 2449-2454.

- [39] GONZALEZ R, GORDON B M, HOROWITZ M A. Supply and threshold voltage scaling for low power CMOS[J]. IEEE Journal of Solid-State Circuits, 1997, 32(8): 1210-1216.
- [40] TANG Q Q, XIE R C, YU F R, et al. Decentralized computation offloading in IoT fog computing system with energy harvesting: a dec-POMDP approach[J]. IEEE Internet of Things Journal, 2020, 7(6): 4898-4911.
- [41] XU M X, BUYYYA R. Managing renewable energy and carbon footprint in multi-cloud computing environments[J]. Journal of Parallel and Distributed Computing, 2020, 135: 191-202.
- [42] SHIN H S, LI T, LEE H I, et al. Sample greedy based task allocation for multiple robot systems[J]. Swarm Intelligence, 2022, 16(3): 233-260.
- [43] JOŠILO S, DÁN G. Decentralized algorithm for randomized task allocation in fog computing systems[J]. IEEE/ACM Transactions on Networking, 2019, 27(1): 85-97.
- [44] ALHAIDARI F, BALHARITH T Z. Enhanced round-robin algorithm in the cloud computing environment for optimal task scheduling[J]. Computers, 2021, 10(5): 63.

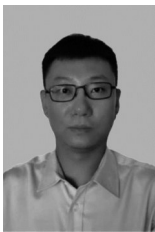
[作者简介]



罗必雄 (1966-), 男, 湖北荆州人, 中国电力工程顾问集团有限公司正高级工程师, 主要研究方向为煤炭清洁高效利用、多能互补、新能源与新型电力系统等能源新技术。



张力 (1966-), 男, 湖南长沙人, 中国电力工程顾问集团有限公司正高级工程师, 主要研究方向为火力发电及污染物治理、核能发电、新能源发电等。



句赫 (1981-), 男, 黑龙江肇东人, 中国电力工程顾问集团有限公司高级工程师, 主要研究方向为人工智能、算电协同等。



李舒涛 (1985-), 男, 江西南昌人, 博士, 中国电力工程顾问集团有限公司正高级工程师, 主要研究方向为数能融合、算电协同等。



黄韬 (1980-), 男, 重庆人, 博士, 北京邮电大学教授, 主要研究方向为网络系统架构、算网融合、确定性网络等。



谢人超 (1984-), 男, 福建南平人, 博士, 北京邮电大学教授、博士生导师, 主要研究方向为信息中心网络、工业互联网、算力网络、边缘计算、无服务器计算等。



唐琴琴 (1994-), 女, 广西桂林人, 博士, 北京邮电大学副研究员、硕士生导师, 主要研究方向为算力网络、网络人工智能、网络孪生等。



彭开来 (1987-), 男, 江苏盐城人, 紫金山实验室研究员, 主要研究方向为标识解析、边缘计算、时间敏感网络、大数据等。



文雯 (2001-), 女, 安徽宿州人, 北京邮电大学博士生, 主要研究方向为算力网络、边缘计算、区块链等。