

算网自智：研究进展与展望

朱海龙^{1,2}, 杨帆¹, 蒋如一¹, 贾庆民², 周晓茂², 谢人超^{1,2}, 黄韬^{1,2}

(1.北京邮电大学网络与交换技术国家重点实验室, 北京 100876; 2.紫金山实验室, 江苏南京 211111)

摘要: 尽管算力网络可以通过网络连接并灵活地调度计算资源, 但在实际落地应用中仍然缺乏智能性和灵活性。针对算力网络亟须提高智能化水平的问题, 首先简要引出了算网自智的背景和概念; 接着提出了一种新型的算网自智基本架构和工作机制, 通过算网感知、智能决策、算力路由、自动化和自优化等关键技术实现算力网络多维资源的智能资源分配和服务调度; 最后, 对算网自智未来的主要研究方向和技术挑战作了展望。算网自智作为未来的必然发展趋势, 可以很好地解决运维流程复杂、过度依赖人工经验、难以自适应动态变化的环境等一系列问题。

关键词: 算网融合; 算网自智; 算力网络; 自智网络

中图分类号: TN91

文献标志码: A

DOI: 10.11959/j.issn.1000-436x.2024174

Computing-network intelligence: research progress and prospects

ZHU Hailong^{1,2}, YANG Fan¹, JIANG Ruyi¹, JIA Qingmin², ZHOU Xiaomao²,
XIE Renchao^{1,2}, HUANG Tao^{1,2}

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

2. Purple Mountain Laboratories, Nanjing 211111, China

Abstract: Although the computing power network (CPN) can connect via the Internet and flexibly schedule computing resources, it still lacks intelligence and flexibility in practical applications. To address the need for enhanced intelligence in computing power networks, the background and concept of the computing-network intelligence (CNI) were first introduced briefly. Then, followed by the proposal of a novel basic architecture and operational mechanism for CNI, intelligent resource allocation and service scheduling of multi-dimensional resources were achieved through key technologies such as computing-network perception, intelligent decision-making, computing power routing, automation, and self-optimization. Finally, future research directions and technical challenges were discussed. CNI is recognized as an inevitable trend for the future, effectively addressing issues such as complex operational processes, excessive reliance on human experience, and difficulties in adapting to dynamically changing environments.

Keywords: computing-network convergence, computing-network intelligence, computing power network, autonomous network

0 引言

随着工业互联网、自动驾驶、智慧城市等技术应用蓬勃发展, 数据总量呈爆发式增长, 推动了数

据传输、分析和存储需求的提升, 传统云计算逐渐难以满足多元业务对低时延、高算力的要求, 边缘计算由于距终端近, 可以提供更低时延、更高效的

收稿日期: 2024-02-26; 修回日期: 2024-06-17

通信作者: 蒋如一, jiangruiyi@bupt.edu.cn

基金项目: 国家自然科学基金资助项目(No.92267301, No.92367104)

Foundation Items: The National Natural Science Foundation of China (No.92267301, No.92367104)

服务, 极大程度地改善用户的体验, 因此计算逐渐向边缘侧下沉。然而, 边缘节点单点算力有限, 在面对大规模用户场景时存在资源短缺、成本较高的问题。在这种情况下, 海量的用户端设备不仅具备一定的计算能力, 还能显著减轻边缘节点的负载和运营成本。因此, 用户也可以根据自身业务的特点来综合选择最佳云、边、端计算资源和网络资源的组合方式, 以实现资源的高效协同。

从云计算衍生而来的边缘计算、雾计算、云边端协同等计算范式是算力网络的重要基础使能技术^[1-3], 引起了广泛的研究。为了屏蔽各种计算范式之间存在的异构性, 文献[1]中引入了区块链、媒体无关切换 (MIH)、软件定义网络 (SDN) /网络功能虚拟化 (NFV) 等先进技术, 提出了一种计算范式的协同解决方案。文献[2]中提出了一种异构分层雾计算的移动通信网络架构, 证明雾计算可以为移动通信网络带来比传统网络更高的通信容量增益。文献[3]中提到了云边端协同的无线分布式计算将在未来 6G 网络中发挥重要作用。

随着各种边缘计算范式和算力融合^[4] (CNC, computing-network convergence) 的发展, 一种新型的网络架构——算力网络^[5-8]应运而生了。目前, 我国在算力基础设施的建设、算力芯片的开发等方向上仍然面临严峻的挑战。2021年5月, 国家发展改革委、中央网信办、工业和信息化部、国家能源局联合印发了《全国一体化大数据中心协同创新体系算力枢纽实施方案》(发改高技〔2020〕1922号), 提出建设全国一体化算力网络枢纽节点, 加快实施“东数西算”重大工程。2023年10月, 美国进一步升级高算力人工智能 (AI) 芯片对华的出口限制, 促使我国加快提升智能芯片的算力水平, 以算力作为基石, 结合 AI 大模型等新技术推动我国一体化算力网络创新发展。

随着人们对算力网络的研究越来越深入, 算力网络原本存在的许多不足也日渐显现^[9-11]。

1) 人力成本消耗过高: 算力网络在运行过程中必须依赖大量的人力参与和人工经验, 缺乏自动化运维的能力。

2) 自适应能力差: 无法根据环境、网络的变化动态对网络策略做出自适应的调整, 也无法根据不同业务需求的特点做出个性化的决策, 网络的弹性、可扩展性差。

3) 网络缺乏系统级的智能能力: 算力网络与 AI 的结合大多是局部的, 如通过 AI 优化路由层的任务调度决策, 但无法实现系统级的智能决策。

为了应对以上挑战, 算力网络的自智能化亟须提上日程。2019年, 国际电信管理论坛 (TM Forum, telemanagement forum) 首先提出了自智网络 (AN, autonomous network)^[11], 旨在通过 AI 等前沿技术将智能引入网络的全生命周期, 实现网络的自动化和智能化, 并将自智网络的等级分为 L1~L5 这 5 个等级。目前, 我国主要网络的自智能化水平还停留在 L3, 仍然过度依赖人工经验, 限制了其向高度自智阶段演进的步伐。因此, 如何最大程度地解除自智网络对人力的依赖, 提高网络的自智能化水平, 也是研究者们目前亟须考虑的问题^[12]。

自智网络对实现算力网络的自动化和智能化具有借鉴意义, 尤其是与自智网络高度相关的意图驱动、数字孪生、知识定义、大模型等 AI 技术, 为算力网络的极简运维、高效优化和深度学习提供了灵感和方法。

从理论上来说, 算网自智是算力网络和自智网络概念的深度结合, 以算力为中心, 以智能为手段, 以网络为根基, 通过算力、网络、数据、安全等多要素融合编排管理实现网络的自动化、自治化以及自优化; 从技术上来说, 算网自智即通过 AI、大数据等技术驱动算力网络的资源感知、算力路由、算力交易、在网计算和算网协同调度等流程自动化和智能化, 逐渐将人力解放出去; 从功能上来说, 算网自智集成了 AI 强大的预测、分析、学习、决策、优化等功能, 可以支撑“规建维优营”全流程, 在未来更有可能解锁故障自定位、自修复和用户意图感知等高度自智能能力; 从发展趋势来说, 算网自智是算力网络由网随算动、算力融合走向算网一体的演进过程中的必然趋势, 自智能化也是算力网络由自动化走向自治化的最高目标。

本文针对算力网络对网络自动化、自适应、系统化、智能化等能力提出的迫切需求, 用自智网络赋能算力网络, 提出了算网自智的概念, 并结合算力感知、智能决策、算力路由、自动化和自优化等关键技术, 设计了一种新型的算网自智基本架构和工作机制, 接着对用到的关键技术进行了概要介绍, 最后探讨了算网自智未来的研究方向和挑战, 对其中的 4 个重要技术方向作了分析和展望。

1 算网自智研究现状分析

算力网络和自智网络是算网自智的关键使能技术, 本节将对算力网络和自智网络的发展和研究现状进行简要概述。

1.1 算力网络发展现状

随着云、边、端泛在的三级算力架构的发展, 产、学以及标准领域对算力网络的关注持续升温, 算力网络也不断延伸出新的术语^[13-15], 如表 1 所示。2021 年在国际电信联盟第 13 研究组 (ITU SG13) 全会上对 CAN、CPN、CFN 三类术语重新界定了研究范围, 并形成了统一术语, 即算网融合 (CNC)。CNC 的主要研究方向包括算力感知、算力路由、算力调度、算力交易等。

在标准方面, 国内外关于算力网络的标准化工作已经逐步展开, 算力网络国际标准体系已初步建立, 如图 1 所示。国际标准组织主要包括 ITU、互联网工程任务组 (IETF) 等, 国内标准组织主要包括中国通信标准化协会 (CCSA)、国际移动通信 (IMT)-2030 (6G) 推进组、中国通信学会等。

在产业方面, 国内各大运营商联合企业发布了一系列的产业白皮书。2019 年 11 月, 中国联通率先发表了《中国联通算力网络白皮书》^[16], 首次倡议算力网络的概念。2021 年 9 月, IMT-2030 (6G) 推进组正式发布首个多方联合发布的 6G 网络架构白皮书^[17], 将算力网络列为 6G 十二大潜在关键技术之一。

在学术方面, 越来越多的研究者开始关注算力网络。目前算力网络的研究主要聚焦在算力网络的架构和关键技术。算力网络的架构经历了由云计算、边缘计算到云边协同、云边端协同的发展历程, Yousafzai 等^[18]提出了基于轻量级进程迁移的计算卸载框架, 相比本地执行节省了近 44% 的执行时间和 84% 的能耗, 显示出云边端分布式计算架构的巨大潜力。算力调度技术是算力网络实现泛在资源分配的关键, Chen 等^[19]提出了一种算力网络计算和传输联合优化方法, 通过虚拟双队列巧妙地简化了计算和传输联合分配的问题。

研究者们也进行了许多面向未来的尝试, 例如, 将算力网络与 SDN、信息中心网络 (ICN) 等网络技术相结合, Król 等^[20]提出的分布式计算平台 CFN 将通用分布式计算框架与 ICN 的网络层相结合, 比基于 Overlay 的计算方式具有更强的可用性。

此外, 算力网络的研究呈现出云原生、确定化、安全化、绿色化等趋势^[21-24]。Alamouti 等^[21]提出了一种云去中心化新型架构 HEC (hybrid edge cloud), 其相比集中式云架构能更充分地利用端设备上的计算资源, 减小网络带宽和时延。Huang 等^[22]设计了基于周期的时敏确定性网络 (C-TSDN) 架构, 将循环排队转发 (CQF) 和循环控制排队转发 (CSQF) 机制相结合来实现端到端的确定性传输。Wang 等^[23]提出了基于区块链的算力

表 1 算力网络术语发展

术语	缩写及全称	定义
计算能力网络	CPN, computing power network	算力网络的统称, 包括集中式、分布式的算网感知和算力路由等能力
计算感知网络	CAN, computing-aware network	强调通过算力感知, 实现算力路由或调度
计算优先网络	CFN, computing first network	强调通过在网络层进行协议扩展, 实现算力路由信息同步
算网融合	CNC, computing and network convergence	ITU 对算力网络的统一定义

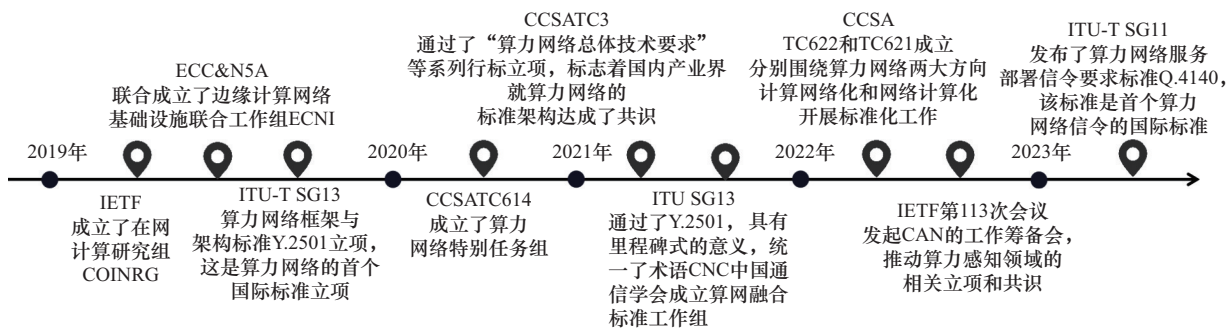


图 1 算力网络国际标准体系

网络联邦学习方法, 兼备安全性和效率。Lu 等^[24]提出了以节能为目标的网络任务卸载和资源分配方案, 有效提高了计算效率并降本节能。

现有的算力网络研究仍然存在许多问题, 如算力资源具有异构性和动态性、网络传输具有不确定性、资源调度算法缺乏灵活性和智能性等。因此亟须研发异构算力资源的统一度量体系, 对算力资源进行实时监控; 优化网络传输协议, 深入结合确定性网络技术研究; 开发基于 AI 和机器学习的自适应智能调度算法。

1.2 自智网络发展现状

随着全球对人工智能技术发展的重视, “AI+网络”^[25-28]的概念愈发火热, 通信领域的标准组织、产业界和学术界开始纷纷布局自智网络领域。TM Forum 作为最早布局自智网络的标准组织之一, 首先以自治网络的概念发表《TM Forum Autonomous Networks 白皮书 1.0》^[11], 后在 3.0 版本中正式确定为自智网络。

在标准方面, TM Forum、ITU-T、IETF、开放网络基金会 (ONF)、CCSA 等标准组织纷纷展开自智网络的标准化工作, 如表 2 所示。

在产业方面, 2023 年 10 月, 在 TM Forum 主办的自智网络产业峰会上, TM Forum 联合 66 家产业伙伴撰写的《自智网络白皮书 5.0》^[29]发布, 标志着自智网络历经 4 年已经从理论迈入了规模部署的阶段, 该白皮书首次提出了自智网络实施框架 ANF, 并对 L4、L5 级自智网络做了展望。

在学术方面, 关于 AI 助力网络的研究^[30-41]也越来越多, 涉及无人机网络^[42]、智能交通^[43]、无线传感器网络^[44]等多个领域。AI 可以赋能网络中难以建模和精确求解的感知、诊断、管理、分类、

预测、优化等任务, 具体介绍如下。

1) 网络感知。未来的网络不仅要传输数据, 还要传输信息和知识, 可以结合带内网络遥测 (INT) 等数据收集方式, 对关联数据之间的特征进行提取和融合。文献[30]提出了一种在动态随机环境中基于稀疏贝叶斯学习的自适应传感框架, 对传感器节点的传感质量和能源效率进行了联合优化。

2) 异常检测和诊断。Chen 等^[31]将变分自编码器 (VAE) 与对抗训练相结合, 提出一种基于 VAE 的无监督网络异常检测方法。

3) 资源管理^[32-34]。将机器学习与 SDN、NFV 等技术结合有助于对网络资源进行智能的全局配置和管理, Guo 等^[34]通过编排服务功能链 (SFC) 和放置虚拟网络功能 (VNF), 提出了一种基于深度强化学习的自调整资源分配算法。

4) 流量分类^[35-36]。流量分类是一种更加细粒度的流量管理方法, 可以通过机器学习对大量流量进行采集、分析和知识提取, 并按需从大象流-老鼠流、应用、服务质量 (QoS) 等多种角度来对流量进行分类。

5) 流量预测^[37-38]。将深度学习模型与流量预测相结合, 如使用循环神经网络 (RNN)、门控循环单元 (GRU)、长短期记忆网络 (LSTM) 来进行时间相关性建模, 使用卷积神经网络 (CNN)、LSTM 等进行空间相关性建模。

6) 路由优化^[39]。机器学习模型经过训练后能够给出近乎最优的路由方案, 是一种高效的路由决策优化方法。

知识定义和意图网络作为自智网络的两大重要技术研究方向也得到了研究者的关注。知识定义网络 (KDN, knowledge-defined networking) 起源于

表 2 自智网络各大标准组织进展

标准组织	标准进展	国际/国内
TM Forum	TM Forum 于 2019 年率先在《TM Forum 自治网络白皮书》中提出自智网络架构中“四层三闭环、Self-X 能力、单域自治跨域协同”等理念和自智网络的分级评估方法	国际
ITU-T	ITU-T SG13 在 2020 年 12 月建立了自智网络焦点组 (FG-AN), 以执行自智网络的预标准活动	国际
IETF	IETF 在 2014 年就建立了自主网络工作组 ANIMA, 目前正在基于 YANG 数据建模语言进行对自主网络北向接口标准和意图接口管理需求的研究工作	国际
ONF	负责 OpenFlow 系列协议制定的 ONF 在 2015 年启动了意图驱动网络 (IBN) 课题, 并在 2016 年发布了基于 SDN 架构和北向接口实现意图驱动网络的指南	国际
CCSA	CCSA 建立了自智网络技术工作委员会 (CCSA TC7) 和推进委员会 (CCSA TC610), 协同推进自智网络标准布局加速生成	国内

Clark 等^[45]发表的互联网知识平面提案, AI 可以通过挖掘知识点之间的深层关系来获取新的知识, 并用这些知识来辅助网络进行智能化决策。文献[46]提出了基于 SDN、AI 算法和知识驱动的网络智能决策方案。意图驱动网络^[47]的概念最早来自 SDN 的意图北向接口, 后又被 Gartner 正式提出, 到现在, IBN 已经被定义为一种将用户提出的网络需求转化为网络具体配置的技术。文献[48]提出了一种意图驱动的网络资源自动化管理框架。

自智网络研究正处于起步阶段, 目前面临着许多挑战, 如 AI 技术的发展和应用尚不成熟, 系统决策存在不稳定性 and 误判风险, 用户的隐私和安全性难以保障, 网络运维过于复杂且尚未建立起统一的标准等。未来需要持续投入对 AI 算法的研究, 对系统决策建立全面的测试验证体系, 并引入实时监控, 以确保决策的稳定性和可靠性; 研究先进的数据加密和匿名化技术; 开发高效的自动化运维工具和算法, 并积极推进行业的统一标准体系建设。

2 算网自智的基本架构与工作机制

相比于基础的算力网络^[49], 自智算力网络的设计从 2 个原则出发: 内生智能和业务智能。内生智能指的是自智算力网络保障系统自身正常运行的自动化、自适应和自优化的能力; 业务智能指的是自智算力网络具备对业务进行解析、自动编排管理资源并形成针对特定业务的智能服务映射模型的能力, 即新型的自智算力网络架构既需要在各个系统层级都能实现自动化和智能化, 还需要在业务层面实现智能化的服务模型编排。基于以上 2 个原则, 本节将对算网自智的基本架构和工作机制做出构想, 并展开深入分析和讨论。

2.1 算网自智基本架构

智能网络架构在产业界和学术界都引起了广泛的讨论和研究。文献[50]引入了内生智能和业务智能的概念, 提出了一种五层的融智算力网络架构。文献[51]从类似的角度给出了智能一代域和智能自主域的概念, 提出了一种双域 6G 智能内生网络架构。相比现有的智能网络架构^[50-51], 本文所提架构的优势在于清晰地展示出自智网络和算力网络的结合, 不仅保留了 TM Forum 自智网络中自治域的概念和“单域自治, 多域协同”的特点, 还融入了意图驱动、数字孪生、算网大脑^[52]等先进技术概念, 清晰展现了分层架构各个层级的、从网元到整个系统不同粒度的智能性, 概括了部分智能化层级的工作机制。该架构参考了自智网络“三层四闭环”的层级化结构特点, 如图 2 所示, 自下而上分别为是算网资源层、算网管理层(包括编排调度层和资源管理层)、算网运营层。

算网资源层一方面作为算网自智的基座, 能够感知并统一管理算网整体的算力、网络、数据等多要素融合资源; 另一方面作为算网基础设施单位的网元, 可以通过内嵌的智能引擎挖掘自身信息, 使得决策更加可靠有效。

编排调度层(即算网大脑)向下拉通各个单域, 通过意图引擎、数字孪生、全域感知、智能分析等技术实现跨域协同。整个算网大脑基于“OODA”模型^[53]设计, 包含全域感知(observe)、智能分析(orient)、统一编排(decision)、调度执行(action)和算网大脑能力开放中心。算网大脑能力开放中心负责向算网运营层开放服务能力目录, 便于对服务能力进行一站式的查询和一体化的管理、调度。

资源管理层是对算网资源层基础设施的能力抽

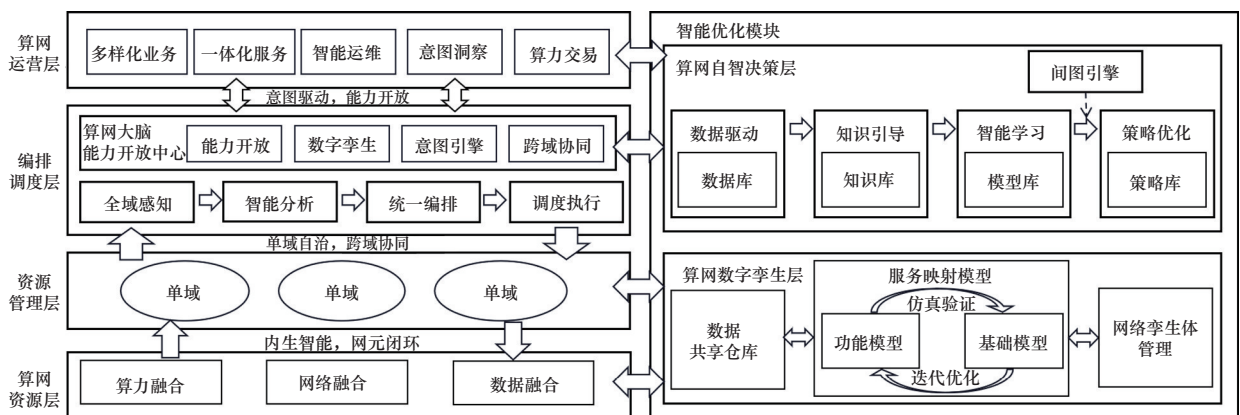


图 2 算网自智基本架构

象和模型化，通过各个域的运维工作台、云管等实现对资源的统一管理。北向对接编排调度层，向算网大脑提供资源、性能等数据和服务化 API 以实现跨域协同，同时各个域执行算网大脑下发的决策，实现单域自治。

算网运营层承载着抽象的多种个性化业务功能，如智能城市、智能家居、自动驾驶等；面对用户开放一体化的服务能力，根据意图自动将合适的服务调度到用户节点；智能运维；结合区块链等技术实现可信的算力交易。

相比于基础的自智网络架构，该架构还引入了智能优化模块，以弥补算网各个层级之间的信息差异。智能优化模块可以综合性地从每个层级收集海量数据，从而对各个层级进行针对性的配置优化，并举例说明了模块中的 2 个层级的运行机制。在算网自智决策层中，数据驱动技术能够深度挖掘和分析网络数据并提取知识，利用专家知识等领域知识进行引导，表征成模型，再根据意图类型从模型库中匹配业务模型，做出决策。算网数字孪生层能够通过建立服务映射模型对网络策略进行仿真验证和迭代优化，最终将优化控制下发到实体网络中。

2.2 算网自智工作机制

基于以上提出的算网自智基本架构，本节对算网自智的基本工作流程做简要阐述，如图 3 所示。

1) 意图感知。算网运营层感知到用户意图，将意图传递给算网大脑能力开放中心；算网大脑能

力开放中心在意图引擎中对业务意图进行解析和建模，再根据意图类型分发业务到统一编排模块。

2) 资源感知。算网资源层对网络、算力、存储等资源持续进行感知，并向上映射成抽象的资源管理层；资源管理层通过单域自治的形式对资源进行统一的管理，并将资源信息向上传递给算网大脑中的全域感知模块。

3) 智能分析。算网大脑的统一编排模块通过智能分析模块根据整体资源信息进行分析，选择用于该业务的算网资源。

4) 统一编排。统一编排模块根据该业务对应的算网资源进行组合编排，使之满足业务需求。

5) 调度执行。调度执行模块接受统一编排模块的指令，向下进行实际的资源编排调度和配置，完成后返回调度结果。

6) 结果反馈。调度结果从调度执行模块返回，依次经由统一编排模块和算网大脑能力开放中心传递给算网运营层，告知业务已开通和反馈结果。

意图感知、资源感知、智能分析、统一编排、调度执行和结果反馈构成了算网自智面向业务的基本工作流程，使得自智算力网络能灵活调度资源以满足业务需求。同时，算网自智也通过网元自治、资源自治、服务自治、业务自治和用户自治 5 个内生闭环不断进行自动化和自优化，这些将在第 3.4 节中详细阐述。接下来将着重介绍算网自智的具体关键技术，包括算网感知、算力路由、算网智能决策、算网自动化和自优化。

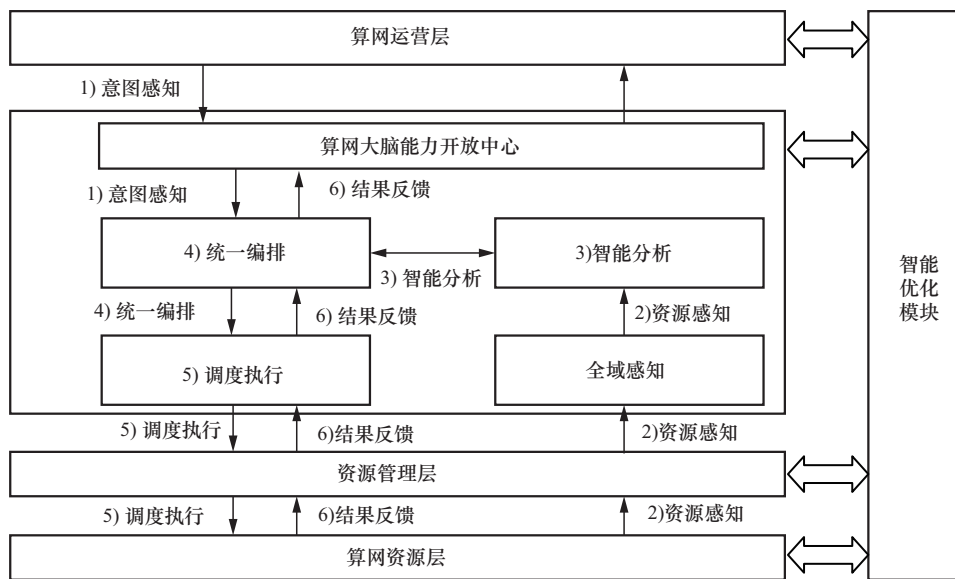


图 3 算网自智的基本工作流程

3 算网自智关键技术

算网自智作为算力网络和自智网络的融合体,不仅需要借鉴算力网络的关键技术,如算网感知、算力路由,还要深入思考如何注入 AI。从业务智能和内生智能 2 个原则出发,分别提出自智算力网络的智能决策、自动化和自优化 2 项技术。本节将对这 4 项关键技术展开详细探讨。

3.1 算网感知

算网感知需要融合算力、网络、业务等多维资源感知,其中算力感知包含算力节点、算力部署状态和算力服务等信息,网络感知包括网络的时延、抖动、丢包率等性能信息,以及网络拓扑、资源配置等状态信息和流量信息等,业务感知包括用户业务、应用的类型和状态。

算网感知的关键在于如何打通计算和网络的边界,通过设计一种自适应的分布式算力通告协议,在数据面实现算、网、业务资源信息的实时通告,为算网协同调度打好基础^[54]。算力通告模式分为集中式、分布式和混合式。图 4 是混合式算力通告模式,其将算力节点分为不同的组,一部分节点通过分布式内部网关协议 (IGP) 和边界网关协议 (BGP) 来实现全网通告,另一部分采用集中式通告来达到更快的响应速度。

实现算网感知的前提是算网度量、建模和标识。不同异构设备的算力不同,采用不同算法需要消耗的算力不同,不同的业务场景对算力的需求也不同,因此业界亟须建立面向算力网络海量异构资

源的统一度量体系。Mcmahon^[55]首次提出了用“浮点运算次数”来表示算力,并引入了单位——每秒浮点运算次数 (FLOPS),但这种单一的度量指标已经逐渐不能满足算力服务的多元需求。目前最主流的算力度量方式是将算力从逻辑计算能力、并行计算能力和神经网络计算能力^[56]3 个维度来进行表征,再将算力资源建模为包含计算、通信和存储等多维度性能指标的模型。此外,要对算力资源进行统一的管理,还需要建立统一的算力标识体系,确保每个算力标识全局唯一。

除了研究算网资源的实时感知方案以外,目前的研究还集中在如何提高算网感知的精度、深度和智能化程度。比如,通过对算网资源状态的信息测量技术,提高对网络和算力节点状态感知的精度^[57];通过机器学习模型和 AI 算法,提取算网资源的有效特征,如建立基于图神经网络 (GNN) 的算网资源感知模型,增加算网感知信息的深度;通过资源状态预测技术^[58],提高算网感知的智能化程度。

3.2 算力路由

中国移动在 2018 年首次提出算力路由^[59],并主导在 IETF 成立了算力路由工作组。算力路由指的是将计算、网络、存储等多维感知的实时状态信息加入路由表,并结合解析的业务需求,选择动态最优的业务调度策略,基于互联网协议第六版 (IPv6) 和基于 IPv6 转发的段路由 (SRv6) 等可编程协议进行最优路径的算力路由转发。算力路由实际上是 IP 技术在算网融合背景下的创新,能够有效降低路由的全局时延,实现系统的负载均衡。

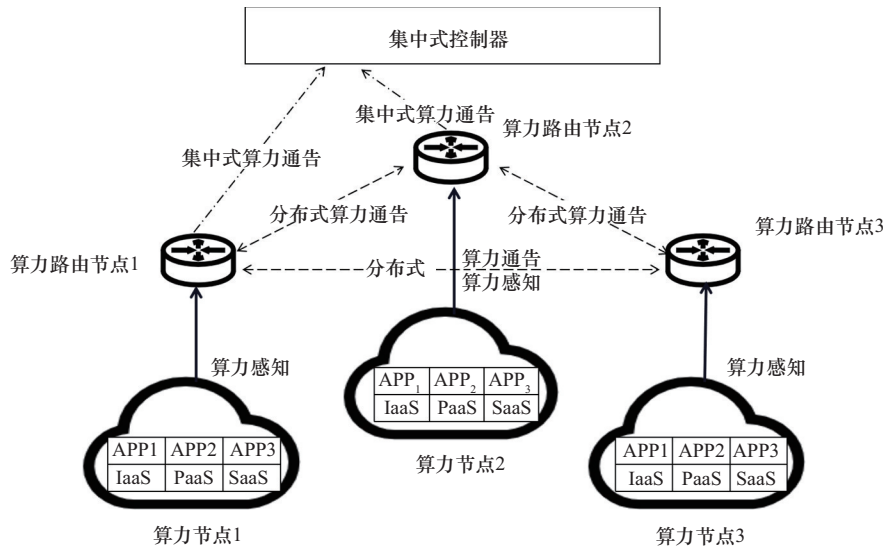


图 4 混合式算力通告模式

算力路由与传统的路由^[60]相比,从单一路由变为了网络、计算等多要素融合的路由,如图5所示。控制平面首先将所有节点、链路的计算、网络信息进行通告,经过算力节点入口收集所有的算网信息和客户端信息,接着进行路由计算,生成综合计算、网络开销的算力路由表,根据用户需求决策出实时最优的算力出口节点和路由路径,最后执行业务调度策略。

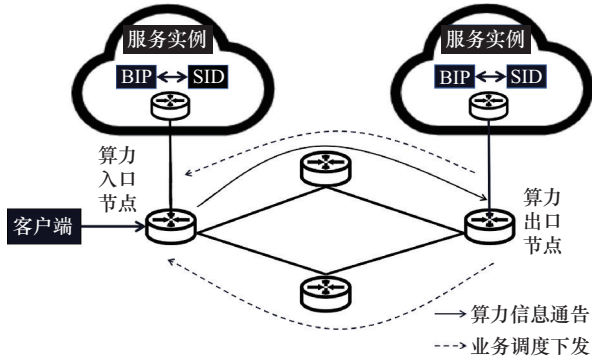


图5 算力路由技术示意

在算力路由的创新上,IPv6/SRv6等多种路由协议可以通过修改扩展报文头来简化协议,降低网络的复杂度;结合任播技术^[61],对不同的算力业务类型分配不同的任播地址,在实现负载均衡的同时提供个性化的服务;结合Overlay和Underlay这2种技术方案^[62],在应用层实现算网感知、算力节点选择、传输路径决策后,再下发到Underlay网络。

3.3 算网智能决策

基于业务智能原则,智能决策将建立在算网度

量和标识、算网感知的基础上,首先借助智能化手段对算网的业务意图、算网状态、资源等信息进行深度分析和理解,并建立起如图6所示的算网业务-服务能力-资源映射关系,再进行智能学习、决策和优化,使得既能实现面向多元业务需求的资源按需分配,又能实现面向不同用户的一站式服务调度,做到用户满意度、服务质量和资源利用率的联合优化。

决策所用到的常见智能手段有机器学习(强化学习、深度学习、监督学习和无监督学习)、优化理论和博弈论等^[63]。智能决策中的“决策”则主要由资源管理和路由规划决策组成。资源管理包括智能运营、管理和维护(OAM)、计算卸载/计算任务调度和服务调度方面,路由规划包括算力节点选择和用于实现任务分发的路径规划。计算卸载决策考虑到算力在网络中不均衡的问题,以提高资源利用率为目的,将任务卸载/调度到网络的不同位置进行,如云、边和终端。服务调度是基于云原生技术,实现上层服务和底层资源解耦的决策,主要指的是基于微服务、容器和服务链的动态服务编排。由于用户的需求是多种多样且随时变化的,因此可以通过容器编排平台K8S^[64]等容器编排工具来按需灵活地激活或关闭部署在容器中的服务,节约资源,实现快速且弹性的资源分配。

算网智能同样可以在排障、运维、预测等决策上发挥作用。在车路协同、无人机协作等移动性强的场景下,算网甚至能通过智能分析自动优化复杂的车辆/无人机移动和切换决策,以最大限度地降低时延和减少切换失败概率^[65]。

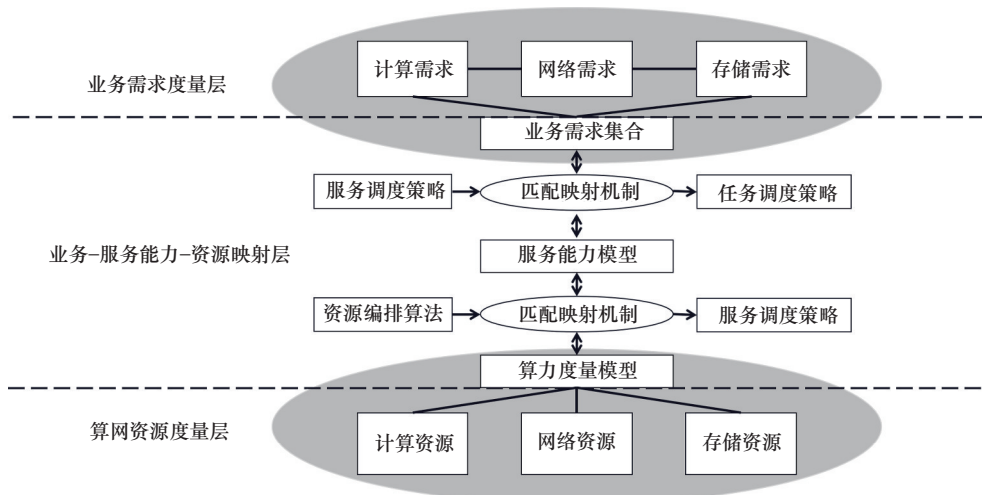


图6 算网业务-服务能力-资源映射关系

3.4 算网自动化和自优化

随着数据量的爆发式增长和新型业务的不断涌现,算力网络规模持续扩大,网络管理和运维变得愈发复杂。传统的网络运维方式主要依赖人工操作,效率低且易出错,无法满足现代网络灵活性和高效运维的需求。因此,提升网络自动化和自优化能力已成为业界的迫切需求^[66]。

基于内生智能原则,算力网络的自动化和自优化通过5个闭环来实现,如图7所示,包括网元自治、资源自治、服务自治、业务自治和用户自治。相较于TM Forum的“三层四闭环”^[67],算网自动化和自动化闭环新增了算网资源层和网元闭环。

1) 网元自治。各网元通过“状态采集、智能分析、执行决策、改善优化”的闭环流程来实现自动化运维,具备数据自动采集、业务预测优化和故障提前预警等功能。

2) 资源自治。将碎片化散布的网元自治进行集成,成为网络架构极简的闭环自治域,并以单域为粒度进行资源自动化管理,为跨自治域协同的网络自治管理奠定基础。

3) 服务自治。以跨域协同的方式通过服务闭环实现网络服务运营自动化,算网大脑相当于TM Forum提出的服务运营层^[11],为资源管理层的多个自治域提供IT服务、网络规划等服务运营能力。

4) 业务自治。面向多样化业务需求提供自动化、个性化的使能和运营能力。为了实现业务自治,通常需要多个服务提供商相互协作。

5) 用户自治。模块中的算网自智决策层和算

网数字孪生层与各个层级和自治闭环进行基于意图的交互和优化,消除各层级间的信息不对称,使系统性能达到全局最优,以支持用户服务的实现。

此外,意图驱动也降低了网络运维复杂度,减轻了运维人员的负担,允许用户只需描述想要实现的结果便可轻松地实现具体的网络配置。

本节对算网自智的4个关键技术做了总结,算网感知是自智算力网络的基础,对多层次异构融合算力资源进行实时感知和监控;算力路由是在自智算力网络中采集并传递实时状态信息的方式,它们共同为智能决策和资源调度提供了准确且实时的数据支持;算网智能决策是算网大脑面向业务需求实现资源编排调度的关键技术,满足了业务智能;算网自动化和自优化通过5个闭环实现了自智算力网络从网元、自治域、服务、业务到用户的闭环控制、管理及优化,系统性地简化了网络管理和运维的流程,实现了内生智能。

4 未来研究方向和技术挑战

算网自智作为算网融合背景下的一种趋势,既面临着算力和网络融合的旧挑战,如算力标识、度量、感知、路由、交易等,又面临着如何使用AI赋能算网融合的新挑战。同时,以机器学习^[68]、自然语言处理^[69]、知识图谱^[70]、专家系统^[71]等为代表的AI技术的飞速发展和突破,也为算网自智带来了更多的可能性。

对于算网自智在未来的研究方向和技术挑战,本节简要选取其中的几个做阐述分析,包括AI大模型技术、算网确定性技术、算网元能力编排技

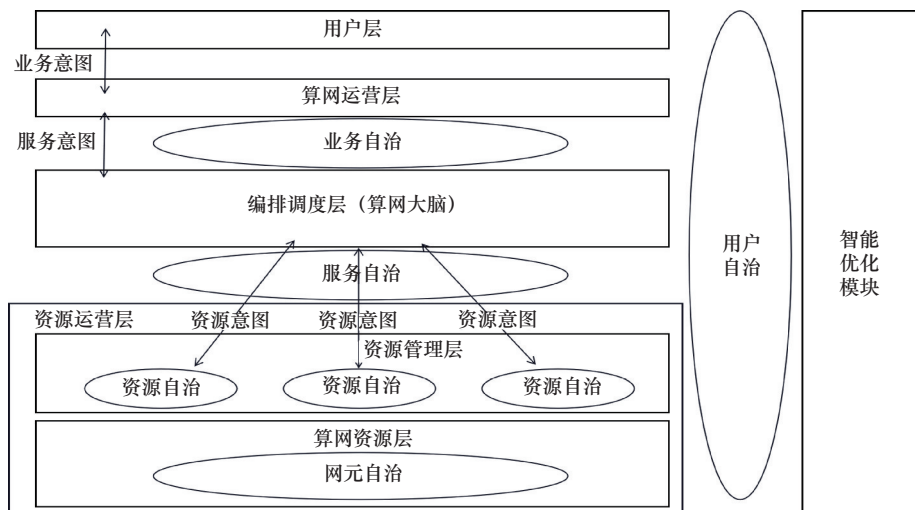


图7 算网自动化和自优化闭环示意

术、知识定义的意图网络技术。这些技术的应用与发展对算网自智的实现至关重要。

4.1 AI大模型技术

Transformer 架构和 ChatGPT 的出现推动了模型即服务 (MaaS, model as a service) 生态的兴起^[72]。AI 大模型在自然语言处理、计算机视觉、语音识别以及多模态领域具有广泛的应用前景,其知识抽取、多模态交流和自学习自进化等重要能力^[73]也让人们看到了大模型赋能通信网络的强大技术潜力。以网络中海量异构的算力资源作为基础, AI 大模型可以通过数据和知识双驱动的方式进行高效训练和精准调优, 推动垂直领域小模型发展, 以大小模型协同进化的方式为复杂的网络业务提供多样化的 MaaS 服务能力, 如图 8 所示。

AI 大模型可以显著提升自智算力网络的性能。首先, AI 大模型拥有处理多模态数据的功能, 能够满足不同来源和类型的网络数据的处理需求, 这需要对算网场景下的多源异构数据进行清洗和融合, 形成用于大模型训练和微调的高质量数据集。其次, AI 大模型拥有知识抽取能力, 能够通过大规模参数对通信网络知识进行关联和深度理解, 并形成新的知识, 从而建立大规模的网络知识库。最后, 大模型可以通过增量学习^[74]来不断进化, 以满足网络的持续进化需求。

在智能学习时, AI 大模型能够精确把握隐藏在大量数据中的规律, 从而助力算网大脑做出更加高效的决策。面对多种多样的网络应用场景与业务, 需要 AI 大模型提供差异化的能力, 可以通过在公开训练数据集的基础上加入私有行业数据来对泛化的

大模型进行微调, 降低大模型幻觉^[75]; 通过上下文学习^[76]方式注入专业领域相关的提示词, 则无须改变通用大模型的权重; 外挂数据库、叠加知识图谱。

虽然 AI 大模型在自智算力网络上的应用前景辽阔, 但是也面临着众多局限, 比如处理多源异构数据, 获得高质量的数据集, 此外, AI 大模型还需要在知识选择、算法设计、资源管理和安全机制等方面深入研究。目前虽然可以使用现成的 AI 算法来解决网络问题, 但效果仍不理想, 因此亟须开发真正适用于网络环境的 AI 模型和算法。

4.2 算网确定性技术

算力网络技术能够高效地统筹调度计算、网络、存储等资源, 但是它更加关注于提升资源的利用率, 面对自动驾驶、智能制造、远程医疗等新型的时间敏感、计算密集型业务, 则缺乏保障任务传输和计算实时、可靠的能力。确定性技术^[77-82]能够与算力网络相结合, 实现满足业务需求的确定性, 对传输而言, 即保障计算任务在节点上处理的实时性, 能及时处理任务, 满足低时延、低抖动、低丢包率、大带宽、高可靠等 QoS 指标; 对计算而言, 即为任务处理提供计算资源保障, 尽量避免任务排队等待。

传统 IP 转发技术无法对分组转发的行为进行控制, 即使在轻载的情况下使用了优先级调度和资源预留等确定性方法, 仍然可能导致多个分组同时到达出端口, 造成排队和微突发。因此需要通过新的确定性技术来控制每个分组的发送/接收的时隙和周期, 避免冲突, 从而控制排队时延, 消除微突发带来的时延和抖动的不确定性。

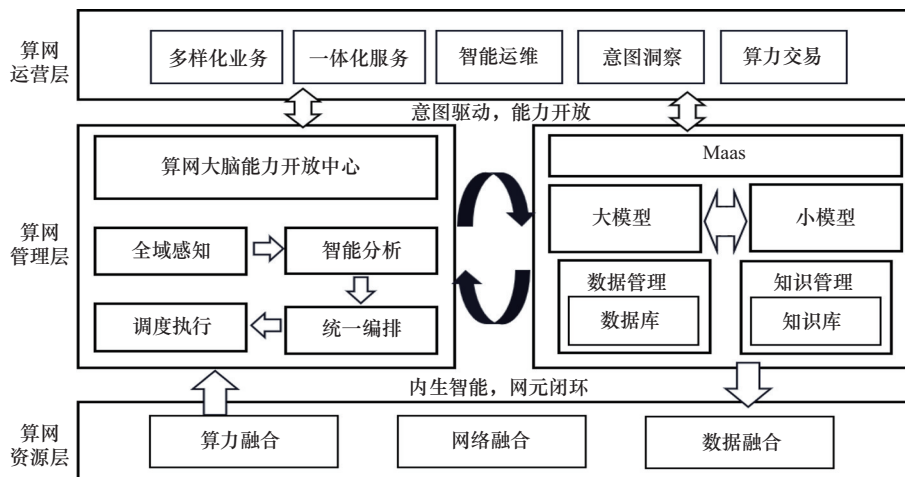


图 8 AI 大模型赋能算网自智

算力网络主要通过结合 L1.5 的灵活以太网 (FlexE) 技术^[83]、L2 的时间敏感网络 (TSN) 技术^[84]和 L3 的确定性网络 (DetNet)^[85]、确定性 IP 网络 (DIP) 技术^[86]来实现。在链路时延确定性方面,使用基于 SRv6^[87-88]的路由方法构建最优的转发路径,在故障发生时能迅速切换备份路径,提供路径确定性保障;在抖动确定性方面,TSN 技术能通过时钟同步、时间感知整形、基于优先级的门控流量调度等保证高优先级任务在确定性的时隙被转发^[89-90];在带宽确定性方面,可以通过 FlexE 技术进行灵活的带宽分配,保障端到端带宽上下界确定性。此外,与 SDN 技术相结合可以使用算力、网络、流等信息进行确定性网络演算^[91],对网络整体的资源预留、路径选择、调度方法等进行集中控制,实现时延上界的确定性保障。

要实现自智算力网络的确定性还有许多复杂的因素要考虑,比如算网状态采集时间的设定(时间过长会影响采集信息的实时性和准确度,时间过短会增大算网的负载)、算力资源的精准感知和度量、任务多并发情况下的调度策略等。

4.3 算网元能力编排技术

算网大脑需要对算网的业务意图、算网状态、资源等信息进行智能分析,根据业务需求进行计算、网络、存储、能耗等多要素编排,并为用户提供高质量的服务,从而建立起算网业务-算网服务能力-算网资源的映射。然而,不管是海量异构的算网基础设施,还是泛在化的服务管理、多样化的业务请求,它们所对应能力的要素广度、粒度、层次都不同,这使得算网的多要素编排变得非常困难。

算网元能力编排技术^[92-95]指的是算网管理层的每个单域在实现自治的同时,通过网络认知和重构技术对单域所具备的服务能力进行抽象和“原子”级别最细粒度的划分,最终重新建模成服务能力——

元服务,每种元服务由一组算网的基本功能元素元能力组成,业务层通过 API 对元服务进行跨域协同的按需调用,以服务链的形式部署在最优的网络路径上,最终形成“元能力-元服务-业务(服务链)”两层映射,具体映射如图 9 所示。

算网元能力编排通常要经过获得全网的元能力实例视图、自适应放置功能实例、解析业务请求并部署服务链、服务链自适应重构 4 个环节,每个环节都面临新的技术挑战。例如,为了灵活地管理和组合元能力,需要对元能力进行聚类,在动态负载变化的情况下,需要研究实时的资源分配策略,以及为了应对多元化的业务需求,需要优化服务链在最短路径上的部署策略等。

自智算力网络在实现元能力编排的过程中,可以通过结合意图识别、AI 等智能化手段对功能实例的负载进行预测,从而预先调整实例数量,合理分配资源。算网元能力编排技术使得算网能更灵活地实现高效的多要素融合编排,提升了网络自适应、可持续演进的能力。

4.4 知识定义的意图网络技术

作为自智网络的一种发展方向,IBN 技术是一种建立在 SDN 基础上,基于意图驱动和 AI 算法实现的智能网络新范式^[96-99]。用户只需输入自然语言形式的命令,IBN 就可以通过解析并转译用户意图,将意图转化为对应的网络策略,实现自动化、智能化的网络管理和运维。然而,如何理解、怎样更深地理解网络状态,并把这种理解用于网络管理和优化中是目前 IBN 面临的一大问题^[48]。知识定义技术^[96]的出现为这一问题提供了解决思路,即通过机器学习算法学习先验知识、历史数据、网络状态等数据,通过挖掘、联系和推理这些知识为根据意图理解生成策略注入智能。知识定义的意图网络策略生成流程如图 10 所示。

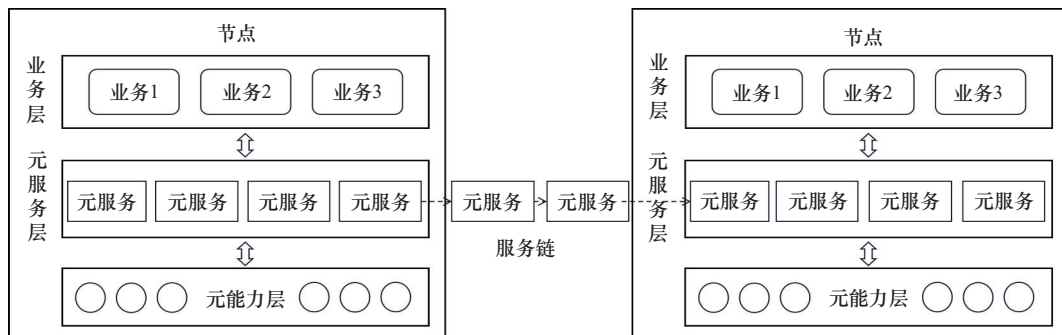


图9 算网元能力编排映射关系示意

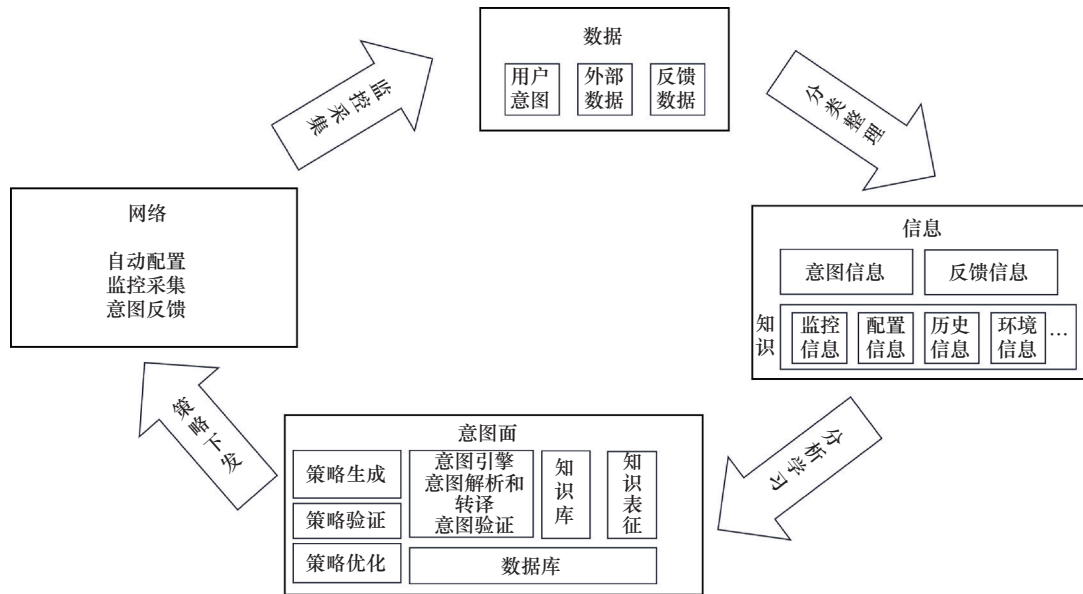


图 10 知识定义的意图网络策略生成流程

知识定义的意图网络技术旨在实现这样的愿景：用户只需通过自然语言的简单交互方式，就能够进行数据、知识双驱动的智能化管理和运维。意图转译主要采用自然语言处理（NLP）的方法进行文本预处理、语义分析等操作。针对网络策略验证成本高、时间长的问题，可以结合数字孪生技术，构建数字孪生体对网络策略进行验证。为了实现及时的意图反馈，使用可编程交换机的INT技术^[100]可以通过对IBN进行实时监控，确保下发策略与用户意图相符。

知识定义与意图网络的融合是未来算力网络的重要研究方向之一。然而，在实际应用中，该方向仍然需要在海量数据处理和分析、用户意图实体的理解和挖掘、传输数据的隐私和安全等关键问题上展开探索和研究，以推动技术的成熟和落地。

5 结束语

AI的快速发展给算力网络带来了无限可能性，算力网络自智作为AI与算力网络深度结合的一种趋势，以算力为中心，以网络为根基，以智能为手段，是一种实现面向多元业务需求的算力、网络等多维资源按需分配，面向不同用户的服务智能一体化调度的算、网、智融合技术。本文首先对算力网络自智的背景、研究现状做了梳理，提出了一种基于内生智能和业务智能的算力网络自智分层架构，该架构清晰地展示了自智网络和算力网络的结合，融合了自治域、意图驱动、数字孪生和算力网络大脑等先进技术特征。

接着介绍了算力感知、算力路由、算力智能决策、算力网络自动化和自优化等关键技术，并对未来AI大模型、算力网络确定性、算力网络能力编排、知识定义的意图网络等重要研究方向做了分析和展望。

目前，受网络智能化和算力网络融合水平影响，我国的算力网络自智仍处于初始阶段，面临着诸多技术和部署上的挑战，如对人工经验的高度依赖和隐私安全问题等。未来，随着AI技术的不断进步，算力网络自智有望成为实现网络高度灵活的自适应调度、自主化管理和自学习的核心技术。

参考文献：

- [1] CAI Q, ZHOU Y Q, LIU L, et al. Collaboration of heterogeneous edge computing paradigms: how to fill the gap between theory and practice[J]. IEEE Wireless Communications, 2024, 31(1): 110-117.
- [2] ZHOU Y Q, TIAN L, LIU L, et al. Fog computing enabled future mobile communication networks: a convergence of communication and computing[J]. IEEE Communications Magazine, 2019, 57(5): 20-27.
- [3] ZHOU Y Q, LIU L, WANG L, et al. Service-aware 6G: an intelligent and open network based on the convergence of communication, computing and caching[J]. Digital Communications and Networks, 2020, 6(3): 253-260.
- [4] TANG S J, YU Y, WANG H, et al. A survey on scheduling techniques in computing and network convergence[J]. IEEE Communications Surveys & Tutorials, 2024, 26(1): 160-195.
- [5] 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.
- [6] 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.
- [7] 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.
- [8] LI M X, CHANG C, TANG X Y, et al. Research on edge resource scheduling solutions for computing power network[J]. Frontiers of Data and Computing, 2020, 2(4): 80-91.
- [9] WANG X F, HAN Y W, LEUNG V C M, et al. Convergence of edge computing and deep learning: a comprehensive survey[J]. IEEE Communications Surveys & Tutorials, 2020, 22(2): 869-904.
- [10] 贾庆民, 郭凯, 周晓茂, 等. 新型算力网络架构设计与探讨[J]. 信息技术与政策, 2022(11): 18-23.
- JIA Q M, GUO K, ZHOU X M, et al. Design and discussion for new computing power network architecture[J]. Information and Communications Technology and Policy, 2022(11): 18-23.
- [11] TM Forum. TM Forum 自治网络白皮书: 赋能电信行业的数字化转型[R]. 2019.
- TM Forum. White paper of autonomous networks: empowering digital transformation for the telecoms industry[R]. 2019.
- [12] 周晓茂, 贾庆民, 胡玉姣, 等. 自智算力网络: 架构、技术与展望[J]. 物联网学报, 2023, 7(4): 1-12.
- ZHOU X M, JIA Q M, HU Y J, et al. Autonomous computing and network convergence: architecture, technologies, and prospects[J]. Chinese Journal on Internet of Things, 2023, 7(4): 1-12.
- [13] HAN X Y, ZHAO Y H, YU K, et al. Utility-optimized resource allocation in computing-aware networks[C]//Proceedings of the 2021 13th International Conference on Communication Software and Networks (ICCSN). Piscataway: IEEE Press, 2021: 199-205.
- [14] XIE Y P, HUANG X Y, LI J C, et al. Computing power network: multi-objective optimization-based routing[J]. Sensors, 2023, 23(15): 6702.
- [15] GONG X, BAI C, REN S, et al. A survey of compute first networking[C]//Proceedings of the 2023 IEEE 23rd International Conference on Communication Technology (ICCT). Piscataway: IEEE Press, 2023: 688-695.
- [16] 中国联通. 中国联通算力网络白皮书[R]. 2019.
- China Unicom. White paper on computing power network of China Unicom[R]. 2019.
- [17] IMT-2030(6G). 6G 网络架构愿景与关键技术展望白皮书[R]. 2021.
- IMT-2030(6G). 6G network architecture vision and key technology prospects white paper[R]. 2021.
- [18] YOUSAFZAI A, YAQOOB I, IMRAN M, et al. Process migration-based computational offloading framework for IoT-supported mobile edge/cloud computing[J]. IEEE Internet of Things Journal, 2020, 7(5): 4171-4182.
- [19] CHEN X Y, XU C Q, WANG M, et al. Augmented queue-based transmission and transcoding optimization for livecast services based on cloud-edge-crowd integration[J]. IEEE Transactions on Circuits and Systems for Video Technology, 2021, 31(11): 4470-4484.
- [20] 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. New York: ACM Press, 2019: 67-77.
- [21] ALAMOUTI S M, ARJOMANDI F, BURGER M. Hybrid edge cloud: a pragmatic approach for decentralized cloud computing[J]. IEEE Communications Magazine, 2022, 60(9): 16-29.
- [22] HUANG Y D, WANG S, HUANG T, et al. Cycle-based time-sensitive and deterministic networks: architecture, challenges, and open issues[J]. IEEE Communications Magazine, 2022, 60(6): 81-87.
- [23] WANG P, SUN W, ZHANG H B, et al. Distributed and secure federated learning for wireless computing power networks[J]. IEEE Transactions on Vehicular Technology, 2023, 72(7): 9381-9393.
- [24] LU Y L, AI B, ZHONG Z D, et al. Energy-efficient task transfer in wireless computing power networks[J]. IEEE Internet of Things Journal, 2023, 10(11): 9353-9365.
- [25] PAN J P, CAI L, YAN S, et al. Network for AI and AI for network: challenges and opportunities for learning-oriented networks[J]. IEEE Network, 2021, 35(6): 270-277.
- [26] SONG L, HU X, ZHANG G H, et al. Networking systems of AI: on the convergence of computing and communications[J]. IEEE Internet of Things Journal, 2022, 9(20): 20352-20381.
- [27] ZHANG T Z, QIU H, MELLIA M, et al. Interpreting AI for networking: where we are and where we are going[J]. IEEE Communications Magazine, 2022, 60(2): 25-31.
- [28] LETAIEF K B, CHEN W, SHI Y M, et al. The roadmap to 6G: AI empowered wireless networks[J]. IEEE Communications Magazine, 2019, 57(8): 84-90.
- [29] TM Forum. 自智网络白皮书: 助力数字化转型——从 2/3 级向 4 级发展[R]. 2023.
- TM Forum. Autonomous networks white paper: empowering digital transformation - evolving from level 2/3 towards level 4. 2023.
- [30] GUPTA V, DE S. SBL-based adaptive sensing framework for WSN-assisted IoT applications[J]. IEEE Internet of Things Journal, 2018, 5(6): 4598-4612.
- [31] CHEN W X, XU H W, LI Z Y, et al. Unsupervised anomaly detection for intricate KPIs via adversarial training of VAE[C]//Proceedings of the IEEE INFOCOM 2019 - IEEE Conference on Computer Communications. Piscataway: IEEE Press, 2019: 1891-1899.
- [32] WANG Y T, YAN M, FENG G, et al. Autonomous on-demand deployment for UAV assisted wireless networks[J]. IEEE Transactions on Wireless Communications, 2023, 22(12): 9488-9501.
- [33] SALHAB N, LANGAR R, RAHIM R, et al. Autonomous network slicing prototype using machine-learning-based forecasting for radio resources[J]. IEEE Communications Magazine, 2021, 59(6): 73-79.
- [34] GUO S Y, DAI Y, XU S Y, et al. Trusted cloud-edge network resource management: DRL-driven service function chain orchestration for IoT[J]. IEEE Internet of Things Journal, 2020, 7(7): 6010-6022.
- [35] LI Y C, LI J X. MultiClassifier: a combination of DPI and ML for application-layer classification in SDN[C]//Proceedings of the 2014 2nd International Conference on Systems and Informatics (ICSAI 2014). Piscataway: IEEE Press, 2014: 682-686.
- [36] XIE J F, YU F R, HUANG T, et al. A survey of machine learning techniques applied to software defined networking (SDN): research issues and challenges[J]. IEEE Communications Surveys & Tutorials, 2019, 21(1): 393-430.

- [37] ZHANG L M, ZHANG H, TANG Q, et al. LNTP: an end-to-end on-line prediction model for network traffic[J]. *IEEE Network*, 2021, 35(1): 226-233.
- [38] QI J Z, ZHAO Z W, TANIN E, et al. A graph and attentive multi-path convolutional network for traffic prediction[J]. *IEEE Transactions on Knowledge and Data Engineering*, 2023, 35(7): 6548-6560.
- [39] YAO H P, MAI T L, JIANG C X, et al. AI routers & network mind: a hybrid machine learning paradigm for packet routing[J]. *IEEE Computational Intelligence Magazine*, 2019, 14(4): 21-30.
- [40] LUONG N C, HOANG D T, GONG S M, et al. Applications of deep reinforcement learning in communications and networking: a survey[J]. *IEEE Communications Surveys & Tutorials*, 2019, 21(4): 3133-3174.
- [41] LIN Y D, GERLA M. Induction and deduction for autonomous networks[J]. *IEEE Journal on Selected Areas in Communications*, 1993, 11(9): 1415-1425.
- [42] BAI Y, ZHAO H, ZHANG X, et al. Toward autonomous multi-UAV wireless network: a survey of reinforcement learning-based approaches[J]. *IEEE Communications Surveys & Tutorials*, 2023, 25(4): 3038-3067.
- [43] YOU L L, HE J S, WANG W, et al. Autonomous transportation systems and services enabled by the next-generation network[J]. *IEEE Network*, 2022, 36(3): 66-72.
- [44] GHOSH S, MANDAL A K, DE S, et al. Light-weight ML aided autonomous IoT networks[J]. *IEEE Communications Magazine*, 2023, 61(6): 51-57.
- [45] CLARK D D, PARTRIDGE C, RAMMING J C, et al. A knowledge plane for the Internet[C]//*Proceedings of the 2003 Conference on Applications, Technologies, Architectures, and Protocols for Computer Communications*. New York: ACM Press, 2003: 3-10.
- [46] MESTRES A, RODRIGUEZ-NATAL A, CARNER J, et al. Knowledge-defined networking[J]. *ACM SIGCOMM Computer Communication Review*, 2017, 47(3): 2-10.
- [47] LEIVADEASA A, FALKNER M. A survey on intent-based networking[J]. *IEEE Communications Surveys & Tutorials*, 2023, 25(1): 625-655.
- [48] 郭令奇, 褚智贤, 廖建新, 等. 意图驱动的自智网络资源按需服务[J]. *北京邮电大学学报*, 2022, 45(6): 82-88.
- GUO L Q, CHU Z X, LIAO J X, et al. Intent-driven demand-aware resource service in autonomous networks[J]. *Journal of Beijing University of Posts and Telecommunications*, 2022, 45(6): 82-88.
- [49] 中国移动. 算力网络白皮书[R]. 2021.
China Mobile. Computing force network white paper[R]. 2021.
- [50] HU Y, JIA Q, SUN Q, et al. Intelligent converged computing network and its functional architecture[J]. *Computer Science*, 2022, 49(9): 249-259.
- [51] SUN W F, DUAN X Y, SHU M, et al. Research on 6G intelligent network architecture and key technologies for intelligent generation and autonomy[C]//*Proceedings of the 2023 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*. Piscataway: IEEE Press, 2023: 1-6.
- [52] 中国移动. 算网大脑白皮书[R]. 2022.
China Mobile. Computing network brain white paper[R]. 2022.
- [53] ZHAO Y, WEN P C, BAI L T, et al. Service-oriented intelligent OODA loop[C]//*Proceedings of the 2023 26th ACIS International Winter Conference on Software Engineering, Artificial Intelligence, Networking and Parallel/Distributed Computing (SNPD-Winter)*. Piscataway: IEEE Press, 2023: 182-187.
- [54] SUN Y K, LEI B, JUNIIN L, et al. Computing power network: a survey[J]. *China Communications*, 2024, PP(99): 1-37.
- [55] MCMAHON F H. The livermore fortran kernels: a computer test of the numerical performance range[R]. 1986.
- [56] LI Y, TANG Q, PENG K, et al. Research on measurement and modeling of service-centric computing power network[J]. *Information and Communications Technology and Policy*, 2023, 49(5): 21-29.
- [57] DEGILA J R, SANSO B. A survey of topologies and performance measures for large-scale networks[J]. *IEEE Communications Surveys & Tutorials*, 2004, 6(4): 18-31.
- [58] CHEN J, WANG Y L. An adaptive short-term prediction algorithm for resource demands in cloud computing[J]. *IEEE Access*, 2020, 8: 53915-53930.
- [59] 中国移动. 算力网络技术白皮书[R]. 2022.
China Mobile. White paper on computing power network technology [R]. 2022.
- [60] VARYANI N, ZHANG Z L, DAI D. QROUTE: an efficient quality of service (QoS) routing scheme for software-defined overlay networks[J]. *IEEE Access*, 2020, 8: 104109-104126.
- [61] TAGHIZADEH S, ELBIAZE H, BOBARSHAD H. EM-RPL: enhanced RPL for multigateway Internet-of-things environments[J]. *IEEE Internet of Things Journal*, 2021, 8(10): 8474-8487.
- [62] RODRIGUEZ-NATAL A, PAILLISSE J, CORAS F, et al. Programmable overlays via open overlay router[J]. *IEEE Communications Magazine*, 2017, 55(6): 32-38.
- [63] YANG H L, ALPHONES A, XIONG Z H, et al. Artificial-intelligence-enabled intelligent 6G networks[J]. *IEEE Network*, 2020, 34(6): 272-280.
- [64] MUDDINAGIRI R, AMBAVANE S, BAYAS S. Self-hosted Kubernetes: deploying docker containers locally with minikube[C]//*Proceedings of the 2019 International Conference on Innovative Trends and Advances in Engineering and Technology (ICITAET)*. Piscataway: IEEE Press, 2019: 239-243.
- [65] YANG H L, XIE X Z, KADOCH M. Intelligent resource management based on reinforcement learning for ultra-reliable and low-latency IoV communication networks[J]. *IEEE Transactions on Vehicular Technology*, 2019, 68(5): 4157-4169.
- [66] ZHANG H K, QUAN W. Networking automation and intelligence: a new era of network innovation[J]. *Engineering*, 2022, 17: 13-16.
- [67] TM Forum IG1218. Autonomous networks business requirements and framework v2.2.0[R]. 2022.
- [68] CERQUITELLI T, MEO M, CURADO M, et al. Machine learning empowered computer networks[J]. *Computer Networks*, 2023, 230: 109807.
- [69] OTTER D, MEDINA J, KALITA J. A survey of the usages of deep learning for natural language processing[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020, 32(2): 604-624.
- [70] HOGAN A, BLOMQUIST E, COCHEZ M, et al. Knowledge graphs [J]. *ACM Computing Surveys*, 2022, 54(4): 1-37.

- [71] ZHOU Z J, HU G Y, HU C H, et al. A survey of belief rule-base expert system[J]. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, 2021, 51(8): 4944-4958.
- [72] VASWANI A, SHAZEER N, PARMAR N, et al. Attention is all you need[J]. *Advances in Neural Information Processing Systems*, 2017, 30: 5998-6008.
- [73] 亚信科技, 清华大学智能产业研究院. AIGC(GPT-4)赋能通信行业应用白皮书[R]. 2023.
AsiaInfo Technologies, Tsinghua University Intelligent Industries Research Institute. a White paper of AIGC (GPT-4) empowering telecom sector[R]. 2023.
- [74] VAN d V G M, TUYTELAARS T, TOLIAS A S. Three types of incremental learning[J]. *Nature Machine Intelligence*, 2022, 4(12): 1185-1197.
- [75] HUANG L, YU W J, MA W T, et al. A survey on hallucination in large language models: principles, taxonomy, challenges, and open questions[J]. *arXiv Preprint*, arXiv: 2311.05232, 2023.
- [76] WANG Z D, JIANG Y F, LU Y D, et al. In-context learning unlocked for diffusion models[J]. *arXiv Preprint*, arXiv: 2305.01115, 2023.
- [77] GUAN Y C, ZHAO L, HU J, et al. Softwarized industrial deterministic networking based on unmanned aerial vehicles[J]. *IEEE Transactions on Industrial Informatics*, 2021, 17(8): 5635-5644.
- [78] IETF. Deterministic networking use cases:RFC 8578[R].2019.
- [79] FINN N. Introduction to time-sensitive networking[J]. *IEEE Communications Standards Magazine*, 2018, 2(2): 22-28.
- [80] 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.
- [81] 黄韬, 汪硕, 黄玉栋, 等. 确定性网络研究综述[J]. *通信学报*, 2019, 40(6): 160-176.
HUANG T, WANG S, HUANG Y D, et al. Survey of the deterministic network[J]. *Journal on Communications*, 2019, 40(6): 160-176.
- [82] 第五届未来网络发展大会组委会. 未来网络白皮书——确定性网络技术体系白皮书(2021版)[R]. 2021.
The Organizing Committee of the Fifth Future Network Development Conference. White paper on future networks - white paper on deterministic network technology system (2021 Edition)[R]. 2021.
- [83] Optical Internetworking Forum. Flex Ethernet Implementation agreement OIF-FLEXE-01.0[R]. 2016.
- [84] FARKAS J, BELLO L L, GUNTHER C. Time-sensitive networking standards[J]. *IEEE Communications Standards Magazine*, 2018, 2(2): 20-21.
- [85] YANG X, SCHOLZ D, HELM M. Deterministic networking (DetNet) vs time sensitive networking (TSN)[R]. 2019.
- [86] KROLIKOWSKI J, MARTIN S, MEDAGLIANI P, et al. Joint routing and scheduling for large-scale deterministic IP networks[J]. *Computer Communications*, 2021, 165: 33-42.
- [87] VENTRE P L, SALSANO S, POLVERINI M, et al. Segment routing: a comprehensive survey of research activities, standardization efforts, and implementation results[J]. *IEEE Communications Surveys & Tutorials*, 2021, 23(1): 182-221.
- [88] ABDULLAH Z N, AHMAD I, HUSSAIN I. Segment routing in software defined networks: a survey[J]. *IEEE Communications Surveys & Tutorials*, 2018, 21(1): 464-486. .
- [89] IEEE. IEEE standard for local and metropolitan area networks-bridges and bridged networks-amendment: IEEE 802.1Qbv[S]. 2015.
- [90] BENZEKKI K, EL FERGOUGUI A, ELBELRHITI ELALAOUI A. Software-defined networking (SDN): a survey[J]. *Security and Communication Networks*, 2016, 9(18): 5803-5833.
- [91] BOUILLARD A, BOYER M, CORRONC E. Deterministic network calculus: from theory to practical implementation[M]. Hoboken: John Wiley & Sons, 2018.
- [92] HENKE C, SIDDIQUI A, KHONDOKER R. Network functional composition: state of the art[C]//*Proceedings of the 2010 Australasian Telecommunication Networks and Applications Conference*. Piscataway: IEEE Press, 2010: 43-48.
- [93] CHENG G Z, CHEN H C, CHEN S Q, et al. How to make network nodes adaptive? [J]. *IEEE Communications Letters*, 2014, 18(3): 515-518.
- [94] CHENG G Z, CHEN H C, WANG Z M, et al. Towards adaptive network nodes via service chain construction[J]. *IEEE Transactions on Network and Service Management*, 2015, 12(2): 248-262.
- [95] 段通, 兰巨龙, 程国振, 等. 基于元能力的SDN功能组合机制[J]. *通信学报*, 2015, 36(5): 160-170.
DUAN T, LAN J L, CHENG G Z, et al. Functional composition in software-defined network based on atomic capacity[J]. *Journal on Communications*, 2015, 36(5): 160-170.
- [96] LU W, LIANG L P, KONG B X, et al. AI-assisted knowledge-defined network orchestration for energy-efficient data center networks[J]. *IEEE Communications Magazine*, 2020, 58(1): 86-92.
- [97] WANG D, ZHANG S H, SU R R, et al. An intent-based network empowered by knowledge graph: enhancement of intent translation and management function for vertical industry[C]//*Proceedings of the 2023 IEEE/CIC International Conference on Communications in China (ICCC Workshops)*. Piscataway: IEEE Press, 2023: 1-6.
- [98] CHANG X T, YANG C G, WANG H, et al. KID: knowledge graph-enabled intent-driven network with digital twin[C]//*Proceedings of the 2022 27th Asia Pacific Conference on Communications (APCC)*. Piscataway: IEEE Press, 2022: 272-277.
- [99] 北京邮电大学. 知识定义的意图网络白皮书[R]. 2022.
Beijing University of Posts and Telecommunications. White paper on knowledge-defined intention networks[R]. 2022.
- [100] TAN L Z, SU W, ZHANG W, et al. In-band network telemetry: a survey[J]. *Computer Networks*, 2021, 186: 107763.

[作者简介]



朱海龙 (1987-), 男, 山东菏泽人, 博士, 北京邮电大学讲师, 主要研究方向为工业互联网、确定性网络、工业以太网、软件定义网络、时间敏感网络和车载网络等。



杨帆 (1981-), 男, 黑龙江哈尔滨人, 博士, 北京邮电大学讲师, 主要研究方向为软件定义网络、高性能路由交换技术等。



周晓茂 (1993-), 男, 安徽阜阳人, 博士, 紫金山实验室研究员, 主要研究方向为边缘智能、算网自智、生成式人工智能等。



蒋如一 (2001-), 女, 湖南邵阳人, 北京邮电大学硕士生, 主要研究方向为算力网络、软件定义网络、确定性网络、时间敏感网络等。



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。



贾庆民 (1990-), 男, 山东泰安人, 博士, 紫金山实验室研究员, 主要研究方向为算力网络、确定性网络、边缘智能、工业互联网等。



黄韬 (1980-), 男, 重庆人, 博士, 北京邮电大学教授、博士生导师, 主要研究方向为路由与交换、软件定义网络、内容分发网络、确定性网络、算力网络等。