



# 计算机科学

COMPUTER SCIENCE

## 移动边缘计算中任务卸载研究综述

高月红, 陈露

引用本文

高月红, 陈露. 移动边缘计算中任务卸载研究综述[J]. 计算机科学, 2022, 49(11A): 220400161-7.

GAO Yue-hong, CHEN Lu. Survey of Research on Task Offloading in Mobile Edge Computing[J].

Computer Science, 2022, 49(11A): 220400161-7.

---

## 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)**

[边云协同计算中成本感知的物联网数据处理方法](#)

Cost-aware IoT Data Processing in Edge-Cloud Collaborative Computing

计算机科学, 2022, 49(11A): 211000101-7. <https://doi.org/10.11896/jsjcx.211000101>

[基于深度确定性策略梯度的服务器可靠性任务卸载策略](#)

Server-reliability Task Offloading Strategy Based on Deep Deterministic Policy Gradient

计算机科学, 2022, 49(7): 271-279. <https://doi.org/10.11896/jsjcx.210600040>

[基于深度强化学习的边云协同资源分配算法](#)

Edge-Cloud Collaborative Resource Allocation Algorithm Based on Deep Reinforcement Learning

计算机科学, 2022, 49(7): 248-253. <https://doi.org/10.11896/jsjcx.210400219>

[多无人机使能移动边缘计算系统中的计算卸载与部署优化](#)

Computation Offloading and Deployment Optimization in Multi-UAV-Enabled Mobile Edge Computing Systems

计算机科学, 2022, 49(6A): 619-627. <https://doi.org/10.11896/jsjcx.210600165>

[D2D辅助移动边缘计算下的卸载策略优化](#)

Optimization of Offloading Decisions in D2D-assisted MEC Networks

计算机科学, 2022, 49(6A): 601-605. <https://doi.org/10.11896/jsjcx.210200114>

# 移动边缘计算中任务卸载研究综述

高月红 陈露

北京邮电大学信息与通信工程学院 北京 100876

**摘要** 随着物联网的普及以及 5G 等无线通信技术的发展,各类新型业务层出不穷,移动数据流量也呈指数级的增长趋势。为了保证服务质量,移动计算模式从传统的云计算转变为了移动边缘计算(MEC),移动边缘计算的主要特点是将网络资源“下沉”到网络边缘,以满足时延敏感型和计算密集型业务的需求,从而给用户提供更好的服务。而任务卸载是移动边缘计算的主要研究问题之一,文中对近几年移动边缘计算中的任务卸载研究现状进行了归纳和总结,首先,对 MEC 的基本概念、架构以及典型应用场景进行了介绍,然后阐述了任务卸载问题,分别从最小化时延、最小化能耗以及最小化时延和能耗的加权和 3 个方面分析和总结了业内的研究现状,最后从数据依赖性、用户移动性、资源公平性和信息安全性 4 个方面对未来的研究方向进行了展望。

**关键词** 移动边缘计算;边缘网络;任务卸载;任务延迟;能量消耗

**中图分类号** TP393

## Survey of Research on Task Offloading in Mobile Edge Computing

GAO Yue-hong and CHEN Lu

School of Information and Communication Engineering, Beijing University of Posts and Telecommunication, Beijing 100876, China

**Abstract** With the popularity of the Internet of things and the development of wireless communication technologies such as 5G, various new services emerge one after another, and mobile data traffic is also growing exponentially. In order to guarantee the quality of service, the mobile computing model has changed from traditional cloud computing to mobile edge computing(MEC). The main feature of mobile edge computing is setting network resources at the edge of the network to meet the needs of delay-sensitive and computation-intensive tasks and provide users with better services. Task offloading is one of the main research problems in mobile edge computing. This paper summarizes the research status of task offloading in MEC in recent years. Firstly, the basic concept, framework and typical application scenarios of MEC are introduced. Then it expounds the problem of task offloading, analyzes and summarizes the existing research results from minimum delay, minimum energy consumption and minimum weighted sum of delay and energy consumption respectively. Finally, the future research directions are prospected from four aspects: data dependency, user mobility, resource fairness and information security.

**Keywords** Mobile edge computing, Edge network, Task offloading, Task delay, Energy consumption

### 1 引言

随着物联网的普及和移动通信技术的发展,各类新型业务不断涌现,如增强现实(Augmented Reality, AR)/虚拟现实(Virtual Reality, VR)、自动驾驶和智慧医疗等智能应用,这些业务在丰富人们生活的同时,也为通信网络带来了巨大的压力。在传统的云计算模式中,用户需将数据上传至云端,待云端服务器处理后再将数据传回,但由于传输距离较远、传输路径不稳定等原因,传统模式无法满足这些时延敏感型、计算密集型业务的需求。而终端设备本身受硬件条件限制,无法处理复杂的计算任务。为了解决这些问题,研究者们提出了移动边缘计算(Mobile Edge Computing, MEC)<sup>[1]</sup>的概念,通过在网络边缘为用户提供云计算能力,从而降低服务交付时延,为用户提供更好的服务体验。

在边缘网络中,用户将计算任务卸载到边缘节点进行处理,可以获得低时延、低能耗的服务体验,但在实际应用中,服务器和终端设备之间的任务卸载通常需要考虑多种因素的

影响,如边缘资源的有限性、终端设备的移动性、计算任务的异构性<sup>[2]</sup>等。如何兼顾各种因素求解出最优的卸载策略是任务卸载问题的研究难点。

本文首先介绍了移动边缘计算的概念、基本架构以及应用场景,然后阐述了移动边缘计算中的任务卸载问题,并对相关研究工作进行了分析和总结,并描述了目前任务卸载中尚需解决的问题,最后对全文进行了总结。

### 2 移动边缘计算概述

#### 2.1 概念

MEC 的概念由欧洲电信标准协会(European Telecommunications Standard Institute, ETSI)于 2014 年正式提出,被定义为“在移动用户附近的无线接入网内为用户提供 IT 服务环境与云计算能力”<sup>[1]</sup>。需要注明的是,在绝大部分文献中,“M”代表“mobile”,意指移动网络环境;少部分文献中“M”表示“multi-access”,这是因为随着研究的不断推进,ETSI 为了将边缘计算扩展到 WiFi 等非 3GPP 场景,把早期的“mobile”

扩展为“multi-access”<sup>[3]</sup>,但由于移动环境下的边缘计算仍然是研究的重点,所以还是将其称为“移动边缘计算”。

MEC的基本原理是在数据源的附近进行计算<sup>[4]</sup>,由边缘节点为用户提供服务,当边缘节点无法满足应用需求时,需要借助云数据中心的辅助<sup>[5]</sup>。云-边-端架构图如图1所示。云中心集中部署了大量的计算资源和存储资源,能够处理复杂的计算任务;边缘节点距离用户较近,且具有一定的计算能力和存储能力,如Cloudlet、微型数据中心、中小型基站等<sup>[6]</sup>;终端包括智能手机、平板电脑、可穿戴设备等。这种架构可以从很大程度上缓解云计算中心和核心网的压力,从架构图也可以看出,MEC的提出并不是为了取代云计算,而是为了弥补云计算的局限性,为用户提供更好的服务。

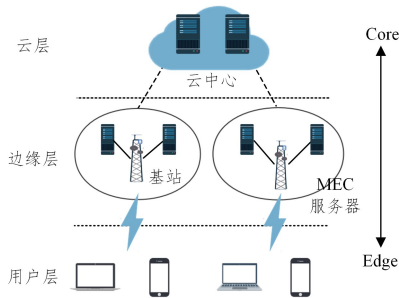


图1 云-边-端架构图

Fig.1 Cloud-edge-end architecture diagram

## 2.2 应用场景

移动边缘计算具有降低终端能耗、延长终端电池寿命、降低交付时延等特点,因此可以应用在不同的场景,如工业物联网、车联网、智能家居、智慧医疗等,下面介绍这几种典型的运用场景。

### 2.2.1 工业物联网

工业物联网(Industrial Internet of Things, IIOT)是一种以生产为导向的工业生态系统,它由大量工业设备组成,设备之间相互连接,利用数据和系统的互操作性,实现灵活的资源分配,以降低生产成本和提高生产率<sup>[7]</sup>。在工业物联网中,通过工业环境中的传感器、执行器等收集数据,然后利用云计算强大的存储能力和计算能力,进行深入的数据分析、挖掘和预测,典型的应用场景有智慧物流、智慧港口、智能仓储等。然而随着智能化的普及,工业场景中对于低时延的要求越来越高,例如在智能仓储中,需要对场景内的暴力分拣情况进行监控,此时需要低时延来保证监控视频分析的结果能够及时反馈回现场<sup>[8]</sup>。MEC靠近终端设备,可以为工业设备提供低时延的计算和存储服务,避免了与云计算中心之间因为较长回传链路造成的高时延;同时,将计算转移到边缘服务器进行,还可以帮助人们延长终端电池寿命,节省终端的硬件成本。

### 2.2.2 车联网

随着交通系统的日益智能化,越来越多的汽车和交通基础设施装备了通信设备,使得车辆与车辆、车辆与道路设施、车辆与路人之间可以相互通信。移动车载终端之间的信息交换,可以帮助人们获取最新的路况信息,以便做出路线调整,除此之外,所有移动车载终端还可以实时上传自己的数据,便于交通流量的监控,同时也有利于车祸时的紧急援助<sup>[9]</sup>。但随着车联网相关应用的发展,任务的计算强度也逐渐增加,而车辆本身的计算能力有限,难以满足任务的低时延要求。

移动车载终端往往与云计算中心相距较远,通信传输时延较长,另一方面,大量的车辆用户同时与云计算中心进行信息交互,会产生对带宽资源的竞争<sup>[10]</sup>。而MEC可以很好地解决这一难题,将MEC服务器部署在路边基站中,可以满足车载应用低时延和高带宽的计算需求,同时也能减少云端服务器的压力,使得云端服务器可以更好地对全局信息进行收集并处理。

### 2.2.3 智能家居

随着生活水平的不断提高,人们对居住环境的舒适和便利也提出了更高的要求,由此产生了家居智能化的需求。智能家居以家庭住宅为平台,结合综合布线技术、自动控制技术、网络通信技术等,实现家居设备的智能化,构建高效智能的住宅管理系统<sup>[11]</sup>。从用户的角度来看,智能家居就是在生活环境中配备各种智能设备,这些设备相互连接,用户可以通过任何可连接到互联网的设备(如智能手机)来控制家中的智能设备<sup>[12]</sup>。然而,随着智能家居的逐渐普及,传统的云计算模式逐渐不再适用。一方面,智能家居中各类传感器收集到的数据需上传至云端进行处理和分析,便于为用户提供更智能更优质的服务,而飞速增长的数据量可能会为传输链路带来巨大的压力,同时造成较高的时延;另一方面,从智能家居中获取的数据包含了个人隐私,较长的传输链路以及云计算的集中处理模式会为个人信息安全带来隐患。而边缘计算将计算移至更靠近用户的网络边缘,将数据就近处理,可以减小链路压力和回传时延,同时降低隐私泄露的风险。因此,边缘计算可以提高人们的家居生活质量。

### 2.2.4 智能医疗

随着人们生活压力的逐渐增大,以及人口老龄化趋势的逐渐显现,慢性病的发病率逐年增高<sup>[13]</sup>,如糖尿病、高血压等。由于慢性病的流行,人们对于医疗资源的需求也逐渐增长,尤其是在人口密度较高的地区,各级医院和诊所的医疗资源十分紧张,很多患者无法获得实时的医疗服务<sup>[14]</sup>。随着各项科学技术的发展,为了使有限的医疗资源可以为更多的患者服务,人们将物联网技术、云计算、大数据处理技术等应用在医疗领域,形成了一套以“感、知、行”<sup>[15]</sup>为核心的解决方案——智慧医疗。“感”即数据获取<sup>[16]</sup>,利用物联网中的传感设备收集患者的数据,再将其传至云端;“知”即数据处理,利用云计算中心强大的处理能力,对数据进行分析;“行”即提供服务,结合数据分析的结果,由智能医疗终端或医护人员为用户提供医疗方案。而将边缘计算引入智能医疗,可以提高服务的实时性。传感器收集到的数据不必传至云端,在边缘节点就可以得到处理,减少了传输时延。同时,不同的医疗机构可以通过边缘服务器实现数据共享,便于进行远程医疗会诊<sup>[17]</sup>。

## 3 移动边缘计算任务卸载研究现状

### 3.1 任务卸载问题

任务卸载,在很多文献中也被称作计算卸载<sup>[18]</sup>、数据卸载<sup>[19]</sup>或计算迁移<sup>[20]</sup>,它是边缘计算领域的主要研究问题之一,也是能否改善用户体验的关键。任务卸载指将终端设备中的时延敏感型业务或计算密集型业务卸载到计算资源更丰富的边缘服务器或云端服务器进行处理<sup>[21]</sup>。早在2000年左右,就有学者提出了类似的概念,以野外觅食做比喻,将任务

卸载称作网络觅食(Cyber Foraging)<sup>[22]</sup>。其核心思想是利用有限的硬件基础设施增加无线终端的计算资源,文献中将这种通过高带宽网络连接到互联网的有线基础设施称为代理(Surrogate),当移动终端进入一个区域时,会检测该区域内的潜在代理,通过近距离的无线点对点技术与其进行通信,当需要处理计算密集型业务时,移动终端将任务发送给代理,代理将为其处理该任务。这一概念在当时受到了学术界的广泛关注,但由于运营管理、技术实现等方面的原因<sup>[23]</sup>,最终没有被产业界采用。然而,近年来随着移动通信技术的发展以及物联网的逐渐普及,云计算中心模式的局限性逐渐暴露出来,为了解决这一难题,有学者提出将云中心的计算资源和存储资源“下沉”至网络边缘,形成“边缘计算”,与之相关的任务卸载、资源分配等问题也成为了重点研究对象。

终端设备受物理因素限制,自身的计算能力和存储能力无法满足某些业务的需求,于是将业务传到计算资源和存储资源相对更丰富的外部服务器进行计算<sup>[24]</sup>,然后再将结果传回,这一过程会涉及到卸载什么、如何卸载等一系列决策问题。这些决策共同决定了服务质量,服务质量通常由时延和能耗两个因素体现。因此,根据卸载目标的不同,任务卸载决策的设计通常可分为3类:以最小化时延为目标、以最小化能耗为目标、以权衡时延和能耗为目标。

### 3.2 相关研究

本节主要从以最小化时延为目标、以最小化能耗为目标、以权衡时延和能耗为目标3个方面对任务卸载策略进行阐述。表1列出了近年来的部分研究进展。

表1 任务卸载研究总结

Table 1 Summary of research on task offloading

卸载目标	参考文献	适用对象	关键研究点
最小化时延	文献[25-32]	时延敏感型应用	<p>采用拉格朗日乘子法和改进的自适应遗传算法获取任务卸载方案</p> <p>提出了一种能够降低时延且提高任务公平性的算法</p> <p>利用马尔可夫理论和一维搜索算法寻找最优解</p> <p>基于深度强化学习提出了一种卸载决策生成算法</p> <p>提出了不同带宽约束条件下的求解算法</p> <p>采用基于信息素机制的粒子群优化算法解决计算资源分配问题</p> <p>利用到达曲线和离开曲线分析时延并提出一种基于梯度的算法</p> <p>提出了一种基于遗传算法和模拟退火算法的启发式算法</p>
最小化能耗	文献[33-41]	耗能型应用	<p>提出了基于遗传算法算子的自适应粒子群优化算法</p> <p>设计了一种启发式算法寻找时延约束下的最优卸载方案</p> <p>将问题分解,利用凸优化和混合整数非线性规划相关理论依次解决问题</p> <p>提出了基于凸优化理论和 Gibbs 抽样的任务卸载算法</p> <p>提出了一种深度强化学习算法来求解能耗最小化的任务卸载策略</p> <p>提出了一种基于动态规划的算法解决任务调度和资源分配联合优化问题</p> <p>基于动态规划分别设计了在线算法和离线算法并进行分析比较</p> <p>基于蚁群系统和 DQN 设计了一种联合迭代算法用于求解任务卸载策略</p> <p>基于 TDMA 提出了一种基于阈值结构的分配策略并将其扩展至 OFDMA 系统</p>
权衡时延和能耗	文献[42-48]	对时延和能耗没有具体要求的应用	<p>将问题转化拆解并使用不同的方法解决各子问题</p> <p>将问题转化为马尔可夫决策过程并基于深度强化学习设计算法</p> <p>提出了基于超密集网络的多基站博弈算法</p> <p>提出了一种多目标进化算法解决任务调度和资源分配联合优化问题</p> <p>提出了一种基于 MOEA/D 的改进算法</p> <p>采用二分搜索和非合作博弈论对传输功率和卸载策略进行优化</p> <p>提出了一种基于李亚普诺夫优化的低复杂度任务分配算法</p>

#### 3.2.1 以最小化时延为目标

文献[25]基于智能电网场景,针对计算资源有限的电力终端处理应用时产生的高时延问题,对任务卸载和资源分配方案进行了研究,构建了混合整数非线性优化(Mixed Integer Nonlinear Programming, MINLP)问题,首先利用拉格朗日乘子法求出了资源分配的最优解,在此前提下,通过改进的自适应遗传算法(improved-AGA),求出了有效的任务卸载方案。仿真实验表明,与几种基线算法相比,该算法能够显著降低终端的总时延。类似地,文献[26]也建立了 MINLP 问题,不同之处在于,本文中任务不仅可以在本地或边缘服务器计算,还可以在其他终端进行计算,经过仿真验证,文中提出的 CTS (Comprehensive Task Scheduling)算法能够提高任务之间的公平性并减少总时延,但该方案的不足之处在于计算复杂度较高。文献[27]利用马尔可夫理论,通过分析任务的平均时延和设备的平均功耗,提出了在功率约束下的时延最小化问题,基于一维搜索算法寻找最优解,有效解决了该问题。文献[28]构建了 WPMECN(Wireless Powered Mobile Edge Computing Network)网络模型,在此模型下,基于深度强化学习

提出了一种卸载决策生成算法,与基线算法相比,该算法在 CPU 的处理延迟方面有较为突出的优势,但在任务平均时延方面与基线算法相差不大。文献[29]针对物联网中的时延敏感型应用,提出了无带宽约束下的近似算法和在线算法,以及带宽约束条件下的启发式算法,以求得最小时延下的卸载决策,并通过仿真证明了所提算法有很好的应用前景。文献[30]通过3个步骤解决了降低总时延的问题,首先对单个边缘节点的计算资源进行优化,在此基础上,采用基于信息素机制的粒子群优化算法对多个边缘节点的计算资源进行分配,最后将任务调度问题转化为半定规划问题进行求解,仿真证明该方法可以有效降低时延。文献[31]针对传统的基于李亚普诺夫方法的调度算法中缺乏时延公式的问题,利用到达曲线和离开曲线对边缘计算中的通信时延和计算时延进行分析,设计了一种基于梯度的算法来求解任务调度问题,与传统算法相比,该算法显著降低了时延。文献[32]研究了资源约束下的时延最小化问题,提出了一种基于遗传算法和模拟退火算法的启发式算法,仿真表明该算法性能优于经典算法,但当终端数量较多时,需要较高的迭代次数才能彰显其优势。

### 3.2.2 以最小化能耗为目标

文献[33]针对基于 DNN(Deep Neural Networks)的智能物联网,提出了基于遗传算法算子的自适应粒子群优化算法(Self-adaptive Particle Swarm Optimization algorithm using the Genetic Algorithm operators, SPSO-GA),用于求解在时延约束下使得能耗最小化的卸载策略,仿真证明,与其他基线算法相比,该算法得出的策略在节省能耗方面具有更好的表现。文献[34]为了进一步减少设备能耗,除了卸载决策外,还考虑了功率控制和计算资源分配的问题,将其建模成联合优化问题,提出了 E2PC(Energy Efficient Power Control)算法来寻找在时延约束下的最优解,仿真证明了该算法在降低能耗方面的优越性,但缺点在于算法复杂度过高。文献[35]将任务卸载问题拆分成优化卸载比例、功率分配以及计算资源分配 3 个子问题,利用凸优化和混合整数非线性规划相关理论,依次解决了几个子问题,求出了基于优先级的最优任务卸载策略。文献[36]研究了具有统计 QoS 保障的任务卸载决策,即允许任务完成的时间在一定范围内超出给定阈值,通过降低 QoS 要求来降低能耗,文中利用凸优化理论和 Gibbs 抽样的方法,提出了任务卸载算法,该算法通过迭代更新卸载策略,能够以较高的概率收敛到全局最优解。文献[37]针对单 MEC 系统和多 MEC 系统下的 AR 应用的任务卸载问题进行了研究,提出了一种基于 MADDPG(Multiagent Deep Deterministic Policy Gradient)框架的深度强化学习算法,通过训练 MADDPG 网络,求出了使得每个用户能耗最小化的任务卸载策略。文献[38]研究了 MEC 系统中的多用户任务卸载问题,以最小化多个移动设备的总能耗为目标,提出了基于动态规划的任务调度和资源分配的联合优化算法,通过仿真证明了算法的有效性。而为了降低算法复杂度,文中又提出了一种近似卸载算法,并仿真证明了近似算法下的总能耗接近最优解,且与最优解之间的差距在一个可确定的范围内。文献[39]以时延约束下的所有用户的总能耗最小化为目标,分别设计了基于动态规划的离线算法和在线算法,离线算法由于获得了信道统计信息和数据到达分布的先验信息,因此性能优于在线算法,但是这些先验信息的获取也会造成一定的开销。文献[40]研究了 MEC 网络中基于正交频分多址(Orthogonal Frequency Division Multiple Access, OFDMA)的多用户协同卸载和资源分配的问题,在该文献中,任务除了可以在本地计算或者边缘节点计算以外,还可以传输给其他终端协同计算。作者基于蚁群系统(Ant Colony System, ACS)和 DQN(Deep Q-network)设计了一种联合迭代算法(Joint Iterative Algorithm, JGA),结果表明算法可以达到较高的任务完成率,但在节省能耗方面的优势不够突出。文献[41]基于时分多址(Time Division Multiple Access, TDMA)和正交频分多址(OFDMA)研究了多用户的资源分配问题,提出了基于阈值结构的资源分配策略。其中,在 TDMA 系统中,在时延约束条件下,将资源的最优分配问题转化为了最小化加权能耗和的凸优化问题,推导了依赖于计算能耗和信道增益的卸载优先级函数,基站对每个移动终端执行二进制卸载决策,即完全卸载或最小卸载(卸载在时延约束下的最小数据量),其中优先级高于给定阈值的终端执行完全卸载,低于该阈值的终端执行最小卸载,随后又将这种基于阈值结构的分配策略扩展到 OFDMA 系统中,设计了一种低复杂度的算法来

解决混合整数优化问题。

### 3.2.3 以权衡时延和能耗为目标

文献[42]将最小化时延和能耗的问题转化成了凸优化问题,将问题拆分为卸载决策、时钟频率配置、传输功率分配和信道速率调度 4 个子问题,每个子问题用不同的方法求解,实验结果表明该分布式算法能够有效降低时延和能耗成本,但在任务数量较少的情况下,算法性能不够突出。文献[43]设计了基于 5G 的车辆感知多址边缘计算网络 VAMECN(Vehicle-Aware Multi-access Edge Computing Network),提出了系统时延和能耗加权和最小化的优化问题,将该混合整数非线性规划问题转化成马尔可夫决策过程,然后基于深度强化学习设计了计算卸载与任务迁移联合优化算法(Joint Computation Offloading and Task Migration Optimization, JCOTM),仿真证明,该算法可以在不同的系统环境下有效减少时延和能耗。文献[44]将移动边缘计算与超密集网络结合,设计了一种分布式边缘计算架构,提出了一种多基站博弈算法,仿真结果证明,相比于基线算法,该算法可以有效降低系统开销,当服务器数目增多时,系统开销会逐渐减小直至收敛。文献[45]提出了一种多目标进化算法(Multi-Objective Evolutionary Algorithm, MOEA),其核心思想是利用交叉、变异和选择等进化原理来寻找帕累托前沿,即平衡能耗和时延的最佳卸载方案。仿真证明,该方案能够有效降低时延和能耗,迭代次数越多,效果越好。类似地,文献[46]针对联合任务卸载、功率分配和资源分配的多目标优化问题,对基于多目标分解的进化算法(Multi-objective Evolutionary Algorithm based on Decomposition, MOEA/D)进行了改进,提出了 MOEAD\_MEC 算法,仿真结果表明,该算法在降低时延和能耗方面明显优于其他基线算法。文献[47]以最小化能耗和时延的加权和为目标,分别通过二分搜索法和非合作博弈论对传输功率和卸载策略进行优化,仿真表明,随着移动设备数量的增加,平均有益卸载用户数和平均计算开销均优于其他基线算法,但收敛的平均迭代次数也会随之增长。文献[48]设计了任务队列,并在时延的计算中考虑了排队时延,以移动终端的时延和能耗的加权和为目标,提出了一种基于李亚普诺夫优化的低复杂度任务分配算法——DLOTA(Dynamic Lyapunov Optimization based Tasks Allocation)算法,通过调节移动设备的 CPU 频率和卸载任务时的传输功率来取得能耗和时延之间的平衡,仿真证明该算法能够实现较高的能效和较低的时延。

综上所述,目前边缘计算中的任务卸载问题的研究通常以最小化某项指标(时延、能耗或是两者的加权和)为目标,建立关于时延和能耗的数学模型,考虑相关约束条件,设计算法,求解出任务卸载策略。然而由于场景需求的差异性、数据的异构性、网络的不稳定性等原因,为了求出最优解,在求解任务卸载策略的同时还需考虑资源分配、功率控制等并发问题。在算法设计中,已有文献大多利用凸优化、李亚普诺夫优化等数学优化算法和遗传算法、蚁群算法等启发式算法来求解任务卸载策略。

## 4 未来研究方向

目前,工业界已经对移动边缘计算中的任务卸载问题进行了大量的研究,并取得了进展,但仍然面临一些挑战,

值得我们进一步去探索。

#### 4.1 数据依赖性

在关于任务卸载问题的研究中,大多都假设计算任务之间是相互独立的,但随着任务类型的日益丰富,这样的假设可能不再适用。例如,人脸识别程序可分为5个子任务:目标获取、人脸检测、信息预处理、特征提取和分类<sup>[49]</sup>,特征提取的输出结果将作为分类任务的输入,那么分类只能在特征提取完成之后才能进行。如果在任务卸载过程中没有考虑计算任务之间的依赖关系,可能会导致整体的时延和能耗的增加<sup>[50]</sup>。文献[51-52]采用有向无环图(Directed Acyclic Graph,DAG)来表示子任务之间复杂的依赖关系,在满足依赖关系的前提下,多个子任务可以进行并行计算,以达到降低时延的目的,但是卸载策略的制定还需考虑众多因素,比如子任务应该在哪个 MEC 服务器进行卸载、任务的执行顺序、以及多用户之间的资源竞争等。数据依赖性势必会对卸载决策、资源分配等问题造成影响,因此如何在依赖关系的约束下设计合理的卸载策略是一个复杂的难题。

#### 4.2 用户移动性

在传统的无线网络中,移动终端在小区之间移动时,需要通过切换来维持服务的稳定性和连续性。类似地,当移动终端从一个边缘节点的覆盖范围移动到另一个边缘节点的覆盖范围时,服务质量也不应该受到影响。在目前关于任务卸载的研究中,某些特定场景下会考虑到终端的移动性,如车联网、无人机等,而大多数文献通过给定的运动模型或者路径预测来对终端移动性进行分析<sup>[53]</sup>,然后将任务迁移到目标服务器,以此维持服务的连续性。其中,如何对真实网络环境下的用户移动性做到精确预测是一项重大挑战。除此之外,用户更换目标服务器后还可能会导致服务器负载均衡或服务小区网络拥塞等问题,从而使服务质量受到影响。因此,如何为移动中的用户提供连续的高质量的服务还需研究人员进行深入研究,寻找解决方案。

#### 4.3 资源公平性

随着移动数据流量的爆炸性增长,大量不同类型的任务涌入边缘节点等待处理,任务和任务之间会对计算、存储和通信资源进行竞争,竞争导致的资源分配不均会使得服务质量下降。然而不同任务对资源的需求是不同的,例如,VR/AR、人脸识别等任务需要进行大量的计算,消耗 CPU 资源,数据检索类业务会消耗服务器的内存<sup>[54]</sup>。除此之外,不同的边缘节点由于 CPU、存储空间等配置不同,所能提供的服务能力也是有差异的。在目前的一些研究<sup>[55-56]</sup>中,根据任务的特性(如迫切程度、信道质量、处理成本等)赋予其相应的优先级,基于优先级进行任务计算资源的分配,可能会导致低优先级任务等待时间过长而无法在截止时间内完成计算。在研究任务卸载问题时,除了关注时延和能耗外,还需考虑资源的公平分配,针对现有资源设计高效的分配方案。

#### 4.4 信息安全性

在现有的任务卸载研究中,信息的安全性问题没有得到充分的重视。绝大多数文献的侧重点在于求解资源成本最小的卸载策略,通常不会考虑卸载的边缘节点是否可信,如何确保用户可以授权边缘服务,以及如何保护从边缘节点回传的数据隐私等安全性问题<sup>[57]</sup>。此外,在智能家居、智能医疗等应用中,通过传感器获取的数据涉及个人隐私,在传输过程

中容易泄漏,对个人隐私安全造成威胁;另一方面,边缘节点分布广泛又相互关联,单个节点若是遭到攻击或是出现故障,极有可能威胁到整个边缘计算系统的稳定性。虽然云计算环境下的安全防护方案不适用于 MEC,但是可以基于 MEC 的分布式、移动性等特性在其基础上进行改造。除此之外,由于 MEC 下的应用场景各有不同,安全防护方案还需满足不同场景对于安全性的需求,同时兼顾效能,这无疑是一个巨大的挑战。

**结束语** 近年来,随着日益增长的计算密集型、时延敏感型应用的涌现,移动边缘计算已受到国内外研究人员的广泛关注,任务卸载作为关键技术之一,也成为了 MEC 领域的重点研究内容。本文介绍了边缘计算的概念、架构及应用场景,以卸载目标作为区分依据,重点介绍了任务卸载问题的研究现状,通过分析和总结可以看出,任务卸载不仅能够解决移动终端计算资源有限的问题,还能缓解云计算中心压力,但用户移动性、信息安全性等问题还需进行更深层次的研究,期待在未来的研究中能有更好的解决方案。

#### 参考文献

- [1] HU Y C, PATEL M, SABELLA D, et al. Mobile edge computing—a key technology towards 5G [J]. ETSI White Paper, 2015, 11(11): 1-16.
- [2] XIE R C, LIAN X F, JIA Q M, et al. Survey on computation offloading in mobile edge computing [J]. Journal on Communications, 2018, 39(11): 138-155.
- [3] SHI W, CAO J, ZHANG Q, et al. Edge Computing: Vision and Challenges [J]. IEEE Internet of Things Journal, 2016, 3(5): 637-646.
- [4] ZHAO S, ZHANG X, CAO P, et al. Design of Robust and Efficient Edge Server Placement and Server Scheduling Policies [C] // 2021 IEEE/ACM 29th International Symposium on Quality of Service (IWQOS). 2021: 1-7.
- [5] LV J N, ZHANG J B, ZHANG Z F, et al. Survey of Mobile Edge Computing Offloading Strategies [J]. Journal of Chinese Computer Systems, 2020, 41(9): 1866-1877.
- [6] SUN Q B, LIU J, LI S, et al. Internet of Things: Summarize on Concepts, Architecture and Key Technology Problem [J]. Journal of Beijing University of Posts and Telecommunications, 2010, 33(3): 1-9.
- [7] NAVEEN S, KOUNTE M R. Key Technologies and challenges in IoT Edge Computing [C] // 2019 Third International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC). 2019: 61-65.
- [8] XU Y N, CAI C, HOU Y L, et al. An Optimization Scheme of 5G Network Slicing Technology in Logistics Warehouse [J]. Designing Techniques of Posts and Telecommunications, 2021(4): 84-87.
- [9] ARTHURS P, GILLAM L, KRAUSE P, et al. A Taxonomy and Survey of Edge Cloud Computing for Intelligent Transportation Systems and Connected Vehicles [C] // IEEE Transactions on Intelligent Transportation Systems. 2021.
- [10] WANG K, WANG X, LIU X, et al. Task Offloading Strategy Based on Reinforcement Learning Computing in Edge Computing Architecture of Internet of Vehicles [C] // IEEE Access. 2020: 173779-173789.

- [11] TONG X Y, FANG B Y, ZHANG Y Y. Internet of Things Smart Home Development Analysis [J]. *Mobile Communication*, 2010, 34(9): 16-20.
- [12] GUPTA N, ANANTHARAJ K, SUBRAMANI K. Containerized Architecture for Edge Computing in Smart Home: A consistent architecture for model deployment[C] // 2020 International Conference on Computer Communication and Informatics (ICCCI). 2020: 1-8.
- [13] CHEN Y, YU L, OTA K, et al. Robust activity recognition for aging society[J]. *IEEE Journal of Biomedical and Health Informatics*, 2018, 22(6): 1754-1764.
- [14] QIU Y, WANG C, QI K Y, et al. A Survey of Smart Health: System Design from the Cloud to the Edge[J]. *Journal of Computer Research and Development*, 2020, 57(1): 53-73.
- [15] NI M X, ZHANG Q, TAN H Y, et al. Smart Healthcare: from IoT to Cloud Computing [J]. *Scientia Sinica Informationis*, 2013, 43(4): 515-528.
- [16] XU J J, CHEN J. Development and applications of smart health care system based on internet of things[J]. *China Medical Devices*, 2017, 32(10): 118-121.
- [17] SACCO A, ESPOSITO F, XMARCHELTO F, et al. On Edge Computing for Remote Pathology Consultations and Computations[C] // *IEEE Journal of Biomedical and Health Informatics*. 2020: 2523-2534.
- [18] ZHANG Y L, LIANG Y Z, YIN M J, et al. Survey on the Methods of Computation Offloading in Mobile Edge Computing[J]. *Chinese Journal of Computers*, 2021, 44(12): 2406-2430.
- [19] DONG J, FENG F. Efficient data offloading method for edge computing with privacy protection[J]. *Application Research of Computers*, 2021, 38(7): 2072-2076.
- [20] ZHU Y K, LE G X, YANG X H, et al. Summary of Edge Computing Migration Research [J]. *Telecommunications Science*, 2019(4): 74-94.
- [21] MAO Y, YOU C, ZHANG J, et al. A Survey on Mobile Edge Computing: The Communication Perspective[C] // *IEEE Communications Surveys & Tutorials*. 2017: 2322-2358.
- [22] SATYANARAYANAN M. Pervasive computing: vision and challenges [J]. *IEEE Personal Communications*, 2001, 8(4): 10-17.
- [23] BALAN R K, FLINN J. Cyber Foraging: Fifteen Years Later[J]. *IEEE Pervasive Computing*, 2017, 16(3): 24-30.
- [24] WONG V, SCHOBER R, NG D, et al. Key Technologies for 5G Wireless Systems[D]. Cambridge: Cambridge University Press, 2017.
- [25] DOU H, XU Z, JIANG X, et al. Mobile Edge Computing Based Task Offloading and Resource Allocation in Smart Grid[C] // 2021 13th International Conference on Wireless Communications and Signal Processing (WCSP). 2021: 1-5.
- [26] LU Y, ZHAO Z, GAO Q. A Distributed Offloading Scheme With Flexible MEC Resource Scheduling[C] // 2021 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/IOP/SCI). 2021: 320-327.
- [27] LIU J, MAO Y, ZHANG J, et al. Delay-optimal computation task scheduling for mobile-edge computing systems[C] // 2016 IEEE International Symposium on Information Theory (ISIT). 2016: 1451-1455.
- [28] YU Y, YAN Y, LI S, et al. Task Delay Minimization in Wireless Powered Mobile Edge Computing Networks: A Deep Reinforcement Learning Approach[C] // 2021 13th International Conference on Wireless Communications and Signal Processing (WCSP). 2021: 1-6.
- [29] LI J. Maximizing User Service Satisfaction for Delay-Sensitive IoT Applications in Edge Computing[J]. *IEEE Transactions on Parallel and Distributed Systems*, 2022, 33(5): 1199-1212.
- [30] NIU X. Workload Allocation Mechanism for Minimum Service Delay in Edge Computing-Based Power Internet of Things[J]. *IEEE Access*, 2019, 7: 83771-83784.
- [31] ZHANG Y, DU P, WANG J, et al. Resource Scheduling for Delay Minimization in Multi-Server Cellular Edge Computing Systems[J]. *IEEE Access*, 2019, 7: 86265-86273.
- [32] WANG H, LI R, FAN L, et al. Joint computation offloading and data caching with delay optimization in mobile-edge computing systems[C] // 2017 9th International Conference on Wireless Communications and Signal Processing (WCSP). 2017: 1-6.
- [33] CHEN X, ZHANG J, LIN B, et al. Energy-Efficient Offloading for DNN-Based Smart IoT Systems in Cloud-Edge Environments[J]. *IEEE Transactions on Parallel and Distributed Systems*, 2022, 33(3): 683-697.
- [34] WU F, LENG S, MAHARJAN S, et al. Joint Power Control and Computation Offloading for Energy-efficient Mobile Edge Networks[J]. *IEEE Transactions on Wireless Communications*, 2022, 21(6): 4522-4534.
- [35] GHOSH S, AGRAWAL D P. Prioritized computation offloading and resource optimization for networks with strict latency[C] // 2021 IEEE 7th World Forum on Internet of Things (WF-IoT). 2021: 686-691.
- [36] LI Q, WANG S, ZHOU A, et al. QoS Driven Task Offloading With Statistical Guarantee in Mobile Edge Computing[J]. *IEEE Transactions on Mobile Computing*, 2022, 21(1): 278-290.
- [37] CHEN X, LIU G. Energy-Efficient Task Offloading and Resource Allocation via Deep Reinforcement Learning for Augmented Reality in Mobile Edge Networks[J]. *IEEE Internet of Things Journal*, 2021, 8(13): 10843-10856.
- [38] ZHAO T, ZHOU S, SONG L, et al. Energy-optimal and delay-bounded computation offloading in mobile edge computing with heterogeneous clouds[J]. *China Communications*, 2020, 17(5): 191-210.
- [39] LABIDI W, SARKISS M, KAMOUN M. Joint multi-user resource scheduling and computation offloading in small cell networks[C] // 2015 IEEE 11th International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob). 2015: 794-801.
- [40] TAN L, KUANG Z F, ZHAO L, et al. Energy-Efficient Joint Task Offloading and Resource Allocation in OFDMA-based Collaborative Edge Computing[J]. *IEEE Transactions on Wireless Communications*, 2022, 21(3): 1960-1922.
- [41] YOU C, HUANG K, CHAE H, et al. Energy-Efficient Resource Allocation for Mobile-Edge Computation Offloading[J]. *IEEE Transactions on Wireless Communications*, 2017, 16(3): 1397-1411.

- [42] YU H, WANG Q, GUO S. Energy-Efficient Task Offloading and Resource Scheduling for Mobile Edge Computing[C]//2018 IEEE International Conference on Networking, Architecture and Storage(NAS). 2018;1-4.
- [43] WU Z, YAN D. Deep reinforcement learning-based computation offloading for 5G vehicle-aware multi-access edge computing network[J]. China Communications, 2021, 18(11):26-41.
- [44] WANG R Y, WU H, CUI Y P, et al. Edge Offloading Strategy for the Multi-Base Station Game in Ultra-Dense Networks[J]. Journal of Xidian University, 2021, 48(4):1-10.
- [45] BOZORGCHENANI A, MASHHADI F, TARCHI D, et al. Multi-Objective Computation Sharing in Energy and Delay Constrained Mobile Edge Computing Environments [J]. IEEE Transactions on Mobile Computing, 2021, 20(10):2992-3005.
- [46] WANG P, LI K, XIAO B, et al. Multi-objective Optimization for Joint Task Offloading, Power Assignment, and Resource Allocation in Mobile Edge Computing[J]. IEEE Internet of Things Journal, 2022, 9(14):11737-11748.
- [47] YU X, SHI X Q, LIU Y X. Joint Optimization of Offloading Strategy and Power in Mobile-Edge Computing[J]. Computer Engineering, 2020, 46(6):20-25.
- [48] GUO S, JIANG Q, DONG Y, et al. TaskAlloc:Online Tasks Allocation for Offloading in Energy Harvesting Mobile Edge Computing[C]//2019 IEEE International Conference on Parallel & Distributed Processing with Applications, Big Data & Cloud Computing, Sustainable Computing & Communications, Social Computing & Networking(ISPA/BDCLOUD/SocialCom/SustainCom). 2019:116-123.
- [49] ZHAO W, CHELLAPPA R, PHILLIPS P J, et al. Face recognition: A literature survey[J]. ACM Computing Survey, 2003, 35(4):399-458.
- [50] GUO S, XIAO B, YANG Y, et al. Energy-efficient dynamic offloading and resource scheduling in mobile cloud computing[C]//IEEE INFOCOM 2016-The 35th Annual IEEE International Conference on Computer Communications. 2016:1-9.
- [51] SHANG Y, LI J, WU X. DAG-based Task Scheduling in Mobile Edge Computing[C]//2020 7th International Conference on Information Science and Control Engineering(ICISCE). 2020:426-431.
- [52] MING Z, LI X, SUN C, et al. Dependency-Aware Hybrid Task Offloading in Mobile Edge Computing Networks [C] // 2021 IEEE 27th International Conference on Parallel and Distributed Systems(ICPADS). 2021:225-232.
- [53] LI B, NIU L, HUANG X et al. Mobility Prediction Based Computation Offloading Handoff Strategy for Vehicular Edge Computing[J]. Journal of Electronics & Information Technology, 2020, 42(11):2664-2670.
- [54] MESKAR E, LIANG B. Fair multi-resource allocation with external resource for mobile edge computing[C]//IEEE INFOCOM 2018-IEEE Conference on Computer Communications Workshops(INFOCOM WKSHPs). 2018:184-189.
- [55] LIAO J X, WU X W. Resource Allocation and Task Scheduling Scheme in Priority-Based Hierarchical Edge Computing System [C]//2020 19th International Symposium on Distributed Computing and Applications for Business Engineering and Science (DCABES). 2020:46-49.
- [56] FAROOQ M O. Priority-Based Servicing of Offloaded Tasks in Mobile Edge Computing[C]//2021 IEEE 7th World Forum on Internet of Things(WF-IoT). 2021:581-585.
- [57] LUO Q, HU S, LI C, et al. Resource Scheduling in Edge Computing: A Survey[J]. IEEE Communications Surveys & Tutorials, 2021, 23(4):2131-2165.



**GAO Yue-hong**, born in 1981, Ph.D, associate professor, Ph.D supervisor. Her main research interests include wireless communication systems and network calculus theory.



**CHEN Lu**, born in 1999, postgraduate. Her main research interests include wireless communication and task scheduling in edge network.