

算力网络研究进展: 架构、关键技术与未来挑战

赵宝康, 时维嘉, 周寰, 孙薛雨

(国防科技大学 计算机学院, 长沙 410073)

摘要: 随着数字经济的蓬勃发展, 算力已成为推动经济增长的核心动力。算力网络通过整合云、边、端的计算资源, 实现资源的高效管理和调度, 以支持多样化的数字应用需求。对算力网络进行全面综述, 首先介绍了算力网络的概念和重要性, 随后分析了其架构设计、算力资源度量和算力网络调度等关键技术。同时, 探讨了算力网络在实现过程中遇到的挑战, 包括技术融合、安全风险和资源优化等问题。最后, 提出了算力网络未来的研究方向, 旨在推动算力网络的进一步发展, 为构建智能世界提供坚实的技术基础。

关键词: 算力网络; 架构设计; 算力资源度量; 算力网络调度; 安全风险

中图分类号: TP 391.41 **文献标志码:** A

Research progress for computing force network: architecture, key technologies, and future challenges

ZHAO Baokang, SHI Weijia, ZHOU Huan, SUN Xueyu

(College of Computer Science and Technology, National University of Defense Technology, Changsha 410073, China)

Abstract: With the vigorous development of the digital economy, computing power has become the core driving force for economic growth. The computing force network integrates computing resources from the cloud, edge, and end to achieve efficient management and scheduling of resources to support diverse digital application requirements. An overview of the computing force network was provided. Firstly, the concept and importance of computing force network were introduced, and then the key technologies such as architecture design, computing resource measurement and computing network scheduling were analyzed. At the same time, the challenges encountered in the implementation of computing force network were discussed, including technology integration, security risks and resource optimization. Finally, the future research direction of computing force network was proposed, aiming to promote the

收稿日期: 2024-10-23

基金项目: 国家自然科学基金联合重点项目 (U22B2005, 62102434)

通信作者: 赵宝康(1981-), 男, 副教授。研究方向: 计算机网络、智能计算、空天网络和网络安全。E-mail: bkzhao@nudt.edu.cn

引文格式: 赵宝康, 时维嘉, 周寰, 等. 算力网络研究进展: 架构、关键技术与未来挑战[J]. 上海理工大学学报, 2024, 46(6): 600-609.

Citation: ZHAO Baokang, SHI Weijia, ZHOU Huan, et al. Research progress for computing force network: architecture, key technologies, and future challenges[J]. Journal of University of Shanghai for Science and Technology, 2024, 46(6): 600-609.

further development of computing force network and provide a solid technical foundation for building an intelligent world.

Keywords: *computing force network; architecture design; computing power resource measurement; computing force network scheduling; security risks*

数字经济时代, 算力已经成为拉动国家经济增长的核心引擎。2023年发布的《中国算力发展指数白皮书》^[1]指出, 当前, 算力的战略地位和支撑性作用正成为普遍共识。算力作为数字经济时代的核心生产力, 正在加速数字经济与实体经济深度融合。全球算力应用多元化、供需不平衡的挑战仍在持续, 以生成式人工智能(artificial intelligence generated content, AIGC)为代表的人工智能大模型等新应用、新需求的崛起, 推动算力规模快速增长、计算技术多元创新、产业格局加速重构。《2021—2022全球算力指数评估报告》^[2]指出, 随着全球数字经济持续稳定增长, 数字经济占比预计到2025年有望达到41.5%。同时, 国家算力指数与GDP的走势呈现出了显著的正相关。15个重点国家的算力指数平均每提高1点, 国家的数字经济和GDP将分别增长3.5%和1.8%, 预计该趋势在2021—2025年将继续保持。并且, 当一个国家的算力指数达到40分以上时, 国家的算力指数每提升1点, 对GDP增长的推动力将增加1.5倍。而当算力指数达到60分以上时, 国家的算力指数每提升1点, 对GDP增长的推动力将提高3倍, 对经济的拉动作用更加显著。

随着算力作用愈发凸显, 算力需求也逐渐高涨。据华为发布的《计算2030》^[3]预测, 2030年, 人类将进入YB数据时代, 全球数据每年新增1YB, 通用算力将增长10倍到3.3ZFLOPS, 人工智能算力将增长500倍, 超过100ZFLOPS。当前, 算力出现云、边、端三级算力架构的泛在演进趋势, 算力的分布将不再集中在数据中心, 而是广泛地分布在边缘或者端侧的任何位置。如果所有的这些算力节点之间没有通过网络互连, 则这些算力资源无法被共享、调度、使用和协同。正如《算力网络白皮书》^[4]中提到的, 水力发展离不开水网, 电力发展离不开电网, 算力发展离不开“算力网络”(computing force network, CFN)。算力网络是新型基础设施, 要打造“一点接入、即取即用”的社会级服务, 最终实现“网络无所不达、算力无所不在、智能无所不及”。

如图1所示, 算力网络通过无处不在的网络, 能统一管理 and 调度来自各方的计算、存储等资源, 以资源池的形式向外提供算力服务, 用户可以灵活地按照业务需求选择合适的算力资源。实际上, 算力网络就是一种根据业务需求, 在云、边、端之间按需分配和灵活调度计算资源、存储资源以及网络资源的新型信息基础设施。

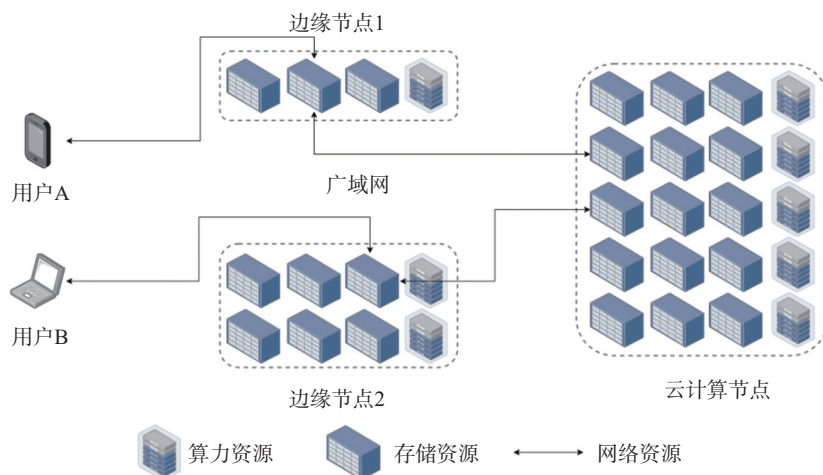


图1 算力网络服务示意图

Fig.1 Schematic diagram of computing force network services

2020年12月与2021年5月,国家发展改革委、国家网信办、工业和信息化部、国家能源局4部委先后联合发布了《关于加快构建全国一体化大数据中心协同创新体系的意见》和《全国一体化大数据中心协同创新体系算力枢纽方案》两份文件,明确指出要大力开展“东数西算”示范工程,强化算力统筹调度。在“东数西算”工程的实施过程中,通过融入算网融合相关技术,将算力请求调度至最佳算力资源节点,可有效促进东西部地区的算力协同,推动国家整体战略布局的稳步发展^[5]。

基于算网融合和“东数西算”工程的背景,中国信息通信研究院、华为技术有限公司和3大运营商牵头,重点研究利用网络来调配分布在各处的算力资源。算力网络基于网络连接,将动态分布的计算资源互联,通过对计算、网络、存储等多维资源的灵活调度,实现连接以及算力在网络中的全局优化,为用户和运营商的计算产业发展带来新的机遇^[6]。未来,全国甚至全球的算力网络,将帮助人类跨入智能世界,通过新型网络技术将地理分布的算力中心节点连接起来,动态实

时感知算力资源状态,进而统筹分配和调度计算任务,传输数据,构成全局范围内感知、分配、调度算力的网络,在此基础上汇聚和共享算力、数据、应用资源。

算力网络作为正在蓬勃发展的新兴事物,早些年工业界和学术界已有部分关于基本概念、基本架构和未来展望的相关研究^[7-8]。然而,近年来技术日新月异,如人工智能等新型技术的出现极大地推动了算力网络的发展,很多综述缺乏对如算力度量这种细节性领域的总结。因此,本文对算力网络的架构、关键技术与未来挑战进行全面的总结和分析。

1 算力网络发展历程

算力网络的概念被提出后,华为和3大运营商牵头,多次召开会议和论坛对算力网络的架构及关键技术进行探讨,有关算力网络讨论的重要时间节点如表1所示。

从表1可以看出,行业内关于计算能力网络的探索活动频率较高,表现出算力网络在未来网

表1 算力网络发展的重要节点

Tab.1 Important milestones in the development of computing force network

组织	时间	标准进程/活动情况	相关单位
CCSA	2019年8月	开展算力网络需求与架构的研究	中国移动、中国联通、中国电信、华为等
GNTC2019	2019年10月	发表《算力网络需求和构想》和《算力网络技术》报告	华为
ITU-T SG13	2019年10月	提出算力网络顶层架构标准“Framework and architecture of Computing Power Network”	中国电信、中国联通和华为等
ECIS2019	2019年11月	发布《算力感知网络技术白皮书》	中国移动
IETF	2020年3月	发布COIN(Compute in Network)系列草案	COINRG(Computing in Network Research Group)
ETSI	2020年5月	“NFV support for network function connectivity extensions”(NFV-EVE2020)	华为、中国电信、Telefónica(西班牙电信)、Verizon(威瑞森通信公司)
中国云网智大会	2020年9月	提出以微服务为代表的轻量化部署架构,提供技术手段	中国联通、中国电信
PTEXPO2020	2020年10月	发布《中国联通算力网络架构与技术体系白皮书》,阐述服务层的关键技术	中国联通、中国电信、华为
CUBE-Net 3.0	2021年3月	讨论算网融合的技术挑战,提出算力服务的控制面和转发面技术	中国联通
中国移动全球合作伙伴大会	2021年11月	发布《中国移动算力网络白皮书》	中国移动
中国通信学会	2022年11月	中国通信学会算力网络委员会成立	中国通信学会
CCF青年精英大会(YEF2023)	2023年5月	YEF 2023专题论坛《算力网络:激发算力服务新动能》	广东联通、中国移动研究院、网络通信与安全紫金山实验室等
世界互联网大会	2024年2月	世界互联网大会国际组织以“算力网络:智能网络赋能智慧世界”为主题举办“算力网络”专题论坛	中国国家互联网信息办公室、全球移动通信系统协会、国际电信联盟、阿拉伯信息通信技术组织、中国电信、中兴通讯等

络领域中的重要地位。算力网络的讨论重点已从最初的需求分析转向架构设计, 目前已进入关键技术的研究阶段。

2 算力网络相关工作

有关算力网络的研究工作也在不断更新, 主要研究方向包括架构设计、算力资源度量及算力网络调度 3 个方面。

2.1 架构设计

算力网络被提出以来, 相关学者就开展了对算力网络的架构设计研究工作。

算力网络的核心思想是将泛在分布的算力节点连接起来, 动态实时感知网络中的计算资源和网络资源状态, 进而统筹协调分配和调度计算任务, 形成一张计算资源可感知、可分配、可调度的网络, 满足新业务新应用对算力的需求, 是一种云、边、网深度融合的新范式, 也是边缘计算向泛在计算网络融合演进的新阶段。基于此核心

思想, 结合边缘计算节点状态的实时感知和边缘计算资源的分布式协同处理及调度这两个设计约束, 贾庆民等^[7]进行了算力网络基本架构的设计, 将算力网络节点、算力网络服务和算力网络适配器 3 个部分作为算力网络的主要构成, 如图 2 所示。该架构可根据计算任务的要求, 结合实时的计算负载和网络状态条件, 动态地将计算任务调度到最匹配的边缘计算节点, 从而提高计算资源利用率和用户体验。但其主要关注点在于边缘计算资源的分布式协同处理和调度, 缺少对云等集中式资源的协同考虑, 以及对“算网协同”的考量。

算力网络研究不仅要考虑算力, 更要关注新型网络, 算力与网络不能只是“算中有网”或“网中有算”的分离式协同, 而是要实现“算力+网络”的融合突破。基于算力网络与新型网络研究相辅相成的目标, 张宏科等^[8]提出了“三层三域”算力网络架构, 如图 3 所示。在该架构中, “三层”包括广义服务层、映射适配层、融

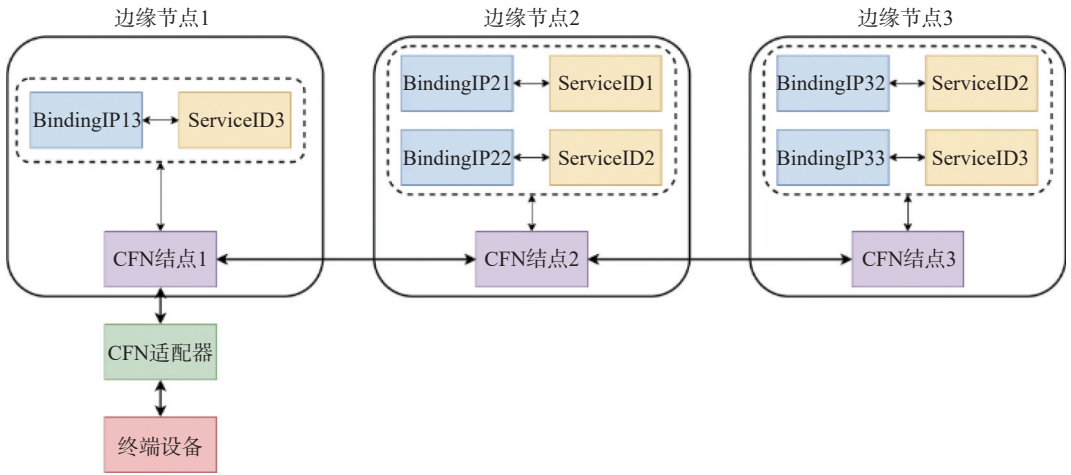


图 2 算力网络基本架构^[7]
Fig.2 Basic architecture of CFN^[7]

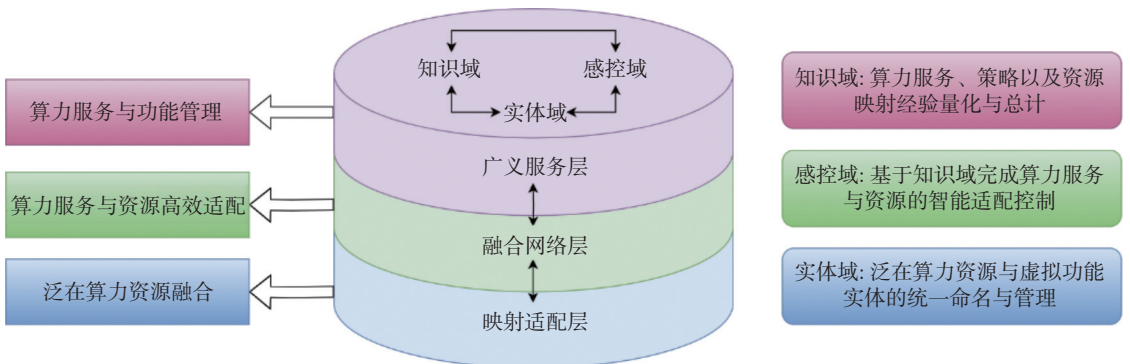


图 3 “三层三域”算力网络架构^[8]
Fig.3 “3 Layer 3 Domain” architecture of CFN^[8]

合网络层，“三域”包括实体域、感控域、知识域，突出了“算网协同”的特点。

除了上述对算力网络基本框架的相关研究，目前也存在一些针对具体场景融合的算力网络架构被提出。Liu等^[9]提出了CFN-dyncast，一种创新的分布式负载均衡技术，旨在优化5G网络中的多接入边缘计算(MEC)环境。该技术突破了单个MEC站点的限制，动态地将客户端请求根据计算站点的实时负载和网络状况分配到最合适的站点，实现了计算资源的全局优化。与传统的基于DNS的负载均衡方法相比，CFN-dyncast不仅减少了平均作业完成时间，还通过会话亲和性机制确保了用户体验的连贯性。尽管这项技术在实验室环境中表现出色，但在更广泛的网络条件下的适应性和实时响应能力仍需进一步研究。Król等^[10]提出的分布式计算框架同样关注边缘计算和分布式计算的网络技术。该框架通过与信息中心网络(information-centric networking, ICN)的紧密结合，实现了对分布式程序的高效管理和优化。CFN利用了ICN的命名和寻址机制，通过CRDTs(conflict-free replicated data types)来维护一个分布式的计算图，这个计算图能够实时反映程序的执行状态和数据流动。此外，CFN采用了RICE(remote method invocation for ICN)作为其执行环境，这使得它能够在不同的网络条件下动态地调度计算任务，以实现负载均衡和性能优化。通过ndnSIM模拟的评估表明，CFN在处理分布式计算任务时，不仅能够提供简洁的编程模型，还能够在保证计算任务的可靠性和容错性的同时，显著提高系统的整体性能。

陈晓等^[11]提出了在微服务集群架构下，通过增强IPv6段路由(SRv6)和虚拟可扩展局域网(VxLAN)技术，实现端到端的算力路由解决方案，以支持算力路由穿透集群L4~L7代理节点并进入服务级颗粒度的场景。该研究在实现与现网平滑兼容的同时，构建了基于通信网络的算力和网络资源全网视图，并以此作为全新的业务交付平台，提升了算力和网络资源效率，为行业提供丰富、高效的算网融合业务能力。然而，虽然该研究提供了技术层面的深入分析，但在实际部署和运营挑战方面的讨论并不够充分。

雷波等^[12]分析了网络虚拟化的特点和发展、各类资源的互通趋势和发展，以及新兴业务对资

源供给的新要求，并在此分析的基础上，提出了一种基于多维资源互通的网络虚拟化架构。即通过引入资源发现和资源交易的过程，将多方异构资源整合为统一的资源平面，然后根据用户需求进行资源切片和虚拟化，以满足新兴业务灵活的资源需求，有效提高资源利用效率，降低运营成本。

以上研究并非对算力网络基本结构的研究，而是针对例如5G网络、信息中心网络和微服务集群等具体场景下的算力网络新型架构进行了相关研究，共同探讨了如何更有效地管理和利用网络资源，以适应新兴业务的需求，并认识到了网络技术发展中的资源泛在化、融合化和智能化趋势。它们为网络技术的研究提供了不同的视角和解决方案，相互之间可以作为补充，共同推动网络技术的发展。

2.2 算力资源调度

算力网络期望能够提供算网一体化的算力服务，统筹调度算力资源和网络资源，优化资源配置，充分利用有限资源为用户提供更高质量的算力服务。但是，不同业务的算力需求不同，不同算力节点的服务能力也各不相同。要想实现灵活高效的资源配置，就需要能够进行动态精准的算力感知，而算力度量就是算力感知的基础之一。

杜宗鹏等^[13]建立了一个全面的算力度量技术体系——“四面三级”算力度量技术体系，如图4所示。该体系从计算、通信、内存和存储能力4个方面，按照三级指标的方式来建模和描述算力网络中异构立体泛在的算力资源，从而实现算力网络中算力资源的合理分配和优化使用。

数字经济时代背景下，算力作为关键生产力的评价指标，对于准确度量计算平台的计算能力并实现算力网络中资源的有效盘活和高效调度具有重要意义^[14]。基于算力的定义，乔楚^[14]分析了算力网络体系架构及其面临的风险，提出了合理的算力度量方案和算力感知与编排的可行思路，所用指标均为静态指标，如逻辑运算能力、并行计算能力、神经计算能力、存力、算法能力、路由协议和算效等。与此不同，孙钰坤等^[15]提出了一种基于Floyd算法的面向6G网络愿景的算力感知网络中的智能任务调度模型，主要度量指标是2项算力节点的动态性能指标——CPU转数和存储剩余量。Zhao等^[16]则使用CPU利用率和存储剩余量这2项动态指标来进行算力度量。陈星延等^[17]

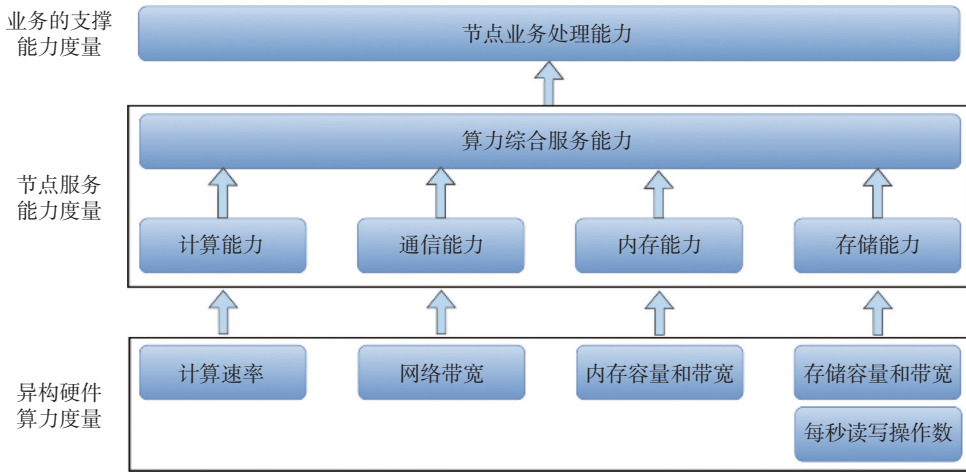


图 4 “四面三级” 算力度量技术体系^[13]

Fig.4 Four-aspect and three-level computing^[13]

提出了一种面向“云-边-端”算力网络的计算和传输联合优化方案, 从应用服务模型、网络状态感知和资源联合优化方面考虑建立一套系统性解决方案。文献 [14-16] 仅使用纯动态或纯静态指标, 无法全面高效地进行算力度量。而柴若楠等^[18]结合静态和动态指标来全面评估算力资源, 提出了一种混合式度量方法 (hybrid metric method, HMM), 通过熵权法^[19]分析了算力节点的静态性能, 利用基于决策树的 CART 算法^[20]对静态性能进行分类, 再通过 n 维欧氏距离^[21]计算用户任务与算力节点之间的距离, 从而选择最符合用户任务需求的算力节点。实验和数据分析表明, 所提度量方法综合考虑了各类经典方法及其相关指标^[22-29], 能有效提升算力资源的利用率和匹配准确率。

Li 等^[30]提出了一种基于极值理论构建算力资源的性能不确定性模型, 该模型可以对算力资源进行度量, 生成适当的阈值来保证应用程序的服务质量, 并借助于 CloudsStorm^[31]采集数据验证了其效果。

2.3 算力网络调度

算力服务为每个用户请求提供具有不同成本的算力资源, 而算力网络调度则是通过各种调度算法来实现用最少资源获得最大用户满意度。假设任务集为 $T = \{t_1, t_2, \dots, t_m\}$, 资源集为 $R = \{r_1, r_2, \dots, r_m\}$, 任务和资源的匹配就是要寻求最佳的二元组集合 $\{(t_i \in T, r_j \in R) | \forall i \in \{1, \dots, n\}; j \in \{1, \dots, m\}\}$, 也就是为每个任务分配最合适的资源, 使得用户满意度^[32-34]和服务资源利用率^[35-37]最大化。

李逸博等^[38]定义任务调度为给排好序的候选任务指定服务资源、资源量、开完工时间等, 并

将其通用模型定义如下:

$$\min \mathbf{F}(\bar{x}) = (f_1(\bar{x}), \dots, f_p(\bar{x}))^T \quad (1)$$

s.t.

$$g_i(\bar{x}) < 0, i \in \{1, \dots, k\} \quad (2)$$

$$h_i(\bar{x}) = 0, i \in \{k+1, \dots, m\} \quad (3)$$

式中: $\mathbf{F}(\bar{x})$ 为目标函数向量, 即需要优化的 p 个目标, 优化目标通常包括任务最大完工时间最短、资源利用率最大和用户满意度最高等; $g_i(\bar{x}) < 0$ ($i \in \{1, \dots, k\}$) 和 $h_i(\bar{x}) = 0$ ($i \in \{k+1, \dots, m\}$) 为约束条件, 通常来源于需求方和资源方, 如需求截止日期、响应时间、需求预算和资源租赁方式等, 共分为需求约束和资源约束两个方面; $\bar{x} \in \mathbb{R}^n$ 为决策变量。

$$Q \propto A(I, t) \quad (4)$$

同时, 李逸博等^[38]还提出了如式(4)所示的计算时间 t (效率) 和求解质量 Q (性能) 之间的关系式, 式中, A 为算力网络调度算法, I 为输入实例。

算力网络调度算法相关的论文发表情况如图 5 所示, 研究方向主要集中在资源负载均衡、资源动态分配和数据传播路由选择等。郝昊等^[39]和 Xie 等^[40]均提出了基于深度强化学习的解决方案。不同的是, 前者在保证时间轴连续和协同多个边缘节点计算资源的前提下, 以服务体验提升率为优化目标, 对云、边、端间任务卸载问题进行建模, 并设计了一种基于深度强化学习的任务卸载方法, 从而更高效地利用算力网络计算资源。而后者则聚焦于边缘计算系统的联合任务调度和负载均衡框架, 通过 twin delayed deep deterministic

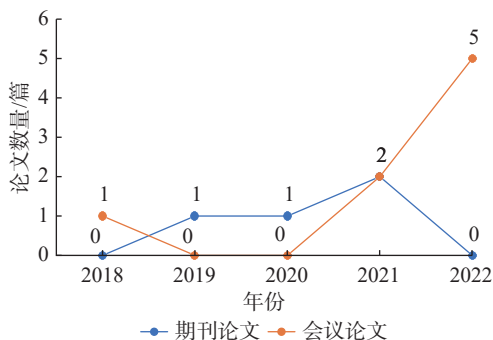


图5 算力网络调度算法论文发表情况

Fig.5 Analysis of the publication situation of computing force network scheduling algorithms

policy gradient(TD3)算法提供最优的调度决策。

在动态资源分配方面, Lei等^[41]提出了一种基于IP扩展的计算网络互联体系结构,通过动态网络编排为用户应用程序分配计算资源和网络连接方式。He等^[42]则将通信场景中的计算资源调度问题建模为混合整数问题,并证明了数字二维通信技术可以提升蜂窝网络的计算潜力。此外, Liu等^[43]和 Zhang等^[44]也分别提出了基于交叉熵的分布式任务调度模型和基于强化学习的分布式任务调度方法,以优化边缘系统性能和跨运营商竞争环境下的服务和计算资源控制。

在数据传播路由选择方面, Pang等^[45]提出了一种具有加权唤醒路由惩罚的选播路由算法,降低了算力网络的能耗。Zhang等^[46]则提出了一种基于“段标识符即服务”的可编程服务系统,通过网络控制层对网络节点和链路的权重进行计算以指导数据包的转发,增强了可编程网络的灵活性。Yao等^[47]通过整合计算信息和网络代价,实现了服务请求的最优联合调度,增强了计算与网络的协同效率。

在算力网络架构与资源优化算法方面, Liu等^[48]提出了一种边缘数据中心与内容联合放置算法,利用 Kuhn-Munkres 算法和 K-Means 算法对边缘数据中心和内容的放置进行优化。Zou等^[49]则提出了一种基于信息中心网络的算力网络架构,并通过强化学习算法进行任务调度。Zhao等^[50]设计的算力网络联邦学习平台在调度时为确保通信安全,采用加密技术保护数据传输的可信度。Wang等^[51]通过激励理论设计定价机制,激励边际价值较高的任务请求更多资源,以提高边缘计算平台的效用。最后, Dong等^[52]提出了一种基于 Bkd-tree 的算力网络服务优化调度算法,以提高用户

满意度和调度效率。

算力网络调度领域的研究贡献显著,通过引入人工智能方法和边际价值等先进的算法和技术,有效解决了算力网络中的资源管理和任务调度难题。特别是深度强化学习等人工智能技术的应用,为算力网络的智能化调度提供了有力支持。未来,随着算力网络规模的不断扩大和复杂度的提升,如何进一步优化资源调度算法、提高资源利用率和服务质量,将是该领域的重要研究方向。同时,加强算力网络在调度过程中的安全性和可靠性,以及探索更加高效、灵活的网络架构和调度策略,也将是未来研究的重点。

2.4 算力网络相关数据集

算力网络相关学者提供了丰富数据集,便于进行相关分析研究。

Shen等^[53]对大型众包边缘平台进行了首次服务质量(QoS)测量,该平台涵盖超过10000台边缘服务器、100000个用户和10000000个用户请求,然后,基于Python实现了完整的建模生成器,并提供了必要的定量数据来支持它,使得用户仅需要输入任意地区人口数量等关键信息即可获得该地区边缘服务器属性及分布、容器化服务属性以及用户请求情况信息。Zhang等^[54]在全国边缘计算平台上从空间和时间维度两个方面深入研究了网络服务质量,通过测量收集了一个多变量的大规模真实延迟数据集。Huang等^[55]则提供了丰富的边缘云服务器工作负载数据并提出了一种具有全局池化和静态内容感知功能的端到端框架 DynEformer,为动态 MT-ECP 提供统一的工作负载预测方案。

3 未来研究方向和挑战

当前,算力网络引起了学术界和产业界的高度关注。尽管关于算力网络方面的研究与实践已经取得了许多进展,但其未来发展依然存在许多未解决的问题和挑战。本节对算力网络未来研究方向和挑战进行简要梳理分析。

算力网络可以与多种技术结合,形成多样化的算力网络^[7]。结合软件定义网络(SDN, software-defined networking)技术^[56],构建软件定义算力网络,将计算资源的感知和计算任务的调度交由SDN控制平台进行集中式处理,使得网络配置和管理

变得非常敏捷高效。结合信息中心网络技术^[57]构建命名算力网络。借助于 ICN 的命名寻址机制,在网络架构层协同融合算力与内容,提升算力服务和内容分发效率。结合区块链技术可以构建可信算力网络,通过区块链的分布式可信机制,将提供计算资源的各异构节点或设备统一纳入算力感知和管理中,从而保证各计算资源提供方的利益以及用户的业务安全。结合确定性网络技术可以构建超低时延算力网络,改变传统的“尽力而为”的传输方式来满足超低时延业务的需求。结合节能优化技术可以构建绿色算力网络,从计算服务实例的放置、合理选择并构建算力网络计算载体,以及对计算量巨大的任务进行分割和高效资源分配等方面开展研究,从而达到节约能耗和优化资源配置的目的。基于 SDN、NFV、AI、云计算等新型技术的试验环境,并且结合 AI 赋能平台,形成了旨在满足 AI 应用需求的算力调度系统,即 AI 算力网络^[58]。

全行业数字化转型的加速对算网一体化基础设施和融合人工智能、大数据等多要素融合服务能力的要求日益提升。当前算网各自编排、分域管理,难以提供算网融合的产品、服务和端到端的质量保障。因此,“算网大脑”通过算网数据感知获取全域实时动态数据,结合算网智能化、多要素融合编排实现要素能力的一体供给和智能匹配,为新型信息基础设施对外提供一体化服务提供能力支撑。但这也同时可能使得“算网大脑”成为整个算力网络的瓶颈,设计并优化“算网大脑”使其能够更好地服务于算力网络是一个重要课题。

传统的算力和网络相对独立,二者仅为简单的连接关系。但随着软件定义网络和网络虚拟化为核心技术的下一代网络规模部署,算力和网络开始在基础设施层面逐步融合。未来,算网将在协议和形态层面进一步融合,网络将深度感知算力,通过在网络协议中引入算力信息将应用请求沿最优路径调度至算力节点。

随着算网融合工作的持续推进,当中存在的一些安全风险也逐渐显现,主要体现在攻击暴露面增加、数据隐私泄露、存证溯源困难和管控复杂度提升等方面^[59]。因此,为确保整个算力网络体系的安全可靠,需要在物理安全防护、系统安全加固、网络访问控制、应用安全防护以及安全

管理等方面进行全面的安全保障,以此来保障整个算力网络自下而上包括物理设施、软件系统、网络架构、系统平台及应用服务等功能组件的安全^[60]。

4 结束语

算力网络作为连接数字经济与智能世界的重要纽带,其发展不仅对计算资源的高效利用至关重要,也是推动社会全面数字化转型的关键。本文全面分析了算力网络的构建、关键技术、面临的挑战及未来发展方向,揭示了其在促进经济发展、提升技术创新能力、优化资源配置和保障网络安全等方面的重要作用。面对未来,算力网络的研究和应用需要全球范围内的合作与创新,以解决跨学科、跨领域的技术难题。我们期待算力网络能够在全球数字化进程中发挥更大的作用,为构建一个更加智能、高效、安全和可持续的未来世界提供坚实的技术支撑。随着技术的不断进步和应用的深入,算力网络必将成为推动全球经济社会发展的重要力量。

参考文献:

- [1] 中国信息通信研究院. 中国算力发展指数白皮书(2023年)[EB/OL]. (2023-09)[2023-09-15]. <http://www.caict.ac.cn/english/research/whitepapers/202311/P020231103309012315580.pdf>.
- [2] 清华大学全球产业研究院. 《2021-2022 全球计算力指数评估报告》正式发布[EB/OL]. (2022-04-01)[2022-04-01]. <https://www.igi.tsinghua.edu.cn/info/1019/1223.htm>.
- [3] HUAWEI. 计算 2030[EB/OL]. https://www-file.huawei.com/-/media/corp2020/pdf/giv/industry-reports/computing_2030_cn.pdf.
- [4] 中国移动通信集团有限公司. 算力网络白皮书[EB/OL]. (2021-11)[2021-11-03]. <https://cmri.chinamobile.com/wp-content/uploads/2021/11/算力网络白皮书.pdf>.
- [5] 雷波. “东数西算”推动网络技术演进的探讨[J]. 通信世界, 2021, (15): 45-48.
- [6] 雷波. 整合多方资源算力网络有望实现计算资源利用率最优[J]. 通信世界, 2020, (8): 39-40.
- [7] 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.
- [8] 张宏科, 权伟, 刘康. 算力网络研究与探索[J]. 中兴通讯技术, 2023, 29(1): 1-5.
- [9] LIU B, MAO J W, XU L, et al. CFN-dyncast: load balancing the edges via the network[C]//Proceedings of 2021 IEEE Wireless Communications and Networking

- Conference Workshops. Nanjing: IEEE, 2021: 1–6.
- [10] KRÓL M, MASTORAKIS S, ORAN D, et al. Compute first networking: distributed computing meets ICN[C]//Proceedings of the 6th ACM Conference on Information-Centric Networking. Macao, China: ACM, 2019: 67–77.
- [11] 陈晓, 黄光平. 微服务架构下的算力路由技术 [J]. *中兴通讯技术*, 2022, 28(1): 70–74.
- [12] 雷波, 王江龙, 赵倩颖, 等. 基于计算、存储、传送资源融合化的新型网络虚拟化架构 [J]. *电信科学*, 2020, 36(7): 42–54.
- [13] 杜宗鹏, 李志强, 陆璐. 算力网络四面三级算力度量技术体系 [J]. *中兴通讯技术*, 2023, 29(4): 8–13.
- [14] 乔楚. 算力度量与算网资源调度思路分析 [J]. *通信技术*, 2022, 55(9): 1165–1170.
- [15] 孙钰坤, 张兴, 雷波. 边缘算力网络中智能算力感知路由分配策略研究 [J]. *无线电通信技术*, 2022, 48(1): 60–67.
- [16] ZHAO Y H, CHONG Z, HAN X Y, et al. Simulation study of routing mechanism in the computing-aware network[C]//Proceedings of the 2021 10th International Conference on Networks, Communication and Computing. Beijing: ACM, 2021: 126–134.
- [17] 陈星延, 张雪松, 谢志龙, 等. 面向“云-边-端”算力系统的计算和传输联合优化方法 [J]. *计算机研究与发展*, 2023, 60(4): 719–734.
- [18] 柴若楠, 郜帅, 兰江雨, 等. 算力网络中高效算力资源度量方法 [J]. *计算机研究与发展*, 2023, 60(4): 763–771.
- [19] 程启月. 评测指标权重确定的结构熵权法 [J]. *系统工程理论与实践*, 2010, 30(7): 1225–1228.
- [20] QUINLAN J R. C4.5: programs for machine learning[M]. Amsterdam: Elsevier, 1993.
- [21] CARBÓ-DORCA R, BESALÚ E. Geometry of n-dimensional Euclidean space Gaussian enfoldments[J]. *Journal of Mathematical Chemistry*, 2011, 49(10): 2244–2249.
- [22] EMMA P G. Understanding some simple processor-performance limits[J]. *IBM Journal of Research and Development*, 1997, 41(3): 215–232.
- [23] SHEKOFTEH S K, NOORI H, NAGHIBZADEH M, et al. Metric selection for GPU kernel classification[J]. *ACM Transactions on Architecture and Code Optimization (TACO)*, 2018, 15(4): 68.
- [24] 王海峰, 陈庆奎. 多指标自趋优的 GPU 集群能耗控制模型 [J]. *计算机研究与发展*, 2015, 52(1): 105–115.
- [25] ZOU Q. An analytical performance and power model based on the transition probability for hard disks[C]//Proceedings of 2011 3rd International Conference on Awareness Science and Technology. Dalian: IEEE, 2011: 111–116.
- [26] MASTELIC T, BRANDIC I, JAAREVIC J. CPU performance coefficient (CPU-PC): a novel performance metric based on real-time CPU resource provisioning in time-shared cloud environments[C]//Proceedings of 2014 IEEE 6th International Conference on Cloud Computing Technology and Science. Singapore: IEEE, 2014: 408–415.
- [27] OWENS J D, HOUSTON M, LUEBKE D, et al. GPU computing[J]. *Proceedings of the IEEE*, 2008, 96(5): 879–899.
- [28] BINH N N, IMAI M, TAKEUCHI Y. A performance maximization algorithm to design ASIPs under the constraint of chip area including RAM and ROM sizes[C]//Proceedings of 1998 Asia and South Pacific Design Automation Conference. Yokohama: IEEE, 1998: 367–372.
- [29] 刘铂熙. 缓存、计算资源受限下无线网络吞吐量优化策略研究 [D]. 武汉: 华中科技大学, 2020.
- [30] LI M J, SU J S, LIU H Y, et al. The extreme counts: modeling the performance uncertainty of cloud resources with extreme value theory[C]//Proceedings of the 20th International Conference on Service-Oriented Computing. Seville: Springer, 2022: 498–512.
- [31] ZHOU H, HU Y, OUYANG X, et al. CloudsStorm: a framework for seamlessly programming and controlling virtual infrastructure functions during the DevOps lifecycle of cloud applications[J]. *Software: Practice and Experience*, 2019, 49(10): 1421–1447.
- [32] KIM J, MOHAN K, RAMESH B. Functional and nonfunctional quality in cloud-based collaborative writing: an empirical investigation[J]. *IEEE Transactions on Professional Communication*, 2014, 57(3): 182–203.
- [33] SITHIPOLVANICHGUL J, CHEN C, LAND J, et al. Enhancing user experiences with cloud computing via improving utilitarian and hedonic factors[J]. *Energies*, 2021, 14(7): 1822.
- [34] SUN Z M, SUN G, LIU Y H, et al. BARGAIN-MATCH: a game theoretical approach for resource allocation and task offloading in vehicular edge computing networks[J]. *IEEE Transactions on Mobile Computing*, 2024, 23(2): 1655–1673.
- [35] SU Q, ZHANG Q H, LI W D, et al. Primal-dual-based computation offloading method for energy-aware cloud-edge collaboration[J]. *IEEE Transactions on Mobile Computing*, 2024, 23(2): 1534–1549.
- [36] LIN Z, LU L M, SHUAI J P, et al. An efficient and autonomous planning scheme for deploying IoT services in fog computing: a metaheuristic-based approach[J]. *IEEE Transactions on Computational Social Systems*, 2024, 11(1): 1415–1429.
- [37] ZAFARI F, BASU P, LEUNG K K, et al. Resource sharing

- in the edge: a distributed bargaining-theoretic approach[J]. *IEEE Transactions on Network and Service Management*, 2023, 20(4): 4369–4382.
- [38] 李逸博, 李小平, 王爽, 等. 面向算力网络的智慧调度综述[J]. *自动化学报*, 2024, 50(6): 1086–1103.
- [39] 郝昊, 杨树杰, 张玮. 算网融合下时间连续的计算任务卸载机制[J]. *计算机研究与发展*, 2023, 60(4): 735–749.
- [40] XIE R C, ZHU H, TANG Q Q, et al. Joint task scheduling and load balancing in computing power network-enabled edge computing systems[C]//Proceedings of 2022 IEEE 8th International Conference on Computer and Communications. Chengdu: IEEE, 2022: 563–568.
- [41] LEI B, ZHAO Q Y, MEI J. Computing power network: an interworking architecture of computing and network based on IP extension[C]//Proceedings of 2021 IEEE 22nd International Conference on High Performance Switching and Routing. Paris: IEEE, 2021: 1–6.
- [42] HE Y H, REN J K, YU G D, et al. D2D communications meet mobile edge computing for enhanced computation capacity in cellular networks[J]. *IEEE Transactions on Wireless Communications*, 2019, 18(3): 1750–1763.
- [43] 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. Shenzhen: ACM, 2018: 1–6.
- [44] ZHANG Y T, DI B Y, ZHENG Z J, et al. Distributed multi-cloud multi-access edge computing by multi-agent reinforcement learning[J]. *IEEE Transactions on Wireless Communications*, 2021, 20(4): 2565–2578.
- [45] PANG R, LI H, JI Y F, et al. Energy-saving mechanism based on tidal characteristic in computing power network[C]//Proceedings of 2021 International Conference on Networking and Network Applications. Lijiang: IEEE, 2021: 150–154.
- [46] ZHANG Y, CAO C, TANG X Y, et al. Programmable service system based on SIDaaS in computing power network[C]//Proceedings of 2022 5th International Conference on Hot Information-Centric Networking. Guangzhou: IEEE, 2022: 67–71.
- [47] YAO H J, DUAN X D, FU Y X. A computing-aware routing protocol for Computing Force Network[C]//Proceedings of 2022 International Conference on Service Science. Zhuhai: IEEE, 2022: 137–141.
- [48] LIU Z, ZHANG J W, GU Z Q. Network and computing-aware edge datacenter placement and content placement in edge compute first networking[C]//Proceedings of 2022 Asia Communications and Photonics Conference. Shenzhen: IEEE, 2022: 1233–1237.
- [49] 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 2022 IEEE 8th International Conference on Computer and Communications. Chengdu: IEEE, 2022: 1615–1620.
- [50] ZHAO L J, TANG X Y, YOU Z P, et al. Operation and security considerations of federated learning platform based on compute first network[C]//Proceedings of 2020 IEEE/CIC International Conference on Communications in China. Chongqing: IEEE, 2020: 117–121.
- [51] WANG P W, SUN Z X, ZHAN Y, et al. Marginal value-based edge resource pricing and allocation for deadline-sensitive tasks[C]//Proceedings of IEEE INFOCOM 2023–IEEE Conference on Computer Communications. New York City: IEEE, 2023: 1–10.
- [52] DONG Y Q, GUAN C C, CHEN Y L, et al. Optimization of service scheduling in computing force network[C]//Proceedings of 2022 International Conference on Service Science. Zhuhai: IEEE, 2022: 147–153.
- [53] SHEN S H, FENG Y C, XU M W, et al. A holistic QoS view of crowdsourced edge cloud platform[C]//Proceedings of 2023 IEEE/ACM 31st International Symposium on Quality of Service. Orlando: IEEE, 2023: 1–10.
- [54] ZHANG H, HUANG S Y, XU M W, et al. How far have edge clouds gone? A spatial-temporal analysis of edge network latency in the wild[C]//Proceedings of 2023 IEEE/ACM 31st International Symposium on Quality of Service. Orlando: IEEE, 2023: 1–10.
- [55] HUANG S Y, WANG Z, ZHANG H, et al. One for all: unified workload prediction for dynamic multi-tenant edge cloud platforms[C]//Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Long Beach: ACM, 2023: 788–797.
- [56] 郭中孚, 张兴明, 赵博, 等. 软件定义网络数据平面安全综述[J]. *网络与信息安全学报*, 2018, 4(11): 1–12.
- [57] XYLOMENOS G, VERVERIDIS C N, SIRIS V A, et al. A survey of information-centric networking research[J]. *IEEE Communications Surveys & Tutorials*, 2014, 16(2): 1024–1049.
- [58] 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.
- [59] 袁长卿, 苏越, 赵伟博. 面向算力网络的安全体系研究[J]. *信息通信技术与政策*, 2023, 49(2): 82–86.
- [60] 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.