



研究与开发

# 面向差异化QoE基于匹配理论的任务卸载与资源分配方案

丁雷<sup>1,2</sup>, 许晨<sup>1</sup>, 刘斯宸<sup>1</sup>, 谢星寰<sup>1</sup>, 唐琳<sup>2</sup>

(1. 北京邮电大学人工智能学院, 北京 100876;

2. 北京微芯区块链与边缘计算研究院, 北京 100876)

**摘要:** 随着6G通信网络和移动边缘计算(mobile edge computing, MEC)技术的迅速发展, 配备边缘服务器的基站部署密度持续提高, 计算任务呈现出日益多样化的趋势。为应对差异化用户体验质量(quality of experience, QoE)以及资源分配不均衡对系统性能的影响, 提出了一种高效的匹配卸载方案, 综合考虑了卸载决策以及计算资源分配, 构建了一个以最大化系统收益为目标的混合整数非线性规划问题。通过分解原问题, 基于双边匹配理论设计了一种迭代优化算法加以求解。仿真实验基于澳大利亚墨尔本中央商务区(central business district, CBD)的公共数据集进行验证, 结果表明, 与现有方案相比, 所提方案在提升系统收益方面表现出显著的优势。

**关键词:** 移动边缘计算; 体验质量; 资源分配; 任务卸载

**中图分类号:** TN929.5

**文献标志码:** A

**doi:** 10.11959/j.issn.1000-0801.2025054

## Differentiated QoE-oriented task offloading and resource allocation scheme based on matching theory

DING Lei<sup>1,2</sup>, XU Chen<sup>1</sup>, LIU Sichen<sup>1</sup>, XIE Xinghuan<sup>1</sup>, TANG Lin<sup>2</sup>

1. School of Artificial Intelligence, Beijing University of Posts and Telecommunications, Beijing 100876, China

2. Beijing Academy of Blockchain and Edge Computing, Beijing 100876, China

**Abstract:** With the rapid development of 6G communication networks and mobile edge computing(MEC) technology, the deployment density of base stations equipped with edge servers has been continuously increasing, and computational tasks show a growing trend toward diversification. To address the impact of differentiated quality of experience (QoE) for users and unbalanced resource allocation on system performance, an efficient matching offloading

收稿日期: 2024-12-25; 修回日期: 2025-02-18

通信作者: 许晨, chen.xu@bupt.edu.cn

基金项目: 国家自然科学基金资助项目(No. 62071179); 国家重点研发计划项目(No. 2022-YFB-2503202); 未来区块链与隐私计算高精尖创新中心课题项目

**Foundation Items:** The National Natural Science Foundation of China (No. 62071179), The National Key Research and Development Program of China (No. 2022-YFB-2503202), Beijing Advanced Innovation Center for Future Blockchain and Privacy Computing



scheme was proposed. The scheme comprehensively considered both offloading decisions and computational resource allocation, and a mixed-integer nonlinear programming problem was formulated with the objective of maximizing system utility. By decomposing the original problem, an iterative optimization algorithm was designed based on bilateral matching theory for solution. Simulation experiments were conducted using public datasets from the central business district (CBD) of Melbourne, Australia. The results demonstrate that, compared with existing schemes, the proposed scheme shows significant advantages in improving system utility.

**Key words:** MEC, QoE, resource allocation, task offloading

## 0 引言

移动边缘计算 (mobile edge computing, MEC) 技术作为 6G 网络的重要组成部分, 凭借其低时延、高效能的特点, 为增强现实、虚拟现实、自适应视频流和在线游戏等计算密集型应用提供了强有力的支持<sup>[1-2]</sup>。然而, 随着物联网环境中计算任务日益复杂多样, 为了更高效地利用计算资源, 任务与网络资源合理匹配变得愈加关键<sup>[3]</sup>。受限于物理尺寸和电池容量, 移动设备 (mobile device, MD) 通常难以满足复杂计算任务的服务需求<sup>[4]</sup>。为应对这一挑战, MEC 通过任务卸载技术, 使 MD 能够将计算任务卸载至邻近基站 (base station, BS) 中的边缘服务器 (edge server, ES)。这种方式有效解决了设备资源受限的问题, 显著提升了计算性能, 并满足了用户对低时延的需求<sup>[5]</sup>。

在实际应用场景中, 随着小型基站密集部署, 如何高效地利用边缘服务器资源以进一步降低时延并提升用户体验质量 (quality of experience, QoE), 已成为当前研究的核心议题之一<sup>[6-8]</sup>。然而, 伴随着移动设备的快速发展以及数据流量的爆发式增长, 边缘服务器的通信资源和计算能力面临巨大压力。在任务卸载决策不当的情况下, 边缘服务器之间可能出现负载不均衡: 部分服务器过载而引发显著的处理时延, 而其他服务器资源闲置。这种失衡不仅降低了系统整体效率, 还可能导致严重的服务中断甚至系统崩溃。如何在多个服务器之间高效卸载任务的同时兼顾用户的差异化 QoE, 仍是尚未得到充分解决的难题。

为了高效利用 ES 资源, 研究人员提出了多种优化方案。例如, 针对低轨卫星在资源受限场景下的挑战, 文献[3]提出了一种综合考虑多维资源容量的策略, 实现了最大化资源利用率和用户服务率。文献[9]提出了一种基于模糊控制机制的分布式动态响应方法实现最大化用户分配率, 旨在应对 MEC 环境中用户资源需求的动态变化问题。此外, 针对边缘服务器与用户之间的空间距离及用户请求的时间分布变化问题, 文献[10]采用了一种基于遗传理论的启发式方法, 最大化用户的整体 QoE。文献[11]联合优化计算卸载与资源分配, 提出了一种基于任务完成时间和能耗加权总和的任务卸载收益最大化方案。文献[12]基于非合作博弈理论设计了一种分布式任务卸载算法, 旨在降低设备卸载时延。针对多租户收益与干扰之间的权衡问题, 文献[13]利用潜在博弈理论提出了一种用户分配策略, 旨在实现更高的资源利用率和服务质量。

上述方案在优化任务卸载和资源分配方面取得了一定成果, 但大多未充分考虑 MEC 网络中任务执行后用户获得 QoE 的差异性问题。在复杂的网络环境中, 受任务服务需求优先级以及计算资源需求复杂性的影响, 用户在处理任务过程中所获得的 QoE 收益存在差异。此外, 在 MEC 网络中, 用户不仅需要决定是否卸载任务, 还需要选择合适的小基站进行处理, 这使得任务卸载决策的复杂性显著增加。为应对上述挑战, 本文结合双边匹配理论<sup>[14]</sup>, 在充分考虑用户 QoE 差异化的基础上, 联合优化任务卸载决策及计算资源分配策略,

旨在实现系统收益最大化。本文的主要贡献如下。

(1) 充分考虑计算任务的多样性, 构建了任务资源需求与用户 QoE 之间的匹配关系。通过将系统收益定义为执行任务获得的用户 QoE 收益与卸载时延的相对改善收益的乘积之和, 提出了一个联合任务卸载与资源分配的混合整数非线性规划问题。

(2) 鉴于该问题是 NP 难 (non-deterministic polynomial-time hard) 问题, 无法在多项式时间内获得最优解。因此, 本文将原问题分解为多个具有独立目标和约束的子问题, 并结合双边匹配理论<sup>[14]</sup>, 提出了一种基于差异化 QoE 的任务卸载和资源分配 (differentiation QoE-task offloading and resource allocation, DQoE-TORA) 方案。

(3) 为验证所提方案的有效性, 本文基于澳大利亚墨尔本中央商务区 (central business district, CBD) 125 个边缘服务器 (基站) 的公共基准数据集<sup>[15-16]</sup>进行了多次仿真实验。实验结果显示, 所提方案在提升系统收益方面优于其他基准方案, 同时能够以较低的计算复杂度获得近似最优解。

## 1 系统模型

系统模型如图 1 所示, 假设该区域内有  $N$  个配备服务器的基站和  $M$  个用户, 分别用集合  $N = \{1, \dots, n, \dots, N\}$ ,  $M = \{1, \dots, m, \dots, M\}$  表示。其中, 用户  $m$  的任务可表示为二元组  $(l_m, w_m)$ ,  $l_m$  为输入数据大小,  $w_m$  为完成任务所需的中央处理器 (central processing unit, CPU) 周期数。

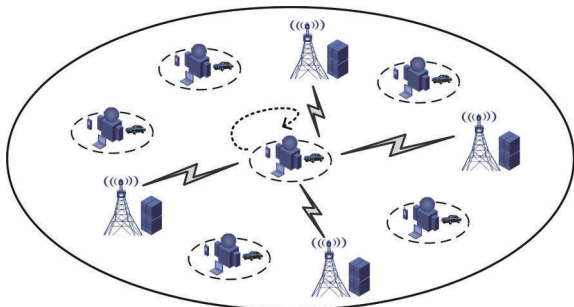


图 1 系统模型

### 1.1 任务模型

MEC 系统通常在蜂窝基站附近分布式部署了众多边缘服务器, 为周边用户提供计算处理和低时延网络服务。然而, 即使使用相同服务的用户, 其计算资源需求也可能存在差异。这种差异可能导致服务质量 (quality of service, QoS) 显著波动<sup>[10]</sup>, 从而直接影响用户体验。例如, 与普通的图像识别任务相比, 高清图形渲染等复杂任务会消耗更多的边缘服务器计算资源。研究表明, QoS 与 QoE 之间存在定量相关性, 即 QoE 会随着 QoS 的变化而波动<sup>[17]</sup>。

由于任务的计算资源需求与 QoS 密切相关, 本文基于任务的计算资源需求特征对任务进行分析<sup>[18]</sup>。其中, 任务均有一个  $q$  维资源属性向量  $\mathbf{attr}_m = [a_{m1}, a_{m2}, \dots, a_{mq}]$ ,  $a_{mq}$  代表一种计算资源, 例如内存等, 最终得到任务属性矩阵  $\mathbf{attr}_s$  为:

$$\mathbf{attr}_s = \begin{bmatrix} \mathbf{attr}_1 \\ \vdots \\ \mathbf{attr}_m \end{bmatrix} = \begin{bmatrix} a_{11} & \cdots & a_{1q} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mq} \end{bmatrix} \quad (1)$$

为降低模型复杂度, 本文选取了与时延相关的 3 个资源需求特征<sup>[7, 18]</sup>, 即 CPU、内存和磁盘空间, 用于构建任务的属性矩阵  $\mathbf{attr}_m = (\mathbf{attr}_m^d)$ , 其中,  $d \in D = [\text{cpu}, \text{diskspace}, \text{memory}]$ 。通过任务的资源需求属性矩阵, 可以预定义任务的 QoS<sup>[19]</sup>。此外, 任务的 QoS 与用户的 QoE 之间存在非线性关联<sup>[17]</sup>: QoE 在初期随着 QoS 的提升增长缓慢, 随后加速增长, 最终趋于饱和并逐渐收敛。这种关系通常采用 S 形函数进行建模<sup>[20]</sup>。

综上所述, 本文定义用户  $m$  的任务执行后获得的 QoE 收益为:

$$Q_m = \frac{L}{1 + e^{-\alpha(x_m - \delta)}} \quad (2)$$

其中,  $L$  为 QoE 的最大值,  $\delta$  为控制 QoE 增长的位置或函数的中点,  $\alpha$  为控制 QoE 水平的增长率 (即从最小到最大 QoE 水平的变化幅度)。  $x_m = \sum_{d \in D} y_m^d / |D|$ , 其中,  $x_m$  表示任务对应的 QoS,  $y_m^d$  表



示用户  $m$  的任务所需资源的规范化量。

## 1.2 通信模型

在 MEC 系统的通信模型中, 为了更真实地模拟实际网络环境, 考虑了用户与小基站之间的信道衰落特性。具体而言, 信道衰落由大尺度和小尺度衰落模型共同构建。小尺度衰落服从归一化的瑞利 (Reyleigh) 分布, 且距离相关路径损耗模型为  $PL(d)[\text{dB}] = PL(d_0) + 10\eta \lg(d/d_0)$ 。其中,  $d$  为信号传播的距离,  $d_0$  为参考距离,  $\eta$  为路径损耗指数,  $PL(d_0)$  为参考距离  $d_0$  处的路径损耗,  $PL(d)$  为距离为  $d$  的路径损耗。此外, 对数正态阴影的标准差为  $\sigma$ 。

为了提高频谱利用率并减少用户间的干扰, MEC 系统中引入了正交频分多址 (orthogonal frequency division multiple access, OFDMA) 技术。该技术将频带  $B$  划分若干大小为  $W = \frac{B}{K}$  [Hz] 的子频带, 从而确保每个小基站的用户在上行传输时保持正交性, 其中,  $K$  表示基站能够分配最大的子频带数。通过为每个用户分配独立的子频带, 保证了每个基站最多可以同时服务  $K$  个用户<sup>[11]</sup>。定义任务决策变量为  $a_{mn} = \{0, 1\}$ ,  $a_{mn} = 0$  表示本地执行,  $a_{mn} = 1$  表示用户  $m$  的任务卸载至小基站  $n$  上执行; 定义任务卸载策略为  $\Gamma = \{a_{mn} | m \in M, n \in N\}$ , 需要满足如下约束条件。

$$\sum_{m \in M} a_{mn} \leq 1, n \in N \quad (3)$$

$$\sum_{n \in N} a_{mn} \leq K, m \in M \quad (4)$$

根据香农公式, 用户  $m$  的任务到小基站  $n$  的信干噪比表示为:

$$\gamma_{mn} = \frac{p_m \|g_{mn}\|^2}{\sigma^2}, \forall m \in M, n \in N \quad (5)$$

其中,  $p_m$  表示用户  $m$  的传输功率,  $\|g_{mn}\|^2$  表示用户  $m$  与小基站  $n$  之间的信道增益,  $\sigma^2$  表示加性高斯白噪声的功率。

用户  $m$  的任务到小基站  $n$  的上行链路数据速率表示为:

$$R_{mn} = W \lg(1 + \gamma_{mn}) \quad (6)$$

## 1.3 计算模型

### (1) 本地计算

用户  $m$  的任务在本地计算模型独立执行, 其计算时延  $t_m^l$  表示为:

$$t_m^l = \frac{W_m}{f_m} \quad (7)$$

其中,  $f_m$  为用户本地计算能力 (即每秒 CPU 周期数)。

### (2) 边缘计算

每个小基站能够并行处理多个用户的计算卸载任务。在这种模式下, 卸载总时延由上行传输时延、任务执行时延和下行传输时延 3 部分组成。由于任务结果通常远小于输入数据, 本文假定下行传输时延忽略不计<sup>[21]</sup>。

用户  $m$  的任务卸载到小基站  $n$  上执行时, 传输时延  $t_m^{\text{up}}$  为:

$$t_m^{\text{up}} = \frac{l_m}{R_{mn}} \quad (8)$$

执行时延  $t_m^{\text{exe}}$  为:

$$t_m^{\text{exe}} = \frac{W_m}{f_{mn}} \quad (9)$$

定义计算资源分配策略  $F = \{f_{mn} | m \in M, n \in N\}$ ,  $f_{mn}$  表示小基站  $n$  分配给用户  $m$  的计算资源,  $F_n$  表示小基站  $n$  的总计算资源 (即每秒 CPU 周期数)。此外, 分配策略需要满足计算资源的约束条件为:

$$\sum_{m \in M} a_{mn} f_{mn} \leq F_n, \forall n \in N \quad (10)$$

用户  $m$  的任务卸载到小基站  $n$  上总时延为:

$$t_m^o = t_m^{\text{up}} + t_m^{\text{exe}} \quad (11)$$

## 1.4 问题公式化

在所考虑的场景中, 定义任务卸载时延的相对改善收益<sup>[22]</sup>为  $\frac{t_m^l - t_m^o}{t_m^l}$ 。因此, 本文系统收益定

义为执行任务获得的用户 QoE 收益与其卸载时延的相对改善收益的乘积之和。

$$E(\Gamma, F) = \sum_{m \in M} (E_m) \quad (12)$$

其中,  $E_m = Q_m \frac{t_m^l - t_m^o}{t_m^l}$ 。本文以最大化系统收益为目标, 基于用户 QoE 的差异化特性, 联合优化任务卸载决策及计算资源分配策略。优化问题表示为:

$$\begin{aligned} & \max_{\Gamma, F} E(\Gamma, F) \\ & \text{s.t. C1: } a_{mn} \in \{0, 1\}, m \in M, n \in N \\ & \text{C2: } \sum_{n \in N} a_{mn} \leq 1, m \in M \\ & \text{C3: } \sum_{m \in M} a_{mn} \leq K, n \in N \\ & \text{C4: } \sum_{m \in M} a_{mn} f_{mn} \leq F_n, \forall n \in N \\ & \text{C5: } f_{mn} \in [0, F_n], \forall m \in M, n \in N \end{aligned} \quad (13)$$

其中, 约束 C1 表示卸载决策为二元变量; 约束 C2 和 C3 表示每个用户最多只能接入 1 个小基站, 而每个小基站最多可同时服务  $K$  个用户; 约束 C4 和 C5 表示小基站分配给用户的总计算资源不超过其上限。

## 2 联合任务卸载和资源分配的优化方案

由于原问题在多项式时间内无法获得最优解, 本文通过固定卸载决策  $a_{mn}$  的方式, 将式 (13) 转化为多个具有独立目标和约束的子问题。基于这一特性, 本文将原本复杂度较高的问题解耦为 2 个更易求解的子问题: 资源分配子问题和卸载决策子问题。该方法有效地降低了问题的复杂性, 同时便于对每个子问题单独进行优化求解。在此基础上, 本文结合双边匹配理论提出了 DQoE-TORA 方案。

### 2.1 资源分配

给定满足式 (13) 中约束的可行卸载决策, 式 (13) 可转化为:

$$E(\Gamma, F) = \sum_{m \in M} \left[ Q_m \left( 1 - \frac{l_m f_m}{w_m W \ln(1 + \gamma_{mn})} \right) \right] - V(\Gamma, F) \quad (14)$$

$$V(\Gamma, F) = \sum_{m \in M} \frac{\Phi_m}{f_{mn}} \quad (15)$$

其中,  $\Phi_m = Q_m f_m$ 。通过分析可知, 与用户  $m$  相关参数  $l_m$ 、 $f_m$ 、 $w_m$ 、 $W$ 、 $\gamma_{mn}$  等为常量。因此, 式 (14) 右侧的第一项为常数, 则  $V(\Gamma, F)$  可视为所有用户的总卸载成本。定义卸载到小基站  $n$  的用户集合:  $U_n^1 = \{m | a_{mn} = 1, \forall m \in M\}$ , 式 (13) 可重新定义为最小化总卸载成本的优化问题, 即:

$$\begin{aligned} & \min_F \sum_{m \in U_n^1} \frac{\Phi_m}{f_{mn}} \\ & \text{s.t. C1: } \sum_{m \in U_n^1} f_{mn} \leq F_n, \forall n \in N \\ & \text{C2: } f_{mn} > 0, \forall m \in U_n^1 \end{aligned} \quad (16)$$

其中, 目标函数关于  $f_{mn}$  的二阶导数为  $\frac{2\Phi_m}{f_{mn}^3}$ :  $\Phi_m > 0, f_{mn} > 0 \Rightarrow \frac{2\Phi_m}{f_{mn}^3} > 0$ , 这表明  $\frac{\Phi_m}{f_{mn}}$  是关于  $f_{mn}$  的严格凸函数, 且约束 C1 是关于  $f_{mn}$  的线性不等式, 线性函数的可行域为凸集, 因此, 该约束满足凸性; 约束 C2 是关于  $f_{mn}$  的非负性约束, 其可行域为正实数空间, 属于凸集, 因此, 该约束也满足凸性。

综上所述, 目标函数与约束条件均满足凸性要求, 所构成的优化问题为凸优化问题。下面通过 KKT 条件对该问题进行求解。

首先将式 (16) 的拉格朗日函数构建为:

$$L = \sum_{i \in U_n^1} \frac{\Phi_m}{f_{mn}} + \lambda \left( F_n - \sum_{i \in U_n^1} f_{mn} \right) \quad (17)$$

对  $\frac{\partial L}{\partial f_{mn}} = 0$ , 其最优计算资源分配为:

$$f_{mn}^* = \frac{F_n \sqrt{\Phi_m}}{\sum_{k \in U_n^1} \sqrt{\Phi_k}}, \forall m \in U_n^1 \quad (18)$$



相应的, 最小总卸载成本为:

$$V^* = \frac{1}{F_n} \left( \sum_{m \in U_n^1} \sqrt{\Phi_m} \right)^2 \quad (19)$$

## 2.2 任务卸载

本节主要探讨给定计算资源分配策略下的任务卸载问题, 原问题可被转化为如下形式。

$$\begin{aligned} & \max_{\Gamma, F} E(\Gamma, F) \\ & \text{s.t. C1: } a_{mn} = \{0, 1\}, m \in M, n \in N \\ & \quad \text{C2: } \sum_{n \in N} a_{mn} \leq 1, m \in M \\ & \quad \text{C3: } \sum_{m \in M} a_{mn} \leq K, n \in N \end{aligned} \quad (20)$$

为获得全局最优解, 采用穷举法寻找最大化系统收益的任务分配方案, 但随着用户和小基站数量的增加, 其复杂度将呈指数级增长。本文基于双边匹配理论, 将式(20)建模为一个多对一的匹配问题<sup>[23]</sup>, 并在每次匹配过程中同步进行计算资源管理。整个过程基于匹配理论的一些核心概念<sup>[6, 24]</sup>, 定义如下。

**定义 1** 设匹配  $\Phi$  是一组匹配对, 其中,  $\Phi \in N \otimes M$ , 满足:

- (1)  $|\Phi_m| = 1, \forall m \in M$ ;
- (2)  $|\Phi_n| \leq K, \forall n \in N$ ;
- (3)  $m \in \Phi_n \Leftrightarrow \Phi_m = n$ 。

其中, (1) 表示每个用户最多匹配 1 个小基站; (2) 表示每个小基站最多匹配  $K$  个用户; (3) 表示如果用户  $m$  的任务匹配到小基站  $n$ , 那么小基站  $n$  的匹配对象也包含用户  $m$ 。

**定义 2** 匹配  $\Phi$  中的成员满足个体理性, 即每个用户和小基站都接受当前的匹配关系且不会主动寻求更高效用匹配时, 则形成无阻碍匹配。

根据以上定义, 为描述用户和小基站之间的偏好, 引入偏好顺序, 记为  $\succ$ ; 用户  $m$  对所有服小基站  $n$  的偏好顺序, 记为  $\succ_m$ , 其效用函数表示为用户  $m$  将任务卸载到小基站  $n$  所获得的卸载效

用:  $U_n^2(n) = E_m$ 。对于任意 2 个小基站  $n, n'$ , 当  $n, n' \in N$  且  $n \neq n'$ , 如果  $U_n^2(n) > U_n^2(n')$ , 该式表示, 用户  $m$  更倾向于选择小基站  $n$  而非  $n'$ ; 同理, 小基站  $n$  对所有用户  $m$  的偏好顺序, 记为  $\succ_n$ , 小基站  $n$  的效用函数定义为其服务的用户集合  $V \subseteq M$  带来的系统收益总和:  $U_n^2(V) = \sum_{m \in M} E_m$ , 对于任意用户集合  $V_1, V_2 \subseteq M$ , 若  $U_n^2(V_1) > U_n^2(V_2)$ , 则偏好顺序满足  $(n, V_1) \succ (n, V_2)$ 。此式说明, 小基站  $n$  更倾向于服务用户集合  $V_1$  而非  $V_2$ , 当且仅当  $V_1$  为小基站  $n$  带来的系统收益更高。

**定义 3** 交换匹配  $\Phi_m^m$  是在保持其他用户匹配关系不变的情况下, 将 2 个用户匹配的小基站进行交换的结果, 具体定义为:

$$\Phi_m^m = \{\Phi \setminus \{(m, n), (m', n')\} \cup \{(m, n'), (m', n)\}\} \quad (21)$$

其中,  $m' \in \Phi_m^m(n)$ ,  $m \in \Phi_m^m(n')$ ,  $m \notin \Phi_m^m(n)$ ,  $m' \notin \Phi_m^m(n')$ 。

**定义 4** 给定匹配  $\Phi$  和用户对  $(m, m')$ , 若  $\exists m \in \Phi(n), m' \in \Phi(n')$ , 满足:

- (1)  $\forall k \in \{m, m', n, n'\}, U_k^2(\Phi_m^m(k)) \geq U_k^2(\Phi(k))$ ;
- (2)  $\exists k \in \{m, m', n, n'\}, U_k^2(\Phi_m^m(k)) > U_k^2(\Phi(k))$ 。

这意味着交换匹配  $\Phi_m^m$  将被允许, 且  $(m, m')$  构成一个交换阻塞对。

综上所述, 稳定匹配意味着不存在参与者能够通过定义 3 中的交换匹配来提高效用, 即当且仅当不存在交换阻塞对时, 匹配才是稳定的。此外, 为确保匹配过程不会在相同结果之间反复循环, 交换匹配仅在至少有一方的效用值提高才会被批准。

## 2.3 联合任务卸载调度与资源分配

本文提出了一种基于交换匹配的联合任务卸载和资源分配算法, 具体的算法步骤详见算法 1。首先, 用户与小基站通过贪婪匹配策略获得初始匹配结果, 然后, 随机挑选 2 名用户, 根据定义 3 检查其当前匹配的小基站是否满足交换条件, 且保持其他用户的匹配关系不变; 最后, 重新分配

这2个小基站上的计算资源，并计算用户与小基站之间的效用值。若这2名用户满足交换条件，即形成交换阻塞对，则执行交换操作。经过多轮迭代，系统最终收敛到双边交换稳定的匹配结果。

#### 算法1 联合任务卸载和资源分配算法

输入 用户与小基站的位置分布

输出  $\Gamma^*, F^*$

计算用户QoE收益值

for  $m \in M$

通过式(2)得到每个用户对应的QoE收益值

end for

联合任务卸载和资源分配

初始化匹配  $\Phi_0$ : 用户与小基站贪婪匹配，满足  $|\Phi(m)| \leq 1$  和  $|\Phi(n)| \leq K$

交换匹配过程

while 没有交换阻塞对存在

for  $m \in M$

for  $m' \in M \setminus \{m\}$

通过KKT条件得到计算资源分配结果；

求解匹配  $\Phi_m^{m'}$  下所有用户和小基站的效用值

if  $(m, m')$  是一个阻塞对，那么

$$\Phi \leftarrow \Phi_m^{m'}$$

end if

end for

end for

end while

假设算法1在  $T$  次迭代后收敛，其复杂度约为  $\mathcal{O}(T(2KN^2 + MN))$ ，其中， $K$  为交换匹配所需的迭代次数。若采用穷举算法获得最优解，则需要  $2^{MN}$  次迭代。因此，算法1的复杂度显著低于穷举算法。

### 3 仿真分析

仿真实验基于边缘计算研究中广泛使用的公

共基准数据集进行验证。为避免用户与小基站分布过于稀疏或密集对结果的影响，本文进行了100次实验，并对结果取平均，以全面评估所提方案的有效性。实验所用数据集来源于澳大利亚墨尔本CBD，包含125个边缘服务器（基站），该数据集已被广泛应用于实验研究<sup>[10, 16]</sup>。在仿真环境中，用户本地计算能力  $f_m$  为1 GHz，小基站的计算能力为  $f_n$  为20 GHz。距离相关路径损耗的计算式为  $L[\text{dB}] = 140.7 + 36.7 \lg d[\text{km}]$ ，对数正态阴影的标准差  $\sigma$  为8 dB。子信道带宽  $W$  为10 MHz，传输功率为20 dBm，噪声功率  $\sigma^2$  为-100 dBm。本文从平均系统收益、用户分配率（即服务用户数与总用户数的比率）等方面来验证所提出DQoE-TORA方案的有效性，并与其他4种方案进行了对比分析。文献[11]和文献[25]以及其他基准方案具体说明如下。

(1) 贪婪距离（greedy distance, GD）方案。GD方案将任务卸载至距离最近的小基站。

(2) 贪婪信道增益的资源分配（greedy channel gain and resource allocation, GCG-RA）方案<sup>[25]</sup>。GCG-RA方案基于信道增益的贪婪分配策略，优先将任务分配至信道增益最高的子频段，直至任务或可用资源耗尽。这种方法能够显著地提升信道利用效率。

(3) 贪婪用户QoE（greedy user QoE, GU-QoE）方案。GU-QoE方案通过优先处理QoE收益值最大的任务来优化用户体验。

(4) 联合任务卸载和资源分配（joint task offloading and resource allocation, JTO-RA）方案<sup>[11]</sup>。JTO-RA方案针对多任务和多服务器的计算密集型任务卸载需求：首先，通过优化通信资源和功率分配获得初始解；随后，利用移除操作或交换操作进行优化；最后，迭代得到次优解。

平均系统收益与用户数量的关系如图2所示，当小基站数量固定为125时，随着用户数量的增



加, 平均系统收益逐步提升, 这主要是由于更多用户的任务得到了有效卸载, 从而提升了整体系统收益。然而, 与其他卸载方案相比, 所提出的DQoE-TORA方案始终实现了最高的系统收益, 这表明该方案在资源分配方面具有显著的优势。用户分配率与用户数量的关系如图3所示, 当小基站数量固定为125时, 随着用户数量的增加, 由于计算资源竞争的加剧, 用户分配率呈现逐渐下降的趋势。相比其他卸载方案, DQoE-TORA方案的下降速度较为缓慢。当用户数量超过600时, 因计算资源严重短缺, 各方案的用户分配率下降趋势明显加剧, 但DQoE-TORA方案仍表现出更好的性能。

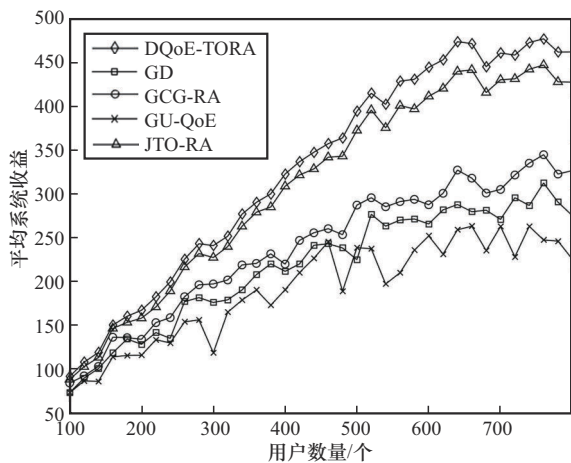


图2 平均系统收益与用户数量的关系

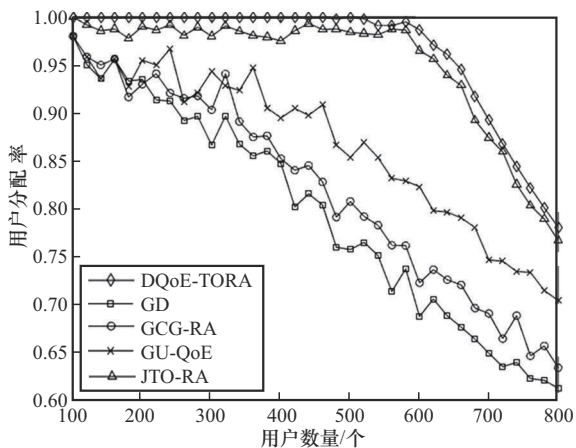


图3 用户分配率与用户数量的关系

平均系统收益与小基站数量的关系如图4所示, 当用户数量固定为500时, 平均系统收益随着小基站数量的增加而逐步提升, 这是因为更多的计算资源能够更好地满足用户任务的需求, 从而提高了整体系统收益。与其他卸载方案相比, 所提出的DQoE-TORA方案在所有小基站数量配置下均获得了最高的系统收益, 且随着小基站数量增加, 其相对优势愈加明显。用户分配率与小基站数量的关系如图5所示, 当用户数量固定为500时, 随着小基站数量的增加, 所有方案的用户分配率逐步上升。然而, 与其他方案相比, DQoE-TORA方案用户分配率上升趋势更显著, 这表明在资源充裕的情况下, DQoE-TORA方案能够更有效地利用小基站资源。

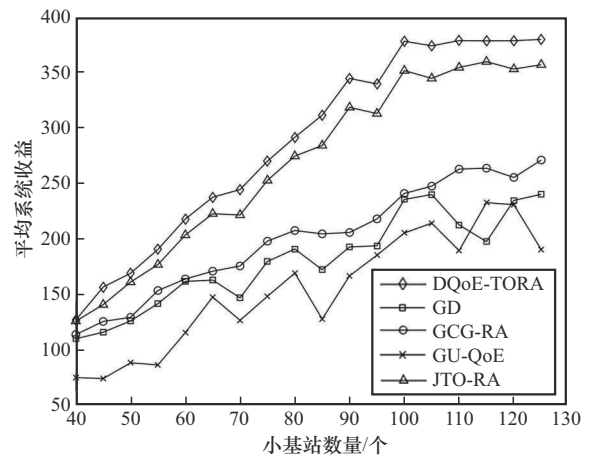


图4 平均系统收益与小基站数量的关系

平均系统收益与传输数据量的关系如图6所示, 当小基站数量固定为125、用户数量为500时, 任务数据量的增加导致传输时延增加, 从而削弱了卸载任务的时延相对改善收益。因此, 随着输入任务数据量的增加, 各方案的平均系统收益呈现下降趋势。然而, 与其他方案相比, DQoE-TORA方案依然能够获得更高的系统收益。

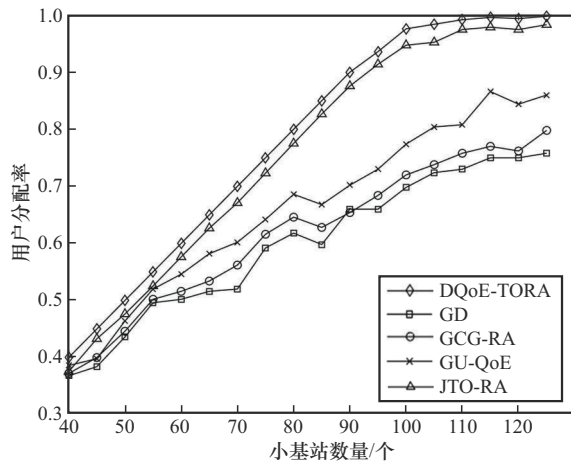


图 5 用户分配率与小基站数量的关系

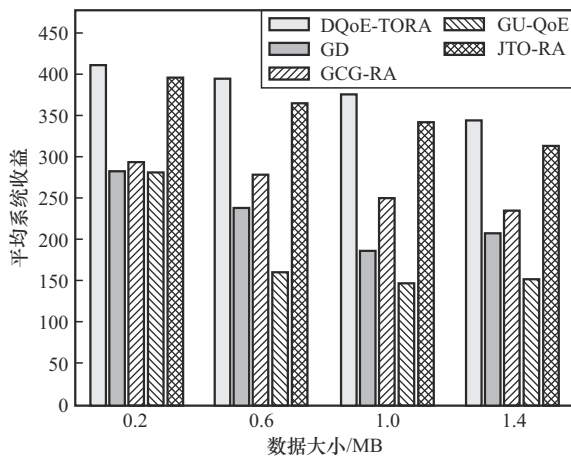


图 6 平均系统收益与传输数据量的关系

## 4 结束语

本文围绕基于 OFDMA 技术的 MEC 系统中的任务卸载与资源分配问题展开研究，旨在实现系统收益最大化。针对差异化用户 QoE 和资源分配不均衡对系统性能的影响，本文提出了一个联合卸载决策与计算资源分配的优化问题。由于该问题属于混合整数非线性问题，求解复杂度较高，本文将其分解为卸载决策和资源分配 2 个子问题，并结合双边匹配理论设计了一种高效的匹配卸载方案，获得近似最优解。仿真实验采用澳大利亚墨尔本 CBD 的公共数据集进行验证，结果表明，所提出的方案在系统收益方面相较于其他基准方

案具有显著优势。未来研究可扩展至在线用户分配场景，探索支持动态任务到 MEC 服务器的实时协同执行方法。

## 参考文献:

- [1] 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.
- [2] RAHMAN W U, HONG C S, HUH E N. Edge computing assisted joint quality adaptation for mobile video streaming[J]. IEEE Access, 2019(7): 129082-129094.
- [3] 周家恩, 赵亚飞, 彭木根. 面向低轨卫星多维资源优化的用户分配技术[J]. 电信科学, 2024, 40(8): 1-10.  
ZHOU J E, ZHAO Y F, PENG M G. User allocation technology for multi-dimensional resource optimization of LEO satellite[J]. Telecommunications Science, 2024, 40(8): 1-10.
- [4] LI Q Y, YAO H P, MAI T L, et al. Reinforcement-learning- and belief-learning-based double auction mechanism for edge computing resource allocation[J]. IEEE Internet of Things Journal, 2020, 7(7): 5976-5985.
- [5] QU B, BAI Y, CHU Y, et al. Resource allocation for MEC system with multi-users resource competition based on deep reinforcement learning approach[J]. Computer Networks, 2022 (215): 109181.
- [6] 邵鸿翔, 孙有铭, 蔡佳昊. 面向用户体验的多小区混合非正交多址接入网络资源分配方法[J]. 电子与信息学报, 2021, 43(4): 1129-1136.  
SHAO H X, SUN Y M, CAI J H. QoE-based resource allocation for multi-cell hybrid NOMA networks[J]. Journal of Electronics & Information Technology, 2021, 43(4): 1129-1136.
- [7] CHEN X F, ZHANG H G, WU C, et al. Optimized computation offloading performance in virtual edge computing systems via deep reinforcement learning[J]. IEEE Internet of Things Journal, 2019, 6(3): 4005-4018.
- [8] 张鸿, 廖彧歆, 王汝言, 等. 面向密集场景的空天地网络资源分配算法[J]. 电子与信息学报, 2024, 46(5): 1968-1976.  
ZHANG H, LIAO Y X, WANG R Y, et al. Resource allocation algorithm of space-air-ground integrated network for dense scenarios[J]. Journal of Electronics & Information Technology, 2024, 46(5): 1968-1976.
- [9] WU C R, PENG Q L, XIA Y N, et al. Online user allocation in mobile edge computing environments: a decentralized reactive approach[J]. Journal of Systems Architecture, 2021(113): 101904.
- [10] ZOU G B, LIU Y, QIN Z, et al. ST-EUA: spatio-temporal edge user allocation with task decomposition[J]. IEEE Transactions



- on Services Computing, 2023, 16(1): 628-641.
- [11] TRAN T X, POMPILI D. Joint task offloading and resource allocation for multi-server mobile-edge computing networks[J]. IEEE Transactions on Vehicular Technology, 2019, 68(1): 856-868.
- [12] HUANG J W, WANG M, WU Y, et al. Distributed offloading in overlapping areas of mobile-edge computing for Internet of things[J]. IEEE Internet of Things Journal, 2022, 9(15): 13837-13847.
- [13] CUI G M, HE Q, CHEN F F, et al. Trading off between multi-tenancy and interference: a service user allocation game[J]. IEEE Transactions on Services Computing, 2022, 15(4): 1980-1992.
- [14] GALE D, SHAPLEY L S. College admissions and the stability of marriage[J]. The American Mathematical Monthly, 1962, 69(1): 9-15.
- [15] CUI G M, HE Q, XIA X Y, et al. OL-EUA: online user allocation for NOMA-based mobile edge computing[J]. IEEE Transactions on Mobile Computing, 2023, 22(4): 2295-2306.
- [16] LAI P, HE Q, GRUNDY J, et al. Cost-effective app user allocation in an edge computing environment[J]. IEEE Transactions on Cloud Computing, 2022, 10(3): 1701-1713.
- [17] LAI P, HE Q, CUI G M, et al. QoE-aware user allocation in edge computing systems with dynamic QoS[J]. Future Generation Computer Systems, 2020, 112: 684-694.
- [18] SHENG Y, XU C, WU R Z. Task-oriented computation offloading and resource management in D2D-assisted heterogeneous networks[C]//Proceedings of the 2022 IEEE/CIC International Conference on Communications in China (ICCC). Piscataway: IEEE Press, 2022: 332-337.
- [19] ALRESHOODI M, WOODS J. Survey on Qoe\Qos correlation models formultimedia services[J]. International Journal of Distributed and Parallel Systems, 2013, 4(3): 53-72.
- [20] KUMAR S, GOSWAMI A, GUPTA R, et al. A cost-effective and QoS-aware user allocation approach for edge computing enabled IoT[J]. IEEE Internet of Things Journal, 2023, 10(2): 1696-1710.
- [21] LIU L Q, YUAN X M, CHEN D C, et al. Multi-user dynamic computation offloading and resource allocation in 5G MEC heterogeneous networks with static and dynamic subchannels[J]. IEEE Transactions on Vehicular Technology, 2023, 72(11): 14924-14938.
- [22] LYU X C, TIAN H, SENGUL C, et al. Multiuser joint task offloading and resource optimization in proximate clouds[J]. IEEE Transactions on Vehicular Technology, 2017, 66(4): 3435-3447.
- [23] BANDO K. Many-to-one matching markets with externalities among firms[J]. Journal of Mathematical Economics, 2012, 48(1): 14-20.

- [24] 绳韵, 许晨, 郑光远. 基于NOMA的超密集MEC网络任务卸载和资源分配方案[J]. 电信科学, 2022, 38(2): 35-46.
- SHENG Y, XU C, ZHENG G Y. Task offloading and resource allocation in NOMA-based ultra-dense MEC networks[J]. Telecommunications Science, 2022, 38(2): 35-46.
- [25] SARDELLITTI S, SCUTARI G, BARBAROSSA S. Joint optimization of radio and computational resources for multicell mobile-edge computing[J]. IEEE Transactions on Signal and Information Processing Over Networks, 2015, 1(2): 89-103.

## [作者简介]



丁雷 (2000-), 男, 北京邮电大学硕士生, 主要研究方向为移动边缘计算、端边协同。



许晨 (1988-), 女, 博士, 北京邮电大学人工智能学院副教授、博士生导师, 主要研究方向为边缘智能、协同通信与计算、智能网络优化。



刘斯宸 (2002-), 男, 北京邮电大学硕士生, 主要研究方向为移动边缘计算。



谢星寰 (2000-), 男, 北京邮电大学硕士生, 主要研究方向为车联网、无人机通信。



唐琳 (1987-), 男, 博士, 北京微芯区块链与边缘计算研究院副研究员、主任研究员, 主要研究方向为数据分析与优化、区块链等技术创新及在政务、金融、食品、数据交易等不同领域的行业解决方案应用。