



计算机科学

COMPUTER SCIENCE

移动边缘计算卸载技术研究综述

桓海盛, 赵鹏, 陈诺, 卡祖铭

引用本文

桓海盛, 赵鹏, 陈诺, 卡祖铭. 移动边缘计算卸载技术研究综述[J]. 计算机科学, 2026, 53(2): 367-378.

HUAN Haisheng, ZHAO Peng, CHEN Nuo, KA Zuming. [Review of Offloading Technologies Research in Mobile Edge Computing](#) [J]. Computer Science, 2026, 53(2): 367-378.

相似文章推荐 (请使用火狐或 IE 浏览器查看文章)

Similar articles recommended (Please use Firefox or IE to view the article)

[基于用户行为的云邮件防御资源分配方法](#)

Cloud Email Defense Resource Allocation Method Based on User Behavior

计算机科学, 2026, 53(2): 442-453. <https://doi.org/10.11896/jsjcx.250300041>

[低轨卫星网络中基于深度强化学习的航空器任务卸载策略](#)

Deep Reinforcement Learning-based Aircraft Task Offloading in Low Earth Orbit Satellite Networks

计算机科学, 2026, 53(2): 406-415. <https://doi.org/10.11896/jsjcx.250200092>

[基于博弈论的UAV辅助MEC系统中飞行路径及任务卸载优化研究](#)

Game Theory-based Optimization of Flight Paths and Task Offloading in UAV-assisted MEC Systems

计算机科学, 2026, 53(2): 396-405. <https://doi.org/10.11896/jsjcx.250300088>

[基于优化模型的MEC网络任务卸载与迁移策略](#)

MEC Network Task Offloading and Migration Strategy Based on Optimization Model

计算机科学, 2025, 52(11A): 241200215-6. <https://doi.org/10.11896/jsjcx.241200215>

[边缘环境下基于兴趣和移动感知的服务迁移路径选择方法](#)

Service Migration Path Selection Method Based on Interest and Mobility Perception in Edge Computing

Environment

计算机科学, 2025, 52(11A): 250200002-8. <https://doi.org/10.11896/jsjcx.250200002>

移动边缘计算卸载技术研究综述

桓海盛 赵鹏 陈诺 卡祖铭

火箭军工程大学作战保障学院 西安 710025

(huanhaisheng01@163.com)

摘要 随着物联网与5G技术的不断演进,数据流量呈现出前所未有的爆炸式增长趋势。在此背景下,传统集中式云计算模式已无法满足终端数据处理低延时和低能耗的需求,能够在数据产生源头提供即时服务的移动边缘计算(Mobile Edge Computing, MEC)逐渐成为解决该问题的最优选。计算卸载作为MEC技术的核心组成部分,其性能受多种因素影响,优化计算卸载以提升性能已成为全球研究者关注的焦点。对此,深入探讨了MEC的计算卸载技术。首先,回顾了MEC的发展历程,并阐述了MEC和计算卸载的概念和架构;其次,从计算卸载技术在不同场景下的实际应用出发,讨论了其研究现状,并介绍了有关计算卸载的优化技术;最后,分析了计算卸载技术面临的挑战,并对未来MEC计算卸载技术的研究方向进行展望,为后续的研究工作指明方向。

关键词: 移动边缘计算;计算卸载;卸载决策;资源分配;计算优化

中图分类号 TP393

Review of Offloading Technologies Research in Mobile Edge Computing

HUAN Haisheng, ZHAO Peng, CHEN Nuo and KA Zuming

College of Operational Support, Rocket Force University of Engineering, Xi'an 710025, China

Abstract With the continuous evolution of the IoT and 5G technologies, data traffic has shown an unprecedented explosive growth trend. Against this backdrop, the traditional centralized cloud computing model can no longer meet the requirements of low latency and low energy consumption for terminal data processing. MEC can provide immediate services at the source of data generation, has gradually become the most optimal solution to this problem. As a core component of MEC technology, the performance of computing offloading is affected by various factors. Optimizing computing offloading to enhance performance has become the focus of global researchers' attention. This paper aims to deeply explore the computing offloading technology of MEC. Firstly, it reviews the development history of MEC and elaborates on the concepts and architectures of MEC and computing offloading. Secondly, starting from the practical applications of computing offloading technology in different scenarios, it discusses the current research status and introduces the optimization technologies related to computing offloading. Finally, it analyzes the challenges faced by computing offloading technology and looks ahead to the future research directions of MEC computing offloading technology, so as to point out the direction for subsequent research work.

Keywords Mobile edge computing, Computing offloading, Offloading decision, Resource allocation, Computing optimization

1 引言

自2006年Google在搜索引擎大会上首次提出“云计算”概念后,全球各国政府与大型企业便纷纷斥巨资,积极开展云计算技术的研发,并大力推进其商业化进程。在过去近十年间,互联网产业迅猛发展,云计算技术不断迭代升级,日臻成熟。如今,云计算已广泛渗透至社会的各个领域,逐步成为全社会通用的重要基础设施,深度融入人们的生产生活。然而,随着物联网(Internet of Things, IoT)技术的持续进步,全球IoT设备数量呈现出显著增长态势。依据Statista预测^[1],2023—2030年全球IoT设备数量将近乎翻倍,从159亿台激增增至321亿台以上。终端设备数量急剧增加,然而传统中心

式云计算架构受限于数据处理模式、传输距离和链路稳定性等,难以满足时延敏感或计算密集业务的需求。而分布式的边缘计算能够有效补充中心式云计算的能力,肩负起推动产业互联网落地的重任,云计算由此正式步入边缘计算时代。

移动边缘计算^[2-3](Mobile Edge Computing, MEC)作为云计算的拓展与补充^[4],通过在网络边缘为用户提供云计算能力的方式,显著降低了服务交付时延,为用户带来更优质的服务体验。然而,在实际应用场景中,服务器与终端设备之间的计算卸载过程需全面权衡多种复杂因素,如边缘服务器资源的有限性、终端设备的移动特性以及计算任务的异构性。如何有效平衡并考虑这些因素,制定出最优的计算卸载策略,是计算卸载问题研究中的核心挑战。

本文首先介绍了 MEC 的产生、概念和基本架构以及计算卸载的概念;接着从应用场景和卸载优化技术两个维度,对计算卸载技术进行深入分析与总结;同时,梳理当前计算卸载领域面临的挑战,并对未来的研究方向进行展望;最后对全文进行系统性总结。期望通过本文的研究,能够更全面、系统地研究计算卸载技术在 MEC 中的应用,推动该领域不断进步与创新。

2 MEC 概述

2.1 MEC 的产生

云计算作为中心式计算的典型代表,依托虚拟化技术构建起“按需使用、按量付费”的创新服务模式。这一模式的出现,犹如为传统企业在 IT 基础设施建设方面的困境找到了突破口,成功化解了传统企业自建 IT 基础设施时面临的高昂建设成本与运营成本难题。企业无需再投入巨额资金用于硬件设备购置、机房搭建与维护、软件系统开发与更新等事项,仅需依据自身实际需求租用云计算资源,能显著提升资源利用率,降低运营风险。

然而,由于其自身架构的固有局限性,以及数据在处理前需进行长距离传输,在物联网蓬勃发展的背景下,云计算的应用面临诸多挑战。例如,在一些对实时性要求极高的物联网场景中,如自动驾驶^[5]、工业自动化控制^[6]等,云计算的长传输时延可能导致严重的后果。

为有效克服上述不足,边缘计算的概念应运而生。作为云计算的拓展与有力补充,边缘计算衍生出了微云、雾计算和 MEC 这 3 种典型模式。其中,微云可被视作一种小型化的云数据中心,部署于网络边缘位置。它具有支持用户移动性的特点,主要功能在于为移动设备提供强大的计算资源支持,并有效降低通信延迟,从而为资源密集型 and 交互式的移动应用提供坚实的技术保障。例如,在移动游戏、移动办公等场景中,微云能够实时处理大量数据,确保应用的流畅运行和快速响应,提升用户体验。

雾计算则部署在终端设备与传统云服务器之间,形象地讲,它犹如一朵接近地面的云,为用户提供计算、存储以及网络服务。雾计算的存在,极大地减轻了 IoT 应用的带宽压力,满足了其对时延的严格要求。在智能交通^[5]、智能工业^[6]等物联网场景中,雾计算能够在本地对大量的传感器数据进行实时处理和分析,减少了数据卸载至云端的流量需求,缩短了网络延迟,提高了系统的实时性和鲁棒性。

由于移动网络技术的迅猛发展,各种新型业务层出不穷。例如,虚拟现实(VR)/增强现实(AR)、高清视频直播、实时在线互动游戏等,对网络的带宽、时延等性能指标提出了前所未有的需求。同时,大量新型业务的接入,进一步加重了网络负载,给网络的稳定运行带来了巨大挑战。为满足移动网络高速发展所需的高带宽、低时延需求,同时缓解网络压力,在移动网络边缘部署计算、存储等资源的理念逐渐成为工业界的广泛共识。在此背景下,MEC 逐步孕育并发展起来。

2.2 MEC 的概念

MEC 并非新生概念,其起源于 2013 年。当时 IBM 和

Nokia Siemens 网络联合推出一款独具创新的计算平台,该计算平台利用在无线基站内部运行应用程序的特性,向移动用户提供服务,这便是 MEC 的雏形。

2014 年,欧洲电信标准协会(European Telecommunications Standards Institute, ETSI)成立了移动边缘计算规范工作组,标志着 MEC 标准化进程的正式开启。其中 ETSI 对 MEC 的定义^[7]是:把边缘服务器带到无线接入侧,以具有在网络边缘为用户提供 IT 服务和云计算的能力。

由于移动边缘计算任务产生的位置处于无线接入网内,更靠近终端设备,因此能够依托超低时延、高带宽的特点来提升服务质量和用户体验。但随着研究的逐步深入,ETSI 将“MEC”中“M”的内涵进行拓展,使它不再局限于传统的移动接入,并将 Wi-Fi 接入、固定接入等其他非 3GPP 的接入方式纳入其中。为了适应内涵的扩充,在 2017 年 3 月,ETSI 把“MEC”中的“M”重新定义为“Multi-Access”,“MEC”的翻译也变更为“多接入边缘计算”。本文采用较为通用的 Mobile Edge Computing 的说法。

MEC 将云计算能力延伸至网络边缘,为终端用户提供了更便捷、更高效的云计算服务,显著改善了用户服务体验^[8]。随着新一代技术的持续发展和应用,MEC 的应用范围将更加广泛,未来将会有越来越多的标准和规范被提出,以提高 MEC 在不同应用场景间的可迁移性和适用性。

2.3 MEC 的架构

早在 2015 年,ETSI 就发布了有关 MEC 的概念验证框架,并对其进行了定义和相关的标准化工作,该框架的参考架构至今仍处于持续更新和完善之中^[7-9]。MEC 的框架被 ETSI 描述为一种生态系统架构,包括所涉及的实体和功能,采用经典 3 层架构,上层是移动边缘系统层,中间层是移动边缘主机层,下层是网络层实体。

在已公布的 MEC 框架里,移动边缘主机层扮演着核心角色,它是整个框架的中枢。该层由移动边缘主机和移动边缘主机层管理两个关键组成部分构成。它们直接承载与运行移动边缘应用,为其提供必要的计算、存储和网络资源,并负责对这些资源进行有效的管理和优化。具体而言,移动边缘主机为了支持移动边缘应用程序的高效执行,不仅提供虚拟化的基础设施,还提供移动边缘平台。底层网络层通过对 3GPP 网络、本地网络和外部网络等多种接入方式的广泛支持,以满足不同场景下的通信需求。依托其提供的网络连接功能,该层确保数据能够在用户设备、移动边缘主机以及外部网络之间实现可靠且高效的传输,从而支持广泛的设备交互。位于框架顶层的移动边缘系统层负责对整个 MEC 系统进行宏观调控,为用户设备和第三方应用提供统一的接口,使其能够与 MEC 系统进行无缝交互,同时隐藏了底层的复杂性,简化了系统的整体操作和管理。ETSI 的 MEC 基本框架如图 1 所示。

在对 MEC 的深入研究中,图 2 呈现了一个更为详细的 MEC 参考架构。该架构基于图 1 的垂直框架,对各个功能实体间的相互关联关系进行了细致入微的定义,并根据不同的关联特性,抽象出 Mp, Mm 和 Mx 这 3 种不同的

参考点类型。其中, M_p 代表与移动边缘平台应用相关的参考点, M_m 代表与管理相关的参考点, M_x 代表和外部实体相关的参考点。这些参考点的明确界定, 为深入理解 MEC 架构的内部运作逻辑以及其与外部的交互机制, 提供了关键线索。

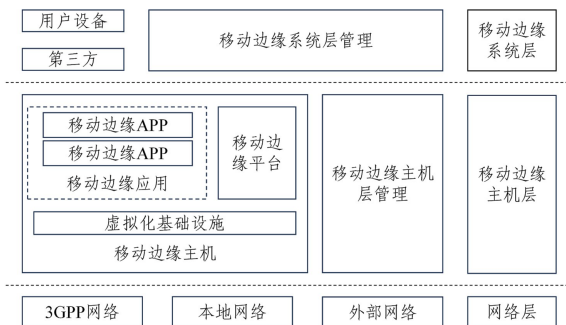


图1 ETSI MEC 基本框架

Fig. 1 ETSI MEC basic framework

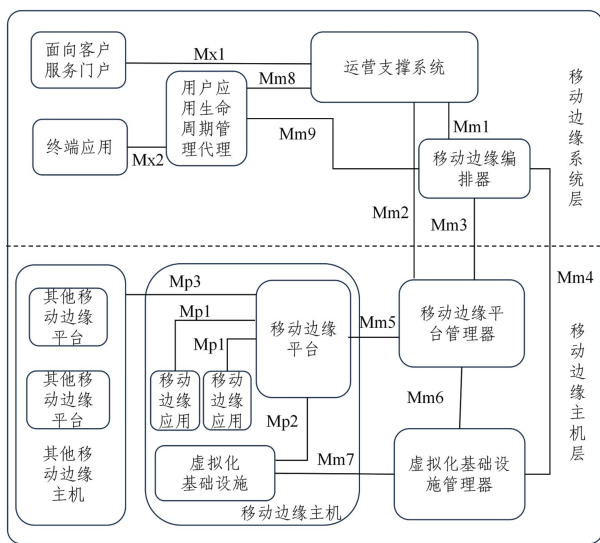


图2 MEC 参考架构

Fig. 2 MEC reference architecture

除此之外, MEC 系统架构还可以用核心网络到终端设备的云-边-端 3 层架构的形式来表现^[10], 分别是云层(云计算)、边缘层(边缘服务器)和终端层(终端设备), 如图 3 所示。

在 MEC 系统架构中, 终端设备与边缘服务器之间并不存在直接的通信链路, 二者的交互需借助基站或无线接入点来实现。具体而言, 用户终端通过无线通信技术与基站或无线接入点建立连接, 而边缘服务器则通常采用有线连接的方式与基站或无线接入点相连。值得注意的是, 在 MEC 系统的部署策略中, 边缘服务器、基站以及无线接入点常被视作一个有机整体, 共同部署于靠近用户终端的网络边缘层。通过这种方式, 系统能够有效降低因将所有任务卸载至云端处理而产生的延迟和能耗, 从而优化终端设备的性能和能源效率。这种架构不仅提升了数据处理速度, 还降低了网络传输的负担, 对于需要低延迟和高带宽的应用场景尤为重要^[11], 被广泛应用于车联网、物联网、移动增强/虚拟现实、医疗保健等延迟敏感领域。

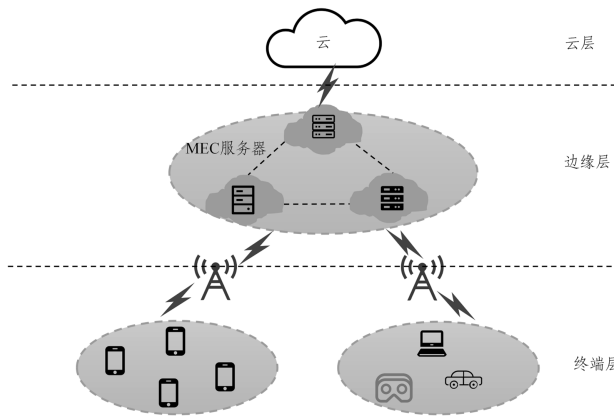


图3 云-边-端三层架构

Fig. 3 Cloud-edge-device three-tier architecture

2.4 计算卸载的概念

计算卸载技术最早于移动云计算领域被提出。在移动云计算架构下, 移动用户可通过无线网络, 将计算密集型任务传输至远端的云服务器进行处理。通过这种方式, 不仅能够减轻移动设备在计算和存储方面所面临的压力, 还能有效延长设备的电池续航时间。值得着重指出的是, 移动云计算并非是对云计算核心概念的拓展, 其本质是云计算应用场景的延伸。这意味着, 尽管移动云计算极大地拓展了云计算技术的应用范畴, 但在根本上, 它并未解决中心化云计算模式长期存在的固有问题。

与之相对应的是 MEC, MEC 是在万物互联的背景下产生的, 是一种与中心化云计算完全不同的分布式计算模式。在 MEC 环境中, 计算卸载技术是指将计算任务从资源受限的终端设备迁移至邻近的一个或多个边缘服务器, 借助边缘服务器的资源完成任务处理的一种技术^[12-13]。计算卸载流程图如图 4 所示。

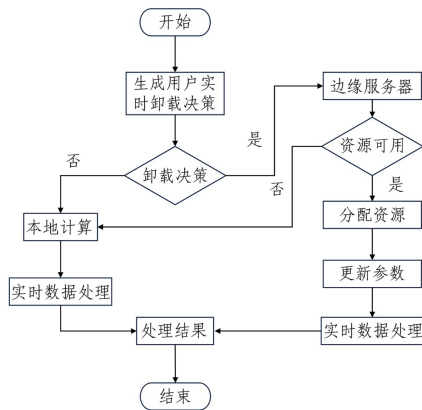


图4 计算卸载流程图

Fig. 4 Flowchart of computational offloading

MEC 中的计算卸载技术主要涉及执行框架、卸载决策和资源分配等方面。

执行框架包括计算卸载流程以及卸载方式。图 5 清晰地展示了计算卸载的主要执行流程。具体地, 终端设备发起计算卸载请求后, 内置资源监测器即刻启动, 全面检测云端资源信息, 获取可用云端服务器的运算能力、当前负载、通信

成本等关键指标。然后,计算卸载决策引擎根据接收到的服务器网络信息,运用专业算法和策略,智能判断任务执行位置。决策明确后,分割模块按结果将整体任务细分为可在不同设备独立运行的子任务。最后,本地执行部分由终端设备自主完成,需云端执行部分经格式转化、数据封装等预处理后,卸载至代理服务器处理。各流程紧密协作,确保计算卸载高效、稳定运行,发挥 MEC 计算卸载技术优势。

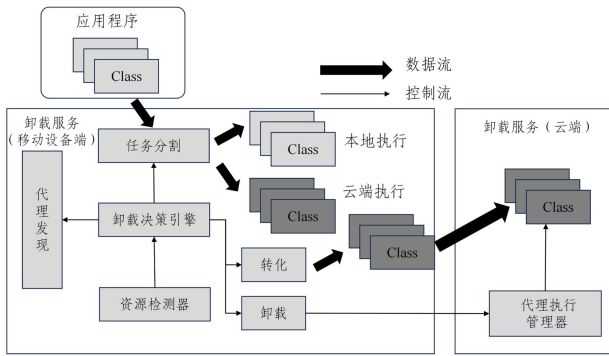


图5 计算卸载主要流程

Fig. 5 Main process of computational offloading

此外,卸载方式可以分为两类:二进制(0-1)卸载和部分卸载。二进制卸载是一种粗粒度的卸载方式,待卸载的任务不可分割,只能作为一个整体选择在终端设备上还是在边缘服务器上执行处理。具体的执行要根据实际情况进行综合考量与评判,如任务特性、设备条件和网络环境等的影响权重。部分卸载则是一种细粒度的卸载方式,通过对任务的合理分割,使服务器能够并行处理分割后的子任务,有效缩短任务完成时间,这种卸载方式主要针对大规模的计算任务场景。

表1 3类卸载决策对比

Table 1 Comparison of three offloading decisions

性能指标	参考文献	适用对象	优势	不足
最小化时延	[14-17]	时间敏感型应用	计算卸载过程耗时最短	只考虑时延,不能确保能耗是否在移动设备可承受范围内
最小化能耗	[18-20]	耗能型应用	计算卸载过程能耗最低	需要在最小化能耗的同时满足应用程序所能容忍的时延限制
权衡时延和能耗	[21-23]	无严苛时延、能耗要求,但需总消耗最小的应用	计算卸载过程总消耗最小	只适用于需求宽容的应用,不适用于有特定严苛要求的应用

当终端设备决定卸载,接下来就要考虑进行合理资源分配。当前计算卸载中资源分配的研究常考虑计算资源和通信资源的分配。随着科学技术的快速发展,出现了大量计算密集型的新型移动终端应用,由于移动终端设备在计算能力方面存在局限性,因此急需为终端设备赋能,但设备所能提供的计算能力是有限的。此外,数据量的不断攀升,不仅增加了通

信成本还加剧了通信链路压力,对通信资源进行合理的分配,可降低传输大量数据时给带宽带来的压力和传输成本。近些年的研究在计算资源^[24-27]、通信资源分配^[28-31]以及联合优化^[32-33]的问题上取得了可喜成果,有效降低了用户的使用成本,提高了用户的服务体验。这3类资源分配的优劣对比如表2所列。

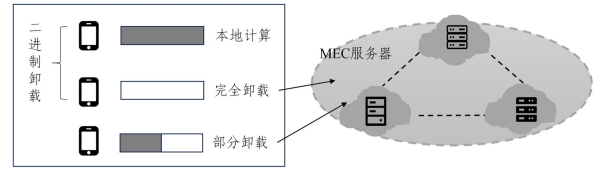


图6 计算卸载的两种方式

Fig. 6 Two methods of computational offloading

计算卸载决策在终端设备的卸载进程中起着关键作用,它决定着终端设备是否执行卸载操作以及卸载的规模大小。终端设备主要由代码解析器、系统解析器和决策引擎3个核心部分构成。卸载决策的执行过程可细分为3个步骤:首先,代码解析器基于应用程序的类型特点以及代码或数据的分区规划,精准判断哪些任务具备可卸载性;接着,系统解析器开始对一系列关键参数进行实时监测,包括可用带宽、待卸载数据的体量大小、本地应用程序执行所消耗的能量与时间等重要指标,这些参数对于全面评估卸载的可行性和效益性至关重要;最后,决策引擎依据前述步骤所获取的信息综合判定是否执行卸载操作。计算卸载的卸载决策依据其优化目标的差异可进一步分类,现有的研究在最小化时延^[14-17]、最小化能耗^[18-20]以及权衡时延和能耗^[21-23]方面取得了诸多成果,有效解决了 MEC 的应用痛点。这3类卸载决策的对比如表1所列。

表2 3类资源分配的对比

Table 2 Comparison of three resource allocation

分配资源类型	参考文献	适用场景	优势	不足
计算资源	[24-27]	终端设备计算资源空闲或多应用并发卸载	提高计算资源利用率,解决终端设备计算能力不足的问题	未考虑通信情况,卸载时可能出现拥塞干扰
通信资源	[28-31]	通信资源不足,防止数据上传堵塞	提高通信网络的频谱效率和系统容量,缓解网络压力	仅考虑数据卸载传输过程,调整决策时,缺少对卸载节点的计算资源的状态的考虑
联合计算和通信资源	[32-33]	请求计算卸载服务和请求通信服务的用户并存的场景或综合资源量有限	实现业务匹配,平衡资源利用,降低终端设备的平均开销	不同小区间协同分配资源时,通信干扰严重

3 计算卸载技术应用场景

计算卸载技术作为移动边缘计算领域的一项核心技术,在多种应用场景中均发挥了重要作用。凭借其靠近终端设备的位置优势以及具备的计算与存储能力,移动边缘计算在车联网、物联网以及虚拟现实/增强现实等多个领域得到应用。在这些应用中,计算卸载技术扮演着至关重要的角色。

3.1 车联网

车联网(Internet of Vehicles, IoV)通过现代通信和网络技术,实现了车与车、人、路以及服务平台之间的信息交换与共享,能够有效提升道路安全水平和交通运行效率,并促进智能驾驶技术^[34-35]的发展。但 IoV 环境复杂,车辆高速行驶会导致网络拓扑频繁变化,故研究重点在于如何在这种动态环境下,结合车辆的实时位置、速度、周边基础设施状况,以及任务的紧急程度、数据量大小等因素,精准、高效地做出计算卸载决策。同时,保障在车辆快速移动过程中,卸载任务的数据传输稳定性与低延迟,避免因网络波动而影响智能驾驶等关键应用。

Huang 等^[36]提出了一种基于图神经网络和深度强化学习的计算卸载方案,利用图神经网络的关系能力来捕捉任务间的依赖关系,同时利用深度强化学习的自适应学习能力来实时优化任务卸载决策,有效提高系统的计算效率和优化能耗表现。为最小化每个时隙中所有任务的总处理延迟,Zhao 等^[37]提出一种利用路侧单元辅助的 IoV 网络中的多跳通信任务卸载策略。Su 等^[38]为解决 IoV 中 DAG 应用的部分卸载问题,综合考虑服务器计算资源分配和计算卸载决策的复杂性,提出一种基于深度强化学习的 DAG 应用部分卸载算法,以综合效用最大化时延和能耗。Shao 等^[39]创新性地提出了一种多跳聚类去重卸载机制,根据任务相似性对车辆进行聚类,合并相似的任务,将延迟优化问题转换为任务量减少问题,有效提升卸载延迟表现。针对高速环境下的延迟优化问题,Zhou 等^[40]提出组优先采样-双决斗深度 Q 网络算法来优化计算卸载,缩短延迟。

虽然现有研究已经解决了实际问题,但可以预料的是,随着自动驾驶的不断发展,车辆自身需要处理的数据将呈指数式增长,因此需要设计更为合理的架构和更为高效的卸载策略,以应对计算密集型 and 延迟敏感性的应用。

3.2 物联网

在学术界,关于物联网的定义存在多种表述,但被广泛接受的定义为:物联网是一种将物品连接起来的互联网形态,并非单纯连接物品的普通网络^[41]。互联网凭借其强大的连接能力,将各类物品相互联通,进而构建起物联网,这一过程使得物品的智能化识别与管理得以实现。

随着 IoT 设备和网络的快速发展,用户对个性化服务的需求与日俱增。然而,IoT 设备受自身计算资源的限制,无法满足延迟敏感的服务请求。对此,人们对边缘协作方案进行了研究。Du 等^[42]通过无人机辅助的方式,提出了一种多智能体联合轨迹与计算卸载方案,在考虑任务类型的多样性、服务质量保障以及服务公平性的前提下,实现了卸载任务量的最大化。此外,Cui 等^[43]利用卫星资源,实现边缘节点和云计

算节点间的负载均衡;Park 等^[44]提出了使用基于多目标的强化学习进行计算卸载,提高边缘计算环境的服务体验质量;Hwang 等^[45]提出了一种应用 IoV 切片和任务卸载机制的参考架构,其在缩短延迟方面极具优势;Chouikhi 等^[46]提出了一种分布式协作计算卸载博弈,以优化能源消耗并确保在容忍的截止日期之前完成任务的体验质量,还使用聚类方法将依赖任务重新分组到同一个集群中,以更好地处理任务间的依赖。

计算卸载技术在物联网中应用广泛且十分重要。随着物联网的发展,受限于计算资源,边缘协作和计算卸载方案应运而生,有效缓解了物联网设备的计算压力,提高了设备的延迟表现和体验质量。但计算卸载技术在物联网中的应用永远在路上,未来研究方向应聚焦于多目标优化,如综合考虑任务多样性、服务质量和公平性;开发更高效的学习算法,如面向优化的多智能体深度强化学习;以及强化不同计算节点间的协同计算。

3.3 虚拟现实/增强现实

虚拟现实(Virtual Reality, VR)与增强现实(Augmented Reality, AR)技术正以一种前所未有的态势深刻影响着人们的生活习惯。这两种技术巧妙融合了不同世界的景象与声音,成功构建出一个全新的感知维度。VR 技术通过对环境、感知、自然技能以及传感器设备等要素的精心模拟,为用户营造出身临其境的实时动态三维立体体验。而 AR 技术则是将虚拟元素天衣无缝地融入真实世界,实现了虚拟与现实场景的和谐共生。

然而,VR/AR 应用程序由于所附着设备的自身硬件条件的限制,无法在本地满足自身对服务质量的严苛要求,尤其是对图形渲染与实时性处理有着极高需求的移动 AR/VR 游戏应用。幸运的是,MEC 中的计算卸载技术为 VR/AR 的发展提供了强有力的技术支持。在 MEC 环境下运用计算卸载技术^[47],能够将复杂的图形渲染和物理模拟任务迁移至云端或边缘服务器进行处理。如此,不仅可以显著提升游戏的画面质量^[48]与运行流畅性^[49],还能减轻终端设备的计算负担^[50],延长电池使用时间^[51],进而从多个方面提升用户的整体体验。

在 VR 方面,Lin 等^[52]提出了一种考虑成本的任务卸载与迁移方案,在该方案下,虚拟终端的视口渲染任务能够在可接受的长期运营成本下,在移动接入点之间进行定期的卸载和迁移。使用其提出的改进异步优势演员评论家算法,能够实现具有高收敛速度的在线决策。Lin 等^[53]提出了一种受量子启发的在线学习算法,以降低虚拟设备能耗并保障视频流畅度,将渲染卸载决策和资源分配这一联合优化问题转换为马尔可夫决策过程,并通过受量子启发的强化学习算法在线学习最优策略。在 AR 方面,Cozzolino 等^[54]提出了 Nimbus——一种多目标任务分配解决方案,该方案通过将移动的实时对象检测模型卸载到边缘云基础设施,来最大限度地缩短延迟。

特别地,利用 MEC 对近距离传输的支持,VR/AR 应用系统可以更好地依托无线技术,摆脱线缆束缚,提升设备的便携性和灵活性。这将极大地推动 VR/AR 系统在更多领域的

广泛应用和深入发展,为未来的科技生活带来更多可能。

3.4 医疗保健

在医疗保健领域的实际工作场景中,计算卸载技术同样得到了极为广泛的应用。当下,数字化浪潮汹涌澎湃、科技发展日新月异,传统医疗模式正稳步朝着智慧医疗模式转型升级。在此过程中,网络作为重要媒介,发挥着打破地域限制的关键作用,促使医护资源得以合理配置和高效利用。

以复杂的手术场景为例,对于经验尚欠丰富的医生而言,远程指导的重要性不言而喻。手术室内,高清摄像头与传感器相互协同,实时采集视频图像以及生理数据,这些数据可通过计算卸载技术传输至边缘或云端服务器^[55-56]进行处理。远端的专家根据实时传输而来的视频与生理数据,结合VR技术开展手术模拟和分析,进而为现场医生提供即时、精准的指导。这种方式不仅跨越了地理空间的限制,更极大地提高了医疗资源的使用效率,有力推动了高质量医疗服务的普及和共享,使更多患者受益。Lin等^[57]创新性地将区块链技术集成到无线VR医疗系统中,达成任务卸载和数据处理全局信息的共识,增强系统安全性,有效抵御恶意攻击,为无线VR医疗系统的信息安全提供了可靠保障,解决了医疗数据安全传输和处理的关键问题。

除了临床应用场景,移动边缘计算中的计算卸载技术在日常生活中也有着出色的应用表现。以高血压患者的健康管理为例,为有效避免患者晕倒,需要对其生理指标进行实时监控与反馈。通过将这类应用数据卸载至边缘服务器执行,可以显著降低延迟,实现对突发状况的快速响应,为患者的生命健康筑牢防线。电磁干扰对医疗传感器带来的挑战是毁灭性的,针对这个问题,Lin等^[58]提出了优化网络性能的资源分配和计算卸载算法,能够为医疗诊断和治疗提供有效存储和查询医疗数据的方法。

4 计算卸载优化技术

计算卸载技术在多元场景中均有广泛应用。但为进一步优化时延与能耗表现,实现用户服务体验的最大化提升,亟待对计算卸载的卸载决策、资源分配以及卸载路径等关键环节展开深度优化。后续内容将着重介绍一系列卸载优化技术,如凸优化、博弈论等前沿方法,深入剖析其在提升计算卸载性能方面的应用原理与实践价值。

4.1 凸优化

凸优化作为数学最优化领域的重要分支,主要聚焦于在凸集上对凸函数进行最小化的研究。本质上,在标准优化问题的范畴内,凸优化问题要求目标函数与约束函数均为凸函数。在凸优化问题的求解过程中,存在一个极为关键的特性,即局部最优解必然等同于全局最优解。这一特性极大地简化了求解过程,使得在求解时能够有效规避陷入局部最优的困境,从而更易找到真正意义上的全局最优解。同时,凸优化问题可借助梯度下降法、牛顿法、内点法等一系列高效算法来实现求解。这些算法具有收敛速度快、计算复杂度低的显著优势,为凸优化问题的高效求解提供了坚实保障。

凸优化在计算卸载中有极其重要的作用。一方面,它可以用于任务分配与资源调度,通过构建精密的凸优化模型,将

任务完成时间、能耗等关键指标设定为目标函数,同时将设备计算能力、网络带宽、电池电量等实际因素作为约束条件,从而精准确定最优的任务卸载方案^[59-60];还可以利用凸优化技术,对边缘服务器的计算资源、网络带宽等核心资源实现合理高效的分配^[61-62]。另一方面,其可降低计算复杂度^[63-65],利用数学性质简化问题求解并提升求解效率,还能提升系统性能与可靠性^[66],优化性能指标并增强稳定性。

4.2 博弈论

博弈论是研究决策主体的行为发生直接相互作用时的决策以及这种决策的均衡问题的学科。在一个博弈中,通常包括参与者、策略、收益等要素。参与者是指参与博弈的各个主体,策略是参与者可以选择的行动方案,收益是参与者采取不同策略时所获得的回报。

博弈论作为一种强大的分析工具,能够有效剖析独立且具有自利属性的参与者之间的互动关系,旨在实现博弈中所有参与者收益的最大化。这一特性在云服务提供商与边缘服务提供商的互动场景中尤为适用^[67-68]。具体而言,斯塔克尔伯格博弈在探讨智能且理性的决策者之间的冲突与合作问题上具有独特优势。在斯塔克尔伯格博弈模型里,参与者被清晰划分为领导者和追随者两个角色^[69-70],领导者率先行动,追随者基于领导者的决策再做出自身的选择。通过这种先后决策的机制,深入研究双方在复杂情境下的策略选择与互动行为。

计算卸载所涉及的博弈主要有3类,用户与服务器、用户与用户,以及用户与服务器间的博弈。其中用户与服务器间的博弈是指用户期望以最低能耗、延迟等成本完成任务,服务器则力求在满足用户需求之际,最大化自身资源利用率与利润。用户策略决定是否卸载及卸载至哪台服务器,服务器策略包括是否接受卸载请求及分配资源量。用户成功卸载且获得低延迟、低能耗服务则收益高,服务器合理接纳任务并高效利用资源则收益理想。通过博弈分析,能在不同网络条件与任务需求下,确定双方最优策略,实现相对满意状态。

多用户间博弈指多个需计算卸载的用户为参与者,各用户需决定卸载时机、数据量等策略。因网络和服务器资源有限,单个用户的卸载行为会影响其他用户。若所有用户过度卸载,会引发网络拥塞、服务器过载,导致任务延迟与能耗上升,收益降低;而通过博弈达到均衡,用户依据其他用户行为调整策略,可实现资源合理分配,使各用户均获得一定收益,如物联网设备同时卸载时,能借此避免恶性竞争,提升系统性能。

多服务器间博弈指多个提供计算卸载服务的服务器参与博弈,它们在价格、服务质量等方面竞争。服务器可降低价格吸引用户,但可能影响利润;也可以提升服务质量,但需增加硬件投入。服务器通过合理的策略吸引足够用户,保障自身利润与资源利用率,从而获取高收益。借助博弈分析,服务器可以在不同市场环境找到最优竞争策略,实现自身利益最大化,推动计算卸载系统更高效稳定的运行。

4.3 机器学习

在MEC环境中,必须快速做出卸载决策。基于机器学习的卸载是一种具有前途的方法,它可以自动映射给定的系统设置,从而得出最佳卸载决策。机器学习算法的差异很大,

可以分为不同的类型,包括监督学习、无监督学习和强化学习。机器学习被定义为从所收集的经验数据中进行学习的计算方法,旨在提高性能或在无需人工干预的情况下做出准确的预测。本质上,机器学习允许用户向计算机算法输入大量数据,然后让计算机分析这些数据,并仅根据输入的数据做出基于数据驱动的推荐和决策。如果算法识别出任何需要修正的地方,它会整合这些修正内容,以便在未来做出更好的决策。深度学习是机器学习的一个子领域。二者的共同之处在于,它们都是通过算法从特定类型任务的有限样本中总结出规则,并且这些规则可以应用于新的未知数据。深度学习更倾向于通过算法从原始数据中自动学习有效的特征。经过训练的模型可以用作计算卸载的决策引擎^[71-72]。

机器学习的另一个子领域是强化学习,它的基础框架是马尔可夫决策过程,其允许智能体在与环境之间的交互中通过试错来训练卸载模型^[73]。常用马尔可夫决策过程来优化卸载决策^[74]。具体而言,强化学习模型是为系统的每个状态选择行动,使累积奖励最大化。强化学习并非从数据中进行学习,相反,它必须从自身的经验中学习。

此外,还有深度学习和强化学习的组合场景,如深度强化学习^[75]。深度强化学习将深度学习的神经网络结构与强化学习的框架相融合。深度学习部分用于对环境状态进行特征提取和表示学习。强化学习部分则负责根据环境反馈的奖励信号,通过优化策略网络或价值网络来学习最优行为策略。深度强化学习中的主要算法包括深度 Q 网络^[76]、策略梯度算法以及深度确定性策略梯度算法。

4.4 元启发式学习

元启发式学习是一种用于解决复杂优化问题的方法,它不依赖于问题的具体数学模型,而是通过模拟自然现象、生物行为或人类经验等,来寻找问题的近似最优解。这类方法具有通用性强、对问题适应性好等特点,能在合理时间内为复杂问题提供可接受的解决方案。元启发式学习算法种类繁多,有基于自然进化的算法,如遗传算法、差分进化算法,以及基于群体智能的算法,如粒子群算法、蚁群算法。

遗传算法通过对问题的解进行编码,模拟生物种群的繁殖、交叉和变异过程,使种群中的个体不断进化,逐渐逼近最优解。由于遗传算法在多目标优化问题中具有良好的全局搜索性能,因此在计算卸载中,遗传算法常被用于决策策略分析^[77]。差分进化算法则是基于群体差异的启发式搜索算法,通过对种群中个体进行差分变异、交叉和选择操作,使种群朝着适应度更高的方向进化。由于其在求解多维问题时具有收敛速度快的优点,适用于连续和离散问题,因此常被应用于优化资源分配^[78]以及无人机轨迹优化^[79]。

粒子群优化算法是通过模拟鸟群、鱼群等生物群体的觅食行为来求解,其中每个粒子代表问题的一个潜在解,粒子根据自身的历史最优位置和群体的全局最优位置来调整自己的速度和位置,从而实现搜索最优解。粒子群优化算法已被广泛用于计算卸载决策制定,例如作为决策策略减少终端设备的能耗^[80],降低计算复杂度^[81],以及获得多目标优化的最优解^[82]。蚁群优化算法模拟蚂蚁群体通过信息素进行通信和路径搜索的行为来获取最优解。蚂蚁在搜索路径过程中会在路径上释放信息素,信息素浓度越高的路径被选择的概率越大,通过信息素的更新和正反馈机制,逐渐找到最优路径。蚁群优化算法常被用于为计算任务找到优化的卸载路径^[83],降低系统成本^[84],优化资源分配和提高效率^[85]。

5 计算卸载技术面临的挑战

MEC 凭借其在显著优化移动终端设备能耗方面的卓越表现,以及为终端设备实时应用卸载提供有力支持的能力,在学术界与工业界均备受瞩目。然而,当前在 MEC 应用场景下,计算卸载技术尚处于发展阶段,成熟度欠佳,仍面临诸多亟待解决的实际问题。

本章聚焦于计算卸载技术面临的挑战,旨在清晰梳理出该技术需要重点攻克的问题,整理计算卸载技术面临的挑战(见表 3),为后续的研究与优化工作指明方向。具体而言,从数据安全、隐私保护、负载均衡,以及硬件间的兼容性与可扩展性等维度,系统且全面地阐述计算卸载技术所面临的挑战。

表 3 计算卸载技术面临的挑战

Table 3 Challenges faced by computation offloading technology

技术挑战	主要问题	相关文献
数据安全	数据的传输和存储环节都可能受到攻击,导致数据完整性受损	[86-88]
隐私保护	终端设备数据涉及隐私问题,卸载过程隐私易被泄露	[89-94]
负载均衡	卸载方案不合理导致部分服务器过载	[95]
兼容性与可扩展性	设备和系统各式各样,卸载方案可能无法部署	—

5.1 数据安全

在传统的集中式云计算模式中,为确保计算的顺畅进行,用户数据被集中存储在云服务器上,这一做法虽便捷,但伴随着隐私泄露的潜在风险^[86]。也因此,作为云计算的有效拓展和补充,数据安全^[87-88]在移动边缘计算中占据着举足轻重的地位。特别是在计算卸载过程中,数据在传输与存储环节面临诸多安全风险,尤其是敏感数据,其保护需求尤为迫切。现有的方法是对卸载任务进行加密处理或实现安全容量最大化。

针对加密卸载问题一般采取两种方法。一种是同态加密卸载,其允许在密文上进行特定的计算,解密后的结果与在明文上进行相同计算的结果一致。另一种是代理重加密卸载,

主要解决数据需要在不同的安全域或用户之间共享时的安全问题。代理重加密技术可以在不泄露数据内容的情况下,将密文从一种加密形式转换为另一种加密形式。

安全容量最大化通过高效利用资源以及增强安全防护来提高安全容量,主要优化技术包括安全缓存机制以及安全多租户技术。安全缓存机制将频繁访问的数据加密并置于本地设备或边缘服务器的安全缓存中,方便获取并减少了加密计算和数据传输量,增加了安全容量。安全多租户技术则通过严格的隔离和访问控制机制,保证各租户之间的数据安全和隐私,并通过优化资源分配和管理,提高资源利用率,实现安全容量的最大化。

虽然现有的方法能够解决计算卸载过程中的数据安全问题,但鉴于 MEC 的分布式特性和所处的复杂网络环境,传统的数据安全方案往往难以直接适用。如何针对 MEC 场景并结合计算卸载技术的独特特点,设计出一套既有效又能妥善应对安全挑战的数据安全方案,成为一个亟待解决的重要课题。

5.2 隐私保护

隐私保护与数据安全二者紧密关联,相辅相成。为实现数据隐私保护与计算处理的并行不悖,工业界目前主要依托隐私计算技术^[89],该技术体系涵盖了差分隐私^[90-91]、同态加密^[92]等诸多细分技术。

当前,联邦学习^[93]技术在满足隐私保护与数据安全需求的同时,能够有效破解数据孤岛困境与数据隐私保护难题,从而在学术界与工业界备受关注。在实际应用场景中,鉴于孤岛数据呈现出不同的分布特征,业内将联邦学习划分为横向联邦学习、纵向联邦学习以及联邦迁移学习三大类别。依据 Lee 等^[94]的研究工作,基于边缘计算的联邦训练模型具备显著特征:训练所需数据不上传至云端,而是在数据产生的源头进行本地训练,随后仅将计算所得的权重上传至云端,用于权重聚合。这一机制既能有效降低隐私泄露风险,又能减少通信开销,同时减轻云端的负载压力。

尽管联邦学习技术为解决计算卸载过程中的隐私保护问题提供了有力途径,但现阶段针对该技术的相关研究仍相对匮乏。开发适用于 MEC 场景的隐私保护策略,不仅要求研究者深入洞悉隐私计算的技术细节,还需充分考量实际应用中的效率与实用性。这无疑是一项复杂且极具挑战性的跨学科研究课题,涉及计算机科学、密码学、统计学等多个领域的知识融合与协同创新。

5.3 负载均衡

随着各类设备日益频繁地将计算任务卸载至远程服务器,服务器负载均衡^[95]已然成为提升系统性能与可靠性的核心要素之一。在移动边缘计算(MEC)环境下,若无法实现计算任务在不同边缘服务器之间的高效、合理分配,极易引发部分服务器的过载现象。这不仅会导致该服务器处理任务的效率大幅下降,还会因连锁反应致使整个系统的服务质量遭受严重影响,如响应延迟显著增加、数据处理出错概率上升等。鉴于此,通过科学的算法与策略确保服务器负载均衡,对保障移动边缘计算系统的稳定和高效运行具有极其重要的意义。

5.4 兼容性与可拓展性

在 MEC 环境中,终端设备与远程服务器之间可能存在操作系统、硬件架构和软件平台的多样性。这种异构性给计算任务的卸载和执行带来了兼容性挑战,因为确保任务能够在不同设备和服务器间无缝迁移和执行需要解决跨平台操作的问题。此外,随着计算卸载技术应用场景的扩展和用户基数的增长,系统必须展现出良好的可扩展性,以适应需求的持续变化。可扩展性要求系统能够轻松地集成新的服务器节点、支持更多类型的设备接入,并能够动态调整资源分配策略。这涉及到系统设计的灵活性和模块化,以便在不牺牲现有功能和性能的前提下,快速适应新的技术发展和业务需求。系统架构的设计应当允许增量升级和功能扩展,以应对未来

技术演进和业务变化所带来的挑战。

6 未来研究方向

尽管计算卸载技术在过去数年中取得了显著进展,但仍存在若干尚未攻克的难题以及极具潜力、值得深入探究的方向。第 5 章已对计算卸载技术所面临的挑战展开了详细分析。在此基础上,除了针对上述挑战进行持续深化研究外,本章将着重提出几个具有重要价值的未来研究方向,旨在为后续相关研究工作提供清晰的指引与导向。

6.1 计算卸载方案

第 4.3 节关于机器学习的阐述充分展现了融入人工智能后带来的显著性能提升,这清晰地表明人工智能赋能的计算卸载技术蕴含着巨大的发展潜力。基于这一趋势,人工智能正逐步从云端向边缘端迁移,智能边缘计算应运而生^[96]。

智能边缘计算巧妙地将人工智能技术与边缘计算架构深度融合,不仅能够充分利用边缘计算所具备的低延迟、高隐私保护等特性,还能全方位发挥人工智能在数据分析、智能决策等方面的独特优势,是未来计算领域的主要发展方向之一。

值得注意的是,当数据集规模较小时,人工智能算法的学习效果与算法复杂度呈现正相关关系。因此,在开展智能计算卸载方案的研究工作时,必须紧密结合实际问题与具体应用场景的特点,精心选取适宜的模型开展。针对训练样本匮乏的情况,借助联邦学习技术,从多个参与方获取数据进行分布式训练,不失为一种有效的解决途径。通过这种方式,一方面能够有效解决数据稀疏问题,确保模型训练数据的丰富性与多样性;另一方面,联邦学习在数据交互过程中采用加密、隐私保护等技术手段,可妥善解决训练过程中面临的数据隐私难题,为智能计算卸载技术在实际应用中的广泛推广提供坚实保障。

6.2 计算卸载技术的可靠性

在数字化浪潮席卷的当下,计算卸载技术已然成为提升系统性能与效率的核心驱动力。其可靠性对于各类应用场景的稳健运行起着决定性作用,尤其是智能交通和医疗保健领域,其对数据传输速率、准确性以及动态响应能力提出了极高的要求,相关应用的容错率必须维持在极低水平。

在实际应用过程中,考虑到网络环境的错综复杂、硬件设备的多样性差异以及软件运行的固有不确定性,要确保计算卸载过程中不出现任何故障,几乎是一项难以攻克挑战。正因如此,计算卸载方案或系统必须具备极高的可靠性,方能契合不同应用场景的严苛要求。

传统方法中,保障计算卸载可靠性主要依托数据备份与恢复机制。在服务器端对卸载任务所产生的数据进行实时备份,如此,一旦遭遇数据丢失、损坏等意外状况,便能迅速从备份中恢复数据,从而有力保障任务的连续性以及数据的完整性。容错机制方面,则借助冗余计算、分布式存储等技术手段,确保即便部分服务器或节点发生故障,系统依旧能够维持正常运行状态,保证计算卸载服务的持续性与稳定性,最大限度降低故障导致的服务中断风险。

随着技术的迅猛发展与应用场景的不断拓展,传统方案逐渐暴露出局限性,难以适应新形势下的多样化需求,亟需针

对现有需求进行深度优化。但目前该领域的研究成果相对匮乏,尚不足以对计算卸载技术的可靠性提供全面、充分且行之有效的技术支撑。因此,开展深入且系统的研究工作以填补这一技术空白,满足实际应用需求,显得尤为迫切。

6.3 云边协同

在云计算与边缘计算的协同架构下,为达成计算资源的精细化、灵活调配,二者将构建起更为紧密、深度的合作纽带。云计算凭借其强大且近乎无限的计算能力,能够承担大规模、复杂的计算任务;而边缘计算则以其低延迟、本地化处理的显著特性,在应对实时性要求高的任务时优势尽显。二者特性互补,相互配合,共同致力于满足多样化、差异化的计算请求。

这种云边协作模式带来的效益是多维度的。一方面,通过优化任务分配与处理流程,显著提升了整体计算效率,缩短了任务的处理时间与减少资源消耗。另一方面,该模式还强化了系统的可靠性与灵活性。在可靠性方面,当一方出现故障或负载过高时,另一方能够迅速接管部分任务,确保系统服务的连续性;从灵活性视角来看,根据不同场景下的计算需求,云边资源可以动态调整,实现资源的高效利用与适配,为各类应用场景提供坚实的技术支撑。

结束语 MEC作为一种新兴的计算范式,因能够满足终端设备的即时计算需求并提供近实时服务,有效迎合用户多样化和高质量的业务需求,而受到广泛关注。本文从宏观视角出发,全面介绍了 MEC 和计算卸载的基本概念和架构,从应用场景和卸载优化技术两个方面详细阐述了计算卸载技术的最新研究进展,并指出了该技术在发展过程中所面临的挑战。同时,展望了计算卸载技术未来的研究方向,强调了其在推动移动边缘计算领域创新发展中的潜力。事实表明,对计算卸载技术的持续探索,对解决日益增长的个性化服务需求极具意义。

参考文献

- [1] STATISTA. IoT Connections Worldwide 2022—2033 [EB/OL]. [2024-11-01]. <https://www.statista.com/statistics/1183457/iot-connected-devices-worldwide/>.
- [2] RADOUANE B, LYAMINE G, AHMED K, et al. Scalable Mobile Computing: From Cloud Computing to Mobile Edge Computing[C]//2022 5th International Conference on Networking, Information Systems and Security: Envisage Intelligent Systems in 5g//6G-based Interconnected Digital Worlds (NISS). IEEE, 2022: 1-6.
- [3] ABBAS N, ZHANG Y, TAHERKORDI A, et al. Mobile Edge Computing: A Survey[J]. IEEE Internet of Things Journal, 2018, 5(1): 450-465.
- [4] AKHLAQI M Y, MOHD HANAPI Z B. Task Offloading Paradigm in Mobile Edge Computing-Current Issues, Adopted Approaches, and Future Directions[J]. Journal of Network and Computer Applications, 2023, 212: 103568.
- [5] JIAO A, LYU Z. Autonomous Driving Task Offloading with Mobile Edge Computing[C]//2021 2nd International Conference on Computing and Data Science (CDS). IEEE, 2021: 416-420.
- [6] ZHU N, XU X, HAN S, et al. Sleep-Scheduling and Joint Computation-Communication Resource Allocation in MEC Networks for 5G IIoT[C]//2021 IEEE Wireless Communications and Networking Conference (WCNC). IEEE, 2021: 1-7.
- [7] ETSI. Mobile Edge Computing A Key Technology Towards 5G [EB/OL]. [2024-11-01]. https://www.etsi.org/images/files/ETSIWhitePapers/etsi_wp11_mec_a_key_technology_towards_5g.pdf.
- [8] SHI H T, DUAN J, LIU Q Y. Understanding Edge Computing in One Book[M]. Beijing, China Machine Press, 2023.
- [9] EDGE COMPUTING CONSORTIUM (ECC) and INDUSTRIAL INTERNET CONSORTIUM (II). Edge Computing Reference Architecture 3.0 [EB/OL]. [2024-11-01]. <http://www.eccconsortium.org/Lists/show/id/334.html>.
- [10] ZHANG Y L, LIANG Y Z, YIN M J, et al. A Survey on Computation Offloading Schemes in Mobile Edge Computing[J]. Chinese Journal of Computers, 2021, 44(12): 2406-2430.
- [11] GONG W, QI L, XU Y. Privacy-Aware Multidimensional Mobile Service Quality Prediction and Recommendation in Distributed Fog Environment [J]. Wireless Communications and Mobile Computing, 2018, 2018(1): 3075849.
- [12] SUN W, ZHAO Y, MA W, et al. Accelerating Convergence of Federated Learning in MEC With Dynamic Community [J]. IEEE Transactions on Mobile Computing, 2024, 23(2): 1769-1784.
- [13] WU L, SUN P, WANG Z, et al. Computation Offloading in Multi-Cell Networks With Collaborative Edge-Cloud Computing: A Game Theoretic Approach [J]. IEEE Transactions on Mobile Computing, 2024, 23(3): 2093-2106.
- [14] GUO X, ZHAO S, ZHU Z, et al. Industrial Internet of Things (IIoT)-enabled Decentralized Computation Offloading in Smart Factory[C]//2024 IEEE 20th International Conference on Automation Science and Engineering (CASE). IEEE, 2024: 3858-3863.
- [15] YANG L, ZHANG X, LI J, et al. Delay-Minimization and Load-Balancing Task Offloading in Mobile Edge Computing [C]//2024 IEEE/CIC International Conference on Communications in China (ICCC Workshops). IEEE, 2024: 72-77.
- [16] HU B, GAO Y, LÓPEZ-BENÍTEZ M, et al. Computation Offloading and Resource Allocation in UAV-Assisted Satellite Network Systems [C]//2024 IEEE/CIC International Conference on Communications in China (ICCC). IEEE, 2024: 382-387.
- [17] HUANG B, ZHOU Y, ZHANG X, et al. Computation Offloading and Resource Allocation for Vehicle-Assisted Edge Computing Networks With Joint Access and Backhaul [J]. IEEE Access, 2024, 12: 110248-110259.
- [18] LIU Y, HOU Z, LIN K, et al. A Deep Reinforcement Learning Approach with Heuristic Optimization for Resource-Efficient Task Offloading in Mobile Edge Computing [C]//2024 6th International Conference on Electronics and Communication, Network and Computer Technology (ECNCT). IEEE, 2024: 500-504.
- [19] CHEN S, DU P, LIU Z, et al. Energy-Efficient Trajectory Design and Computing Offloading in UAV-Aided IoT Networks [C]//2024 IEEE/CIC International Conference on Communications in China (ICCC Workshops). IEEE, 2024: 1-6.

- [20] HAN B, YE Y, SHI L, et al. Energy-Efficient Computation Offloading for MEC-Enabled Blockchain by Data Compression [C]//2024 IEEE/CIC International Conference on Communications in China(ICCC). IEEE, 2024; 1970-1975.
- [21] JI S, LI J, JIN H, et al. Resource Aware Multi-User Task Offloading In Mobile Edge Computing[C]// 2024 IEEE International Conference on Web Services (ICWS). IEEE, 2024; 665-675.
- [22] CHEN X. Federated Reinforcement Learning-Driven Offloading Strategy for Edge Computing[C]//2024 6th International Conference on Electronics and Communication, Network and Computer Technology(ECNCT). IEEE, 2024; 467-470.
- [23] FU S, ZHAI X B, YI C, et al. Energy Efficient and Low Latency Computation Offloading via Truly PPO in Mobile Edge Computing Networks with Multi-UAV[C]// 2024 IEEE International Conferences on Internet of Things (iThings) and IEEE Green Computing & Communications (GreenCom) and IEEE Cyber, Physical & Social Computing (CPSCom) and IEEE Smart Data (SmartData) and IEEE Congress on Cybermatics. IEEE, 2024; 98-104.
- [24] TANG X, WANG Z, CAI X, et al. Research on Heterogeneous Computation Resource Allocation Based on Data-Driven Method [C]//2024 6th International Conference on Data-driven Optimization of Complex Systems(DOCS). IEEE, 2024; 916-919.
- [25] JU S, ZHANG Z, SONG W. Dynamic Resource Allocation Strategy Based on K-Means[C]//2024 IEEE 2nd International Conference on Image Processing and Computer Applications(ICIP-CA). IEEE, 2024; 262-265.
- [26] GONG L L, ZHANG Y, ZHANG J Y, et al. Optimization of Task Offloading and Resource Allocation Based on Deep Reinforcement Learning [J]. Computer Technology and Development, 2024, 34(4): 116-123.
- [27] ZHANG W Z, SONG X, QIANG L L, et al. Joint Optimization of Computation Offloading and Resource Allocation under Different Delay Constraints[J]. Journal of Chinese Computer Systems, 2024, 45(8): 2017-2025.
- [28] ZHANG Z, MO R, HOU X, et al. Resource Allocation of Large-Scale Communication Network based on Machine Learning [C]//2024 International Conference on Integrated Circuits and Communication Systems(ICICACS). IEEE, 2024; 1-5.
- [29] FANG J, LIU Z Z, CHEN S Q, et al. Optimization of IRS-Assisted NOMA-MEC Communication Resource Allocation Based on Deep Reinforcement Learning[J]. Journal of Beijing University of Technology, 2024, 50(8): 930-938.
- [30] PEI E R, LOU Y H, LI Y G, et al. Research on Communication Network Resource Allocation and Trajectory Optimization for Multi-Task UAV-Assisted Communication Networks[J]. Journal of Electronics & Information Technology, 2024, 46(7): 2748-2756.
- [31] WANG J B, ZHANG J, DING C, et al. Joint Optimization of Transmission Bandwidth Allocation and Data Compression for Mobile-Edge Computing Systems [J]. IEEE Communications Letters, 2020, 24(10): 2245-2249.
- [32] REN J, HE Y, YU G, et al. Joint Communication and Computation Resource Allocation for Cloud-Edge Collaborative System [C]// 2019 IEEE Wireless Communications and Networking Conference(WCNC). IEEE, 2019; 1-6.
- [33] IMRAN M, ALI M N, DIN M S U, et al. An Efficient Communication and Computation Resources Sharing in Information-Centric 6G Networks[J]. IEEE Internet of Things Journal, 2024, 11(16): 27275-27294.
- [34] PAPARELLA F, OLIVIERI G, VOLPE G, et al. A Deep Reinforcement Learning Approach for Route Planning of Autonomous Vehicles* [C]// 2024 IEEE International Conference on Systems, Man, and Cybernetics(SMC). IEEE, 2024; 2047-2052.
- [35] ASHA S, SREE K D, SANDHYA S A. Federated Learning-Based Network Slicing in 6G Systems for Enhanced Performance in Autonomous Vehicle Communication[C]// 2024 First International Conference on Software, Systems and Information Technology(SSITCON). IEEE, 2024; 1-6.
- [36] HUANG Y. Multimedia tasks-oriented edge computing offloading scheme based on graph neural network in vehicular networks[J]. IEEE Access, 2025, 13: 9780-9791.
- [37] ZHAO W, CHENGY, LIU Z, et al. Asynchronous DRL-based multi-hop task offloading in RSU-assisted IoV networks[J]. IEEE Transactions on Cognitive Communications and Networking, 2025, 11(1): 546-555.
- [38] SU S, YUAN P, DAI Y. Reliable computation offloading of DAG applications in internet of vehicles based on deep reinforcement learning[J]. IEEE Transactions on Vehicular Technology, 2025, 74(2): 2116-2128.
- [39] SHAO Z, LIAO Z, TANG X. A Task Dependency-based Duplicated Task Offloading Mechanism in Vehicular Edge Computing[C]// 2024 IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA). IEEE, 2024; 1400-1405.
- [40] ZHOU P, XING X, WANG X, et al. Adaptive Task Offloading in High-Speed Vehicular Networks Using Enhanced D3QN Algorithm[C]// 2024 IEEE International Symposium on Parallel and Distributed Processing with Applications (ISPA). IEEE, 2024; 607-614.
- [41] SHEN S B, LIN C. Foreword to the Special Topic: Opportunities and Challenges in Internet of Things Research [J]. Journal of Software, 2014, 25(8): 1621-1624.
- [42] DU X, LI X, ZHAO N, et al. A Joint Trajectory and Computation Offloading Scheme for UAV-MEC Networks via Multi-Agent Deep Reinforcement Learning[C]// ICC 2023 - IEEE International Conference on Communications. IEEE, 2023; 5438-5443.
- [43] CUI G, DUAN P, XU L, et al. Latency optimization for hybrid GEO-LEO satellite-assisted IoT networks[J]. IEEE Internet of Things Journal, 2023, 10(7): 6286-6297.
- [44] PARK J, CHUNG K. RL-based Computation Offloading Scheme for Improving QoE in Edge Computing Environments[C]// 2023 IEEE 9th World Forum on Internet of Things(WF-IoT). IEEE, 2023; 1-6.
- [45] HWANG J, NKENYEREYE L, SUNG N, et al. IoT service slicing and task offloading for edge computing[J]. IEEE Internet of Things Journal, 2021, 8(14): 11526-11547.
- [46] CHOUIKHI S, ESSEGHIR M, MERGHEM-BOULAHIA L.

- Clustering-Based Cooperative Computation Offloading Game for Dependent Tasks in Industrial Internet of Things Systems[C]// ICC 2023—IEEE International Conference on Communications. IEEE,2023;2159-2165.
- [47] LI S,LIN P,SONG J,et al. Computing-Assisted Task Offloading and Resource Allocation for Wireless VR Systems[C]// 2020 IEEE 6th International Conference on Computer and Communications(ICCC). IEEE,2020;368-372.
- [48] GUPTA S,CHAKARESKEI J,POPOVSKI P. Millimeter Wave meets Edge Computing for Mobile VR with High-Fidelity 8K Scalable 360° Video[C]// 2019 IEEE 21st International Workshop on Multimedia Signal Processing(MMSP). IEEE,2019;1-6.
- [49] GAO N,ZHOU J,WAN G,et al. Low-Latency VR Video Processing-Transmitting System Based on Edge Computing[J]. IEEE Transactions on Broadcasting,2024,70(3):862-871.
- [50] KUMAR P J,KANTH M K,NIKHIL B,et al. Edge Computing in 5G for Mobile AR/VR Data Prediction and Slicing Model [C]// 2023 World Conference on Communication & Computing (WCONF). IEEE,2023;1-4.
- [51] MEDEIROS A,DI MAIO A,BRAUN T,et al. Service Chaining Graph: Latency- and Energy-aware Mobile VR Deployment over MEC Infrastructures [C] // GLOBECOM 2022 — 2022 IEEE Global Communications Conference. IEEE,2022;6133-6138.
- [52] LIN P,LIU Y,ZHANGZ,et al. Cost-aware task offloading and migration for wireless virtual reality using interactive A3C approach[J]. IEEE Transactions on Vehicular Technology,2024,73(7):10850-10855.
- [53] LIN P,SONG Q,WANG D,et al. Resource Management for Pervasive-Edge-Computing-Assisted Wireless VR Streaming in Industrial Internet of Things[J]. IEEE Transactions on Industrial Informatics,2021,17(11):7607-7617.
- [54] COZZOLINO V,TONETTO L,MOHAN N,et al. Nimbus: towards latency-energy efficient task offloading for AR services [J]. IEEE Transactions on Cloud Computing,2023,11(2):1530-1545.
- [55] ZHANG J,SHI C,LI B. Optimizing Edge Computing for Telemedicine: A Deep Reinforcement Learning Approach to Task Offloading and Resource Allocation[C]// 2024 6th International Conference on Communications, Information System and Computer Engineering(CISCE). IEEE,2024;481-485.
- [56] KOVACEVIC I,UL HAQ R I,OKWUIBE J,et al. Reinforcement Learning based Cloud and Edge Resource Allocation for Real-Time Telemedicine [C] // 2023 IEEE 17th International Symposium on Medical Information and Communication Technology(ISMICT). IEEE,2023;1-6.
- [57] LIN P,SONG Q,YU F R,et al. Task offloading for wireless VR-enabled medical treatment with blockchain security using collective reinforcement learning[J]. IEEE Internet of Things Journal,2021,8(21):15749-15761.
- [58] LIN D,TANG Y. Edge computing-based mobile health system: network architecture and resource allocation[J]. IEEE Systems Journal,2020,14(2):1716-1727.
- [59] ZHANG C,LIN B,HAN Y,et al. Secure Computation Offloading for Integrated Communication and Computing Maritime Networks[C]// 2024 IEEE/CIC International Conference on Communications in China(ICCC). IEEE,2024;1846-1850.
- [60] DIAO X,ZHENG J,CAI Y,et al. Fair data allocation and trajectory optimization for UAV-assisted mobile edge computing[J]. IEEE Communications Letters,2019,23(12):2357-2361.
- [61] CHEN X,ZHENG S. Resource Allocation and Task Offloading Strategy Base on Hybrid Simulated Annealing-Binary Particle Swarm Optimization in Cloud-Edge Collaborative System[C]// 2022 IEEE 5th Advanced Information Management, Communications, Electronic and Automation Control Conference(IMCEC). IEEE,2022;379-383.
- [62] ZHENG Y,ZHANG T,HUANG R,et al. Computing Offloading and Semantic Compression for Intelligent Computing Tasks in MEC Systems[C]// 2023 IEEE Wireless Communications and Networking Conference(WCNC). IEEE,2023;1-6.
- [63] LI Y,XU G,YANG K,et al. Energy efficient relay selection and resource allocation in D2D-enabled mobile edge computing[J]. IEEE Transactions on Vehicular Technology,2020,69(12):15800-15814.
- [64] XIA J,LIU Y,TAN L. Joint optimization of trajectory and task offloading for cellular-connected multi-UAV mobile edge computing[J]. Chinese Journal of Electronics,2024,33(3):823-832.
- [65] WANG J,WANG L,ZHU K,et al. Lyapunov-based joint flight trajectory and computation offloading optimization for UAV-assisted vehicular networks[J]. IEEE Internet of Things Journal,2024,11(12):22243-22256.
- [66] LI N,HAO W,ZHOU F,et al. Smart grid enabled computation offloading and resource allocation for SWIPT-based MEC system[J]. IEEE Transactions on Circuits and Systems II: Express Briefs,2022,69(8):3610-3614.
- [67] MAHN T,WIRTH M,KLEIN A. Game Theoretic Algorithm for Energy Efficient Mobile Edge Computing with Multiple Access Points[C]// 2020 8th IEEE International Conference on Mobile Cloud Computing, Services, and Engineering (MobileCloud). IEEE,2020;31-38.
- [68] ZHANG J,HUANG X,YU R. Optimal task assignment with delay constraint for parked vehicle assisted edge computing: a stackelberg game approach[J]. IEEE Communications Letters,2020,24(3):598-602.
- [69] ZENG F,CHEN Q,MENG L,et al. Volunteer Assisted Collaborative Offloading and Resource Allocation in Vehicular Edge Computing[J]. IEEE Transactions on Intelligent Transportation Systems,2021,22(6):3247-3257.
- [70] WANG R,ZENG F,YAO L,et al. Game-theoretic algorithm designs and analysis for interactions among contributors in mobile crowdsourcing with word of mouth[J]. IEEE Internet of Things Journal,2020,7(9):8271-8286.
- [71] YU S,CHEN X,YANG L,et al. Intelligent edge: leveraging deep imitation learning for mobile edge computation offloading [J]. IEEE Wireless Communications,2020,27(1):92-99.
- [72] EBRAHIM M A,EBRAHIM G A,MOHAMED H K,et al. A deep learning approach for task offloading in multi-UAV aided mobile edge computing [J]. IEEE Access,2022,10:101716-101731.
- [73] SUTTON R S,BARTO A G. Reinforcement learning: an intro-

- duction[J]. IEEE Transactions on Neural Networks, 1998, 9(5):1054-1054.
- [74] SEOL J, KIM J, KANCHARLA A. DRL Model for Distributed Agent-based IoT on Multi-Access Edge Computing for Accident Forecast[C] // 2023 IEEE/ACIS 8th International Conference on Big Data, Cloud Computing, and Data Science(BCD). IEEE, 2023:154-161.
- [75] WANG B, YUAN X, CHAI Y. Multi-Agent Deep Reinforcement Learning Based on Actor-Critic for Task Offloading[C] // 2024 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting(BMSB). IEEE, 2024:1-6.
- [76] WANG C, ZUO M, LIU S, et al. Research on Persistent Advantage Learning Based Deep Q Network for Computational Offloading in Edge Computing[C] // 2023 Eleventh International Conference on Advanced Cloud and Big Data(CBD). IEEE, 2023:206-210.
- [77] XU X, LIU Q, LUO Y, et al. A computation offloading method over big data for IoT-enabled cloud-edge computing[J]. Future Generation Computer Systems, 2019, 95:522-533.
- [78] WANG Y, CHEN C R, HUANG P Q, et al. A new differential evolution algorithm for joint mining decision and resource allocation in a MEC-enabled wireless blockchain network[J]. Computers & Industrial Engineering, 2021, 155:107186.
- [79] LIAO Y, LIU L, SONG Y, et al. Joint communication-caching-computing resource allocation for bidirectional data computation in IRS-assisted hybrid UAV-terrestrial network[J]. Chinese Journal of Electronics, 2024, 33(4):1093-1103.
- [80] BI J, YUAN H, DUANMU S, et al. Energy-optimized partial computation offloading in mobile-edge computing with genetic simulated-annealing-based particle swarm optimization[J]. IEEE Internet of Things Journal, 2021, 8(5):3774-3785.
- [81] PHAM HG T, PHAM Q V, PHAM A T, et al. Joint task offloading and resource management in NOMA-based MEC systems: a swarm intelligence approach[J]. IEEE Access, 2020, 8:190463-190474.
- [82] LUO Q, LI C, LUAN T H, et al. Minimizing the delay and cost of computation offloading for vehicular edge computing[J]. IEEE Transactions on Services Computing, 2022, 15(5):2897-2909.
- [83] ANNU, RAJALAKSHMI P. Joint Scheduling of Communication and Computation Resources for Efficient Computation Offloading in MEC-based V2X Systems[C] // 2023 IEEE 9th World Forum on Internet of Things(WF-IoT). IEEE, 2023:1-8.
- [84] GUO Y, ZHAO Z, ZHAO R, et al. Intelligent offloading strategy design for relaying mobile edge computing networks[J]. IEEE Access, 2020, 8:35127-35135.
- [85] ZHAO M, YANG Z, HE Z, et al. Multi-round stackelberg game-based pricing and offloading in containerized MEC networks[J]. IEEE Transactions on Green Communications and Networking, 2025, 9(1):191-206.
- [86] WANG T, ZHOU J, CHEN X, et al. A three-layer privacy preserving cloud storage scheme based on computational intelligence in fog computing[J]. IEEE Transactions on Emerging Topics in Computational Intelligence, 2018, 2(1):3-12.
- [87] JIANG N, WANG J, YU X, et al. Computer Data Security and Encryption Algorithm Based on Mobile Edge Computing[C] // 2023 International Conference on Networking, Informatics and Computing(ICNETIC). IEEE, 2023:582-586.
- [88] LI X, CHEN T, CHENG Q, et al. Smart Applications in Edge Computing: Overview on Authentication and Data Security[J]. IEEE Internet of Things Journal, 2021, 8(6):4063-4080.
- [89] ZHENG Z M, HE J F, TANG L X, et al. Key Theories and Frontier Applications of Privacy-computing [J]. Bulletin of National Natural Science Foundation of China, 2024, 38(4):603-611.
- [90] SUN J M, ZHAO M X. A Survey on the Application Research of Differential Privacy in Edge Computing [J]. Computer Science, 2024, 51(S1):896-904.
- [91] NIU S, XU X, WU T, et al. UAV-Assisted Federated Edge Learning Framework Based on Hierarchical Differential Privacy and Model Segmentation[C] // 2024 43rd Chinese Control Conference(CCC). IEEE, 2024:8792-8797.
- [92] LIU J S, WANG X A, YU D, et al. A Key Target Face Recognition Scheme Based on Homomorphic Encryption and Edge Computing [J]. Journal of Information Security Research, 2024, 10(11):1004-1011.
- [93] YANG Q. AI and Data Privacy Protection: The Way to Crack Federated Learning [J]. Journal of Information Security Research, 2019, 5(11):961-965.
- [94] LEE B S, KIM J W, CHOI M J. Federated Learning Based Network Intrusion Detection Model[C] // 2023 24st Asia-Pacific Network Operations and Management Symposium(APNOMS). IEEE, 2023:330-333.
- [95] CONG R, ZHAO Z, ZHANG L, et al. Cost-Effective Server Deployment for Multi-Access Edge Networks: A Cooperative Scheme[J]. IEEE Transactions on Parallel and Distributed Systems, 2024, 35(9):1583-1597.
- [96] ZHANG X D, ZHANG C K, ZHAO J J. Research Progress of Edge Intelligence[J]. Journal of Computer Research and Development, 2023, 60(12):2749-2769.



HUAN Haisheng, born in 1997, post-graduate. His main research interests include mobile edge computing and optimization of computing offload technology.



ZHAO Peng, born in 1979, Ph.D, associate professor. His main research interests include intelligent information processing, recommendation systems and distributed computing.