中国科学: 信息科学 2024 年 第 54 卷 第 5 期: 991–1024
SCIENTIA SINICA Informationis
《中国科学》杂志社
SCIENCE CHINA PRESS
6G 专项研究进展专刊 · 评述

6G 智能业务网络: 愿景、架构与关键技术

廖建新, 戚琦*, 王敬宇*, 王晶, 庄子睿

北京邮电大学网络与交换技术全国重点实验室, 北京 100876
* 通信作者. E-mail: qiqi8266@bupt.edu.cn, wangjingyu@bupt.edu.cn

收稿日期: 2023–10–31; 修回日期: 2024–03–04; 接受日期: 2024–03–28; 网络出版日期: 2024–04–30

国家重点研发计划 (批准号: 2020YFB1807800) 和国家自然科学基金 (批准号: UB23B2001, 62171057, 62101064) 资助项目

摘要 移动通信业务网络负责基础网络与业务应用之间的适配. 5G 以前的业务网络是在建好基础网络以后, 通过叠加逻辑网络方式构建. 6G 时代, "场景驱动" 将成为移动通信发展的主要演进范式, 业务网络的重要性更加突出. 6G 业务网络应与基础网络同步规划和建设, 由 6G 全场景驱动设计, 实现全场景全域按需服务. 传统业务网络仅从独立视角设计安全认证、网络控制和能力开放等机制, 难以实现细粒度、动态资源管控, 无法高效适配 6G 全场景业务应用. 为此, 本文基于国家首批 6G 重点研发计划专项 "全场景按需服务关键技术" 项目成果与当前学术研究进展, 梳理和总结了 6G 网络按需服务的挑战、6G 智能业务网络的愿景和总体架构设计, 设计了涵盖资源接入、能力开放与场景解析 3 个层次以及安全内生、智能内生和协作共享 3 个关键管控平面的 6G 智能业务网络, 实现对全域资源的细粒度智能管控. 最后结合大模型、元宇宙、绿色节能等技术与通信网络深度融合的趋势, 展望了 6G 智能业务网络更加广阔的应用发展空间.

关键词 业务网络, 按需服务, 全场景, 智能内生, 协作共享, 大模型, 6G

1 引言

从第一代移动通信 (1G) 到第五代移动通信 (5G) 的演进表现在两个方面: 一是网络技术, 向更高速率、更低延时、更多连接、更加智能的方向演进, 二是业务应用, 从仅提供电话、短信等个人基础通信业务, 向提供大量通信新业务, 支持移动互联网丰富应用以及千行百业的垂直业务应用发展. 同时, 2G∼5G 的标准化进程中, 国际标准化组织第三代合作伙伴计划 (3rd generation partnership project, 3GPP) 侧重于网络技术, 将移动通信网络横向分为无线接入网络和核心网络, 而国际电信联盟 (International Telecommunication Union, ITU) 侧重于业务应用, 将移动通信网络纵向分为基础网络与业务网络, 业务网络负责基础网络与业务应用之间的适配 [1]. 兼容并发展 2G∼5G 两类移动通信

引用格式: 廖建新, 戚琦, 王敬宇, 等. 6G 智能业务网络: 愿景、架构与关键技术. 中国科学: 信息科学, 2024, 54: 991–1024, doi: 10.1360/SSI-2023-0328
Liao J X, Qi Q, Wang J Y, et al. 6G intelligent service networking: vision, architecture, and key technologies (in Chinese). Sci Sin Inform, 2024, 54: 991–1024, doi: 10.1360/SSI-2023-0328

© 2024《中国科学》杂志社

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

网络技术理念, 2020 年 6G 重点研发计划专项 “全场景按需服务关键技术” 启动, 聚焦 6G 服务随心所想、网络随需而变、资源随愿共享的目标, 历时三年, 初步构建了面向全场景、全域知识定义的网络管控体系结构.

基础网络多代并存、庞大复杂、持续演进, 业务应用动态多变、差异巨大、难以预测, 实现两者的动态全局适配极具挑战. 传统业务网络技术存在诸多问题, 例如, 竖井型业务网络为每个通信新业务需求建设一个专用业务系统, 导致网络改造频繁, 资源难以复用, 无法支撑移动通信的大规模高速发展. 3GPP 提出开放业务平台 (open service architecture, OSA) 架构 [2], 将网络能力抽象为非定制化应用接口对外开放, 以避免网络频繁改造, 但其无法满足快速涌现、动态多变的移动互联网应用需求. 针对此问题, 作者所在课题团队提出了智能业务网络的概念 [3], 从业务应用出发, 自顶向下以全局视角研究通信网络机理, 让网络更好地理解业务需求, 使持续升级换代的基础网络能够快速、经济、灵活、高效地适配多领域不断涌现的新业务应用.

纵观智能业务网络的发展历程, 2G~3G 时代, 基于电路交换的第一代智能业务网络, 以将业务控制与呼叫控制解耦的移动智能网[1)]技术为特征, 支撑了预付费、移动虚拟专网、个性化回铃音、来电提醒等业务 [4]; 4G~5G 时代, 基于全 IP 的第二代智能业务网络, 以外挂式智能与能力开放为特征 [5], 开放了移动通信网络的中间号隐私保护、一键登录、多媒体验证码、来去电身份提示等能力, 支撑了移动互联网应用的繁荣以及大量视频类通信新业务.

在 5G 以前, 业务网络跟随基础网络演进, 目前已经落地的业务网络均是建设基础网络以后通过叠加逻辑网络的方式构建, 存在一定局限性, 导致无法进行各类资源的全局优化. 进入 6G 时代, “场景驱动” 成为移动通信发展的主要演进范式, 业务网络的重要性更加突出. 6G 业务网络应与通信网络同步规划与建设, 由 6G 全场景驱动设计, 形成对细粒度全域资源的智能管控, 实现全场景全域按需服务. 6G 全场景不只是某个场景对单一维度性能与功能的需求, 而是要求网络同时满足所有场景对资源的多维度指标的实时动态需求 [6]. 此外, 6G 空天地一体化接入, 以及人工智能应用带来算力与感知等资源需求, 将使业务特征更为复杂 [7]. 传统网络管控机制, 以及从独立视角设计的安全认证、网络管控和能力开放, 难以实现细粒度、动态资源管控. 场景驱动的关键技术之一是业务感知, 6G 网络管控基于业务场景的定义, 利用流量识别、网络探测、用户行为追踪等方法感知业务需求、状态, 挖掘业务特征与规律, 进行实现由业务场景驱动的各类管控优化策略 [8]. 为此, 本文基于国家首批 6G 重点研发计划专项 “全场景按需服务关键技术” 项目成果与当前学术研究进展, 设计了涵盖资源接入、能力开放与场景解析 3 个层次以及安全内生、智能内生和协作共享 3 个关键管控平面的 6G 智能业务网络, 指引未来 6G 网络管控的发展方向.

本文介绍了 6G 智能业务网络的演进历程、愿景和总体设计思路, 以重点研发计划项目 “6G 全场景按需服务关键技术” 核心成果为基础, 对 6G 智能业务网络涉及的安全内生、智能内生和协作共享 3 个关键技术的研究进展情况进行了梳理和总结. 首先, 第 2 节归纳了当前 6G 网络按需服务面临的挑战; 基于上述挑战, 第 3 节介绍了 6G 智能业务网络的愿景以及按需服务实现的技术路径; 第 4 节, 阐述了 6G 智能业务网络的架构设计与理论模型; 第 5 节介绍了 6G 智能业务网络的智能内生技术; 第 6 节介绍了 6G 智能业务网络的协作共享技术; 第 7 节介绍了 6G 智能业务网络的安全内生技术; 第 8 节介绍了原型系统对 6G 智能业务网络机理的初步验证结果; 第 9 节展望了 6G 智能业务网络应用发展前景; 最后, 第 10 节总结回顾了全文内容.

1) 移动智能网: 2G 时代智能网技术在移动网中的应用. 在以电路交换为基础的移动网上叠加以计算机系统实现的智能网, 后者通过与移动交换机的信令交互执行预定义的业务逻辑, 实现 “增值业务”.

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

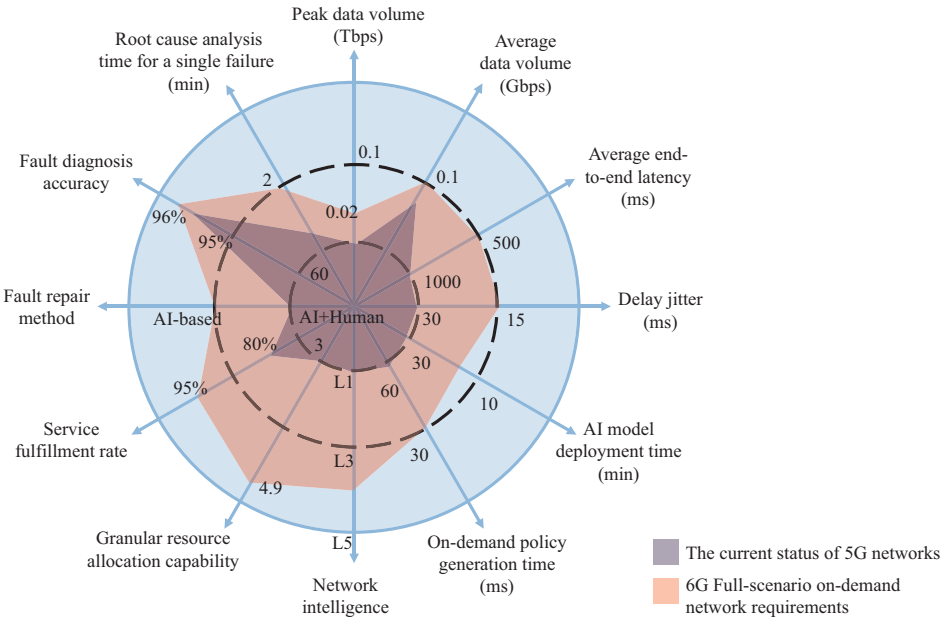


图 1 (网络版彩图) 6G 全场景按需服务能力要求

Figure 1 (Color online) 6G full-scenario on-demand service requirements

2 6G 网络按需服务的挑战

2023 年, ITU-T 联合 IMT 2023 推进组在 “6G 全球愿景框架与建议书” 中从通信增强扩展场景、覆盖增强新增场景、业务扩展新增场景 3 个方面, 确立了 6G 的六大应用场景, 即沉浸式通信、超大规模连接、超可靠低时延通信、泛在连接、通信 AI 一体化、通信感知一体化, 并提出改善数据速率、区域流量容量、连接密度、时延和可靠性 [9] 的性能指标. 为了支持上述六大场景, 6G 需要实现峰值数据速率、用户体验数据速率、频谱效率、区域流量容量、连接密度、移动性、时延、可靠性与安全性、隐私性与弹性等增强能力以及覆盖、定位、感知相关、人工智能、可持续性与互操作性等新增能力. 上述六大场景与性能指标对 6G 网络提出了巨大的挑战, 需要 6G 在理论、技术等方面进行全面突破.

通信网络持续升级换代, 多领域新业务应用需求不断快速变化, 必须在同时满足现网稳定运行、已有业务不受影响、新业务应用大规模快速低成本提供的约束条件下, 实现通信网络与新业务应用的智能适配. 目前欧盟 Hexa-X 计划与美国 6G 联盟, 也在 6G 研究计划与白皮书中提出 6G 按需服务与场景的研究计划与构想. 欧盟提出基于意图的网络与服务管理、服务解耦以及动态切片等思路, 期望以便捷、灵活、高效的方式实现资源适配; 美国 6G 联盟提出沉浸式、个性化的服务需求以及自主系统与资源的最佳匹配问题. 与欧洲和美国战略目标一致, 以业务网络视角重新审视 6G 网络按需服务, 则发现要满足全场景按需服务的要求, 需要更高级的智能化管控手段以及细粒度的全局资源适配.

针对当前的 6G 愿景与 6G 性能指标要求 [10], 要满足全场景按需服务, 对于 6G 网络管控提出了指标能力要求, 如图 1 所示, 包括峰值速率、平均带宽、端到端时延、时延抖动等网络性能指标, 以及 AI 模型部署时间、按需资源调配策略的生成时间、单次故障根因分析时间、故障诊断正确率、业务服务满足率等网络智能化等定量指标, 以及网络智能性 (L1~L5)、资源细粒度分配能力 (1~5 分)、故障修复方式 (AI+ 人工、AI 为主). 上述指标涉及多个维度, 要实现全场景按需服务, 面临以下 3 项挑战:

(1) 高度实时性. 6G 网络中用户大量增长、新业务不断涌现导致网络资源复杂度剧增, 为实现业

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

务与网络实时适配, 发展大规模、支持极致性能的业务网络理论是必要的. 6G 业务相对于 5G 而言, 沉浸式与交互类业务均要求网络提供高度实时性, 例如远程无人机控制, 需要时延低于亚毫秒级, 沉浸式 XR 业务需要端到端实验低于 20 ms. 6G 边缘学习是保障高度实时性与全面智能化的重要支撑, 将统筹集成感知、通信、计算等资源, 对全网调度决策的实时性提出更高要求 [11, 12]. 传统马氏决策过程适用于轮转式的交互环境, 难以准确刻画高动态节点的行为, 提供实时决策理论. 需要支持机器学习分布式、自主演化的新理论, 在高度实时性约束下, 实现业务网络原生智能的在线学习、自我成长.

(2) 复杂关联性. 6G 存在通信、计算、存储、感知、控制等多维度资源, 6G 智能业务网络要有机融合移动通信网络与群体智能, 需要继续发现合作机制、探索网络潜力, 发展自组织智能业务网络. 传统博弈论已经广泛应用于各类竞争场景的激励结构描述, 但是大规模业务网络的智能节点的行为会导致维度灾难与有限控制资源之间的矛盾, 群智感知和博弈论新理论亟待突破. 例如, 从易辛模型入手将数量巨大互相作用的多体问题转化成每一个节点在弱周期场中的单体问题.

(3) 高维动态性. 6G 空天地一体化的业务场景中, 终端、网络与业务都具有不同程度的移动性, 同时五类资源相互影响. 6G 智能业务网络中网络拓扑和资源的动态变化, 需要兼顾智能管控成本与智能适配效果, 发展通用网络认知理论. 当业务网络的节点具有多维度的连接关系, 且动态多变的情况时, 业务网络行为具有不确定性, 难以使用传统的图模型理论难以用于对其进行分析. 因此, 在全维关系中动态认知与抽取识别节点关系, 解决业务网络相关连接的复杂拓扑结构问题, 捕获连接的时空动态性, 挖掘节点和边的有效信息, 感知网络节点之间信息特征的传递, 面向业务划分子网, 适应动态拓扑以及可认知、可适应关系的动态变化, 传统图理论亟待突破.

3 6G 智能业务网络愿景

3.1 基于意图的需求理解

6G 移动通信的技术愿景是服务随心所想、网络随需而变、资源随愿共享, 支持六大应用场景, 以及未来场景的扩展需求. 6G 智能业务网络的愿景则是通过逐层映射与跨层管控, 自主理解用户需求, 自动发现网络故障, 优化全域资源调配, 实现全网无人值守的运维. 6G 智能业务网络从业务应用的角度出发, 自顶向下以全局视角研究通信网络机理, 让网络更好地理解业务需求, 使持续升级换代的通信网络能够快速、经济、智能地适配多领域不断涌现的新业务应用. 针对 6G 沉浸化、交互式与智慧化的场景需求, 智能业务网络应支持意图驱动的业务适配与应用提供.

意图是指对网络资源与功能各类需求, 来源于个人用户、企业用户、网络运维人员以及全网相关模块的接口调用等, 可定义为在高级操作和业务目标中编写的一组特定策略类型. 意图驱动的主要思想是用户只需给出特定的业务目标 (希望网络达到何种状态), 而不需详细说明如何实现这些目标, 使用户对需求的输入更加明确和简化. 意图表达的过程无需知晓底层硬件, 并支持跨网元、跨网络移植. 业务网络通过学习意图的上下文, 解析意图, 将其翻译为网络策略与设备指令, 部署非冲突的业务应用并提供相应的网络资源, 满足用户个性化的性能需求.

目前, 已有诸多文献对意图网络 (intent-based networking, IBN) 给出了不尽相同的定义 [13, 14], 即: 向用户提供一个简单的意图接口, 用户通过这个接口描述他们需要网络达到的状态 (而不需描述如何实现这个状态), 网络自动将其转化为网络策略和行为, 满足用户的需求. 业界提出了多种意图语言、框架和编译器, 以在网络设备和中间件中实现基于意图的配置, 并支持高层策略的组合, 辅助网络管理员实现管理行为的抽象 [15]. 文献 [16] 将自然语言意图转化为网络配置, 使用一种名为 Nile 的结

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

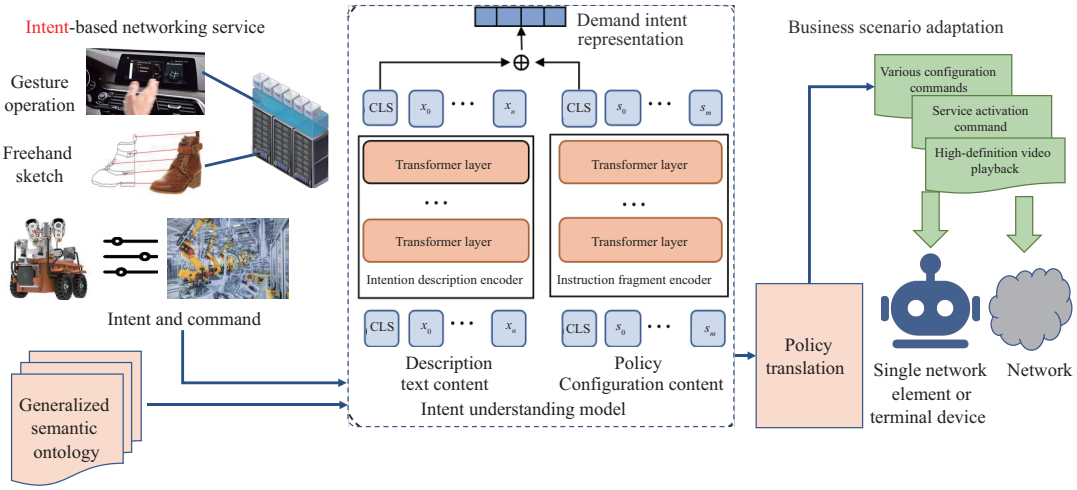


图 2 (网络版彩图) 基于意图的 6G 智能业务网络愿景

Figure 2 (Color online) Intent-based 6G intelligent service network vision

构化语言, 利用深度神经网络将提取的实体转化为高级结构化的网络意图程序, 再根据目标网络将提取的意图程序编译成网络策略. 清华大学徐明伟教授团队也采用意图驱动的网络故障监控 [17]、配置生成 [18]、路由策略变更 [19] 等解决云网络基础设施上的运维问题. 3GPP SA5、欧洲电信标准化协会 (European Telecommunications Standards Institute, ETSI)、体验式网络智能组 (Experiential Networked Intelligenc, ENI) 等标准化组织将持续这方面工作 [20,21], 例如定义合理精准的意图表达、自动化机制、意图的生命周期管理等.

6G 智能业务网络将对 IBN 进一步深化, 丰富意图的输入形式, 扩展更符合 6G 场景的输入, 即通过文字、语言、动作、绘画或机器指令表达需求. 例如, 如图 2 所示, 通过手势表达车辆自动驾驶的网络需求, 利用草图绘制表达网络规划需求, 通过语音或文字输入某个应用的网络带宽需求等. 同时, 支持网络自主习得的潜在意图, 即当感知到网络运行为次优状态、预测异常或出现故障时, 对网络进行配置优化, 以无人值守的方式实现网络实时优化. 然而, 意图是一种抽象、高层次策略, 而部署到网络则需要 “描述如何管控网络资源” 的低层次策略. 基于意图的需求理解从高层次意图中抽取需求, 参照网络状态、已有网络策略、业务场景知识等, 准确完整地生成新策略. 面向 6G 全场景按需服务, 作者所在团队前期工作中提出了基于意图的网络管控策略, 对策略建立统一的模型, 在不完备的自然语言意图中提取高层次策略, 基于网络管控知识细化推理补全为低层次策略, 实现了 “不完备意图 – 声明性策略 – 确定性策略 – 命令性策略” 流程, 克服了自然语言意图的语义意图不完备性, 实现云网络自智 [22]. 6G 网络随需而用愿景的实现, 需要基于潜在意图对网络进行动态部署与配置, 例如, 基础网络中随时加入新设备时, 智能业务网络负责高效准确地下发设备配置, 将其纳入全域资源. 团队前期与华为公司合作提出了网络配置管理框架 Nassim 与网络配置语义模型 NetBERT, 自动化现有网络 “消化、理解、吸收” 新设备的过程, 让智能业务网络 “读懂” 设备配置说明书, 把运维工程师从繁琐易出错的工作中解放出来, 大幅提升网络运维的效率 [23].

基于意图的网络管控, 将在今后十年持续演进和成熟. 基于意图的需求理解长期来看是 6G 智能业务网络的愿景, 短期而言则可为运营商在网络服务开通、切片资源优化、切片性能保障、网络规划、网络容量管理、网络功能部署等方面, 进一步减少管理复杂度和对底层设备的知识需求, 同时在跨多厂商的场景提升网络管控效率.

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

3.2 按需服务的动态切片

由于新兴服务种类繁多、相互冲突且需求极其多样化, 从 5G 移动通信发展之初, 传统的 “一刀切” 网络方法已不再有效. 受益于软件定义网络 (software defined networking, SDN) 和网络功能虚拟化 (network function virtualization, NFV) 技术的发展, 网络切片作为当前 5G 系统的一项关键技术 [24], 为低时延、高带宽与大连接三大场景的业务分配通信与网络资源. 网络切片通过将一个物理网络划分为多个专用、虚拟化、相互隔离的逻辑网络, 同时为不同的应用场景提供按需定制的服务. 网络切片包括接入网切片 [25]、边缘网切片 [26]、核心网切片 [27], 以及端到端切片等 [28]. 对于给定的需要提供切片的服务, 网络编排系统需要建立跨越无线接入网、传输网与核心网的端到端网络切片. RAN 侧的网络切片主要以划分无线频谱的方式为不同服务提供切片资源, 传输网侧的网络切片主要以划分传输带宽的方式提供切片资源, 而核心网侧的网络切片主要以虚拟网络功能 (virtual network function, VNF) 编排和调度的方式提供切片资源. 在端到端切片资源调度方面, 文献 [29] 提出了包含流量预测、切片接纳控制和切片流量调度等机制的网络流量切片框架. 文献 [30] 提出了一种优化、快速的实时资源切片框架, 在考虑租户资源需求不确定性的前提下, 利用深度强化学习 (deep reinforcement learning, DRL) 进行实时快速资源分配, 使网络提供商的长期收益最大化.

全场景按需服务的愿景下, 6G 智能业务网络应支持六大场景下更为广泛的垂直应用 (例如智慧城市、工业 IoT、车联网以及 AR/VR), 实现网络资源高效、动态、自主的调配, 并保证满足服务水平协议 (service level agreement, SLA). 与传统的服务不同, 新兴应用具有高度多样化和极致性能要求 (例如吞吐量、时延和可靠性), 基础网络需要在可扩展性、可用性和成本效益方面适应服务需求. 由于与场景相关的网络服务可能具有完全不同的性能指标与功能特性, 且需要在多个维度同时满足要求, 将极大地增加业务网络的复杂性, 必须对其进行精细化管理以保证其能够提供可接受的服务水平. 6G 时代将进一步深化网络切片技术, 使其满足 6G 场景下不同网络服务和应用的多样化需求. 例如, IEEE Fellow 沈学民院士团队提出 6G 人工智能业务的 AI 原生网络切片, 适用于 AI 业务的实例构建和 AI 业务的高效资源管理, 并将 AI 技术集成到下一代无线网络的网络切片中 [31].

网络切片作为网络管理的一项重要功能, 通常具有静态和长周期的特点, 网络控制 (如路由) 一般是指实现通信网络控制平面指令下发的功能, 通常具有动态实时特性. 由于 6G 大量新兴业务应用涌现, 网络管理的周期越来越短. 例如, 低时延高可靠通信 (ultra-reliable low-latency communications, uRLLC) 由于其显著的突发特性, 对数据传输提出了严格的时延和可靠性要求, 需要细粒度的资源管理周期而不是粗粒度的管理 [32]. 因此, 网络管理和网络控制之间的界限越来越模糊, 开始出现相互融合的趋势. 然而, 在过去的网络资源管理中网络的管与控是孤立的, 限制了 6G 网络的统一管控能力, 导致其难以应对网络高度动态变化的资源调度以及满足 6G 网络中新兴用例的严格要求. 作者所在团队在前期工作中, 提出一种将网络切片 (网络资源管理功能) 与路由 (网络控制功能) 相结合的联合优化框架 [33], 同时创建一组服务的多个网络切片, 根据 SLA 将基础设施的资源分配给切片; 同时, 每个网络切片中数据包/流量将根据实时网络状态被路由到一组传输路径中, 以最大限度地发挥网络基础设施的效用.

为了进一步优化切片性能, 在多种具有低时延需求的场景下部署边缘切片, 然而, 边缘网络基础设施的发展与业务需求之间存在较大差距, 有限的边缘网络资源很难完全满足所有用户. 当服务提供商同时响应其租户的多种切片服务时, 其独立资源决策不可避免地会导致多个服务提供商之间的资源竞争甚至冲突, 直接导致切片部署失败, 从而影响网络切片服务的交付. 为此, 团队设计了并行切片资源协调框架, 实现多服务提供商网络切片部署, 提高资源效率 [26].

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

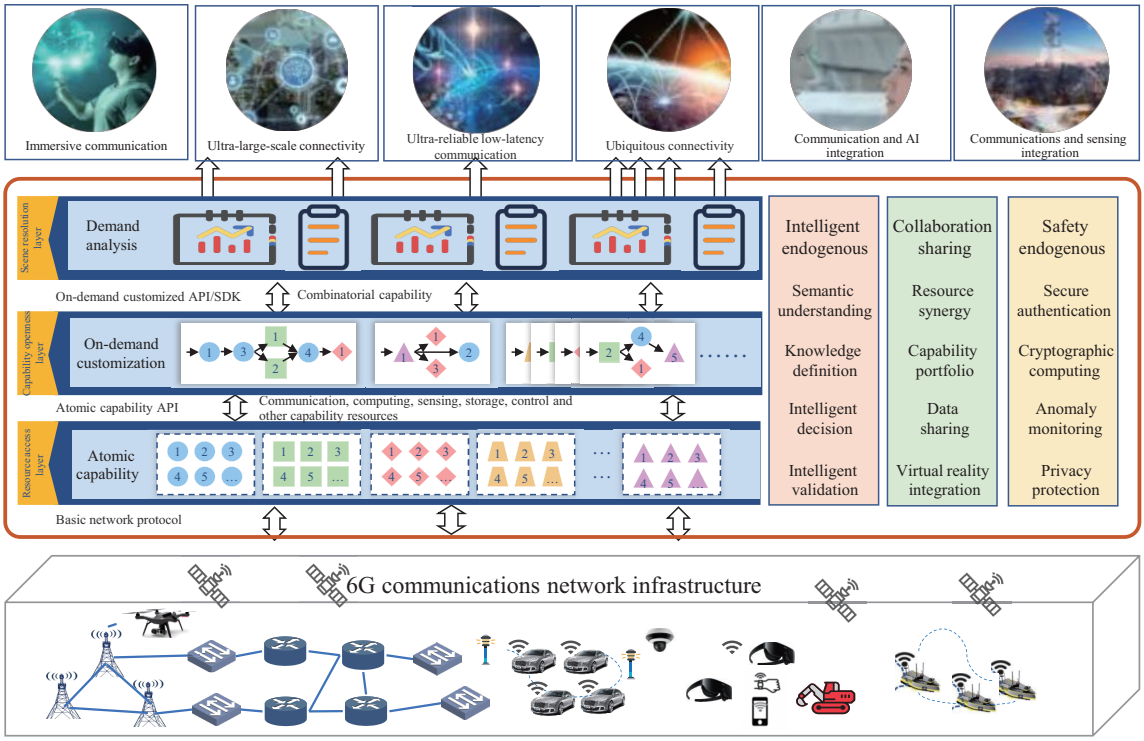


图 3 (网络版彩图) 6G 智能业务网络体系结构

Figure 3 (Color online) 6G intelligent service network architecture

针对较为复杂的 6G 应用场景, 例如自动驾驶场景与工业控制场景, 涉及多个不同特性的业务, 需设计多个隔离的切片, 且不过度分割网络资源, 满足场景需求. 团队针对 V2X 场景, 设计了动态按需服务切片, 包括基础道路安全服务切片、增强道路安全服务切片和无关道路安全服务切片, 分别用于实现基本驾驶信息传递、高水平自动驾驶以及提高驾驶舒适性和效率 [34]. 考虑到单车计算能力有限和云计算传输时延较长, 且单一的计算制式难以满足每个切片的指标需求, 提出了利用多层计算的系统级架构设计自动驾驶网络切片, 将计算和通信能力分布在云和终端用户之间的任何地方. 考虑车辆由于切片的固有特性使其对多维资源的需求呈现锯齿状, 团队基于业务流量与物理资源之间的时空相关性提出了一种双时间尺度的智能资源管理方案.

4 6G 智能业务网络架构

4.1 6G 智能业务网络三层架构

6G 基础网络将向立体空间扩展, 覆盖空、天、地、海等所有人类活动的空间区域, 并针对任意具有通信需求的情景, 利用空域、时域、频域等资源以及接入域、传输域、业务域、控制域、管理域等功能层面. 在 AI 赋能网络理念的推动下, 6G 网元异构性强, 智能化程度高 [35]. 面向具有复杂巨系统特性的复杂 6G 网络 [36], 以叠加网络形式呈现的 6G 智能业务网络, 应屏蔽基础网络差异, 为 6G 全场景业务提供感知、通信、认知、计算、控制一体化资源的按需服务. 6G 智能业务网络实现了全域立体接入资源的收敛、抽象、封装、编排组合和统一开放, 包括网络资源接入层、业务能力开放层、场景解析层的架构, 可开放更丰富、更细粒度的能力, 实现资源的充分共享, 提供多粒度的能力开放接口, 快

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

速低成本地满足多领域新业务应用对能力的差异化需求. 智能业务网络的体系架构分为 3 层, 如图 3 所示, 分别为:

(1) 资源接入层. 实现 6G 移动通信网络异构网元资源统一接入, 以轻量级网络协议连接网元, 屏蔽其具体部署和实现细节, 将其提供的通信、计算、感知、存储、控制等资源封装为丰富的原子能力调用接口. 原子能力抽象出了独立、稳定、与场景无关的细粒度的能力, 具有很好的可重用性, 适合快速提供规模不大的通信业务. 同时, 资源接入层还提供基础网络状态感知与监控功能.

(2) 能力开放层. 将细粒度的基础网络原子能力提供给能力开放平台及运营商自有通信业务平台使用, 并且接受能力开放平台的能力调用任务. 基于资源接入层获取的实时状态数据和历史数据, 通过一系列多态性认知服务接口, 根据不同场景, 智能调度与配置原子能力, 实现全域网络资源按需调配. 支持复合能力动态编排, 将原子能力编排组合成复合能力 (可称为业务能力), 在业务运行态, 根据网络状态实时动态组合原子能力, 满足多样化的业务应用需求. 例如, 自动驾驶场景中, 针对道路场景识别的推理任务, 不同车速下, 多态性服务接口支持不同响应时间的推理.

(3) 场景解析层. 对接互联网、行业企业第三方应用平台以及运营商自有通信新业务平台, 以及个人用户、家庭用户等, 将 6G 六大场景按照具体业务需求 (包括功能与性能), 分解为网络能力调用关系, 针对具体应用场景和特定需求, 构建半结构化的能力需求说明, 包括对调用下层原子能力或组合的具体功能描述以及性能要求, 可实现在原子能力基础上组合编排而成的粗粒度的定制化能力, 具有动态灵活、简单易用、针对性强的优点, 适于广泛的普通个人用户与行业企业等领域第三方应用使用.

为了实现 3 层结构动态映射与智能适配, 6G 智能业务网络具备 3 个内在纵向管控平面, 以数据获取和策略部署方式与 3 个层次交互. 智能平面包括流量感知、知识定义、智能决策与智能验证等功能; 协作共享平面包括资源协同、能力组合、数据共享、虚实融合等功能; 安全平面包括安全认证、加密计算、异常检测、隐私保护. 纵向管控平面的功能可以系统插件形式部署于智能业务网络, 具备扩展与后向兼容特性. 各项策略由 3 个纵向管控平面管控功能根据收集到的网络状态信息动态生成, 部署于资源接入、能力开放与场景解析层, 最终网元控制指令下发至基础网络的各节点中. 对于庞大的 6G 网络而言, 复杂 AI 模型的训练和推理过程也将耗费大量的时间, 采用分布式 6G 网元部署管控功能, 例如边缘智能节点, 以实现对网络资源的实时精准控制. 此外, 6G 智能业务网络可作为全网统一的能力开放门户, 连接无人机、工业互联网等多个自主网络接入平台, 根据多样化的场景需求对原子能力进行封装和动态编排组合, 形成灵活的复合能力接口, 统一对外开放, 同时面向全场景应用, 实现鉴权、计费、安全管理等功能, 保证能力被安全可控地调用.

6G 智能业务网络具有以下特征: (1) 逐层映射, 将网络功能、业务系统进一步解耦, 按照接入、网络与业务逐层抽象资源 (通信、计算、感知、存储、控制), 再将资源封装为能力. (2) 多粒度封装, 通过虚拟机、容器以及函数级服务提供不同粒度的服务能力, 并支持将原子能力动态组合为复合能力, 便于资源快速流动与精准按需适配. (3) 精准需求理解. 分析挖掘业务场景特征, 针对业务场景提供按需服务, 例如, 捕捉用户视角, 根据目光位置调整局部直播码率为高清. (4) 实时资源调配, 资源调配的时间周期短, 提供超低时延管控, 新策略生成时间短.

4.2 6G 智能业务网络建模

6G 业务网络的 3 层架构逐层抽象, 涉及网络切片与多层虚拟化技术, 网络智能管控粒度的不断细化, 且具有多层基础网络、多层虚拟机/容器/微服务调用链网络 (overlay) 等多层管控视图, 各个层级的业务应用、网络虚拟资源和网络物理节点与连接的关系也变得更加复杂. 6G 业务网络 3 层架构的功能模块共享基础网络的配置环境, 具有强依赖性, 业务网络行为模式复杂和业务特征难以预测, 同

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

时, 不同的管控任务中上下层之间以不同的资源粒度交互关联, 传统管控以单独一层的视角建模分析. 例如, 当网络性能突然下降时, 传统有向无环图的建模方式, 管控任务割裂, 无法将跨层的故障传播过程当作整体进行建模. 6G 业务网络需要具有跨域能力和全局视角, 且适合数据与知识驱动的建模方式, 从而动态适应动态扩容更新的网络环境. 超图是图论中的一个扩展概念, 允许边连接任意数量的节点, 而不仅仅是两个节点之间的连接, 从而能够更好地描述复杂的关系和依赖结构. 超图在通信网络已有一些应用, 文献 [18] 采用决策树和超图相结合将基于深度学习的网络控制策略转化为可解释的规则控制, 文献 [37] 基于超图结构建模分布式集群中的网络、存储以及 CPU 与 GPU 等种资源的调度过程. 使用超图 (hypergraph) 对 6G 智能业务网络与基础网络进行建模, 可以将不同层级、不同模态的网络状态信息进行融合并统一形式化表示. 例如, 网络路径信息和不同节点上部署的网络功能这两种模态的信息, 可以使用超图进行建模. 一旦超图构建完成, 还可以灵活地从超图中提取不同的子图, 以针对不同的网络管控任务提供基础的图结构信息.

超图可表示为 $G(V,E)$, 其中 $V=\{v_1,\ldots,v_n\}$ 表示超图中的 n 个节点集合, $E=\{E_1,\ldots,E_m\}$ 表示超图中的 m 条超边集合, E_α 表示超边 α, 是一个无序节点集合. 超边可以连接多个顶点, 而不仅仅是两个顶点之间的连接. $|E_\alpha|$ 为超边 α 的尺寸, 即集合中点的个数. 当任意超边 α 均符合 $|E_\alpha|=2$ 时, 则超图退化为普通图. 6G 智能业务网络中, 超边中的节点集合, 可以有效表示多层叠加网络的虚拟节点或原子能力.

超图的关联矩阵 H 定义如下:

$$h_{ia}=\begin{cases}1, & i\in E_\alpha,\\ 0, & \text{otherwise.}\end{cases}\tag{1}$$

超图的邻接矩阵 A 可定义为 $A=HH^{\mathrm{T}}$, 其中 A_{ij} 表示节点 i 和 j 共享超边的数量. 同时可以定义超边矩阵 C 为 $C=H^{\mathrm{T}}H$, 其中 $C_{\alpha\beta}$ 表示超边 α 和超边 β 共有的节点数量, 超图可以通过团式或者星式展开的方法转化为普通的成对图. 用 V_i 和 v_i 表示其第 i 个节点和相关属性向量, 用 E_j 和 e_j 表示其第 j 个超边和相关属性向量. 因此, 一个带属性的超图可以用一个五元组 $G=(V,E,\overline{V},\overline{E},g)$ 表示, 其中

- $V=v_i,1\leqslant i\leqslant n$, 表示节点的集合;
- $E=E_j,1\leqslant j\leqslant m$, 表示超边的集合;
- $\overline{V}=v_i,1\leqslant i\leqslant n$, 表示与每个节点相关的属性向量的多重集合;
- $\overline{E}=e_j,1\leqslant j\leqslant m$, 表示与每个超边相关的属性向量的多重集合;
- g 表示与整个超图相关的全局属性向量.

定义1 (K-均匀超图 K-uniform hypergraph) 每个超边与恰好 K 个节点相关联. 6G 智能业务网络中, 可将基础网络资源, 建模为 K-均匀超图, 每个超边包含 K 个相同功能、不同性能的原子能力.

定义2 (无向超图 undirected hypergraph) 每个超边是 V 的子集, 例如, $E_1=\{V_1,V_2,V_3\}$;

定义3 (有向超图 directed hypergraph) 每个超边是包含来自 V 的元素的元组, 例如, $E_1=(V_1,V_2,V_3)$.

图对齐 (graph alignment, GA), 也被称为网络对齐 (network alignment, NA), 是图数据挖掘中的一项重要任务, 旨在寻找两个图中的节点对应关系. 例如, 给定两个图 $G_1=(V_1,E_1)$ 和 $G_2=(V_2,E_2)$, 其中 V_1 和 V_2 分别表示两个图的节点集合, E_1 和 E_2 分别表示两个图的边集合, 图对齐的目标是找到一个节点映射函数 $f:V_1\to V_2$, 使得两个图中节点匹配的相似度最大化. 传统图对齐通常用于知识图谱融合、事实检验、社交网络分析等任务中 [38]. 由于传统图对齐任务中图节点规模较大, 但边的属性

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

不强, 6G 智能业务网络是通过 3 层 3 面架构提供全场景按需服务的过程, 可建模为超图的图对齐任务, 寻找复杂 6G 基础网络资源与上层业务场景需求之间的最佳映射.

定义4 (超图对齐 hypergraph alignment) 给定两个超图 $G_1 = (V_1, E_1)$ 和 $G_2 = (V_2, E_2)$, 超图对齐任务的目标是找到一个映射函数 $f_V : V_1 \to V_2$, 将 G_1 的节点映射到 G_2 的节点, 以及一个映射函数 $f_e : E_1 \to E_2$, 将 G_1 的边映射到 G_2 的边, 从而最大程度地匹配两个超图的对应节点和对应边.

超图对齐通常被当作最优化问题, 一般包括 3 个步骤. 第一, 特征提取: 从每个超图中提取节点和边的特征表示, 包括节点和边的属性、结构信息等. 第二, 相似度计算: 计算两个超图中节点和边之间的相似度或匹配程度. 第三, 匹配优化: 使用优化算法或匹配算法, 根据节点和边之间的相似度进行匹配优化. 通常会转化成一个组合优化问题, 通过最大化匹配的相似度得到最佳的超图对齐结果.

当前最有效的一种做法, 是将超图匹配问题转化为节点分类问题, 利用深度神经网络解决. 为了匹配超图 G_1 和 G_2, 首先在其上构造一个关联超图 G_a, 并将匹配问题转换为 G_a 上的节点分类问题. 然后通过预定义的流程来解决分类问题, 包括几个超图神经网络模块 (编码器模块、核心模块和解码器模块). 首先, 编码器模块对每个属性使用不同的映射函数将 G_a 从输入空间转换为嵌入空间. 然后应用核心模块多次更新嵌入空间中 G_a 的状态, 其中超边更新函数、节点更新函数、全局属性更新函数依次应用. 最后, 解码器模块利用节点的映射函数, 将关联超图从嵌入空间转换到期望的概率空间. 另一种有效的方法是将进化博弈论运用到关联超图 G_a 上, 将匹配问题转换为收益最大化博弈. 博弈进化过程是一种典型的选择机制, 一方面为参与方保留了盈利的策略, 即保留并优化最优的对齐测率, 另一方面则导致有害的参与方灭绝, 在迭代过程中筛选掉错误的对齐方案. 博弈进化的最终结果达到纳什均衡, 使所有博弈参与者的整体收益最大化, 使整体匹配方案收益最高. 6G 智能业务网络利用超图对齐的最终目标也是在尽最大可能满足业务需求的同时, 使全域网络获得最佳收益.

5 6G 业务网络的智能内生

人工智能具有强大的学习能力和推理能力, 这使得网络能够在不需要人为干预的情况下学习和适应支持不同的业务. 任何传统上可以用策略规划和优化问题公式来解决的任务, 从理论上都可以用强化学习解决. 目前深度强化学习已广泛应用于各层面的决策控制问题. 深度强化学习通过探索与反馈, 可学习如何解决复杂的网络管理和控制问题, 例如, 流量调度和资源分配等. 6G 全场景需求以及全域复杂网络环境, 依赖人工智能模型准确识别用户需求、场景语义和用户特征, 对服务需求及其变化进行细致的感知, 并通过规则匹配或近似推理进行决策及自适应动态调整, 即通过人工智能实现网络对用户需求的精确感知和与网络资源的精准匹配. 因此, 人工智能模型在 6G 时代将渗入网络的各个层次与各类机制中, 不同算力的网元将部署不同量级的智能模型, 提供决策与服务.

6G 智能业务网络从网络管控理念入手, 突破网络管理域限制, 深入网络控制层面, 按需提供有效资源, 以确保用户可感受到基础网络通信的极致性能. 内生智能平面, 利用人工智能方法在 6G 网络全场景全生命周期闭环自治中, 实现网络的自配置、自适应、自修复和自演进. 同时, 在当前数据共享与跨层协作的思想下, 数据、算力与存储需求较高的人工智能模型, 将以分布式学习与推理架构部署于 6G 网元中, 对外提供 AI 服务, 并支持动态高效的 AI 服务适配、AI 服务迁移、AI 服务协作和 AI 服务演化, 实现真正的内生智能.

5.1 知识增强的全域流量感知

流量感知是网络管控的基础, 对于 6G 智能业务网络资源调配至关重要. 基于网络测量技术, 智能

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

化的全域流量准确感知, 通过历史数据和采集的网络信息, 提取流量特征用于网络管控, 可有效节省网络测量代价. 深度学习具有从复杂原始数据中学习抽象特征的能力, 适用于处理不规则、大规模的非线性问题, 在许多复杂问题上表现出无可比拟的优势. 在 6G 网络资源智能调配中, 通过建立基于深度学习的通用流量感知模型实现对网络流量的精准感知, 为后续的网络知识获取和资源调配策略生成提供知识输入.

目前已有工作利用各类深度神经网络对网络流量特性进行建模. 例如, RouteNet [39] 基于图神经网络 (graph neural networks, GNN), 从输入的网络拓扑与流量矩阵中, 学习路径性能与链路性能, 预测流级别的端到端性能指标, 具有较高的精度与泛化性. RouteNet-Fermi 针对训练过程中不可见的网络拓扑, 进行精准的预测, 支持多种流量模式, 如恒定比特速率、突发流量模式、ON-OFF 模式, 以及设备端口上的各种转发策略. MimicNet [40] 模型聚焦于数据中心网络的性能指标评估, 将深度神经网络的模拟范围从整个网络缩小到集群维度, 减小了模型的黑盒特性, 并增加了可解释性. 利用深度神经网络的泛化性, MimicNet 实现了网络拓扑规模的泛化性, 其在较大规模的网络拓扑下进行一次性能评估的速度相比于网络模拟器要快出两个数量级. 与此同时, 该模型针对尾时延、往返时延、流完成时间、吞吐量等流级别指标进行了预测. 在此基础上, DeepQueueNet 模型 [41] 利用 Transformer 模型与大量的真实流量特征训练模型, 实现了与网络模拟器相似的精度, 并具有更快的预测速度. DeepQueueNet 可预测网络设备在各种转发策略与不同流量输入下的性能指标, 包括数据包级别的时延、抖动、丢包率、吞吐等, 并整合多个 Transformer 模型进行联合预测, 评估整个网络在各个维度上的性能指标.

6G 智能业务网络承载的业务应用纷繁复杂, 针对不同的业务属性和内容, 需要依据多种类型、多个维度的流量特征进行识别. 因此, 知识增强的全域流量感知将建立通用多任务流量识别感知模型, 同时支持不同业务场景与性能指标, 减少模型部署对计算资源的消耗, 提高感知效率. 依据特征之间的相互关系, 判断特征辨别力强弱, 通过识别准确率进一步检验特征的流量识别效果, 从而迭代地对特征进行选择. 图 4 所示为通用流量特征选择的具体流程, 通过控制器获取多个流量的基本特征数据, 进而获取多种特征的特征值, 经过人工标注流量类别后获取训练数据. 在训练数据中每条流量的特征可组成一个特征向量, 如流 $x_1 = [f_1, f_2, L, f_n]$. 对不同类型流组成的流量特征矩阵进行分解, 计算出每种特征 f_i 的辨别力能力值, 依据每个特征对应的辨别力能力值的大小进行排序, 并随机选择辨别力能力值的上下限值来选取特征作为被选择的特征集输入. 辨别力能力值的上下限值通过随机梯度下降 (stochastic gradient descent, SGD) 计算, 以提高流量识别准确率为目标进行求解.

针对通用流量感知模型的建立, 依据输入的流量特征, 利用深度学习获取不同的特征表达模式进而进行流量识别. 由于不同的流量类别会共享部分相同的流量特征表达模式, 因此, 多维度流量之间存在较强的相关关系. 将流量特征输入深度神经网络, 利用多维度流量类型进行学习, 进而获取多个维度的识别结果, 相较于利用多个分类器识别可以提高在线识别的效率.

5.2 知识可增量学习的网络知识获取

在 6G 智能业网络的资源调配中, 知识的获取是一个重要的过程. 将按需服务的相关网络规律、机理、策略凝练为知识, 构建全域资源调配的知识空间, 实现特征共享、模型公用、策略互通. 同时, 网络通过不断学习, 让知识空间更加丰富. 6G 知识空间中的知识有两种来源: 一类是来自于历史积累的人类经验知识, 例如, 手册、说明书、文档, 另一类则是网络运行过程中产生的大量动态数据及其抽象形式, 例如, 流量分类模型、网络异常检测模型. 知识获取或知识生成过程, 通过人工或自动方式收集大量网络管控相关的数据, 按照既定的方式对这些数据进行分析处理, 并按照既定的表征方式存储到知识空间. 基于输入数据及获取知识类型的不同, 知识获取的方法及表征方式也有所不同, 但这两种

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

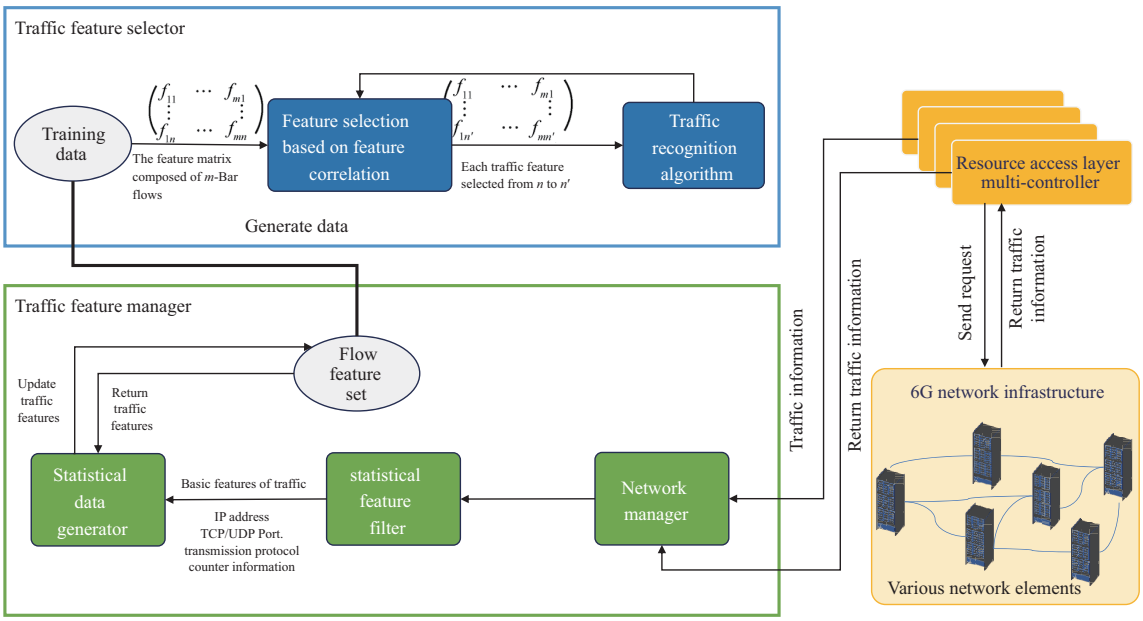


图 4　(网络版彩图) 通用流量特征选择的具体流程

Figure 4　(Color online) Specific flow of general traffic feature selection

知识都以统一知识图谱的形式加载到智能业务网络中, 用于 6G 网络管控.

用户文本输入的意图理解可以用自然语言处理方法, 例如, 语言模型、语义检索等, 获得意图关键词, 并对该意图关键词添加对应的标签, 生成结构化的意图知识. 网络覆盖特征的获取则是基于大数据网络覆盖特征的挖掘与分析方法, 通过构建网络覆盖特征数据模型, 获得网络组网知识. 对网络流量数据的处理则一般采用深度学习的方式, 为知识获取、流量感知、资源调配而构建的各类通用模型, 在经过预训练后, 将成为知识空间的重要组成部分.

网络拓扑和用户习惯会随时间推移发生变化, 网络资源类型及服务种类的不断丰富也会为网络带来新的流量类别. 因此, 利用可以支持增量学习的网络知识获取模型, 使模型可自动感知、自主适应知识发生的迁移. 网络知识增量学习面临的主要问题是未知知识的出现. 以网络流量感知为例, 对于新流量类别的出现, 深度学习方法需要对大量现有流量和特定新流量数据进行训练, 才能在新的流量类别识别中取得良好的效果. 然而, 在网络中出现一种未知流量类别时, 其标记样本很少甚至没有, 无法针对既有模型进行有效训练, 从而难以对新类别进行准确识别. 在深度学习领域, 为应对此类识别未知情况的问题, 大量解决方案被提出, 包括终身学习、迁移学习、领域适应、零样本学习、小样本学习等.

作者所在团队前期工作中, 结合网络知识的特点, 利用元学习方法, 通过少量样本的训练识别出新的类别, 以实现网络知识的增量学习 [42]. 如图 5 所示构建基于元学习的增量学习框架. 给定一个输入网络流量, 将网络流量划分为共享相同 5 元组信息 (源 IP 地址、目标 IP 地址、源端口号、目标端口号和传输层协议) 的流, 使用特征提取器和特征选择器提取流特征. 其中, 流特征包括统计特征和时间序列特征. 随后, 应用元学习算法建立小样本的自适应类别检测模型, 该模型由元训练阶段和元测试阶段组成. 元训练阶段通过训练大量的小样本类自适应类别检测任务来学习适应新的类. 元测试阶段将使用预先训练好的模型通过几个迭代步骤来调整新类.

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

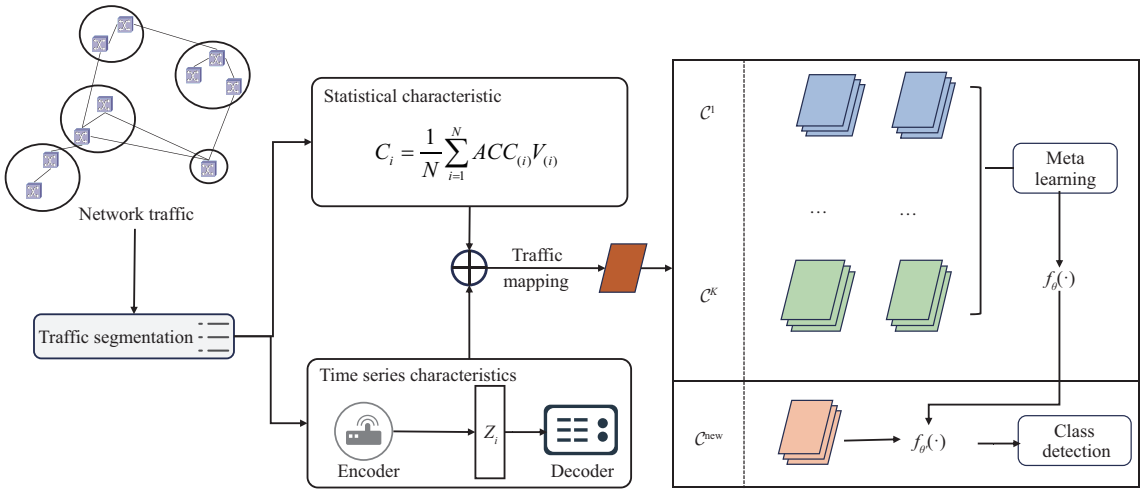


图 5 (网络版彩图) 基于元学习的增量学习框架

Figure 5 (Color online) Incremental learning framework based on meta-learning

5.3 知识定义的资源调配策略生成

资源调配策略生成是一个由资源调配需求和当前网络环境生成资源调配策略的过程. 在 6G 全场景应用的视角下, 知识定义的调配策略生成依据场景需求动态获取网络状态, 预测未来网络中流量的趋势, 设计更合理的资源分配机制. 知识定义的资源调配策略生成可包含基于知识定义的编排器、用于网络流量测量收集的模块、处理网络流量数据的算法, 以及生成资源调配策略的机器学习算法. 根据具体资源调配任务的不同, 编排器为各模块配置相应的调配参数, 如收集数据的节点集合、流量收集的持续时长及数据聚合的层级、机器学习算法的特定参数等 [43].

不确定和随机环境下, 大多数决策问题都可以用马尔可夫决策过程 (Markov decision process, MDP) 建模, 并采用动态规划算法, 如数值迭代以及强化学习求解. 与之前的移动网络相比, 6G 网络规模庞大、结构复杂, 节点更加分散、网络自组织, 使得传统的 MDP 方法难以应对其资源调配决策问题. 深度强化学习以一种通用方式结合了深度学习的感知功能和强化学习的决策功能, 通过环境的感知和策略奖励过程的循环迭代生成最优策略. 然而, 随着网络规模增大, 网络中的链路数量随之增长, 使用 DRL 算法输出层神经元控制每条链路的权重也将引起维度灾难问题. 文献 [44] 为降低输出动作维度, 将链路权重进行了离散化处理, 从而限定了输出动作空间, 并采用多智能体强化学习的方法, 对每一条链路对应的权值选择过程单独采用一个强化学习模型进行处理, 进一步减小了每个强化学习模型的决策难度和探索空间. 文献 [45] 提出了一种基于迁移学习的强化学习加速方法, 利用网络专家的经验知识 (经验来源于专家以动态开放的控制方式与接入网络进行的实时交互式采样) 设计奖励函数, 使用网络关键性能指标的加权和表示奖励函数, 同时利用训练时产生的奖励权重组合作为专家策略知识, 选择期望奖励收敛误差最小的专家策略来加速强化学习的收敛速度, 解决 6G 无线接入网络资源分配问题.

作者所在团队前期工作中, 针对网络流量状态知识, 通过多智能体分布式 DRL 算法与自关注机制结合, 从网络状态元素中获取语义, 帮助资源分配代理了解每个状态知识对无线边缘网络资源决策的重要性, 在资源决策过程中能够合理充分利用网络状态语义信息, 从而缓解网络拥塞, 最大限度地利用网络资源 [46]. 针对决策类知识, 团队前期工作中利用生成式对抗网络学习适用于不同网络拓扑

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

环境下网络资源分配的域不变特征, 以在其他环境中训练好的模型参数作为决策先验知识实现迁移学习, 加快了 SDN 网络中的路由决策过程 [47].

5.4 基于知识的网络资源调配策略验证

随着 6G 基础网络规模和复杂度的日益增长, 越来越容易由于错误配置与软硬件问题等产生如转发环路、黑洞、可达性问题等网络错误. 这些错误在现实生活中可能带来严重后果, 导致巨大的经济损失以及社会影响. 对于智能化生成的网络策略, 在将网络资源调配策略部署到网络中之前, 需要对策略的可执行性进行验证, 以确保策略实际部署后能够按照预期执行. 网络策略验证最常用的是形式化方法, 即基于数学方法描述和建模系统行为. 通过为复杂系统构建数学模型, 设计人员不仅可以更彻底地验证系统的属性, 还可以使用数学证明作为系统测试的补充, 以确保系统行为的正确. 近些年, 形式化方法已经成为网络验证领域的主要方法 [48], 主要有 4 类: 模型检查、定理证明、符号执行及可满足性理论/可满足性模理论 (satisfiability theories/satisfiability modulo theories, SAT/SMT) 求解器.

网络策略验证通常包括转发策略验证与控制策略验证, 大多采用形式化方法建模数据转发与控制行为, 反映网元配置方面的综合影响. 转发策略验证通过给定一个拓扑结构和网络数据平面快照, 由逻辑公式来模拟整个网络, 然后验证从指定不变量导出的逻辑公式. 指定不变量指定了网络中转发行为的正确性条件, 包括无环、包可达性和双向转发. 当网络转发策略更新的速度相对较慢时, 现有验证方法如 APKeep [49] 等可实现亚毫秒级的单次更新验证, 但由于端到端建立、中断或恢复等网络事件, 大量数据平面更新可能会在短时间内到达, 即出现更新风暴/突发更新. 此时, 即使 Flash 方法 [50] 可通过避免冗余计算和工程优化加快更新验证速度, 但由于每次更新涉及全模型修改, 无法应对大量更新风暴. 为进一步加速数据平面验证, 作者所在项目团队提出了基于边缘谓词的快速网络更新风暴/突发更新验证 [51]. 边缘谓词在每条边缘上使用单个谓词表示本地流量信息, 从而避免了以往工具无法避免的全模型变化问题. 基于具有本地化特性的 EP 网络模型, 实现并行验证加速, 依托于交换机进行验证任务的划分, 可以在更新到达后再划分子任务, 从而解决了以往划分方法中子任务分配不均的问题. 同时, 在可扩展性方面设计了不改变网络模型的整体粒度的规则扩充方案, 通过为每个设备添加功能模块来在转发规则之外完成对访问控制列表 (access control list, ACL) 和网络地址转换 (network address translation, NAT) 规则的建模扩充, 从而在不增加任何额外开销的情况下实现对数据包过滤和数据包重写的支持.

卷积神经网络、循环神经网络以及图神经网络在路由、切片、拥塞控制、资源调度等任务中取得广泛应用, 但随着神经网络层数的增加, 模型愈加复杂, 可解释性越差, 神经网络模型越接近于 “黑盒”, 其潜在的威胁和隐患则越难以发现. 边界输入条件的微扰动会导致神经网络输出错误结果, 由此在输入边界条件下, 策略的正确性和可预测性至关重要. 最近, 学术界围绕如何有效验证神经网络属性、如何有效测试和发现神经网络存在的缺陷等问题积极探索, 并取得了一定进展. Reluplex [52], Marabou [53], MIPVerify [54], DNNV [55] 等框架提供了各自的神经网络验证算法, 这些框架基于可满足性模理论或混合整数线性规划搜索引擎实现用户对神经网络期望行为的验证, 提升神经网络决策的正确性与鲁棒性. 网络智能管控任务中的神经网络验证是亟待关注的新兴方向 [56, 57], 文献 [56] 通过对有非平凡属性的代理进行形式化分析, 使神经网络可以快速应对对抗攻击, 文献 [57] 则关注到了深度强化学习的可扩展性验证, 利用系统状态转移图与形式化表达式判断模型的泛化性与可靠性.

与其他机器学习的问题领域不同, 网络问题 (如路由、调度) 具有明确的结构和约束, 这些结构和约束传统上激发了可解释的特定领域 (尽管不是最优的) 解决方案. 然而, 这种原理方法与深度神经网络不透明模型不兼容, 阻止了深度学习模型和这些网络特定领域知识之间的协同作用 [58]. 因此, 研究

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

深度学习网络的可解释性, 将其策略转化为基于规则的可解释控制器, 是实现基于深度学习模型的策略在 6G 网络资源调配领域实际部署的重要前提, 也为全面结合网络领域知识实现对模型策略的验证提供了可能性. 目前针对网络领域的深度学习网络可解释性的研究还非常少. 文献 [18] 设计了通用框架 Metis, 将当前网络系统抽象为本地系统和全局系统, 为基于深度学习的网络提供可解释性. 本地系统收集本地信息并为交换机上的一个实例调度程序做出决策. 全局系统通过网络聚合信息, 并对多个实例 (如 SDN 中的控制器) 进行全局规划. 对于本地系统, Metis 使用决策树实现可解释性; 对于全局系统, Metis 则使用超图提供可解释性. Metis 可作为基于深度学习的 6G 网络资源调配策略验证的重要参考.

6 6G 业务网络的协作共享

6G 业务网络的协作共享旨在针对现有智能业务网络中, 业务能力互联和组织存在单一化、粗粒度、不灵活的问题, 实现动态实时、细粒度、个性化的服务能力提供. 通过网络孪生支撑下的用户数字资产共享, 保护用户隐私、提供全场景信息的虚实映射, 满足全场景、沉浸式的按需服务. 6G 智能业务网络的协作平面, 支持不同级别的个性化定制服务, 根据用户需求动态配置 5 类资源, 使业务网络的每个层次向上提供服务. 例如, 自动驾驶场景中, 根据用户驾驶需求, 生成业务能力组合, 包括场景识别、行人识别、转向控制、人机对话、地图服务等; 每项服务映射为一系列网元来执行, 如边缘节点、路侧单元、云等, 并基于网络孪生技术生成自动驾驶的共享数据, 支持自动驾驶服务.

6.1 感知、通信、认知、计算和控制能力的一体化协同

通过不同类型资源的一体化协同, 包括资源之间的转化、全类型资源的统一调配、多种资源组合联合优化等方式, 可满足不同场景业务共享基础网络的需求. 5G 利用边缘计算技术将计算与存储资源引入移动通信系统, 在边缘网络设备中增加计算、存储等功能, 将无线接入、数据缓存与云计算等不同层面的技术有机融合, 就近选择本地化计算与存储资源, 极大降低了网络资源消耗, 满足低时延业务需求. 6G 网络立体覆盖, 且网络具有泛在智能性, 存在大量通信能力之外, 可承担感知、计算与控制等能力的节点, 其功能角色可相互转换作为替代资源. 因此, 6G 智能业务网络将基础网络资源抽象为感知、通信、认知、计算和控制能力, 探索者这 5 种能力的联合优化, 使得各个资源得到协调统一的利用, 实现资源的一体化协同. 例如, 自动驾驶场景可实现细粒度的动态计算与控制资源协同, 车辆慢行时, 在保障安全的情况下, 可降低计算资源配置, 反之高速场景应提高算力, 保证超低时延控制.

具体约束场景下, 多种资源的联合优化通常是 NP 难 (NP-hard) 问题, 传统研究通过各种近似数学方法与启发式算法对其进行求解, 例如, 文献 [59] 将高维 NP 难问题分解为许多小规模 NP 难问题求解. 目前网络智能化趋势下, 较多采取深度强化学习方法提供优化资源协同分配策略. 例如, 文献 [60] 提出了一个联合缓存和计算决策优化问题, 在每个移动设备的时延、缓存和能量约束下, 最小化所需的传输带宽. 文献 [61] 考虑移动边缘网络中的同质双向计算任务模型, 边缘计算服务器和移动设备都具有计算和缓存功能, 在时延、缓存大小和平均功率约束下, 对双向计算任务模型的计算和缓存策略进行联合优化, 实现平均带宽最小化; 作者在齐次情形下给出了最优策略和最小带宽的闭式表达式. 从分析和数值两方面阐述了通信、计算和缓存之间的折中. 文献 [62] 针对车载网中通信、缓存和计算资源的联合优化问题, 提出采用深度强化学习方法求解资源分配策略.

边缘智能是 6G 的一项重要特征, 利用泛在的边缘计算节点, 部署 AI 模型, 以分布式、跨层的方式实现模型训练、微调与推理, 实现感知、控制、计算与通信资源的有机协同. 文献 [63] 设计了 “In-Edge

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

AI" 框架, 提出将深度强化学习和联邦学习与移动边缘系统相结合的方式, 以优化移动边缘计算、缓存和通信, 智能化地利用设备与边缘节点之间的协作交换学习参数, 从而更好地训练和推理模型, 在减少不必要的系统通信负载的同时, 进行动态系统级优化和应用级增强. 文献 [64] 针对边缘智能的计算资源进行了更为细粒度的调配, 在满足严苛的时延限制的条件下, 尽可能提高模型准确度, 根据应用时延限制分配图形处理器 (graphics processing unit, GPU) 资源, 并每个应用所使用的 GPU 资源在重训练和推理之间进行分割. 文献 [65] 概述了分布式边缘节点对学习 (计算)、通信以及感知的相互作用, 针对语义通信应用, 构建了面向目标的源熵数学优化问题, 并从信息论视角, 设计了分布式学习与计算的通信速率区间特征识别方法. 此外, 边缘智能的网络可采取机器学习的效果为目标, 而不再以面向通信的优化为目标, 例如, 文献 [7] 以最小化分类误差为目标研究无线资源分配方法, 而不是传统的通信网络容量等目标. 作者所在团队前期工作中, 研究了无人机应急场景下, 具有能耗约束时无人机通信能力与飞行控制能力的有机协同, 通过动态调整飞行角度与速度, 规划无人机路径, 灵活部署无人机网络, 使其作为空中基站提供应急灵活的通信覆盖服务 [66], 恢复受灾区域通信, 加速救灾进程.

6.2 业务能力动态组合的服务个性化定制

自 2G 和 3G 时代智能网基于独立于业务的构造块 (service independent building block, SIB) 进行业务逻辑开发开始 [4], 在网络中提供原子级的服务能力, 通过服务能力的静态或动态组合满足用户服务需求成为了实现灵活、可扩展的服务提供的基本方式. 传统移动通信网络中, 面向服务的开发 (service oriented development, SOD) 或面向服务的计算 (service oriented computing, SOC) 提供的原子级服务能力自身往往由多种资源 (如通信、计算等) 组合而成, 且具备一定的逻辑执行过程. 然而, 传统智能网中在进行业务设计时, 通过 SIB 的静态搭建完成业务逻辑流程, 由集中的业务控制点统一提供, 在业务部署和实现方面均缺乏灵活性. 针对面向服务的研究中, 探讨了服务组件分布于网络中, 通过对服务组件的静态或动态组合实现业务提供的方法.

服务组合定义为通过组合多个服务组件创建新服务的过程, 以增加服务价值并以复合服务的形式完成更复杂的任务. 产生复合服务的活动包括描述服务需求 (用户需求分解)、发现相关服务、选择最佳服务和执行新的 (复合) 服务 [67]. 从用户需求分解的实现方式的角度, 可以将服务组合分为静态服务组合和动态服务组合. 其中, 静态服务组合表示在组合计划实施前创建一个抽象的过程模型 (服务链), 模型中包括服务的集合及服务间的数据依赖关系, 每个服务对应一组同类型的候选服务, 在复合服务执行的过程中, 在这些候选服务中选择出满足复合服务 QoS 要求的服务组合为具体的执行服务链. 动态服务组合自动地选择、绑定服务, 更重要的是自动地创建过程模型, 即根据用户的需求目标, 动态化地构建复合服务的服务链. 由于自动化服务组合更适用于网络服务动态变化的环境, 大大扩展了服务选择算法可优化的空间, 因此当前的研究更多集中于自动化服务组合. 4G 与 5G 时代, 移动通信网络中的 NFV 技术基于服务组合方法, 采用静态与动态方式, 将不同功能的 VNF 组合为服务功能链 (service function chain, SFC), 提供面向场景的更为复杂的网络能力. 5G 时代基于 NFV 技术, 业务网络为电信运营商提供了中间号、一键登录等复杂网络能力 [5].

6G 全域资源的全场景按需服务, 需要将服务能力的粒度细化至具体资源类型, 以网元具有单一功能的资源类型作为原子级服务能力, 以最大限度地满足 6G 服务按需、灵活提供的要求. 通过分析 6G 网络中的资源类型, 将其抽象为感知、通信、认知、计算和控制能力, 并参照现有移动网络中资源组合优化的思路, 对这些能力进行一体化协同建模及优化设计. 为了灵活和高效地满足业务的个性化需求, 6G 智能业务网络的能力开放层需要根据场景与用户需求, 自动将原子能力组合为所需的复合业务. 基于 "感知、通信、认知、计算和控制一体化协同" 的原子业务实现, 设计从用户意图表达到业务能力组

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

合的自动化编译和运行架构, 实现全场景全域服务按需自动定制. 通过赋予服务认知能力, 实现对目标行为、场景语义以及用户特征的精确认知, 并通过全网统一的服务描述方法对网络服务进行自适应动态调整. 具体而言, 网络能力动态组合的服务个性化定制包括两个方面的能力: 将复合业务的个性化需求转化为原子业务转发路径需求的能力, 可抽象编程语言设计, 由场景解析层实现; 最优的原子业务放置能力, 即原子能力编译器, 由能力开放层实现.

利用抽象化的高层次语言来表达复合网络能力动态组合的个性化需求, 将具体复合业务的需求转化为对原子能力的转发路径的需求. 如某节点 A 发起的复合业务 s, 由原子业务功能 $\{f_1, f_2, \ldots, f_n\}$ (如视频转码、防火墙、深度包检测等) 构成, 即可以将该业务视为从节点 A 出发, 经过功能节点 $\{f_1, f_2, \ldots, f_n\}$, 到达目的节点 B 的路径. 同时, 该复合业务还对性能指标 (如网络时延、业务指标等) 有一定要求, 可以使用该抽象编程语言将该复合业务 s 表述如下, 其中 path.delay.threshold 和 QoE.threshold 表示相应的性能约束:

$$\text{minimize}(A, f_1, f_2, \ldots, f_n, \text{path.delay.threshold}, \text{QoE.threshold}).$$

由于同一复合业务可能对应为不同的原子业务转发路径, 而不同原子业务的转发路径具有不同的业务性能, 并会给网络整体带来不同的影响, 如何选择原子业务转发路径, 以及在确定原子业务转发路径后, 如何规划原子业务之间的路径是实现服务个性化定制的关键, 可通过设计原子业务编辑器完成上述任务. 原子业务编译器根据各个业务通过抽象编程语言表达出的个性化需求和网络的实时状态, 自动地将上层业务的个性化需求编译为下层网络中的具体实施策略. 例如, NFV 中原子业务编译器的两大任务就转换为 VNF 的编排 (即如何根据复合业务需求选择位置合适的 VNF) 和流调度 (合理调度原子业务之间的路径使复合业务的完成时间满足需求, 并最优化网络效用) 问题. 编排和调度的目标是满足复杂业务的性能需求, 并且最优化网络的利用, 可利用网络智能化的决策方法, 如深度强化学习, 获得最优编排策略.

6.3 网络孪生支撑下的用户数字资产共享

数据共享是网元协作、业务互联的基础, 是智能化管控的必备条件. 6G 智能业务网络按需服务所产生和需要运用的数据范围包括网络运行状态、用户画像以及场景意图等. 6G 时代需要探索高效实现数据共享的新机制. 为了合法有效地利用共享数据, 提出具备数据权属的数字资产概念. 数字资产有利用价值, 是有效计量的数据集合体 [68]. 6G 智能业务网络的协作平面利用数字孪生技术, 构建用户、网元与业务应用集成的网络孪生体作为可流通交易的数字资产, 实现按需服务的数据共享.

数字孪生是一项新兴技术, 也是 ITU 关于未来网络的 12 个代表性用例之一 [69]. 数字孪生指建立物理对象的逻辑副本 (也被称为虚拟镜像或软件化副本), 即 “数字孪生体”, 反映其原始物理对象在整个生命周期内运行和生活的所有动态、特征、关键组件和重要特性 [70]. 将数字孪生技术应用于 6G 网络, 创建 6G 网络中用户、网元与业务应用的虚拟镜像, 即可构建 6G 网络的数字孪生网络. 通过物理网络和孪生网络的实时交互构建而成的网络孪生体系 [71], 能够帮助物理网络实现低成本试错、智能化决策和高效率业务应用创新.

为在 6G 网络中实现按需服务, 6G 智能业务网络将网络孪生作为基础服务之一, 利用数字孪生体记录人或物在物理空间和网络空间中的属性和行为数据. 数字孪生体记录的数据既包括反映物理对象记忆及重要特征的静态数据, 也包括随着时间推移不断产生的动态数据. 在 6G 智能业务网络中, 随着业务网络的持续运行、用户对业务能力的不断使用及各种新业务能力的不断涌现, 不但动态数据的数据量将以极快的速度不断增长, 其数据的格式也将表现出极大的动态性和不确定性. 由此可见, 由数

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

字孪生体构成的数字资产空间必须具备对大规模异构数据的高效存储以及对高维数据的编码和分析能力 [72].

基于数字孪生体构建数字资产空间, 建立用户孪生体和其相关数据的存储和自主控制, 可以实现对 6G 智能业务网络中用户数字资产的统一管理. 6G 智能业务网络以数字资产的视角, 初步构建基于网络孪生的用户数字资产共享模型. 物理空间包含用户以及基础网络中归属于用户的设备 (终端、传感设备等)、基站、边缘云、核心云等通信基础设施, 以及 IoT 终端、V2X 终端等. 孪生数据感知层是由物理空间的数字孪生体组成的数字化空间. 通过与基础网络之间的虚实映射, 获取物理对象的各种数据. 这些数据既包括数字孪生体建立时已经存在的表示物理对象特征的描述性数据、历史运行数据, 也包括在数字孪生体建立以后因物理对象的动态变化 (如, 人的移动及对外部世界的感知、人对网络服务的使用、网络的运行等) 而产生的运行时数据. 数据管理功能包括数据处理、数据融合、数据存储等典型的数据处理流程, 并建立用户孪生体和其相关数据的控制关系, 实现对用户数字资产的统一管理. 逻辑关联图层与定制服务相关, 基于网络孪生的群组功能, 通过在逻辑相关的虚拟体之间进行结网, 满足对业务个性化的定制需求. 该框架最上层为抽象的网络能力, 在用户的自主控制下, 实现用户数字资产在不同用户间的共享, 也可提供给运营商用于进行网络优化, 或提供给第三方用于 AI 训练或交易, 或用于提升用户的体验等.

可共享数据的一方可以将数字资产共享给其他用户、运营商或第三方, 用于提升用户体验、进行网络优化或模型训练等, 用户也可以从中获得收益. 区块链技术因其去中心化、不变性和可审计性等特性, 已成为实现数字资产可信共享最有潜力的技术 [73]. 然而, 6G 智能业务网络数字资产规模庞大、数字资产的隐私保护等问题, 给通过区块链实现用户数字资产的可信共享带来了挑战 [74]. 此外, 运营商之间的数字资产共享具有与其他用户数字资产共享不同的特性, 需要单独考虑. 总之, 对网络孪生中的用户数字资产统一管理并实现数字资产可信共享是 6G 智能业务网络的关键技术要求之一. 设计基于区块链的数字资产可信共享及资产交易架构、隐私保护机制, 以及可信共享资源服务提供机制, 解决中心化服务收益分配过程不透明且隐私易泄露而引发的公平性问题, 可助推可信网络数字空间的有序建设.

6.4 多模态全场景信息的智能虚实映射

6G 业务场景的体验存在多模态, 包括看见物体、听到声音、感觉质地、闻到气味、品尝味道等, 同时, 用户与网络交互也存在多模态, 即发出声音, 做出表情, 写下文字, 采取行动. 多模态机器学习旨在建立能够处理和关联来自多模态的信息的模型 [75]. 通过多模态机器学习技术, 多模态全场景信息的智能虚实映射意图为用户建立物理世界向数字空间的虚实映射模型, 构建个性化、全方位的数字孪生体, 系统性实现全场景随需而变的业务需求与网络动态资源映射.

针对 6G 全场景下表情、动作、声音等多模态信息展现的个性化需求和异构资源智能化按需管控的要求, 需要感知用户相关的多模态全场景信息, 将感知数据映射到资源表征空间, 实现对目标行为、场景语义以及用户特征的精确认知, 以统一描述方法自适应动态调整网络服务, 实现资源的按需、随愿提供. 例如, 虚拟现实场景中, 手势姿态的捕获和识别尤为重要, 可利用无线信号、视觉与深度点云等多模态信息, 获取手部关节点的精细坐标, 作为用户意图表达的输入或相关沉浸式业务应用的交互手段. 多模态全场景信息的智能虚实映射模型设计如图 6 所示.

如图 6 所示, 模型对不同来源、不同类型的多模态信息进行感知及数据预处理后, 通过跨模态信息处理模块 (如由编码器 – 解码器结构组成的掩码和匹配等预训练模块) 将多模态信息映射到同一个表征空间中, 实现由 “实” (物理空间中的多模态信息) 到 “虚” (数据空间中的用户意识) 的映射; 后续

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

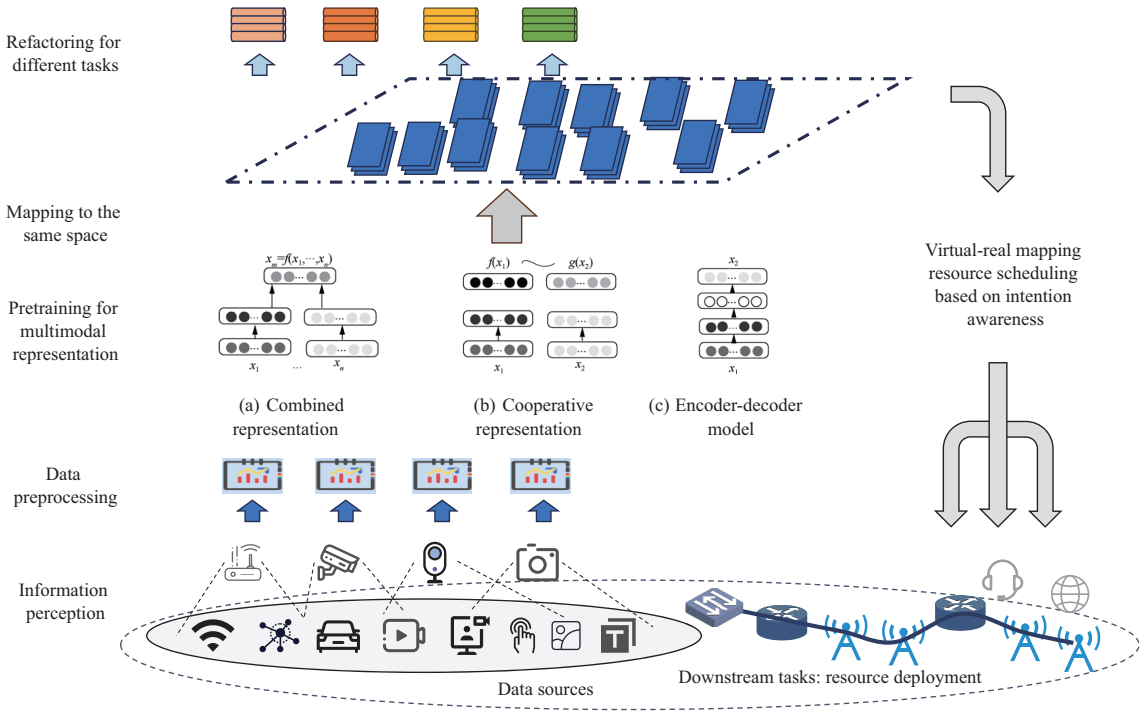


图 6 (网络版彩图) 多模态全场景信息的智能虚实映射模型

Figure 6 (Color online) Intelligent virtual-real mapping model for multimodal full-scenario information

根据不同任务进行重构, 在 6G 全场景网络的资源调配策略演化中注入用户意识, 建立用户意图的语义解析和网络资源调配的对应关系, 形成基于意识感知的资源调度方法, 对下游任务进行资源调配.

跨模态信息处理模块将多模态信息映射到同一个表征空间中, 即实现多模态信息的表征. 来自不同模态信息的特征向量最初位于不相等的子空间中, 因此与相似语义相关联的向量表示是完全不同的. 这种现象被称为异质性差距. 异质性差距将阻碍多模态数据被后续机器学习模块综合利用 [76]. 针对该问题, 多模态表征 (或称 “多模态表示”) 将异构特征投影到一个公共子空间中. 多模态表征学习的主要目标是缩小公共子空间中的分布差距, 同时保持模态特定语义的完整性. 近年来, 基于深度学习的多模态表征学习因其强大的多层次抽象表征能力而备受关注. 多模态表征学习方法可分为以下 3 类 [77].

第一类方法, 可采用联合表征将单模态表征投影到一个多模态空间中. 最简单的联合表征方法是将单模态特征串联起来, 也称为早期融合 [78], 并且为了避免串联特征, 特征以张量形式表示 [79]. 更能体现多模态联合表征特点的研究则遵循表征学习的两条主线 [80]: 概率图模型 [81] 和神经网络模型. 这两条主线的根本区别在于, 对每一层描述为概率图还是计算图, 或者隐藏层的节点是潜在的随机变量还是计算节点. 然而, 联合表征倾向于保留模态间的共享语义, 而忽略模态特定的信息, 因此不能自动保证多模态的互补性属性. 为了更具表现力, 需要融合不同形式的互补语义. 解决方案为优化目标添加额外的正则化项 [82]. 例如, 文献 [83] 提出的方法对网络权重进行迹范数正则化, 以揭示多模态特征的隐藏相关性和多样性.

第二类是协调表征, 在某些约束下为每个模态学习单独但协调的表征, 而不是将多模态一起投影到联合空间. 由于不同模态中包含的信息是不同的, 学习分离表征有助于保持独特和有用的模态特

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

征 [84]. 基于约束类型, 协调表征方法可分为两类: 基于跨模态相似性的协调表征方法和基于跨模态相关的协调表征方法. 目前, 深度神经网络已成为构建协调表征的一种流行方法, 其优势在于可以以端到端的方式共同学习协调表征.

第三类是编码器 – 解码器模型, 可用于多模态学习领域中一种模态到另一种模态的映射方法. 例如, 文献 [85] 提出的跨乐器翻译音乐的模型中包括几个解码器, 共享编码器负责提取与域无关的音乐语义, 每个解码器将在目标域中再现一段音乐; 文献 [86] 提出的图像到图像的翻译模型中包含一个内容编码器和一个样式编码器. 多模态全场景信息的智能虚实映射模型可采用编解码器实现多模态表征学习, 大多数编解码器模型包含一个编码器和一个解码器, 但一些变体也可由多个编码器或解码器组成. 编码器将源模态映射为一个潜在向量, 解码器基于该向量一个新的目标模态样本.

在多模态全场景信息的智能虚实映射过程中, 将多模态信息映射为对用户意图的统一表达后, 再结合语义解析功能即可建立用户意图和网络资源调配的对应关系, 从而形成基于意识感知的资源调度方法. 通过编解码器模型的设计, 特别是实现了跨模态语义嵌入的编解码器, 可以将通过多模态表达的用户意识转换为具体的资源调配需求, 实现基于意识感知的资源调度. 在此基础上, 结合多模态信息获取方法和信息类型的进一步丰富, 以及未来更为高级的多模态表征方法, 可将多种模态信息以更为全面、灵活的方式映射到虚拟空间中, 与数字孪生空间相融合, 为用户数字孪生体提供动态、立体的多模态信息, 从而构建全方位数字孪生体, 以此为基础实现全场景、全面沉浸式、极度个性化的用户体验.

7 6G 业务网络的安全内生

在 6G 时代, 人们将面临的不仅是一个广泛的万物互联社会, 而且是一个多样化且富有智能的世界. 在这样的环境中, 各种业务和场景对业务网络安全的需求将更加严格. 邬江兴院士指出 [87], 内生安全问题具有不可分割性, 只可能演进转化或和解而不可能彻底消除. 产业和学术界都认为, 传统的 “外挂式” 和 “补丁式” 网络安全防护机制已经无法应对未来 6G 网络可能面临的广泛攻击和不确定性安全风险. 因此, 需要改变安全发展的模式, 从 6G 网络设计的初期就必须要考虑整体的安全问题, 将安全设计融入到网络设计和构建的过程中. 中国 IMT-2030 (6G) 推进组在 2021 年 6 月首次发布的《6G 总体愿景与潜在关键技术白皮书》[88] 指出, 6G 需要支持多模信任的网络内生安全, 6G 的安全架构应奠定在一个更具包容性的信任模型基础之上, 具备韧性且覆盖 6G 网络全生命周期, 内生承载更健壮、更智慧、可扩展的安全机制. 因此, 6G 业务网络的安全内生技术考虑支撑智能业务网络的整个生命周期, 涉及到的关键技术应该包含融合身份认证、安全计算、异常监控和隐私保护几个方面 [89], 囊括了面向 6G 业务网络内生安全的赋能技术包括零信任安全、区块链、同态加密、人工智能安全等.

7.1 跨域一体化融合的用户身份认证

基于 6G 网络的全域覆盖性, 6G 智能业务网络应对全域网络平面和智能管控平面进行一体化对齐, 用户的身份认证要满足跨域一体化融合的要求, 在多个网络管控域中只需要进行统一的用户身份识别即可完成用户的识别. 文献 [90, 91] 指出, 零信任安全是新一代的网络安全防护理念, 核心思想是 “持续验证, 永不信任”. 这种安全模型默认不信任业务网络内外的任何人、设备和系统, 而是基于身份认证和授权重新构建访问控制的信任基础. 通过身份认证技术可以验证服务用户所具有的属性, 判断用户是否真实. 目前, 根据身份认证方式的不同, 一般可以将身份认证技术分为基于口令的身份认证、基于公钥基础设施 (public key infrastructure, PKI) 的身份认证和基于生物特征的身份认证.

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

在早期的移动通信系统中, 用户设备与应用服务器之间的身份验证通常基于用户名/密码、令牌/证书等凭证进行. 在此模型中, 会话密钥等凭证将同时在用户设备和应用服务器中保留, 用于保护数据的完整性和机密性. 基于凭证的身份验证方法的共同点在于都依靠预先共享的密钥或证书管理. 然而, 对于业务能力提供商来说, 管理大量的预先共享的密钥或证书可能是一项巨大的挑战, 存在数据泄露等方面的风险. 因此, 3GPP 标准化组织提出了基于 5G 的应用程序身份验证和密钥管理 (authentication and key management for applications, AKMA) 技术. 文献 [92] 中详细地介绍了一种基于 3GPP 证书的应用程序身份验证和密钥管理方案. 在这个方案中, 用户在接入 5G 网络之前, 需要先从可信任的第三方机构获取 3GPP 证书. 3GPP 证书包含了用户的身份信息以及一组用于数据加密的密钥. 该证书由可信任的第三方机构 (例如运营商或者证书颁发机构) 颁发, 用户在接入 5G 网络之前需要从这些机构获取这个证书. 在接入网络后, 网络会使用该证书对用户进行身份验证, 并且利用证书中的信息生成一组新的加密密钥, 用于保护用户的数据通信. 这些密钥在网络中的设备之间进行安全地分发和更新, 确保了数据传输的机密性和完整性.

PKI 是一种用于管理加密密钥和证书的系统, 使用了公钥和私钥来加密和解密数据. PKI 的核心组件是证书颁发机构 (certificate authority, CA), 负责颁发证书给实体, 并在它们之间建立信任关系. 文献 [93] 提出了基于区块链的 PKI 原型系统, 如图 7 所示. 基于区块链的 PKI 原型系统通过使用区块链技术来建立透明的数字证书机制, 以实现安全通信. 该机制旨在解决传统 PKI 中存在的中心化和不透明的问题, 并防止 CA 的不当行为. 在传统的 PKI 体系中, CA 是负责验证实体身份并颁发数字证书的机构. 然而, 由于其集中化的设计, CA 可能存在不当行为的风险, 例如发布伪造证书. 为了解决这个问题, 基于区块链的 PKI 采用了分散化的设计, 利用区块链的去中心化和加密特性来确保证书的安全和透明. 基于区块链的 PKI 的实现通常使用智能合约技术, 智能合约是在区块链上执行的可编程代码. 智能合约可以确保证书的发行和管理过程是透明的, 并提供了防止 CA 不当行为的机制. 通过区块链的验证和存储功能, 只有有效的证书才能被验证并存储在区块链上, 任何不当行为都将被区块链拒绝, 从而防止 CA 的不当行为. 基于区块链的 PKI 还提供了其他安全性保障, 例如防止恶意内部人员的威胁和数据泄露, 保护实体的隐私和伪匿名性, 确保实体拥有其凭证的所有权, 以及防止 Sybil 攻击、欺骗攻击和中间人攻击等. 基于区块链的 PKI 可以消除在集中式 PKI 模型中只能信任 CA 的风险, 同时增强业务网络身份认证的可扩展性和可靠性. 尽管基于区块链的 PKI 具有许多安全特性和优势, 但也存在一些限制. 目前的研究主要集中在独立的 PKI 模型上, 未来的研究可以考虑设计和测试层次、网状和混合 PKI 模型的框架和原型.

此外, 基于生物特征的身份验证技术是一种利用人体独有特征进行身份识别的技术, 包括指纹识别、面部识别、虹膜扫描和声纹识别等. 指纹识别是最常见的生物特征识别技术, 通过扫描和比对指纹图像验证身份. 这种技术已经广泛应用于智能手机、门禁系统和银行业务等领域, 可以提高身份验证的安全性. 面部识别技术则是使用摄像头捕捉用户的面部特征, 然后通过分析和比对这些特征验证身份. 这种技术在手机解锁、身份证明和视频会议等领域得到广泛应用. 虹膜扫描和声纹识别是更高级的生物特征识别技术, 分别利用眼睛的虹膜图像和声音特征进行身份验证, 这些技术在高安全性需求的环境中使用, 如银行业务和政府身份认证. 虽然基于生物特征的身份验证技术在提高安全性方面有着很大的优势, 但在实际应用中也存在一些问题. 例如, 指纹识别技术在应用中存在一些安全隐患, 如指纹被复制、泄露或被强制要求解锁等. 同时, 面部识别技术也面临着光线、角度、表情等因素的影响, 以及伪造和攻击的风险. 因此, 在实际应用中, 需要结合其他安全措施和技术手段来提高基于生物特征的身份验证技术的可靠性和安全性. 文献 [94] 指出了生物特征认证系统中两种类型的漏洞: 内在局限性和对手攻击, 讨论了生物特征系统中可能发生的不同类型的故障以及这些故障的潜在原因. 作

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

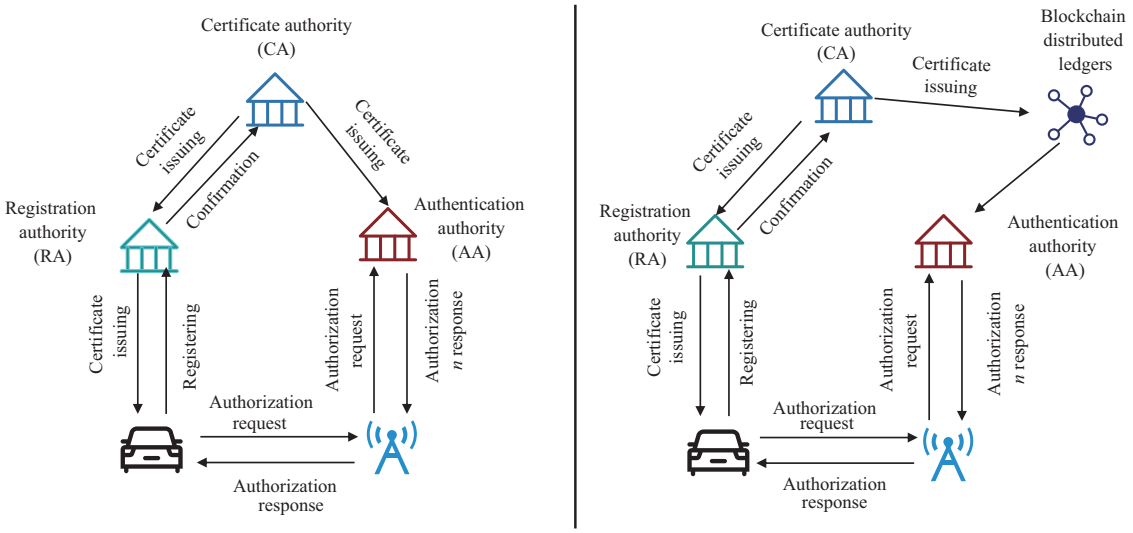


图 7　(网络版彩图) 传统 PKI 和基于区块链的 PKI

Figure 7　(Color online) Traditional PKI and blockchain-based PKI

者提到了可以针对生物特征系统不同组件的各种攻击, 并建议采取对策来防止或最小化其影响, 还进一步讨论了存储在系统数据库中的生物特征模板的安全性以及开发模板保护方案方面的挑战, 强调了方案中非逆向性、可区分性和可撤销性的要求. 文章最后提到了活动检测技术的重要性以及与模板数据库泄漏和欺骗攻击相关的风险. 文献 [95] 提出了一种隐私保护的轻量级生物特征系统, 该系统专门设计用于资源有限的物联网设备以节省内存和计算成本. 作者提出的系统采用基于块逻辑操作的算法来以简单高效的方式减小生物特征的大小, 并且生成新的 (缩短的) 二进制特征并不在 XOR 操作中保留所选候选特征, 能有效地给所产生的特征添加不确定性, 使得原始特征数据难以检索.

项目组在前期工作中, 面向全域资源协同管控提出一种分散端系统资源跨域认证机制 [96], 针对全域资源协同调度过程中涉及到的用户身份和位置等敏感信息, 提出一种网、链协同的分布式身份安全模型, 实现 6G 网络节点的敏捷接入与跨域认证, 降低数据隐私泄露和用户被追踪等安全性风险, 可实现多设备间跨域认证和数据协同, 对保障 6G 大规模分散自主节点间的资源共享的安全性和可靠性有重要作用.

7.2　基于同态加密技术的业务安全计算

安全计算是指在保护数据本身不对外泄露的前提下实现数据分析计算的一类信息技术, 提供了一种打通不同实体间信息孤岛的方式, 用加密计算来保护每个层级的数据处理, 保证数据隐私的同时最大程度地应用与发挥数据的价值. 目前, 同态加密被视为最有前景的安全计算技术之一, 原始数据经过加密后, 可以在加密状态下进行计算, 计算结果经过解密后与原始数据的计算结果相同. 这意味着, 即使在加密状态下进行计算, 也不会泄露原始数据的信息. 同态加密算法在源头上对数据进行处理, 采用密文进行安全计算, 只要特定的假设条件满足, 可以在不暴露敏感数据的情况下进行数据计算和分析. 根据允许对加密数据进行的操作数量, 有 3 种同态加密方案: 完全同态加密 (fully homomorphic encryption, FHE)、些许同态加密 (somewhat homomorphic encryption, SHE) 和部分同态加密 (partially homomorphic encryption, PHE). 对于一个完全同态的密码系统, 它应该支持任意数量的任意计算. 实际应用中 FHE 开销巨大 [97~99], 计算成本非常高, SHE 方案实际上更可行. 但 SHE 方案只允许对加

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

密数据进行某些操作, 并且由于噪声导致密文大小随着每一步计算增加限制了计算数量 [100]. PHE 方案只允许任意次数的一种对加密数据的操作, 即只对加密数据进行加法或乘法运算, 而些许同态加密方案可以同时支持这两种操作.

文献 [101] 概述了同态加密的各种潜在应用, 文献 [102] 提出了一种基于 Paillier 密码系统的模型, 并使用 PHE 和重新加密的代理服务器在云中处理银行数据, 以提高安全性. 文献 [103] 中开发了一种通信协议, 用于确保云环境中分布式系统上测量数据的真实性、保密性和完整性. Gai 等 [104] 指出, FHE 是确保云计算中数据隐私不泄露的有效方法, 但作者也强调了目前尚不可行的原因是由于其高延迟时间和误差率. GSW (Gentry, Sahai, Waters) 是一种完全同态的加密技术, 加法和乘法运算都是允许的. 文献 [105] 提出了一种使用 GSW 加密方案保护物联网设备生成的数据的方法. 由于物联网设备的计算能力有限, 这些计算在云服务器上执行. 一旦用户将查询发送到云端, 就会进行计算. 数据集的大小和相关的安全参数会影响计算时间, 在噪声值过高的情况下, 会产生不正确的结果. 文献 [106] 提出了一种基于 Paillier 算法的物联网系统隐私保护方案, 该方案有助于确保物联网环境中数据所有者、云服务器和数据用户的隐私.

同态加密是一个突破性的技术, 6G 智能业务网络中存大量加密传输的数据, 可依据场景需求, 智能化选择采取的加密方法进行进行业务安全计算. 但这一方案的实施、需要借助标准化组织和政府机构的力量实现同态加密的标准化, 其中一个重要部分是对相应安全级别的参数设置达成一致. 虽然学术研究已经进行了充分的研究, 建立了全面的参数设置方法 [107,108], 但目前还没有形成统一的参数设置方式.

7.3 基于智能协同防治的业务异常检测

在 6G 智能业务网络的业务能力开放环境中, 恶意使用业务能力的风险急剧增加, 然而业务组合的复杂性和时序规模的扩增使得传统的人工异常判断方式捉襟见肘. 需要结合人工智能方法对业务网络上的异常行为流量进行检测和定位, 联合多个系统进行协同异常防治, 以增强业务应用的安全性. 其中, 业务时间序列是业务能力使用过程中产生的按照时间发生先后顺序进行排列的一组数据集合. 在业务的异常行为中, 业务时间序列通常表现出相比于正常模式的偏移. 因此, 业务异常的检测和防治通常是基于时间序列分析技术进行的. 文献 [109] 提出了一种通用的深度学习框架, 结合了分解方法、深度神经网络和进化计算来更好地实现物联网中的异常检测. 文献 [110] 提出了混合机器学习的框架, 利用关键绩效指标捕获先验知识, 使用工业过程子流程分层异常检测, 实现了更细粒度的异常分类. 文献 [111] 提出了基于扩展卷积变压器的生成对抗网络模型提高模型的精度和泛化能力, 利用多个生成器和一个鉴别器来缓解模态崩溃问题. 文献 [112] 通过提取时空双重特征捕获多个变量之间的时间依赖性和空间相关性, 分别是采用深度自编码器捕获多变量时间序列数据的空间特征和采用多尺度稀疏变压器网络提取时间特征. 文献 [113] 设计了基于 Transformer 的多维时间序列预测器, 对指标序列进行并行建模和预测, 通过动态更新 K 值和阈值区间, 设计验证分数自适应滑动窗口方法, 检测上下文异常, 定量合理地评价网元指标异常.

移动通信网络中的欺诈行为威胁着用户隐私与财产安全. 作者所在团队前期工作中设计了一种基于时空融合特征的欺诈检测模型, 将传统的静态统计特征扩展为动态形式, 使用混合隐马尔可夫模型对用户的呼叫序列进行建模, 同时使用图注意力网络捕捉用户之间的交互关系, 从而实现时空特征的深度融合, 并输出用户潜在的欺诈可能性, 从而有效识别采用虚拟号码、团伙合作等伪装手段的隐蔽诈骗者 [114]. 基于该方法主持研制了通讯信息反欺诈系统, 对智能业务网络中的业务异常检测技术实现了落地应用. 通讯信息反欺诈系统包括多个专项反欺诈平台, 实现了对诈骗的模式识别和预测、平台业

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

务流量数据的异常检测与根因分析等. 在中国移动现网应用中, 与公安、银行、其他运营商、互联网企业等多方协作, 建立了跨部门一体化闭环反欺诈联合防控机制, 实现了对通讯信息诈骗的全流程立体防控.

在向 6G 发展的过程中, 智能业务异常处理技术也需要不断演进, 在应对新技术和新的网络形态引入的新型异常及通信风险的同时, 也要充分利用新的技术和模型提升自身的异常发现和异常处理能力. 如通讯信息反欺诈系统中对诈骗分子自然人的生物识别, 物联网、区块链技术的应用, 对新型接入网络引入的欺诈模式的预判和监控等.

团队在前期工作中, 针对可编程网络基础设施, 提出了基于深度强化学习的异常检测机制 [115], 可以迭代式地推断异常链路. 针对入侵检测, 提出基于概率学习的半监督异常检测机制 [116], 实现了有效特征的抽取和概率关系学习, 可以应对多种类、非平衡输入数据下的网络入侵异常检测. 针对边缘网络的管控, 提出了基于时空学习的人工智能模型 [117], 可以从时序和图结构数据中更好地挖掘关联关系, 实现更加优异的网络异常检测性能. 上述工作有助于实现快速、低开销地智能业务网络的异常检测和定位.

7.4 智能业务管控的用户数据隐私保护

业务管控的数据采集过程对管控数据的收集、可移植性和存储访问的隐私保护提出了挑战. 为了减轻过度数据收集的风险, 大多数隐私立法建议数据最小化, 即个人数据处理应充分、相关且仅限于预期目的 [118]. 收集个人资料和私有设备相关的数据后, 可减少个人资料的数量, 以进一步降低过度处理个人资料的风险. 某些个人资料信息的属性, 包括名称和地址, 在大多数应用程序中与模型训练无关, 因此, 首先应该采取的手段是删除或不收集这类数据. 在某些个人资料信息 (如性别和年龄) 是 AI 模型重要的学习特征的应用中, 应该使用去识别和匿名化方法. 例如, 文献 [119] 提出通过去除某些特征或泛化它们来降低输入特征的粒度. 为了处理减少收集的数据和对模型准确性施加最小影响之间的权衡, 作者使用模型中编码的知识来产生泛化. 根据大多数隐私法规, 集成适当的同意管理框架对于数据收集是必要的. 未经数据主体同意而收集和使用个人数据是对其隐私的侵犯. 为了遵守隐私保护的原则, 应该为数据主体提供一种可互操作的标准数据格式 [120]. 例如, 文献 [121] 设计了一个框架, 通过医疗设备的协作呈现医疗数据的互操作性和标准化. 互操作性体现在可以从不同来源的大型系统中接收数据, 进行数据集成和处理. 文献 [122] 研究了数据标准化所需的基础设施和资源, 以使数据更易于访问、互操作和可重用. 数据控制者应为数据主体提供适当的文档和教程, 以促进数据的可移植性. 管理数据访问的另一个预防性解决方案是使用高安全性协议和访问管理. 例如, 数据目录可以作为一种合适的协议和访问管理工具, 它提供了可用数据库中数据资产的结构化列表, 以促进可访问性和安全性 [123]. 数据目录使用元数据来帮助组织管理数据, 并根据数据的重要性组织数据, 从而执行数据治理. 此外, 差分隐私本地化 [124] 是一种有效的解决方案, 可以降低各种应用程序的识别风险. 在这种情况下, 用户在与数据收集器共享数据之前会对其数据进行扰动, 在数据中添加少量随机噪声, 以保持总体统计属性.

业务管控数据采集无效和技术故障 (包括传感器、数据记录器故障) 可能导致数据丢失、失真或不准确. 不准确和不完整的数据可能导致业务管控模型的准确性风险, 并影响有关数据主体的决策. 智能业务管控算法的开发可以在数据准备阶段集成先发制人的解决方案, 以减轻私有数据识别的风险. 在这种情况下, 私有数据的去身份化是必要的. 去身份化是指从个人记录中删除姓名、电话号码等可识别信息, 以保护数据主体的隐私. 为此, 文献 [125] 证明了在隐私保护算法中使用降维可以降低重新识别的风险. 关联和再识别是对数据的攻击, 而重构攻击是对预处理阶段提取的特征的隐私威胁. 重

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

构攻击可能针对数据集中可用的特征向量或在模型构建过程中发生重构攻击. 文献 [126] 提出了同态加密等加密模型来降低重构攻击的风险. 同态加密的主要优点是可以使用加密的数据开发 AI 模型, 而无需解密, 这种方法可以保证数据的私密性. 为了降低对抗性攻击的风险, 还可以在特征工程中使用生成式对抗网络进行对抗性特征脱敏 [127].

此外, 智能业务管控模型的构建阶段应该能够主动处理身份识别风险. 最常见的重构攻击是利用人工智能模型中的特征向量. 因此, 存储显式特征向量的算法, 包括支持向量机和 K 近邻方法, 更容易受到这些攻击 [128]. 通过在存储前对特征向量进行编码, 可以减轻重构攻击. 一些攻击, 如隶属推理和模型反演, 依赖于模型预测输出和算法的结果, 即分类中的类标签或回归方法中的预测值. 这类攻击可以通过限制入侵者对系统结果的了解来降低危害 [129]. 目前, 各种形式的差分隐私方法已经被开发出来, 以增强人工智能模型抵御对抗性攻击的能力. 尽管差分隐私有很多好处, 但是文献 [130] 证明差分隐私并不能保证全面应对所有的攻击. 因此, 有必要多次运行候选算法并对其进行测试, 以检测特定差分隐私算法的实现效果. 如前所述, 同态加密下的数据可以用于构建不需要解密的模型, 从而在隐私保护方法中提供补充保证. 在协作应用中, 当不同的利益相关者想要达到共同的目标, 但不共享他们的数据时, 由于数据可用性有限, 训练人工智能算法是困难的. 在中心学习模型的情况下, 数据集可以保留在每个设备的局部, 不需要在边缘之间传输. 联邦学习的变体, 包括水平学习、垂直学习和迁移学习, 已被提出用于解决隐私和安全问题 [131]. 尽管联邦学习有很多好处, 但训练过程中的数据局部性并不能保证集中式算法中的数据隐私性. 因此, 为了全面防御对智能业务网络的对抗性攻击, 有必要整合不同类型的隐私保护算法.

项目组在前期工作中, 针对泛在的地面 – 非地面融合网络, 提出基于差分隐私的分别学习机制 [132], 在支撑智能模型训练的同时, 保障智能业务网络中不同数据所有人对于数据的隐私性要求. 项目组还进一步对上述机制实现了联邦化的改进 [133], 使其更适用于异构化的智能业务网络环境.

8 原型系统验证

基于上述 6G 智能业务网络思想, 国家重点研发 6G 专项 “全场景按需服务关键技术” 项目组研制了按需服务的 6G 网络原型系统, 采用知识定义的内生智能思想, 建立以可信自主接入、知识定义调配、能力协同互联为核心的 6G 网络管控技术体系, 挖掘知识定义和可信共享的内在潜力. 原型系统构建方法如图 8 所示. 通过部署 3 个典型业务, 即沉浸式全景 VR、无人机监控以及云游戏业务, 验证 6G 智能业务网络架构各层功能, 以及 3 项关键技术接入层的可信认证、接入管控、自主组网管控, 网络层的切片管控、路由管控、核心网管控, 业务层的业务能力管控等, 以及平台级功能, 包括满足按需服务的意图解析、策略生成、验证、执行、系统监控, 异常发现、诊断, 以及资源管控. 基于按需服务的 6G 网络原型系统, 开发了以 6G 为网络基础设施的元宇宙应用支撑平台, 利用知识定义网络, 准确理解用户意图, 快速生成管控策略, 并利用图对齐理论精准调配全网资源, 实现了多场景精准人机意图交互与全景三维展示的按需服务. 支持混合现实视角下的人体和物体的交互实时控制, 实现了多场景异构网元按需配置, 加速超过 9 倍, 1000 个以上 6G 网元的资源调配时间低于 500 ms, 沉浸式交互时延低于 100 ms, 业务交互吞吐量超过 5 Gbps, 人体手部姿态估计误差低于 7 mm. 原型系统实现了虚实融合应用新形态的网络管控, 具有智能性高、实时性强、个性化友好、可管可控等突出优势, 为工业互联网、智慧城市等应用提供多场景业务电信级极致性能保障, 为我国下一代互联网与中国版元宇宙的创新发展提供重要基础设施平台.

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

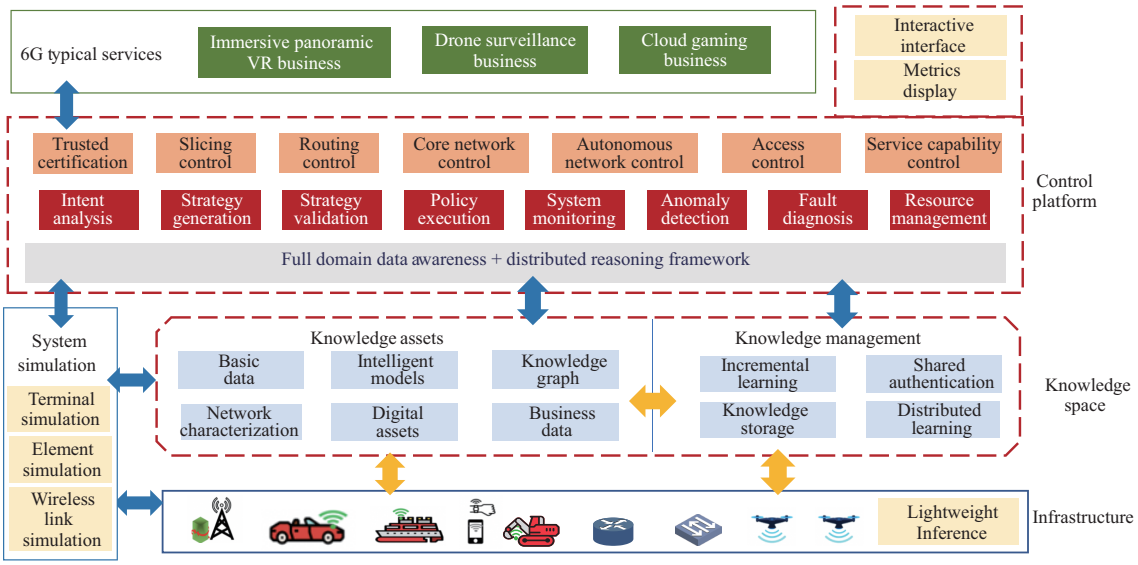


图 8　(网络版彩图) 6G 按需服务网络管控原型系统

Figure 8　(Color online) 6G on-demand service network management and control prototype system

9　展望

业务网络的智能化随着 3G, 4G 的一路演进与发展, 在 5G 时代得到学术界和产业界的共识: 人工智能技术作为新一代智能化的基石技术, 将与移动业务网络深度融合, 为行业的发展、为移动智能业务网络的管理和运营, 注入蓬勃的生机. 特别的, 在下一代 6G 移动网络研发的关键时刻, 结合人工智能领域在大模型等方向上的最新进展, 6G 智能业务网络具有更加广阔的发展空间, 其对行业的促进和助推作用也有无限的潜力.

9.1　大模型赋能 6G 业务网络管控

6G 业务网络将是面向多种场景、多种终端、多种业务, 随愿组合、随愿可享的网络, 这意味着其网络管理将涉及具有不同类型、且具有复杂关联关系的业务应用场景. 从多样化的应用场景、极致的业务优化目标和高动态的终端需求来看, 这要求 6G 业务网络具有显著的可塑性 [10]. 在智能化的业务网络管控方面, 现有 3GPP MDAF [134], ETSI ENI [135], TMF OSS [136] 等方案, 对于具体模块的数据获取、数据分析、规则匹配和配置生成等流程进行智能化, 并针对如切片管理、故障检测、用户体验优化、能耗优化等单一目标问题能取得良好效果. 然而, 一方面, 不同的目标问题之间可能存在或是促进、或是制约的关联性, 针对单一问题重复构建智能模型在负担了冗余的计算资源消耗的同时可能丢失了这些关联信息; 另一方面, 更加动态且细粒度的业务网络管控意味着管控决策本身需要能够挖掘不同层级的关联性, 力争达到更加优越的联合优化目标. 人工智能大模型的出现, 将有助于实现统一的智能业务网络管控, 通过集中化的大型基础模型 (foundation model, 简称为大模型) 共享网络管控知识 [137], 在集约化智能模型构建的同时, 容纳了业务网络管控的不同侧面.

与此同时, 业务网络业务支撑系统 (business support system, BSS) 域的智能化与运营支撑系统 (operation support system, OSS) 域的智能化也同等重要. 对于运营商来说, BSS 域的智能化主要涉及市场营销、客户销售、客户服务和计费等方面 [138]. 针对不同的业务与客户, 通过人工智能模型可以构

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

建出不同的营销策略, 并且利用集中化的人工智能能力为业务本身的智能功能进行支撑, 还可以结合智能计费系统对业务配置进行修订以帮助业务获取最佳收益. 目前, 上述智能化功能多分散而独立地在不同的业务上进行实现, 而大模型化的业务网络智能化, 将进一步地将上述智能化功能集中化, 一方面可以利用多种业务的信息, 提升智能化能力; 另一方面, 利用大模型化的智能模型, 也有助于帮助新兴业务的接入和发展, 促发业务创新与生态环境的生命力.

业务网络智能化发展的另一大趋势是 OSS 域与 BSS 域智能化的融合发展 [139]. OSS 域管控的网络性能与 BSS 域管控的客户策略是高度相关的, 而基于大型基础模型的 OSS 域管控和 BSS 域管控, 将有助于二者的进一步深度融合, 在给定网络资源的情况下, 创造更加优异的客户服务体验, 提升业务服务收益.

9.2 6G 业务网络支撑元宇宙应用

元宇宙 (metaverse) 开创了新时代的社交网络组织形式, 参与者可以突破物理空间的限制而获得真实的交互体验 [140]. 然而为了达成元宇宙极致沉浸式的用户体验, 创造极限接近现实世界的虚拟声光氛围, 元宇宙也对网络提出了无缝连接性、超低时延、超高容量、超高可靠性、严格安全性等方面上的要求 [141]. 6G 智能业务网络将从网络编排管控和智能能力提供等方面上助力支撑元宇宙这类新型应用的落地与发展.

元宇宙业务在对网络性能具有上述极高要求的同时, 还因为用户之间的实时交互, 在用户间会产生高动态而大带宽的数据流量, 这需要对网络和计算资源的动态编排 [142]. 6G 智能业务网络将能够利用人工智能技术辅助动态的网络切片和流量调度 [33], 一方面使网络切片满足元宇宙业务应用对于网络的基础需求, 另一方面利用路由策略更新对于用户交互数据流量进行动态调度, 提升元宇宙业务应用的体验.

元宇宙的沉浸式体验还来源于基于网络孪生的数字副本的构建和映射, 而这一过程需要把海量、异构的物理世界数据抽象成简约的数字世界特征, 这对元宇宙系统的智能能力提出了极高的要求, 但是这样超高的智能能力是难以在最终终端用户侧直接得到满足的, 往往需要在边缘网络部署智能能力以进行协助. 6G 智能业务网络将针对元宇宙业务的数据传输需求以及智能能力需求, 结合网络拓扑以及不同层级的边缘智能能力部署与利用情况, 以及异构的终端智能能力, 对网络数据传输以及元宇宙业务应用向边缘设备的任务卸载进行综合性的调度与优化.

9.3 6G 业务网络的绿色与可持续性

移动通信网络的发展在带来超高吞吐量、超低时延等性能指标上的不断革新之外, 也带来了对于更高能源消耗的担忧. 因此, 绿色可持续的通信网络将是 6G 以及未来移动通信网络追求的目标之一. 一方面需要达到更加严苛的性能指标要求, 另一方面又需要提升能源效率, 而更加动态且细粒度的监测与调整是人力所不能及的. 目前学界与业界的共识是人工智能技术是达成上述目标的最具希望的手段 [143].

在未来的移动网络中, 以 CPU, GPU 为代表的计算能力将普适性地存在于云端网络和边缘网络中, 6G 智能业务网络将可以利用已知的业务请求信息和计算能力部署信息, 针对不同设备的能源利用效率和潜在的传输路径, 对任务和数据进行规划, 以执行最优化的任务卸载和数据卸载策略, 以提升整个系统综合的能源利用效率.

6G 智能业务网络还可以为海量的物联网设备规划数据发送策略 [144, 145], 以减少数据包的碰撞与重传, 进而减少在重传等修正过程中额外消耗的能源. 类似地, 6G 智能业务网络通过优化的数据流量

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

调度决策可以协调不同的业务应用, 以整体化的视角优化降低不同业务应用之间的冲突, 促进达成更加优异的资源利用效率以及能源利用效率.

10 总结

移动通信网络是国家的重要基础设施, 对于数字经济与网络强国具有重要意义. 业务网络负责基础网络与业务应用之间的适配, 智能业务网络从业务应用出发, 自顶向下以全局视角研究通信网络机理, 使持续升级换代的基础网络能够快速、经济、灵活、高效地适配多领域不断涌现的新业务应用, 是移动通信各代际持续发展的必备技术, 是业务应用生态建设的核心. 本文详细阐述了 6G 智能业务网络演进历程、愿景和总体设计思路, 分析了当前 6G 网络按需服务面临的挑战, 提出了 6G 智能业务网络按需服务实现的技术路径, 设计了 3 层 3 面的 6G 智能业务网络架构, 提出超图对齐的理论模型, 并对其涉及到的智能内生、协作共享与安全内生 3 个关键技术的国内外相关研究工作和最新研究成果进行了梳理和总结. 随着 6G 研究与标准化工作的推进, 作为移动通信基础网络演进与业务应用场景落地的中间桥梁, 6G 智能业务网络将面临更多严峻挑战, 具有重要的研究价值和广泛的应用前景.

参考文献

1 ITU-T. Functional requirements and architecture of the NGN. ITU-T Y.2012, 2010. https://www.itu.int/rec/T-REC-Y.2012-201004-I/en

2 3GPP. Open service architecture (OSA) application programming interface (API) — part 1. TS 29.198, 2015. https://www.3gpp.org/ftp/Specs/archive/29_series/29.198-01/29198-01-900.zip

3 Liao J, Wang J, Wu B, et al. Toward a multiplane framework of NGSON: a required guideline to achieve pervasive services and efficient resource utilization. IEEE Commun Mag, 2012, 50: 90–97

4 Wang J. Research on service network intelligence and its key technologies. Dissertation for Ph.D. Degree. Beijing: Beijing University of Posts and Telecommunications, 2008 [王晶. 业务网络智能化及其关键技术研究. 博士学位论文. 北京: 北京邮电大学, 2008]

5 Liao J X, Zhang H, Xu T, et al. Intelligent and open new generation telecom service architecture and its key technologies. Telecommun Sci, 2020, 36: 32–43 [廖建新, 张昊, 徐童, 等. 智能开放的新一代电信业务体系及关键技术. 电信科学, 2020, 36: 32–43]

6 Liao J X, Fu X Y, Qi Q, et al. 6G-ADM: knowledge based 6G network management and control architecture. J Commun, 2022, 43: 3–15 [廖建新, 付霄元, 戚琦, 等. 6G-ADM: 基于知识空间的 6G 网络管控体系. 通信学报, 2022, 43: 3–15]

7 Wang S, Wu Y C, Xia M, et al. Machine intelligence at the edge with learning centric power allocation. IEEE Trans Wireless Commun, 2020, 19: 7293–7308

8 Zhou Y, Liu L, Wang L, et al. Service-aware 6G: an intelligent and open network based on the convergence of communication, computing and caching. Digital Commun Netw, 2020, 6: 253–260

9 IMT-2030 (6G) Promotion Group. 6G Usage Scenarios and Key Capabilities White Paper. 2022 [IMT-2030 (6G) 推进组. 6G 典型场景和关键能力白皮书. 2022] https://www.imt2030.org.cn/html/default/en/Publications/Whitepaper/index.html

10 You X, Wang C X, Huang J, et al. Towards 6G wireless communication networks: vision, enabling technologies, and new paradigm shifts. Sci China Inf Sci, 2021, 64: 110301

11 Xu W, Huang Y M, Wang W, et al. Toward ubiquitous and intelligent 6G networks: from architecture to technology. Sci China Inf Sci, 2023, 66: 130300

12 Zhu G X, Lyu Z H, Jiao X, et al. Pushing AI to wireless network edge: an overview on integrated sensing, communication, and computation towards 6G. Sci China Inf Sci, 2023, 66: 130301

13 Janz C, Davis N, Hood D, et al. Intent NBI-definition and principles. Open Networking Foundation, 2015. https://opennetworking.org/wp-content/uploads/2014/10/TR-523_Intent_Definition_Principles.pdf

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

14 3GPP. Telecommunication management; Study on scenarios for Intent driven management services for mobile networks: R17. TS 28.812, 2020. https://www.3gpp.org/ftp//Specs/archive/28_series/28.812/28812-h10.zip

15 Prakash C, Lee J, Turner Y, et al. PGA: using graphs to express and automatically reconcile network policies. SIGCOMM Comput Commun Rev, 2015, 45: 29–42

16 Jacobs A S, Pfitscher R J, Ferreira R A, et al. Refining network intents for self-driving networks. In: Proceedings of the Afternoon Workshop on Self-Driving Networks, 2018. 15–21

17 Zhou Y, Zhang D, Gao K, et al. Newton: intent-driven network traffic monitoring. In: Proceedings of the 16th International Conference on Emerging Networking EXperiments and Technologies, 2020. 295–308

18 Meng Z, Wang M, Bai J, et al. Interpreting deep learning-based networking systems. In: Proceedings of the Annual Conference of the ACM Special Interest Group on Data Communication on the Applications, Technologies, Architectures, and Protocols for Computer Communication, 2020. 154–171

19 Lei Y R, Zhou Y, Lin Y S X, et al. DOVE: diagnosis-driven SLO violation detection. In: the 29th International Conference on Network Protocols (ICNP), 2021. 1–11

20 3GPP. Management and orchestration; Architecture framework, Release 17, v17.0.0. TS 28.533, 2021. https://www.3gpp.org/ftp/Specs/archive/28_series/28.533/28533-h00.zip

21 ETSI. Experiential networked intelligence (ENI)-intent aware network autonomicity (ITANA). 2021. https://www.etsi.org/deliver/etsi_gr/ENI/001_099/008/02.01.01_60/gr_ENI008v020101p.pdf

22 Guo L Q, Zhu Z X, Liao J X, et al. Intent-driven demand-aware resource service in autonomous networks. J Beijing Univ Posts Telecommun, 2022, 45: 82–88 [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. 北京邮电大学学报, 2022, 45: 82–88]

23 Chen H X, Miao Y K, Chen L, et al. Software-defined network assimilation: bridging the last mile towards centralized network configuration management with NAssim. In: Proceedings of the ACM SIGCOMM Conference, 2022. 281–297

24 3GPP. Service requirements for the 5G system, Release 16. TS 22.261, 2020. https://www.3gpp.org/ftp//Specs/archive/22_series/22.261/22261-gc0.zip

25 Filali A, Nour B, Cherkaoui S, et al. Communication and computation O-RAN resource slicing for URLLC services using deep reinforcement learning. 2022. ArXiv:2202.06439

26 Han R, Chen D, Guo S, et al. Multi-SP network slicing parallel relieving edge network conflict. IEEE Trans Parallel Distrib Syst, 2023, 34: 2860–2875

27 Dong T, Qi Q, Wang J, et al. Standing on the shoulders of giants: cross-slice federated meta learning for resource orchestration to cold-start slice. IEEE ACM Trans Netw, 2022, 31: 828–845

28 Liu Q, Han T, Moges E. Edgeslice: slicing wireless edge computing network with decentralized deep reinforcement learning. In: Proceedings of the 40th International Conference on Distributed Computing Systems (ICDCS), 2020. 234–244

29 Sciancalepore V, Costa-Perez X, Banchs A. RL-NSB: reinforcement learning-based 5G network slice broker. IEEE ACM Trans Netw, 2019, 27: 1543–1557

30 van Huynh N, Thai Hoang D, Nguyen D N, et al. Optimal and fast real-time resource slicing with deep dueling neural networks. IEEE J Sel Areas Commun, 2019, 37: 1455–1470

31 Wu W, Zhou C, Li M, et al. AI-native network slicing for 6G networks. IEEE Wireless Commun, 2022, 29: 96–103

32 Rasti M, Taskou S K, Tabassum H, et al. Evolution toward 6G multi-band wireless networks: a resource management perspective. IEEE Wireless Commun, 2022, 29: 118–125

33 Dong T, Zhuang Z, Qi Q, et al. Intelligent joint network slicing and routing via GCN-powered multi-task deep reinforcement learning. IEEE Trans Cogn Commun Netw, 2021, 8: 1269–1286

34 Liu Y, Zhuang Z, Qi Q, et al. Slice sandwich: jagged slicing multi-tier dynamic resources for diversified V2X services. IEEE Trans Mobile Comput, 2024, 23: 4285–4302

35 Liu G, Li N, Deng J, et al. The SOLIDS 6G mobile network architecture: driving forces, features, and functional topology. Engineering, 2022, 8: 42–59

36 Zhang P, Zhang J H, Qi Q, et al. Ubiquitous-X: constructing the future 6G networks. Sci Sin Inform, 2020, 50: 913–930 [张平, 张建华, 戚琦, 等. Ubiquitous-X: 构建未来 6G 网络. 中国科学: 信息科学, 2020, 50: 913–930]

37 Zhao Y, Liu Y, Peng Y, et al. Multi-resource interleaving for deep learning training. In: Proceedings of the ACM SIGCOMM Conference, 2022. 428–440

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

38 Heimann M, Shen H, Safavi T, et al. Regal: representation learning-based graph alignment. In: Proceedings of the 27th ACM International Conference on Information and Knowledge Management, 2018. 117–126

39 Ferriol-Galmés M, Paillisse J, Suárez-Varela J, et al. RouteNet-Fermi: network modeling with graph neural networks. IEEE ACM Trans Netw, 2023, 31: 3080–3095

40 Zhang Q Z, Ng K K W, Kazer C, et al. MimicNet: fast performance estimates for data center networks with machine learning. In: Proceedings of the ACM SIGCOMM Conference, 2021. 287–304

41 Yang Q Q, Peng X, Chen L, et al. Deepqueuenet: towards scalable and generalized network performance estimation with packet-level visibility. In: Proceedings of the ACM SIGCOMM Conference, 2022. 441–457

42 Feng T T, Qi Q, Wang J Y, et al. Few-shot class-adaptive anomaly detection with model-agnostic meta-learning. In: Proceedings of IFIP Networking Conference (IFIP Networking), 2021. 1–9

43 Wang J Y, Zhou C, Zhang L, et al. Knowledge-defined intent-based network autonomy. Telecommun Sci, 2021, 37: 1–13 [王敬宇, 周铖, 张蕾, 等. 知识定义的意图网络自治. 电信科学, 2021, 37: 1–13]

44 Xu Q, Zhang Y F, Wu K, et al. Evaluating and boosting reinforcement learning for intra-domain routing. In: Proceeding of the 16th International Conference on Mobile Ad Hoc and Sensor Systems (MASS), 2019. 265–273

45 Nagib A M, Abou-Zeid H, Hassanein H S. Accelerating reinforcement learning via predictive policy transfer in 6G RAN slicing. IEEE Trans Netw Serv Manage, 2023, 20: 1170–1183

46 He B, Wang J Y, Qi Q, et al. RTHop: real-time hop-by-hop mobile network routing by decentralized learning with semantic attention. IEEE Trans Mobile Comput, 2023, 22: 1731–1747

47 Dong T, Qi Q, Wang J, et al. Generative adversarial network-based transfer reinforcement learning for routing with prior knowledge. IEEE Trans Netw Serv Manage, 2021, 18: 1673–1689

48 Qadir J, Hasan O. Applying formal methods to networking: theory, techniques, and applications. IEEE Commun Surv Tut, 2014, 17: 256–291

49 Zhang P, Liu X, Yang H K, et al. APKeep: realtime verification for real networks. In: Proceedings of the 17th USENIX Symposium on Networked Systems Design and Implementation (NSDI 20), 2020. 241–255

50 Guo D, Chen S S, Gao K, et al. Flash: fast, consistent data plane verification for large-scale network settings. In: Proceedings of the ACM SIGCOMM Conference, 2022. 314–335

51 Zhao C Y, Guo Y B, Wang J Y, et al. EPVerifier: accelerating update storms verification with edge-Predicate. In: Proceedings of the 21st USENIX Symposium on Networked Systems Design and Implementation (NSDI 24), 2024

52 Katz G, Barrett C, Dill D L, et al. Reluplex: an efficient SMT solver for verifying deep neural networks. In: Computer Aided Verification. Berlin: Springer, 2017. 97–117

53 Katz G, Huang D A, Ibeling D, et al. The marabou framework for verification and analysis of deep neural networks. In: Computer Aided Verification. Berlin: Springer, 2019. 443–452

54 Tjeng V, Xiao K, Tedrake R. Evaluating robustness of neural networks with mixed integer programming. 2019. ArXiv:1711.07356

55 Shriver D, Elbaum S, Dwyer M B. DNNV: a framework for deep neural network verification. In: Computer Aided Verification. Berlin: Springer, 2021. 137–150

56 Dethise A, Canini M, Narodytska N. Analyzing learning-based networked systems with formal verification. In: Proceedings of IEEE Conference on Computer Communications, 2021. 1–10

57 Eliyahu T, Kazak Y, Katz G, et al. Verifying learning-augmented systems. In: Proceedings of the ACM SIGCOMM Conference, 2021. 305–318

58 Zheng Y, Liu Z Y, You X Y, et al. Demystifying deep learning in networking. In: Proceedings of the 2nd Asia-Pacific Workshop on Networking, 2018. 1–7

59 Yang Y, Gong Y, Wu Y C. Intelligent-reflecting-surface-aided mobile edge computing with binary offloading: energy minimization for IoT devices. IEEE Int Things J, 2022, 9: 12973–12983

60 Sun Y P, Chen Z Y, Tao M X, et al. Modeling and trade-off for mobile communication, computing and caching networks. In: Proceedings of IEEE Global Communications Conference (GLOBECOM), 2018. 1–7

61 Sun Y P, Zhang L, Chen Z Y, et al. Communications-caching-computing tradeoff analysis for bidirectional data computation in mobile edge networks. In: Proceedings of the 92nd Vehicular Technology Conference (VTC2020-Fall), 2020. 1–5

62 He Y, Zhao N, Yin H. Integrated networking, caching, and computing for connected vehicles: a deep reinforcement

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

learning approach. IEEE Trans Veh Technol, 2017, 67: 44–55
63 Wang X, Han Y, Wang C, et al. In-edge AI: intelligentizing mobile edge computing, caching and communication by federated learning. IEEE Network, 2019, 33: 156–165
64 Shubha S S, Shen H. AdaInf: data drift adaptive scheduling for accurate and SLO-guaranteed multiple-model inference serving at edge servers. In: Proceedings of the ACM SIGCOMM Conference, 2023. 473–485
65 Xu W, Yang Z, Ng D W K, et al. Edge learning for B5G networks with distributed signal processing: semantic communication, edge computing, and wireless sensing. IEEE J Sel Top Signal Process, 2023, 17: 9–39
66 Chen D, Qi Q, Zhuang Z, et al. Mean field deep reinforcement learning for fair and efficient UAV control. IEEE Int Things J, 2020, 8: 813–828
67 Muhamad W, Bandung Y. A research challenge on mobile and cloud service composition. In: Proceedings of International Conference on Information Technology Systems and Innovation (ICITSI), 2018. 347–352
68 CAICT. White Paper on Data Elements. 2023 [中国信息通信研究院. 数据要素白皮书. 2023] http://www.caict.ac.cn/english/research/whitepapers/202311/t20231103_465068.html
69 ITU-TY. 3000-series — representative use cases and key network requirements for Network 2030. ITU-T Y Suppl.67, 2020. https://www.itu.int/rec/T-REC-Y.Sup67-202007-I
70 Pires F, Cachada A, Barbosa J, et al. Digital twin in industry 4.0: technologies, applications and challenges. In: Proceedings of the 17th International Conference on Industrial Informatics (INDIN), 2019. 721–726
71 Tan B, Qian Y, Lu H, et al. Toward a future network architecture for intelligence services: a cyber digital twin-based approach. IEEE Netw, 2021, 36: 98–104
72 Barricelli B R, Casiraghi E, Fogli D. A survey on digital twin: definitions, characteristics, applications, and design implications. IEEE Access, 2019, 7: 167653
73 Hu S S, Liang Y C, Xiong Z H, et al. Blockchain and artificial intelligence for dynamic resource sharing in 6G and beyond. IEEE Wirel Commun, 2021, 28: 145–151
74 Kumar P, Kumar R, Najmul Islam A K M, et al. Distributed AI and blockchain for 6G-assisted terrestrial and non-terrestrial networks: challenges and future directions. IEEE Netw, 2023, 37: 70–77
75 Baltrusaitis T, Ahuja C, Morency L P. Multimodal machine learning: a survey and taxonomy. IEEE Trans Pattern Anal Mach Intell, 2018, 41: 423–443
76 Peng Y, Qi J. CM-GANs: cross-modal generative adversarial networks for common representation learning. ACM Trans Multimedia Comput Commun Appl, 2019, 15: 1–24
77 Guo W, Wang J, Wang S. Deep multimodal representation learning: a survey. IEEE Access, 2019, 7: 63373–63394
78 D'mello S K, Kory J. A review and meta-analysis of multimodal affect detection systems. ACM Comput Surv, 2015, 47: 1–36
79 Cheng L, Wu Y C, Poor H V. Scaling probabilistic tensor canonical polyadic decomposition to massive data. IEEE Trans Signal Process, 2018, 66: 5534–5548
80 Bengio Y, Courville A, Vincent P. Representation learning: a review and new perspectives. IEEE Trans Pattern Anal Mach Intell, 2013, 35: 1798–1828
81 Cheng L, Tong X, Wang S, et al. Learning nonnegative factors from tensor data: probabilistic modeling and inference algorithm. IEEE Trans Signal Process, 2020, 68: 1792–1806
82 Wang S, Zhang H, Wang H. Object co-segmentation via weakly supervised data fusion. Comput Vision Image Understanding, 2017, 155: 43–54
83 Jiang Y G, Wu Z, Wang J, et al. Exploiting feature and class relationships in video categorization with regularized deep neural networks. IEEE Trans Pattern Anal Mach Intell, 2017, 40: 352–364
84 Peng Y, Qi J, Yuan Y. Modality-specific cross-modal similarity measurement with recurrent attention network. IEEE Trans Image Process, 2018, 27: 5585–5599
85 Mor N, Wolf L, Polyak A, et al. A universal music translation network. 2018. ArXiv:1805.07848
86 Huang X, Liu M Y, Belongie S, et al. Multimodal unsupervised image-to-image translation. In: Proceedings of the European Conference on Computer Vision, 2018. 172–189
87 Wu J X. Cyberspace's endogenous safety and security problem and the countermeasures. Sci Sin Inform, 2022, 52: 1929–1937 [邬江兴. 论网络空间内生安全问题及对策. 中国科学: 信息科学, 2022, 52: 1929–1937]
88 IMT-2030 (6G) Promotion Group. White Paper on 6G Vision and Candidate Technologies. 2021 [IMT-

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

2030 (6G) 推进组. 6G 总体愿景与潜在关键技术白皮书, 2021] http://www.caict.ac.cn/kxyj/qwfb/ztbg/202106/P020210604552573543918.pdf
89 Nguyen V L, Lin P C, Cheng B C, et al. Security and privacy for 6G: a survey on prospective technologies and challenges. IEEE Commun Surv Tut, 2021, 23: 2384–2428
90 Shore M, Zeadally S, Keshariya A. Zero trust: the what, how, why, and when. Computer, 2021, 54: 26–35
91 Meng L, Huang D, An J, et al. A continuous authentication protocol without trust authority for zero trust architecture. China Commun, 2022, 19: 198–213
92 3GPP. Authentication and key management for applications (AKMA) based on 3GPP credentials in the 5G system (5GS). TS 33.535, 2021. https://www.3gpp.org/ftp/Specs/archive/33_series/33.535/33535-h40.zip
93 Rashid A, Masood A, Abbas H, et al. Blockchain-based public key infrastructure: a transparent digital certification mechanism for secure communication. IEEE Net, 2021, 35: 220–225
94 Jain A K, Nandakumar K. Biometric authentication: system security and user privacy. Computer, 2012, 45: 87–92
95 Yang W, Wang S, Zheng G, et al. A privacy-preserving lightweight biometric system for Internet of Things security. IEEE Commun Mag, 2019, 57: 84–89
96 Wang M, Rui L, Yang Y, et al. A blockchain-based multi-CA cross-domain authentication scheme in decentralized autonomous network. IEEE Trans Netw Serv Manage, 2022, 19: 2664–2676
97 Brakerski Z, Gentry C, Vaikuntanathan V. (Leveled) Fully homomorphic encryption without bootstrapping. ACM Trans Comput Theor, 2014, 6: 1–36
98 Chillotti I, Gama N, Georgieva M, et al. Faster packed homomorphic operations and efficient circuit bootstrapping for TFHE. In: Proceedings of International Conference on the Theory and Application of Cryptology and Information Security, 2017. 377–408
99 Chillotti I, Gama N, Georgieva M, et al. Faster fully homomorphic encryption: Bootstrapping in less than 0.1 seconds. In: Proceedings of International Conference on the Theory and Application of Cryptology and Information Security, 2016. 3–33
100 Fan J, Vercauteren F. Somewhat practical fully homomorphic encryption. 2012. https://eprint.iacr.org/2012/144.pdf
101 Archer D, Chen L, Cheon J H, et al. Applications of homomorphic encryption. In: Proceedings of Crypto Standardization Workshop, Microsoft Research, 2017
102 Altaee M M S, Alanezi M. Enhancing cloud computing security by paillier homomorphic encryption. Int J Electr Comput, 2021, 11: 1771–1779
103 Oppermann A, Grasso-Toro F, Yurchenko A, et al. Secure cloud computing: communication protocol for multithreaded fully homomorphic encryption for remote data processing. In: Proceedings of IEEE International Symposium on Parallel and Distributed Processing with Applications and IEEE International Conference on Ubiquitous Computing and Communications (ISPA/IUCC), 2017. 503–510
104 Gai K K, Qiu M K, Li Y J, et al. Advanced fully homomorphic encryption scheme over real numbers. In: Proceedings of the 4th International Conference on Cyber Security and Cloud Computing (CSCloud), 2017. 64–69
105 Caudhari A, Bansode R. Securing IoT devices generated data using homomorphic encryption. In: Intelligent Computing and Networking. Berlin: Springer, 2021. 219–226
106 Ren W, Tong X, Du J, et al. Privacy-preserving using homomorphic encryption in Mobile IoT systems. Comput Commun, 2021, 165: 105–111
107 Marcolla C, Sucasas V, Manzano M, et al. Survey on fully homomorphic encryption, theory, and applications. Proc IEEE, 2022, 110: 1572–1609
108 Sathya S S, Vepakomma P, Raskar R, et al. A review of homomorphic encryption libraries for secure computation. 2018. ArXiv:1812.02428
109 Djenouri Y, Djenouri D, Belhadi A, et al. Emergent deep learning for anomaly detection in Internet of everything. IEEE Int Things J, 2021, 10: 3206–3214
110 Terbuch A, O'Leary P, Khalili-Motlagh-Kasmaei N, et al. Detecting anomalous multivariate time-series via hybrid machine learning. IEEE Trans Instrum Meas, 2023, 72: 1–11
111 Li Y, Peng X, Zhang J, et al. DCT-GAN: dilated convolutional transformer-based GAN for time series anomaly detection. IEEE Trans Knowl Data Eng, 2021, 35: 3632–3644

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

112 Wang F W, Yan M, Li Q R, et al. A multivariate time series anomaly detection model based on spatio-temporal dual features. In: Proceedings of International Conference on Networking and Network Applications, 2023. 416–421

113 Zheng J, Feng D, Yang Z, et al. TransKS: an anomaly detection method for telecommunication networks based on deep learning. IEEE Access, 2023, 11: 118048

114 Chu G, Wang J, Qi Q, et al. Exploiting spatial-temporal behavior patterns for fraud detection in telecom networks. IEEE Trans Dependable Secure Comput, 2023, 20: 4564–4577

115 Zhan P, Qin G Y, Qian X X, et al. Deep reinforcement learning based fast anomaly detection and localization for programmable networks. In: Proceedings of IEEE International Conference on Communications, 2023

116 Wu J H, Zhang L, Liu C, et al. Probability correlation learning for anomaly detection based on distribution-constrained autoencoder. In: Proceedings of Asia-Pacific Network Operations and Management Symposium, 2022

117 Qi Q, Shen R, Wang J, et al. Spatial-temporal learning-based artificial intelligence for IT operations in the edge network. IEEE Netw, 2021, 35: 197–203

118 Principle (C): data minimisation. 2022. https://ico.org.uk/for-organisations/guide-to-data-protection/guide-to-the-general-data-protectionregulation-gdpr/principles/data-minimisation/

119 Goldsteen A, Ezov G, Shmelkin R, et al. Data minimization for GDPR compliance in machine learning models. AI Ethics, 2021, 2: 477–491

120 Pandit H J, Debruyne C, O'Sullivan D, et al. An exploration of data interoperability for GDPR. Int J Stand Res, 2018, 16: 1–21

121 Jaleel A, Mahmood T, Hassan M A, et al. Towards medical data interoperability through collaboration of healthcare devices. IEEE Access, 2020, 8: 132302

122 Bezuidenhout L. Being fair about the design of FAIR data standards. Digit Gov: Res Pract, 2020, 1: 1–7

123 Kasrin N, Qureshi M, Steuer S, et al. Semantic data management for experimental manufacturing technologies. Datenbank-Spektrum, 2018, 18: 27–37

124 Wang T, Zhao J, Hu Z, et al. Local differential privacy for data collection and analysis. Neurocomputing, 2021, 426: 114–133

125 Jaidan D N, Carrere M, Chemli Z, et al. Data anonymization for privacy aware machine learning. In: Machine Learning, Optimization, and Data Science. Berlin: Springer, 2019. 725–737

126 Naehrig M, Lauter K, Vaikuntanathan V. Can homomorphic encryption be practical? In: Proceedings of the 3rd ACM Workshop on Cloud Computing Security Workshop, 2011. 113–124

127 Bashivan P, Bayat R, Ibrahim A, et al. Adversarial feature desensitization. In: Advances in Neural Information Processing Systems. New York: Curran Associates, 2021. 10665–10677

128 Al-Rubaie M, Chang J M. Privacy-preserving machine learning: threats and solutions. IEEE Secur Privacy, 2019, 17: 49–58

129 Goodfellow I, McDaniel P, Papernot N. Making machine learning robust against adversarial inputs. Commun ACM, 2018, 61: 56–66

130 Ding Z Y, Wang Y X, Wang G H, et al. Detecting violations of differential privacy. In: Proceedings of the ACM SIGSAC Conference on Computer and Communications Security, 2018. 475–489

131 Yang Q, Liu Y, Chen T J, et al. Federated machine learning: concept and applications. ACM Trans Intel Syst Tec, 2019, 10: 1–19

132 Wu M, Cheng G, Li P, et al. Split learning with differential privacy for integrated terrestrial and non-terrestrial networks. IEEE Wireless Commun, 2024. doi: 10.1109/MWC.015.2200462

133 Wu M, Cheng G, Ye D, et al. Federated split learning with data and label privacy preservation in vehicular networks. IEEE Trans Veh Technol, 2024, 73: 1223–1238

134 3GPP. Study on enhancement of management data analytics (MDA), v17.0.0. TR 28.809. 2021. https://www.3gpp.org/ftp/Specs/archive/28_series/28.809/28809-h00.zip

135 ETSI. Experiential networked intelligence (ENI); Overview of Prominent control loop architectures: GR ENI 017, v2.1.1. 2021. https://www.etsi.org/deliver/etsi_gr/ENI/001_099/017/02.01.01_60/gr_ENI017v020101p.pdf

136 MCELLIGOTT T. Future OSS: towards an open digital architecture. https://inform.tmforum.org/research-and-analysis/reports/future-oss-towards-an-open-digital-architecture

137 Wang J, Zhang L, Yang Y, et al. Network meets ChatGPT: intent autonomous management, control and operation.

https://engine.scichina.com/doi/10.1360/SSI-2023-0328

J Commun Inf Netw, 2023, 8: 239–255
138 TM Forum. GB921 business process framework (eTOM) suite. https://www.tmforum.org/resources/suite/gb921-business-process-framework-etom-suite-v23-5/
139 Ou Y Y, Wang L L, Yang A D, et al. Next decade of telecommunications artificial intelligence. Telecommun Sci, 2021, 37: 1–36 [欧阳晔, 王立磊, 杨爱东, 等. 通信人工智能的下一个十年. 电信科学, 2021, 37: 1–36]
140 Peng H R, Chen P C, Chen P H, et al. 6G toward metaverse: technologies, applications, and challenges. In: Proceedings of IEEE VTS Asia Pacific Wireless Communications Symposium, 2022. 6–10
141 Tang F, Chen X, Zhao M, et al. The roadmap of communication and networking in 6G for the metaverse. IEEE Wireless Commun, 2023, 30: 72–81
142 Aloqaily M, Bouachir O, Ridhawi I A, et al. Realizing the metaverse in the 6G era with AI-enabled network orchestration. IEEE Netw, 2023, 37: 78–85
143 Mao B, Tang F, Kawamoto Y, et al. AI models for green communications towards 6G. IEEE Commun Surv Tut, 2021, 24: 210–247
144 Alenezi M, Chai K K, Alam A S, et al. Unsupervised learning clustering and dynamic transmission scheduling for efficient dense LoRaWAN networks. IEEE Access, 2020, 8: 191495
145 Verma S, Kaur S, Khan M A, et al. Toward green communication in 6G-enabled massive Internet of Things. IEEE Int Things J, 2021, 8: 5408–5415

6G intelligent service networking: vision, architecture, and key technologies

Jianxin LIAO, Qi QI*, Jingyu WANG*, Jing WANG & Zirui ZHUANG

State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing 100876*, China*
* Corresponding author. E-mail: qiqi8266@bupt.edu.cn, wangjingyu@bupt.edu.cn

Abstract The mobile communication service network is responsible for the adaptation between the basic network and applications. After building the basic network, the service network before 5G was constructed by overlaying logical networks. Entering the 6G era, "scenario-driven" has become the main evolution paradigm for mobile communication development, and the importance of service networks has become more prominent. The 6G service network should be planned and constructed synchronously with the communication network, driven by the 6G full scenario design, to achieve full scenario, full domain, and on-demand services. Traditional service networks only design mechanisms such as security authentication, network control, and capability openness from an independent perspective, making it difficult to achieve fine-grained and dynamic resource control, and unable to efficiently adapt to 6G full-scenario services. Therefore, this article is based on the achievements of the first 6G group of the National Key R&D Program of China (Key Technologies for Full-Scenario On-Demand Services) and current academic research progress. First, the challenges of 6G on-demand services are summarized and sorted out. The vision and overall architecture design of the 6G intelligent service network consists of three levels, including resource access, capability opening, and scenario analysis, as well as three key control planes, including security endogeneity, intelligent endogeneity, and collaborative sharing is proposed to achieve fine-grained intelligent management of global resources. Finally, combined with the trend of deep integration of large models, metaverse, green energy conservation, and communication networks, a broader development space for 6G intelligent service networks is envisioned.

Keywords service network, on-demand services, full-scenario, native intelligence, collaboration and sharing, large foundation model, 6G

https://engine.scichina.com/doi/10.1360/SSI-2023-0328