

移动边缘计算中的 资源管理

Resource Management in Mobile Edge Computing



游昌盛 / YOU Changsheng

(新加坡国立大学, 新加坡 117576)
(National University of Singapore, Singapore 117576, Singapore)

摘要: 通过对移动边缘计算 (MEC) 网络的基本原理、应用场景, 以及通信和计算的研究模型的阐述, 提出了针对单用户和多用户 MEC 系统的绿色节能频谱和计算资源综合管理方案。通过分析当前 MEC 技术的局限和挑战, 认为 MEC 和人工智能技术的有机结合, 能够有效提高未来网络的计算性能。

关键词: 移动边缘计算; 无线通信; 资源管理

Abstract: For mobile edge computing (MEC), the principles, use cases, and its communication-and-computation modelling are introduced. Then, a set of energy-efficient joint radio-and-computational resource management is proposed for single-user and multi-user MEC systems. Finally, in view of existing limitations and challenges in MEC systems, an outlook on its seamless integration with artificial intelligence (AI) for improving the computing performance of future networks is provided.

Keywords: mobile edge computing; wireless communication; resource management

DOI: 10.12142/ZTETJ.202004002
网络出版地址: <https://kns.cnki.net/kcms/detail/34.1228.TN.20200714.0933.002.html>

网络出版日期: 2020-07-14
收稿日期: 2020-06-04

1 研究背景

随着 5G 技术和物联网 (IoT) 的发展, 无线设备的数目呈指数级增长, 物联网应用场景也越来越多样化。这其中包括大量计算密集型和时延敏感型的应用, 如虚拟现实、在线游戏等, 这类应用需要强大的计算能力支持来满足超低时延的要求。为了满足这种需求, 近年来, 传统的云计算网络架构正悄然向移动边缘计算 (MEC) 网络发生转变: 原本位于核心网云数据中心的计算服务和功能正在往网络边缘下沉, 通过离用户更近的基站和无线接入点向用户提供无处不在的计算、

存储、通信等服务, 从而有效降低用户的计算时延和能耗, 并大大提高整个网络的资源利用率^[1-4]。

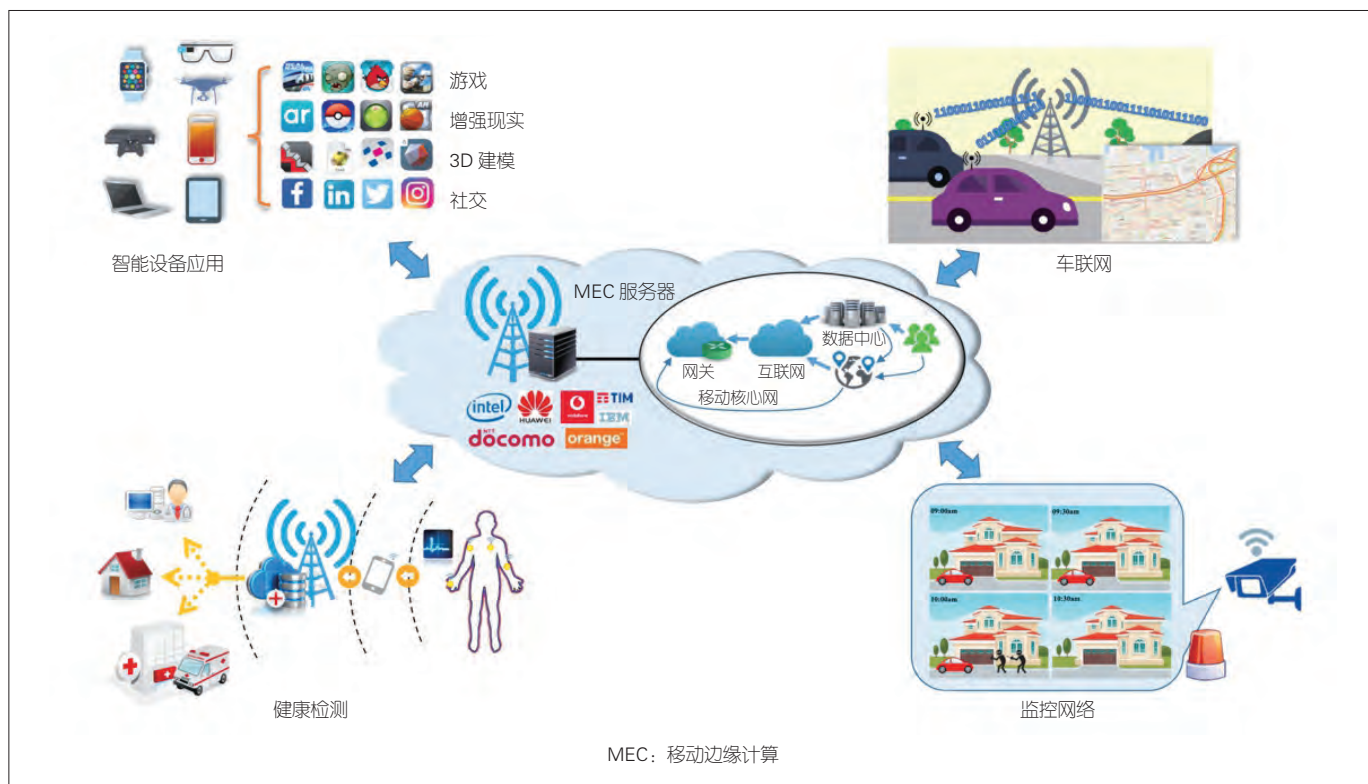
如图 1 所示, 在 MEC 系统中, 用户可将本地计算密集型的任务卸载 (迁移) 到 MEC 服务器中, 让它代为计算并向用户反馈计算结果。与传统的基于数据中心的移动云计算 (MCC) 相比, MEC 拥有如下几方面明显优势:

1) 更低的时延: 由于边缘云离用户更近且计算数据迁移到云的过程中不涉及在核心网中的数据传输, MEC 可以大大降低 MCC 中的数据传播和核心网回程链路时延。另一方面, 通过广泛部署 MEC 服务器, 每台 MEC 服

务器仅需向周边少量用户提供计算服务, 从而达到较低的计算时延。因此, 相较于 MCC 所需要的 100 ms 量级时延, MEC 可满足 1~10 ms 量级的超低时延要求。

2) 更低的能耗: MEC 用户可选择将高能耗型的计算任务迁移到边缘云中, 从而避免本地计算带来的巨大能耗。另一方面, 由于离 MEC 服务器更近, MEC 用户可以大大降低计算数据传输中的能量消耗。

3) 更优的情境感知: 利用近距离优势, MEC 服务器可以通过用户的定位信息等更加准确地预测和判断用户的计算行为和需求, 从而提供更及



▲图1 MEC系统架构与应用

时有效的计算和存储服务。

4) 更强的安全保护: 和 MCC 相比, MEC 服务器的用户数目更少, 且用户数据信息不需要经过复杂的核心网到达数据中心。这样可以有效缓解数据在多跳网络传输中的信息泄漏问题。

2 研究模型

为了研究 MEC 系统的计算性能, 我们首先介绍 MEC 的基本研究模型。

1) 计算任务模型: 总的来说, MEC 的计算卸载模型包括全部卸载和部分卸载。其中, 全部卸载计算模型适用于数据不可分割的高集成计算任务, 它要求用户只能选择全部本地计算或者全部卸载到 MEC 服务器。这类计算任务的关键参数包括: 计算数据量 (比特数)、计算强度 (每比特数据需要的中央处理器时钟数), 以及计算时延要求。这些参数与具体的计算任务相关, 可以通过对计算任务的

剖析和建模得出。另一方面, 部分卸载模型适用于两类计算任务: 一类是数据可任意分割的计算 (如数据压缩等); 另一类是包含多个子任务的计算, 不同任务间往往具有一定的运算顺序和联系, 如图 2 所示。与全部卸载模型相比, 部分卸载模型有更多的设计自由度, 可以更有效地卸载部分数据或子任务来减少用户的计算时延和能耗。

2) 计算时延和能耗模型: 对于用户的本地计算, 计算时延与计算所需的中央处理器 (CPU) 时钟数成正比, 与 CPU 主频成反比; 因此, 我们可以通过提高 CPU 的主频来降低本地计算时延, 但这样做同时也会增加本地计算的能耗。本地计算的能耗主要来自于 CPU 的功耗, 而 CPU 的功耗又与 CPU 主频的平方成正比; 因此, CPU 主频越高, 本地计算能耗越高且增长越快。对于 MEC 服务器 (或边缘

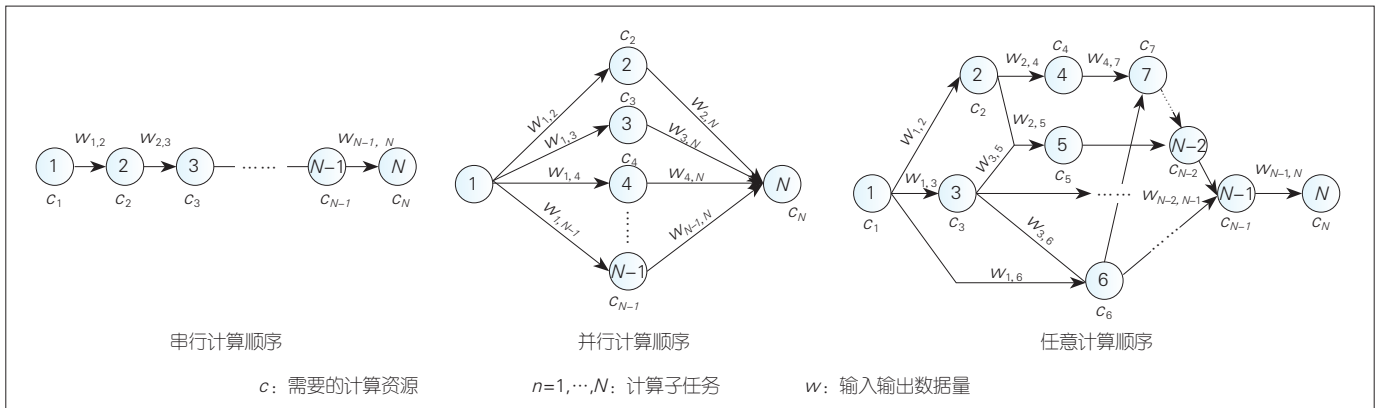
云) 来说, 它的计算时延包括 CPU 的运算时延和多计算任务下的队列时延, 能耗包括 CPU 的计算能耗以及服务器的开机运行能耗。

3) 无线数据传输模型: 5G 通信的各种技术, 如毫米波通信、非正交复用接入多址等都可以被有效用于提高计算卸载时数据传输的可达速率 (通过香农公式建模), 从而降低计算数据的传输时延。同时, 用户还可以利用设备到设备 (D2D) 的通信技术来实现低时延的用户间数据传输和计算卸载^[4]。

3 MEC 系统中的综合资源管理

无线和计算资源的综合管理是 MEC 系统设计中的重要组成部分。针对不同的 MEC 系统设置, 我们需要解决不同的综合资源管理问题。

首先考虑单用户情况下基于全部卸载计算模型的 MEC 资源管理。其中, 最重要的设计问题是如何做卸载决策,



▲图2 移动边缘计算任务不同的计算顺序

即是否进行计算卸载和如何设计卸载策略。为了研究这个问题，文献[6]提出了一个新型的无线供能下的 MEC 系统，并设计它最优的计算卸载方案。为了满足计算时延要求并最小化用户的计算耗能，我们分别优化了本地计算和计算完全卸载两种模式下的设计：对于本地计算，通过优化用户 CPU 的主频来降低计算能耗，同时满足计算耗能不大于获得的无线能量的条件；对于计算完全卸载，提出最优的时间分割方案，使用户能够先获取充足的能量然后进行数据传输和计算迁移，同时最大化用户的剩余能量。最后，基于本地计算和完全卸载两种模式的不同能耗，提出最优的本地计算/完全卸载的决策。这个工作后续被拓展到更加复杂的 MEC 系统，如基于能量收集的 MEC 系统^[7]和基于无线供能的多用户 MEC 系统^[8]。

对于多用户下的 MEC 系统而言，它的综合资源管理更加复杂。在文献[9]中，我们考虑部分卸载的计算模型并假设所有用户需要在相同的时间段内完成不同强度的计算任务。为了最小化所有用户的总计算能耗（包括每个用户的本地计算和卸载能耗），利用凸优化工具我们提出了一套最优的综合资源管理设计方案。具体来说，首先计算得到一个（计算）卸载优先

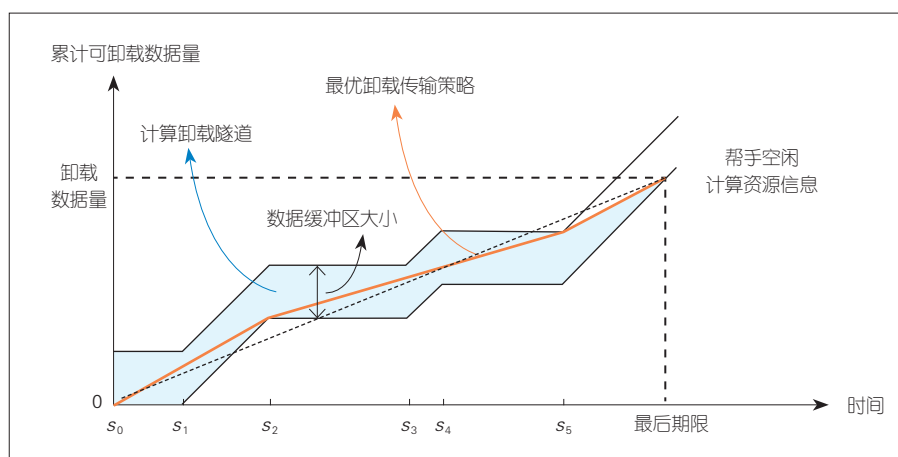
级函数，它与用户的信道增益和本地计算耗能成正比；因此，对于每个用户来说，他的无线信道越好或者本地耗能越大，用户的卸载优先级就越高。基于这个优先级函数，我们证明了最优的综合资源管理方案是一种基于门限的资源分配：对每个用户来说，如果他的卸载优先级函数值高于一定门限，他将选择把计算任务全部卸载到 MEC 服务器；反之，他将尽量在本地完成所有运算。与文献[9]中的集中式资源管理不同，文献[10]研究了分布式的计算资源分配。作者考虑了完全卸载的计算模型，并采用游戏理论来解决不同用户计算卸载与否的非确定性（NP）难问题。文献[10]中的研究证明，当用户受到的信号干扰强度低于一定门限时，他应该将计算任务卸载到云端。因为在这种情况下，无线传输可达速率较大，完成计算数据传输的能耗比本地计算更小。

文献[11]提出将基于边缘基站的 MEC 系统拓展到用户间计算卸载的 MEC 系统，从而有效降低边缘基站的计算和通信负载压力，并提高整个 MEC 系统的计算资源利用率。具体来说，主要利用用户周边的移动设备（如电脑等）的计算资源来支持用户的计算卸载。与基于边缘基站的 MEC 系统相比，用户周边的移动设备（简称帮手）

呈现出时有时无的空闲的计算资源。这是因为帮手只有在自己没有计算任务时，才能给周边的用户提供空闲的计算资源。利用这个特性，我们提出了一种基于帮手空闲计算资源的变速率计算卸载算法。该算法的核心在于首先在以横坐标为时间、纵坐标为累计可卸载数据量的坐标轴上构建一个“计算卸载隧道”，隧道的顶部和底部形状与帮手的缓存区大小和空闲计算资源的存量有关，具体如图3所示。为了最小化用户能耗，用户的计算卸载速率可以利用这个计算卸载隧道和几何方法得到。直观来看，如图3所示，这个方法就是在隧道的两端拉一条绷紧的线，不同线段的斜率反应了不同时间段内计算卸载的数据传输速率。这个方法可以被进一步拓展到多用户间的计算卸载场景。

4 MEC 未来工作展望

用人工智能算法设计 MEC 策略。当前的 MEC 策略设计主要有两种方法：一种是利用凸优化等优化理论来设计最优或次优的计算卸载策略，但是对于大规模 MEC 系统或优化问题本身是 NP 难的情况，用优化理论来设计 MEC 策略的方法可能需要很长的时间，这与 MEC 致力于缩短计算时延的初衷相违背；另一种方法是用启发式算法



▲图3 用户间计算卸载策略

来设计低复杂度的 MEC 策略,但这类方法往往缺乏一定的理论支撑,可能无法达到较好的 MEC 计算性能。为了解决这个问题,一个有效的方法是利用人工智能技术来实现快速高效的计算卸载策略设计。例如,我们可以将 MEC 策略优化问题转化为相应的深度学习问题:神经网络输入是用户的计算模型信息,神经网络的输出是计算卸载的策略。通过大量的计算卸载策略采样,我们可以训练出一个智能神经网络。这样一来,在实际的计算卸载决策中,我们只需将即时的计算模型信息输入到神经网络中,就可以快速得到一个有效的计算卸载策略方案。对于大规模 MEC 系统来说,基于凸优化理论的策略采样可能无法实现,但这时候我们可以利用小规模 MEC 系统进行神经网络训练,然后通过迁移学习的方法得到大规模 MEC 系统的策略采样。如何设计神经网络的训练是未来研究工作中的一个重要方向。

针对人工智能算法的 MEC 建模和设计。当前的 MEC 计算模型主要考虑普适性的计算,即计算数据量的大小与计算强度通常是一个固定的线性关系,但这个简单模型并不一定适用于具体的人工智能算法。例如,深度学习的计算复杂度除了和数据相关外,

还与神经网络的深度、每层网络的节点数、网络的类型(如卷积/自回归网络)等息息相关。因此,如何对具体人工智能算法进行计算模型建模是一个亟待探索和研究的重大问题。除此之外,当前的 MEC 策略设计主要关注计算时延和能耗,但这些性能指标并不完全是人工智能应用中最关心的问题。

5 结束语

MEC 将无线设备终端上计算密集型的运算任务迁移到边缘云中,从而有效降低用户的计算时延和能耗。在本文中,我们阐述了 MEC 系统的基本原理和模型,并提出如何紧密结合无线通信和计算机技术来设计不同 MEC 系统下的计算卸载策略。在无线网络智能化的关键时期,如何将 MEC 和人工智能技术有机结合来提升未来无线网络的性能需要继续探索。

参考文献

- [1] MAO Y Y, YOU C S, ZHANG J, et al. A survey on mobile edge computing: the communication perspective [J]. IEEE communications surveys & tutorials, 2017, 19(4): 2322–2358. DOI: 10.1109/comst.2017.2745201

- [2] YU W, LIANG F, HE X F, et al. A survey on the edge computing for the Internet of Things [J]. IEEE access, 2018, 6: 6900–6919. DOI: 10.1109/access.2017.2778504
- [3] 马洪源. 面向 5G 的边缘计算及部署思考 [J]. 中兴通讯技术, 2019, 25(3): 77–81. DOI: 10.12142/ZTETJ.201903011
- [4] QIN M, CHEN L, ZHAO N, et al. Power-constrained edge computing with maximum processing capacity for IoT networks [J]. IEEE Internet of things journal, 2019, 6(3): 4330–4343. DOI: 10.1109/ijot.2018.2875218
- [5] PU L J, CHEN X, XU J D, et al. D2D fogging: an energy-efficient and incentive-aware task offloading framework via network-assisted D2D collaboration [J]. IEEE journal on selected areas in communications, 2016, 34(12): 3887–3901. DOI: 10.1109/jsac.2016.2624118
- [6] YOU C S, HUANG K B, CHAE H. Energy efficient mobile cloud computing powered by wireless energy transfer [J]. IEEE journal on selected areas in communications, 2016, 34(5): 1757–1771. DOI: 10.1109/jsac.2016.2545382
- [7] MAO Y Y, ZHANG J, LETAIEF K B. Dynamic computation offloading for mobile-edge computing with energy harvesting devices [J]. IEEE journal on selected areas in communications, 2016, 34(12): 3590–3605. DOI: 10.1109/jsac.2016.2611964
- [8] BI S Z, ZHANG Y J. Computation rate maximization for wireless powered mobile-edge computing with binary computation offloading [J]. IEEE transactions on wireless communications, 2018, 17(6): 4177–4190. DOI: 10.1109/twc.2018.2821664
- [9] YOU C S, HUANG K B, CHAE H, et al. Energy-efficient resource allocation for mobile-edge computation offloading [J]. IEEE transactions on wireless communications, 2017, 16(3): 1397–1411. DOI: 10.1109/twc.2016.2633522
- [10] CHEN X, JIAO L, LI W Z, et al. Efficient multi-user computation offloading for mobile-edge cloud computing [J]. ACM transactions on networking, 2016, 24(5): 2795–2808. DOI: 10.1109/tnet.2015.2487344
- [11] YOU C S, HUANG K B. Exploiting non-causal CPU-state information for energy-efficient mobile cooperative computing [J]. IEEE transactions on wireless communications, 2018, 17(6): 4104–4117. DOI: 10.1109/twc.2018.2820077

作者简介



游昌盛,新加坡国立大学博士后;主要研究方向为移动边缘计算、边缘学习、无人机通信、智能反射面通信等;2019年获IEEE通信学会亚太地区优秀论文奖;已发表SCI/EI论文20余篇。