



研究与开发

基于 NOMA 的超密集 MEC 网络任务卸载和资源分配方案

绳韵^{1,2}, 许晨², 郑光远³

(1. 华北电力大学电气与电子工程学院, 北京 102206;

2. 北京邮电大学人工智能学院, 北京 100876;

3. 华南理工大学电子与信息学院, 广东 广州 510641)

摘要: 为了提高移动边缘计算 (mobile edge computing, MEC) 网络的频谱效率, 满足大量用户的服务需求, 建立了基于非正交多址接入 (non-orthogonal multiple access, NOMA) 的超密集 MEC 系统模型。为了解决多个用户同时卸载带来的严重通信干扰等问题, 以高效利用边缘服务器资源, 提出了一种联合任务卸载和资源分配的优化方案, 在满足用户服务质量的前提下最小化系统总能耗。该方案联合考虑了卸载决策、功率控制、计算资源和子信道资源分配。仿真结果表明, 与其他卸载方案相比, 所提方案可以在满足用户服务质量的前提下有效降低系统能耗。

关键词: 移动边缘计算; 非正交多址接入; 任务卸载; 资源分配

中图分类号: TN929.5

文献标志码: A

doi: 10.11959/j.issn.1000-0801.2022033

Task offloading and resource allocation in NOMA-based ultra-dense MEC networks

SHENG Yun^{1,2}, XU Chen², ZHENG Guangyuan³

1. School of Electrical and Electronic Engineering, North China Electric Power University, Beijing 102206, China

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

3. School of Electronic and Information Engineering, South China University of Technology, Guangzhou 510641, China

Abstract: In order to improve spectrum efficiency and meet the service demands of massive users, a non-orthogonal multiple access (NOMA)-based ultra-dense mobile edge computing (MEC) system was considered. In order to solve the serious communication interference caused by simultaneous offloading of multiple users and make efficient use of edge server resources, a joint task offloading and resource allocation scheme was proposed to minimize the system energy consumption while meeting the quality of service (QoS) of all users. Offloading decision, power control, computation resource and subchannel resource allocation were jointly considered in the proposed scheme. Simulations results show that the proposed scheme can efficiently lower system energy consumption compared to the other offloading schemes.

Key words: MEC, NOMA, task offloading, resource allocation

收稿日期: 2021-10-12; 修回日期: 2022-02-09

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); 北京市科技新星计划项目 (No.Z191100001119124); 中央高校基本科研业务费专项资金资助项目 (No.2019MS009)

Foundation Items: The National Natural Science Foundation of China (No.62071179), The Beijing Nova Program of Science and Technology (No.Z191100001119124), The Fundamental Research Funds for the Central Universities (No.2019MS009)



0 引言

随着各种智能设备和网络服务的迅速发展,人脸识别、自动驾驶、虚拟现实等智能应用越来越普及。这些应用无一例外地需要庞大的计算资源对数据进行实时处理。然而,由于设备物理尺寸和电池容量受限,一般的智能设备无法提供强大的计算能力满足这些应用的需求。为了应对这一挑战,欧洲电信标准化协会于2014年提出了移动边缘计算(mobile edge computing, MEC)的概念,旨在无线网络边缘侧(如基站、无线接入点等)部署具有强大计算能力的服务器就近为用户提供所需的服务^[1]。借助 MEC,用户可以将自己的计算任务卸载到边缘服务器上执行,从而降低任务的计算时延、减少用户的能耗,并满足用户的服务质量(quality of service, QoS)要求^[2-3]。然而,随着移动终端数量的指数级增长,由于频谱资源有限,传统的蜂窝 MEC 网络无法满足大规模用户同时卸载的服务需求。

超密集网络(ultra-dense network, UDN)和非正交多址接入(non-orthogonal multiple access, NOMA)是未来无线网络中提高频谱利用率和系统吞吐量的两种新兴技术^[4-5]。UDN 的基本原理是在宏蜂窝内部署小基站,不同小小区通过复用宏小区的信道资源提升网络的频谱利用率^[6]。由于同层干扰和跨层干扰的存在,如何有效地进行干扰管理和资源分配在超密集网络中是非常重要的。子信道分配是超密集网络中进行干扰管理的重要手段。不同于传统的正交多址接入(orthogonal multiple access, OMA)技术, NOMA 技术通过功率域复用,在发送端使用叠加编码的方式使多个用户占用相同的频域/时域资源进行信号传输,然后在接收端采用串行干扰消除(successive interference cancellation, SIC)技术识别不同用户的信号^[7],以获得更高的频谱效率和更大的系统吞吐量。近年来,许多研究已经证明

了 NOMA 和超密集网络的结合在提升频谱效率和系统吞吐量方面的优势^[8-9]。

目前,已存在一些工作研究了 NOMA 和 MEC 技术的融合。文献[4]将 NOMA 技术应用在 MEC 网络的上行和下行链路,并证明了 NOMA 技术的引入可以进一步降低卸载过程的时延和能耗。文献[10]提出了一种联合计算卸载和资源分配的方案以实现任务执行时延的最小化。文献[11]提出了一种联合优化传输功率、传输时间分配和任务卸载分区的方案,达到最小化用户总能耗的目的。关于 MEC 和超密集网络的结合,文献[5]设计了计算卸载机制使设备的总能耗最小化。文献[12]综合考虑了任务执行时间和能量消耗,提出了一个任务虚拟机匹配算法。最近也有一些工作研究了基于超密集网络中的 NOMA-MEC 系统^[13-15]。文献[13]研究了超密集网络中 NOMA-MEC 系统的资源分配问题,以最小化系统总开销为目标,联合优化了用户分簇、发射功率和计算资源分配。文献[14]考虑了一种两层的异构网络,用户可以通过 NOMA 技术将计算任务卸载到宏基站或者小基站执行,并提出了一种高效的边缘卸载决策算法最小化计算成本。以最大化系统能效为目标,文献[15]研究了基于 NOMA 的 MEC 在超密集网络中的用户接入、传输功率和计算资源分配的联合优化问题。

从现有的研究工作中看,部分研究者对超密集网络中 NOMA-MEC 系统的任务卸载和资源分配联合优化进行了研究,然而忽略了子信道资源的高效分配问题。在超密集网络中,信道资源分配是一种降低干扰、提高频谱效率和能效的重要手段。在 NOMA-MEC 超密集网络中,多个用户同时进行任务卸载会产生严重的干扰,从而造成传输时延的增加和发射功率的浪费,影响用户的服务体验。而信道资源分配是一种降低干扰、提高频谱效率和能效的重要手段。此外,在基于 NOMA 的异构 MEC 网络中,卸载决策和用户间

的计算资源竞争会影响信道分配方案的设计, 这比传统的 NOMA 异构网络^[8-9]更具有挑战性。因此, 本文提出了一种联合任务卸载和多维度资源分配的优化方案, 综合考虑了任务卸载决策、用户功率控制、计算资源和子信道资源分配, 在满足用户 QoS 的前提下最小化系统能耗。本文的主要贡献如下。

(1) 构建了一个基于 NOMA 的超密集 MEC 网络, 其中每个小基站都部署一个 MEC 服务器并使用 NOMA 技术服务其相关用户。之后, 本文提出了一个联合任务卸载决策、功率控制、计算资源和子信道资源分配的优化问题, 在保证每个任务 QoS 的前提下最小化所有用户的能耗。

(2) 由于所提优化问题是一个混合整数非线性规划, 属于 NP 难问题, 无法在多项式时间内获得最优解, 因此本文将原混合问题解耦为多个子问题, 并提出一个基于迭代优化算法的解决方案, 获得原问题的近似最优解。

(3) 通过模拟仿真实验, 将提出的方案与不同的卸载方案和正交接入方案进行了对比。仿真结果表明, 与其他卸载方案相比, 本文所提方案可以实现更低的系统能耗, 并且能够以较低的复杂度获得接近穷举方案得到的结果。

1 系统模型与问题描述

基于 NOMA 的超密集 MEC 系统模型如图 1 所示, 该系统包含一个宏基站和多个小基站, 每个小基站配备一个 MEC 服务器为其小区内用户提供计算卸载服务。小小区采用 NOMA 技术支持不同用户占用同一信道进行任务卸载, 且不同小小区可以复用同一信道。假设该系统中有 M 个小基站和 K 个子信道, 分别用集合 $M = \{1, \dots, m, \dots, M\}$ 和 $K = \{1, \dots, k, \dots, K\}$ 表示。每个小基站服务 U 个用户, 用集合 $U_m = \{1, \dots, u, \dots, U\}$ 表示小小区 m 中的用户。为了避免信道复用产生严重干扰导致用户的通信质量降低, 假设占用同

一子信道的小小区的最大数目为 $q_{\max}$ 。定义小基站 m 占用子信道 k 的指示变量为 $x_m^k = \{0, 1\}$, $x_m^k = 1$ 时表示小基站 m 占用子信道 k ; 否则 $x_m^k = 0$ 。

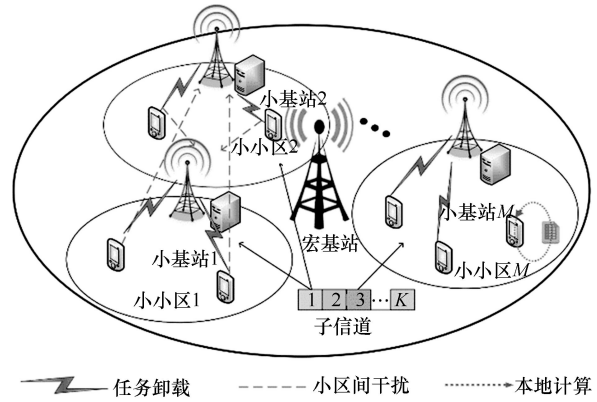


图1 基于 NOMA 的超密集 MEC 系统模型

假设每个用户都有一个时延受限的任务, 用 $T_{u,m} = (L_{u,m}, \rho_{u,m}, D_{u,m}^{\max})$ 表示小小区 m 中用户 u 的任务, 其中 $L_{u,m}$ 为输入数据的大小, $\rho_{u,m}$ 为完成单位任务量所需的 CPU 周期数, $D_{u,m}^{\max}$ 为最大容忍时延 (以 s 为单位), 即任务的 QoS 要求。相应地, 完成任务所需要的 CPU 周期数可以用 $C_{u,m} = L_{u,m} \rho_{u,m}$ 表示。本文考虑一种二元卸载模式, 即每个用户的任务可以在本地执行或者卸载到相应的小基站上执行。定义用户卸载决策参数为 $a_u^m \in \{0, 1\}$, $a_u^m = 1$ 表示用户 u 将其整个任务卸载到小基站 m 上执行; $a_u^m = 0$ 表示小小区 m 中用户 u 在本地执行其任务。

1.1 通信模型

当不同用户通过 NOMA 技术将其任务卸载到同一个小基站时, 用户之间存在小区内干扰。可以在小基站上采用 SIC 技术解码同一子信道上不同用户发送的信号。根据上行传输场景中的 SIC 方案, 小基站首先解码具有较高信道增益的用户信号, 并把其他用户的信号视作干扰, 然后将解码信号从其他用户的干扰项中去除^[11]。在子信道 k 上, 对于小基站 m 服务的用户集 U_m , 假设其信道增益遵循 $g_{1,m}^k > g_{2,m}^k \dots > g_{U,m}^k$ 的顺序。那么,



小小区 m 中第 u 个解码用户在子信道 k 上的信干噪比 (signal to interference plus noise ratio, SINR) 表示为:

$$\gamma_{u,m}^k = \frac{P_{u,m}^k g_{u,m}^k}{I_{u,in}^{k,m} + I_{co}^{k,m} + \sigma^2} \quad (1)$$

其中, $P_{u,m}^k$ 是小小区 m 中第 u 个用户在子信道 k 上的发射功率。 $g_{u,m}^k$ 是用户 u 和小基站 m 在子信道 k 上的信道增益。 $I_{u,in}^{k,m} = \sum_{i=u+1}^U a_i^m P_{i,m}^k g_{i,m}^k$ 表示小小区内的干扰, $I_{co}^{k,m} = \sum_{s \in M \setminus \{m\}} (x_s^k \sum_{n \in U_s} a_n^s P_{n,s}^k g_{n,m}^k)$ 表示小区间干扰。 σ^2 表示加性高斯白噪声的功率。

根据香农公式, 小小区 m 中用户 u 在子信道 k 上的上行链路数据速率表示为:

$$R_{u,m}^k = B \ln(1 + \gamma_{u,m}^k) \quad (2)$$

其中, B 表示每个子信道的固定带宽。

1.2 计算模型

(1) 本地计算

用 $f_{u,m}^l$ 表示小小区 m 中用户 u 的计算能力 (即每秒 CPU 周期数)。因此, 任务 $T_{u,m}$ 的本地执行时延表示为:

$$T_{u,m}^{\text{loc}} = \frac{C_{u,m}}{f_{u,m}^l} \quad (3)$$

相应能量消耗为:

$$E_{u,m}^{\text{loc}} = \kappa (f_{u,m}^l)^2 C_{u,m} \quad (4)$$

其中, κ 为能耗参数^[14]。

(2) 边缘计算

用户将计算任务上传到相应的基站上, 由基站配备的 MEC 服务器执行该任务。任务上传过程中, 小小区 m 中用户 u 在子信道 k 上的传输时间表示为:

$$T_{u,m}^{k,\text{up}} = \frac{L_{u,m}}{R_{u,m}^k} \quad (5)$$

上传过程中的能耗为:

$$E_{u,m}^{k,\text{up}} = P_{u,m}^k T_{u,m}^{k,\text{up}} \quad (6)$$

用 F_m 表示小基站 m 的计算能力 (即每秒 CPU

周期数)。任务 $T_{u,m}$ 在 MEC 服务器上的执行时间表示为:

$$T_{u,m}^{\text{exe}} = \frac{C_{u,m}}{F_m^e} \quad (7)$$

其中, F_m^e 表示小基站 m 分配给用户 u 的计算资源。这里忽略了在 MEC 服务器上执行任务的能耗, 因为小基站通常由电网供电, 执行任务所消耗的能量是足够的^[15]。此外, 由于计算结果较小, 忽略了从 MEC 服务器向用户发送结果的时间和能耗^[1]。因此, 用户 u 的任务在小基站 m 上执行的总时延为:

$$T_{u,m}^{\text{off}} = \sum_{k \in K} x_m^k T_{u,m}^{k,\text{up}} + T_{u,m}^{\text{exe}} \quad (8)$$

能耗为:

$$E_{u,m}^{\text{off}} = \sum_{k \in K} x_m^k E_{u,m}^{k,\text{up}} \quad (9)$$

综上所述, 小小区 m 中用户 u 完成任务的能耗表示为:

$$E_{u,m} = a_u^m E_{u,m}^{\text{off}} + (1 - a_u^m) E_{u,m}^{\text{loc}} \quad (10)$$

1.3 问题描述

本文以最小化用户总能耗为目标, 联合考虑任务卸载决策、功率控制、计算资源和子信道资源分配, 同时保证每个任务的服务质量要求。优化问题描述为:

$$\begin{aligned} \min_{A, X, F, P} \quad & \sum_{m \in M} \sum_{u \in U_m} E_{u,m} \\ \text{s.t.} \quad & C_1: a_u^m T_{u,m}^{\text{off}} + (1 - a_u^m) T_{u,m}^{\text{loc}} \leq D_{u,m}^{\text{max}}, \forall m \in M, u \in U_m \\ & C_2: \sum_{u \in U_m} a_u^m F_{u,m}^e \leq F_m, \forall m \in M \\ & C_3: F_{u,m}^e \geq 0, \forall m \in M, u \in U_m \\ & C_4: \sum_{k \in K} x_m^k \leq 1, \forall m \in M \\ & C_5: \sum_{m \in M} x_m^k \leq q_{\text{max}}, \forall k \in K \\ & C_6: x_m^k \in \{0, 1\}, \forall m \in M, k \in K \\ & C_7: a_u^m \in \{0, 1\}, \forall m \in M, u \in U_m \\ & C_8: 0 \leq a_u^m P_{u,m}^k \leq P_{\text{max}}, \forall k \in K, m \in M, u \in U_m \end{aligned} \quad (11)$$

其中, $A = \{a_u^m, m \in M, u \in U_m\}$, $X = \{x_m^k, m \in M,$

$k \in K\}$, $\mathbf{F} = \{F_{u,m}^e, m \in M, u \in U_m\}$, $\mathbf{P} = \{P_{u,m}^k, k \in K, m \in M, u \in U_m\}$ 。约束条件 C_1 保证了每个用户的任务完成时间小于其最大容忍时延。 C_2 和 C_3 表示分配给卸载用户的计算资源不超过每个 MEC 服务器的总资源。 C_4 和 C_5 保证了每个小基站最多复用一个小信道, 而每个子信道最多被 $q_{\max}$ 个小基站复用。 C_6 和 C_7 表示卸载决策和子信道分配是二元变量。 C_8 是所有卸载用户的功率约束。

式 (11) 是一个混合整数非线性规划问题, 它涉及二元变量和连续变量, 且目标函数是非凸的。此外, 卸载决策与资源分配的耦合使得问题更具挑战性。一般情况下无法在多项式时间复杂度内获得该问题的最优解。因此, 本文使用块坐标下降法将原问题解耦为资源分配子问题和卸载决策子问题^[16], 通过迭代求解这两个子问题直到目标值收敛, 即可获得原问题的近似解。

2 联合任务卸载和资源分配的优化方案

本节提出了一种联合任务卸载和资源分配的方案, 通过迭代求解资源分配和卸载决策两个子问题, 得到所提问题的联合解。

2.1 资源分配

本节主要解决给定卸载决策下的资源分配问题, 该问题包含功率控制、计算资源和子信道资源的联合优化。用 $U_m^e = \{u \mid a_u^m = 1, \forall u \in U_m\}$ 表示小小区 m 中选择任务卸载的用户集合, 则优化问题的表示如下:

$$\begin{aligned} \min_{\mathbf{X}, \mathbf{F}, \mathbf{P}} & \sum_{m \in M} \sum_{u \in U_m^e} E_{u,m}^{\text{off}} \\ \text{s.t. } C_1: & R_{u,m} \geq \frac{L_{u,m}}{D_{u,m}^{\max} - T_{u,m}^{\text{exe}}}, \forall m \in M, u \in U_m^e \\ C_2: & \sum_{u \in U_m^e} x F_{u,m}^e \leq F_m, \forall m \in M \\ C_3: & F_{u,m}^e > 0, \forall m \in M, u \in U_m^e \\ C_4: & \sum_{k \in K} x_m^k \leq 1, \forall m \in M \end{aligned}$$

$$\begin{aligned} C_5: & \sum_{k \in K} x_m^k \leq q_{\max}, \forall k \in K \\ C_6: & x_m^k \in \{0, 1\}, \forall m \in M, k \in K \\ C_7: & 0 \leq P_{u,m}^k \leq P_{\max}, \forall k \in K, m \in M, u \in U_m^e \end{aligned} \quad (12)$$

式 (12) 也是一个涉及二元变量和连续变量的混合整数非线性规划问题, 且目标函数是非凸的, 不同优化变量之间具有强耦合性。为了解决该问题, 将其解耦为两个子问题, 即计算资源分配问题, 以及联合子信道分配和功率控制问题, 通过解决这两个子问题获得式 (12) 的解。

2.1.1 计算资源分配

由式 (5) ~ 式 (9) 可知, 分配给用户的计算资源越多, 任务执行过程中花费的时间越少, 那么在总时延约束不变的条件下, 留给任务上传的时间越长, 则用户的上传能耗越小^[17]。因此, 从计算资源分配的角度出发, 能耗优化问题可以表述为最小化每个小基站服务的卸载用户的任务执行时延, 即:

$$\begin{aligned} \min_{\mathbf{F}_m} & \sum_{u \in U_m^e} \frac{C_{u,m}}{F_{u,m}^e} \\ \text{s.t. } C_1: & \frac{C_{u,m}}{F_{u,m}^e} < D_{u,m}^{\max} \\ C_2: & \sum_{u \in U_m^e} F_{u,m}^e = F_m \\ C_3: & F_{u,m}^e > 0, \forall u \in U_m^e \end{aligned} \quad (13)$$

其中, $\mathbf{F}_m = \{F_{u,m}^e, u \in U_m^e\}$ 。由于 C_1 是反比例函数, 它在 $0 \sim +\infty$ 是凸函数, C_2 是仿射函数, C_3 是直线也为凸函数, 因此式 (13) 的约束条件是一组线性凸约束。且目标函数也为凸函数, 该问题为标准的凸优化问题, 可以用拉格朗日乘数法求解^[18], 其计算式为:

$$\begin{aligned} L(\mathbf{F}_m, \lambda, \mathbf{v}_m) = & \sum_{u \in U_m^e} \frac{C_{u,m}}{F_{u,m}^e} + \\ & \lambda \left(\sum_{u \in U_m^e} F_{u,m}^e - F_m \right) - \sum_{u \in U_m^e} v_{u,m} F_{u,m}^e \end{aligned} \quad (14)$$

其中, λ 和 $\mathbf{v}_m = \{v_{u,m}, u \in U_m^e\}$ 是拉格朗日乘数。然后, 利用 KKT (Karush-Kuhn-Tucker) 条件找到



每个用户的最优计算资源分配结果, 表示如下:

$$F_{u,m}^{e,*} = \frac{F_m \sqrt{C_{u,m}}}{\sum_{s \in U_m^e} \sqrt{C_{s,m}}}, \forall u \in U_m^e \quad (15)$$

相应地, 可以得到每个任务的最小上传速率要求为:

$$R_{u,m}^{\min} = \frac{L_{u,m}}{D_{u,m}^{\max} - \frac{C_{u,m}}{F_{u,m}^{e,*}}}, \forall m \in M, u \in U_m^e \quad (16)$$

2.1.2 子信道分配

给定计算资源分配条件下, 原问题 (12) 可以转化为如下通信资源分配问题:

$$\begin{aligned} \min_{X,P} \quad & \sum_{m \in M} \sum_{u \in U_m^e} E_{u,m}^{\text{off}} \\ \text{s.t.} \quad & C_1: R_{u,m}^k \geq R_{u,m}^{\min}, \forall k \in K, m \in M, u \in U_m^e \\ & C_2: \sum_{k \in K} x_m^k \leq 1, \forall m \in M \\ & C_3: \sum_{k \in K} x_m^k \leq q_{\max}, \forall k \in K \\ & C_4: x_m^k \in \{0, 1\}, \forall m \in M, k \in K \\ & C_5: 0 \leq P_{u,m}^k \leq P_{\max}, \forall k \in K, m \in M, u \in U_m^e \end{aligned} \quad (17)$$

其中, 约束条件 C_1 是每个任务的最小上传速率要求。式 (17) 是一个涉及二元变量和连续变量的混合问题, 目标函数是非凸的。此外, 子信道分配和功率控制共同影响用户的能耗。为了获得问题的全局最优解, 需要通过穷举法找出能够实现能耗最小的子信道分配结果, 然而随着用户数和子信道数的增加, 这种方法的复杂度呈指数级上升。为了降低复杂度, 将该问题建模为一个小基站之间存在外部效应的多对一匹配问题^[19]。同时, 在每次临时匹配的过程中进行功率控制。基于匹配理论的一些概念^[20], 有以下定义。

定义 1 多对一匹配函数 Φ 是从集合 $M \cup K$ 到无序元素集合 $M \cup K \cup \{0\}$ 的一个函数, 满足:

(1) $|\Phi(m)| = 1, \forall m \in M$; (2) $|\Phi(k)| \leq q_{\max}, \forall k \in K$; (3) $\Phi(m) = k$ 当且仅当 $m \in \Phi(k)$ 。

其中, $|\Phi(\bullet)|$ 表示结果 $\Phi(\bullet)$ 的基数。条件 (1) 表示每个小基站只能和一个子信道匹配; 条件 (2)

表示每个子信道最多匹配 $q_{\max}$ 个小基站; 条件 (3) 表示如果小基站 m 匹配到子信道 k , 那么子信道 k 的匹配对象也包含小基站 m 。

对于每一个小基站, 其效用函数可以定义为服务的所有卸载用户能耗总和的倒数:

$$U_m(k) = \frac{1}{\sum_{u \in U_m^e} E_{u,m}^{k,\text{up}}} \quad (18)$$

而对于任意子信道 k , 效用函数可以定义为占用该信道的所有卸载用户总能耗的倒数:

$$U_k(\Phi(k)) = \frac{1}{\sum_{i \in \Phi(k)} \sum_{u \in U_i^e} E_{u,i}^{k,\text{up}}} \quad (19)$$

为了更好地描述一方对另一方的喜好程度, 引入小基站和子信道对彼此的偏好顺序, 记为 $\succ$ 。任一小基站 m 对所有子信道都存在一个偏好顺序, 记为 $\succ_m$ 。具体地, 给定两个子信道 k 和 k' 满足 $k, k' \in K$ 且 $k \neq k'$, 任意两个匹配 $\Phi(m) = k$ 和 $\Phi'(m) = k'$ 满足:

$$(k, \Phi) \succ_m (k', \Phi') \Leftrightarrow U_m(k) > U_m(k') \quad (20)$$

式 (20) 表示, 小基站 m 更倾向于匹配 Φ 中的子信道 k 而不是 Φ' 中的 k' , 当且仅当小基站 m 复用子信道 k 比子信道 k' 可实现更低的能耗。类似地, 任意子信道 k 对所有小基站也存在一个偏好顺序, 记为 $\succ_k$ 。具体地, 对于任意两个小基站集合 S 和 S' 满足 $S, S' \subseteq M$ 和任意两个匹配 $\Phi, \Phi', S = \Phi(k), S' = \Phi'(k)$, 有:

$$(S, \Phi) \succ_k (S', \Phi') \Leftrightarrow U_k(S) > U_k(S') \quad (21)$$

式 (21) 表示, 与 S' 相比, 子信道 k 更倾向于选择 S , 当且仅当子信道 k 被集合 S 中的小基站复用时总能耗更低。

由于每个小基站只能占用一个子信道, 而每个子信道最多能与 $q_{\max}$ 个小基站匹配, 因此子信道分配问题是一个多对一匹配。由于存在小区间干扰 (如式 (1) 所示), 每个小基站的偏好列表不仅取决于与之匹配的子信道, 还取决于匹配同一子信道的其他小基站。因此, 这是一个小基站

之间存在外部效应的多对一匹配问题^[19-20]。与偏好值由参与者自身决定的传统匹配相比, 获得该问题的稳定匹配结果将更具挑战性^[20]。受多对一住房分配问题的启发^[21], 本文采用交换匹配的方案获得信道资源分配的结果。

定义 2 交换匹配定义为 $\Phi_m^{m'} = \{\Phi \setminus \{(m, k), (m', k')\} \cup \{(m, k'), (m', k)\}\}$, 其中 $m' \in \Phi_m^{m'}(k), m \in \Phi_m^{m'}(k'), m \notin \Phi_m^{m'}(k), m' \notin \Phi_m^{m'}(k')$ 。

两个小基站交换它们匹配的子信道, 同时保持其他小基站匹配不变。交换阻塞对定义如下。

定义 3 给定匹配函数 Φ 和一对小基站 (m, m') , 如果存在小基站 $m \in \Phi(k)$ 和小基站 $m' \in \Phi(k')$ 满足:

- (1) $\forall t \in \{m, m', k, k'\}, U_t(\Phi_m^{m'}(t)) \geq U_t(\Phi(t))$;
- (2) $\exists t \in \{m, m', k, k'\}, U_t(\Phi_m^{m'}(t)) > U_t(\Phi(t))$

那么, 交换匹配 $\Phi_m^{m'}$ 将被批准, 且小基站对 (m, m') 是匹配 Φ 中的一个交换阻塞对。

上述定义表明, 如果一个交换匹配批准通过, 那么所有参与此次交换匹配的小基站和子信道的效用值不会降低, 且至少有一个参与者的效用值会提高。

定义 4 一个匹配函数 Φ 双边交换稳定的, 当且仅当不存在交换阻塞对时。

基于上述定义, 本文提出了一种基于交换匹配的子信道分配和功率控制算法。首先, 将每个小基站随机匹配一个子信道, 得到一个初始匹配结果。在交换匹配过程中, 随机选择两个小基站, 并临时交换它们匹配的子信道, 其他小基站的匹配保持不变。然后对两个子信道上的用户执行功率控制, 得到基站和子信道的效用值。这两个小基站形成交换阻塞对, 则执行交换操作。交换匹配过程持续进行, 直到不存在交换阻塞对, 最终获得双边交换稳定的匹配结果。

2.1.3 功率控制

在给定子信道分配的情况下, 可以独立地解决同一子信道上卸载用户的功率控制问题。文献[22]

证明了卸载用户的能量消耗相对于传输功率是单调增加的。此外, 在计算资源分配后, 传输时间受到限制。因此, 功率控制问题的目标为最小化占用相同子信道的用户发射功率之和。子信道 k 上卸载用户的功率优化问题可以表示为:

$$\begin{aligned} \min_{\mathbf{P}^k} & \sum_{m \in M} \sum_{u \in U_m^e} x_m^k P_{u,m}^k \\ \text{s.t. } & C_1: R_{u,m}^k \geq R_{u,m}^{\min} \\ & C_2: 0 < P_{u,m}^k \leq P_{\max}, \forall m \in M, u \in U_m^e \end{aligned} \quad (22)$$

其中, $\mathbf{P}^k = \{P_{u,m}^k, m \in M, u \in U_m^e\}$ 是占用子信道 k 的卸载用户的功率集。

由于约束条件 C_1 的非凸性, 式 (22) 不是凸优化问题, 因此, 首先通过收紧非凸约束将原问题转化为凸优化问题, 然后基于逐次凸逼近 (successive convex approximation, SCA) 算法获得原问题的解^[23]。

利用以下不等式收紧约束条件 C_1 :

$$\text{lb}(1 + \gamma) \geq \alpha \text{lb}(\gamma) + \beta \quad (23)$$

其中, $\alpha = \tilde{\gamma} / (1 + \tilde{\gamma})$ 和 $\beta = \text{lb}(1 + \tilde{\gamma}) - \tilde{\gamma} \text{lb} \tilde{\gamma} / (1 + \tilde{\gamma})$ 。当 $\gamma = \tilde{\gamma}$ 时, 式 (23) 的等号成立。

根据不等式 (23), 可以得到用户 u 的数据速率下限:

$$R_{u,m}^k \geq \tilde{R}_{u,m}^k = B\alpha_{u,m}^k \text{lb}(\gamma_{u,m}^k) + B\beta_{u,m}^k \quad (24)$$

因此, 式 (22) 中约束 C_1 可以改写为:

$$\tilde{R}_{u,m}^k \geq R_{u,m}^{\min}, \forall m \in M, u \in U_m^e \quad (25)$$

令 $\mathbf{P}^k = \mathbf{e}^{S^k}$, 式 (22) 可以重新表述为以下近似问题:

$$\begin{aligned} \min_{S^k} & \sum_{m \in M} \sum_{u \in U_m^e} x_m^k e^{S_{u,m}^k} \\ \text{s.t. } & C_1: \tilde{R}_{u,m}^k \geq R_{u,m}^{\min}, \forall m \in M, u \in U_m^e \\ & C_2: e^{S_{u,m}^k} \leq P_{\max}, \forall m \in M, u \in U_m^e \end{aligned} \quad (26)$$

该问题是凸优化问题, 可利用凸优化工具直接求解^[18]。因此, 可以利用 SCA 算法迭代解决式 (26) 获得原式 (22) 的解^[23]。



2.1.4 资源分配算法

通过求解上述两个子问题(式(17)和式(22)),得到最终的任务卸载用户的资源分配算法,见算法1。

算法 1 任务卸载用户的资源分配算法

- 1) 计算资源分配
- 2) for $m \in M$
- 3) for $u \in U_m^e$
- 4) 通过式(15)得到计算资源分配结果
- 5) 通过式(16)计算最小数据速率要求
- 6) end for
- 7) end for
- 8) 联合子信道分配和功率控制
- 9) 初始化匹配 Φ_0 : 小基站和子信道随机匹配并满足 $|\Phi(k)| \leq q_{\max}$ 和 $|\Phi(m)| \leq 1$
- 10) 交换匹配过程
- 11) Repeat
- 12) for $m \in M$
- 13) for $m' \in M \setminus \{m\}$ do
- 14) 通过 SCA 算法^[19]解决式(22)计算匹配 $\Phi_m^{m'}$ 下所有用户和子信道的效用值
- 15) if (m, m') 是一个阻塞对, 那么
- 16) $\Phi \leftarrow \Phi_m^{m'}$
- 17) end if
- 18) end for
- 19) end for
- 20) until 没有交换阻塞对存在

引理 1 每次交换操作后, 系统能耗降低。

证明 假设小基站 m 和 m' 以及它们匹配的子信道 k 和 k' 之间的交换操作使匹配 Φ 变为 $\Phi_m^{m'}$ 。根据定义 3, 交换操作后, k 和 k' 的效用满足 $U_k(\Phi_m^{m'}(k)) \geq U_k(\Phi(k))$ 和 $U_{k'}(\Phi_m^{m'}(k')) \geq U_{k'}(\Phi(k'))$, 其中至少有一个等式不成立。由于其他子信道的效用保持不变, 因此有不等式 $U_{\text{total}}(\Phi_m^{m'}) > U_{\text{total}}(\Phi)$ 和 $E_{\text{total}}(\Phi_m^{m'}) < E_{\text{total}}(\Phi)$ 。其中, $U_{\text{total}}(\Phi) = \sum_{k=1}^K U_k(\Phi(k))$ 是匹配 Φ 下所有子

信道的总效用, $E_{\text{total}}(\Phi)$ 表示匹配 Φ 下所有卸载用户的能耗。由此可得, 每次交换操作后, 系统能耗都会降低。

定理 1 算法 1 在有限次交换匹配过程中收敛。

证明 因为用户数量有限, 所以计算资源分配过程的循环次数也有限。算法 1 的收敛性主要取决于交换匹配过程。根据引理 1, 每次交换操作后, 系统的总能耗都会降低。此外, 由于有限的通信和计算资源, 系统能耗是有界的。因此, 算法 1 在有限次迭代后必然会收敛。

算法 1 的复杂度主要取决于联合子信道分配和功率控制过程。在子信道分配过程中, 每一次迭代最多需要 M^2 次交换次数^[19]。每次交换过程需要执行功率控制算法获得临时匹配结果的效用值。假设利用 SCA 算法解决式(26)的复杂度为 $\mathcal{O}(X)$, 那么, 算法 1 的整体复杂度约为 $\mathcal{O}(2NXM^2)$, 其中 N 为交换匹配所需的迭代次数。

2.2 卸载决策优化

通过第 2.1 节, 可以得到用户在卸载过程中的能耗。由于穷举法复杂度极高, 需要迭代 2^{UM} 次得到最优卸载决策方案, 在用户规模较大的场景中难以实现。因此, 本文提出一种低复杂度的卸载决策方案, 该方案通过比较每个用户在本地计算和边缘计算过程中的能耗确定其卸载决策, 表示如下:

$$a_u^m = \begin{cases} 1, & E_{u,m}^{\text{loc}} > E_{u,m}^{\text{off}} \\ 0, & E_{u,m}^{\text{loc}} \leq E_{u,m}^{\text{off}} \end{cases} \quad (27)$$

2.3 联合优化方案

本节提出一个联合任务卸载和资源分配算法, 见算法 2。

算法 2 联合任务卸载和资源分配算法

输入 初始化 $a_u^m = 1, \forall m \in M, u \in U_m^e$

输出 A^*, X^*, F^*, P^*

1) Repeat

2) 通过表 1 计算任务卸载用户的能耗

3) 通过式 (27) 更新用户的卸载决策

4) until 没有用户改变自身卸载决策

假设算法 2 需要 T 次迭代收敛, 则算法的整体复杂度约为 $\mathcal{O}(T(2NXM^2 + UM))$ 。而穷举算法需要 $K^M 2^{UM}$ 次迭代获得最优的子信道分配和卸载决策方案; 此外, 每次迭代过程还需要进行计算资源优化和功率控制。因此, 算法 2 的复杂度远低于穷举算法。

3 仿真分析

本节通过仿真验证所提联合优化算法的性能, 仿真参数见表 1。假设一个半径为 1 000 m 的宏小区, 小小区随机分布在该区域中。小基站的数量为 4~8 个, 子信道数量为 4 个, 并假设每个基站的计算能力可以满足关联用户的计算需求。为了限制同信道干扰, 降低硬件复杂度和处理时延, 假设每个 NOMA 簇包含 2 个用户, 每个小基站分配一个子信道^[24]。详细仿真参数设置见表 1, 这些参数是根据现有研究^[5,25]设置的。在仿真中, 将所提出的基于 NOMA 的任务卸载和资源分配方案 (NOMA-task offloading and resource allocation, N-TORA) 与其他 4 种方案的能耗性能进行了比较。为了便于理解, 对文献[3, 26-27]中的对比方案和其他基准方案进行了说明, 如下所示:

(1) 全部本地执行方案 (entire local processing, ELP): 每个用户的任务都在本地执行, 即 ELP 方案。

(2) 全部 MEC 执行方案^[3] (entire MEC processing, EMP): 结合文献[3]中的 EMP 方案, 即所有用户的任务都卸载到关联基站上执行, 通过算法 1 获得发射功率、计算资源和子信道分配的结果。

(3) 基于 OMA 的任务卸载和资源分配方案^[26] (OMA-task offloading and resource allocation, O-TORA): 为了研究子信道分配对系统性能的影响, 将所提 N-TORA 方案与文献[26]中的方案进行了对比。文献[26]中用户通过正交频分多址接入

方式将任务卸载到关联基站上, 即每个小小区的带宽平均分配给关联用户。

(4) 联合功率控制和子信道分配的任务卸载和资源分配方案^[27] (joint power allocation and sub-channel assignment-task offloading and resource allocation, JPS-TORA): 由于复杂度较高, 在仿真中很难通过穷举算法找到最优的任务卸载和子信道分配方案, 因此, 采用文献[27]中的联合功率控制和子信道分配算法获得最优的子信道, 并结合所提出的任务卸载和资源分配方案以实现全局最优解。

表 1 仿真参数

仿真参数	数值
用户最大传输功率/dBm	23
带宽/MHz	1
噪声功率/dBm	-80
子信道数量	4
小基站数量	4-8
小基站计算能力/(吉周期/s)	2.5
用户最大计算能力/(吉周期/s)	[0.1, 1]
单位比特任务所需 CPU 周期数/(周期/bit)	[100, 200]
输入数据大小/kbit	[100, 200]
任务最大容忍时延/ms	100

图 2 为总能耗与算法 2 的迭代次数的关系。从图 2 可以看出, 随着小基站数量的增加, 总能耗也逐渐增加。此外, 随着迭代次数的增加, 系统能耗会逐渐降低并最终收敛于一个定值。图 2 中收敛所需的迭代次数为 2 次, 说明所提 N-TORA 方案的收敛速度是非常快的。

图 3 为不同算法下总能耗随小基站数量的变化关系曲线 ($K=4$)。可以看出, 系统能耗随着小基站数量的增加而增大。相较于其他卸载方案, 所提 N-TORA 方案可以实现更低的能耗, 并且接近最优的 JPS-TORA 方案的结果。另外, 在节省用户能耗方面, 基于 NOMA 的 MEC 网络性能优于基于 OMA 的网络。

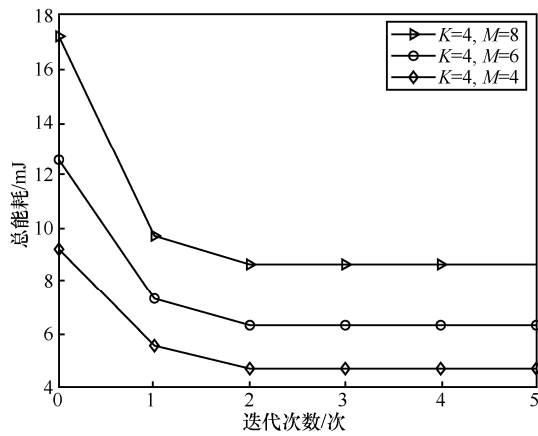


图2 总能耗与迭代次数关系

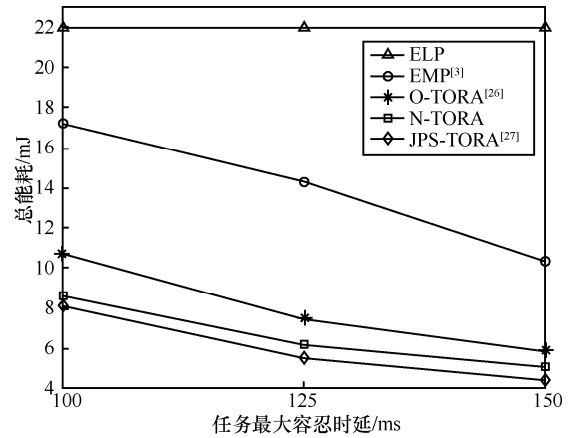


图4 总能耗与任务最大容忍时延关系

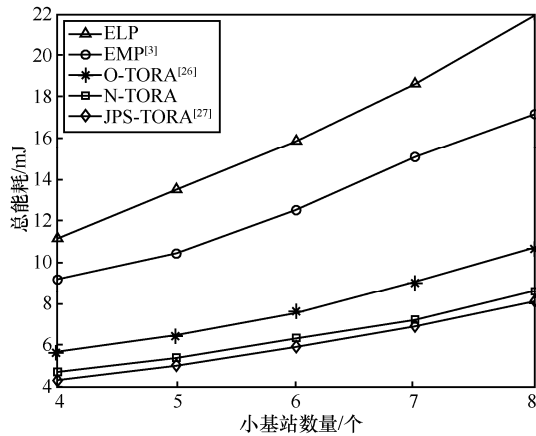


图3 总能耗与小基站数量关系

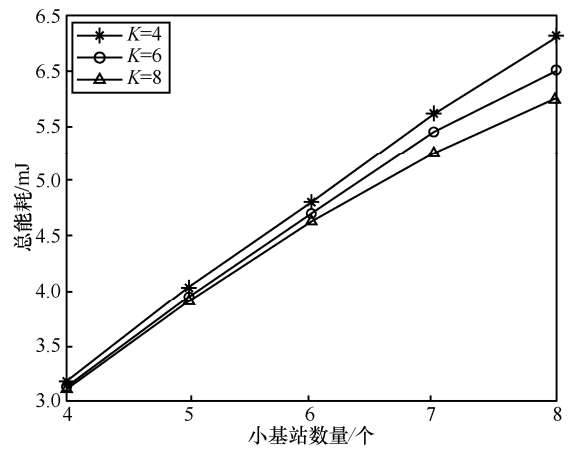


图5 总能耗与小基站数量关系

图4为不同算法下所有用户的总能耗与任务最大容忍时延的关系。由图4可知,随着任务时延要求放宽,能耗逐渐降低。这是因为随着任务卸载传输时间延长,发射功率要求降低,从而减小了任务上传过程中的能耗。此外,所提N-TORA方案在节约能耗方面具有明显优势。

图5为子信道数量不同时,总能耗与小基站数量的关系。由图5可知,随着小基站数量的增加,总能耗是逐渐增加的。另外,子信道数量减少会导致系统总能耗增大,这是因为当子信道较少时,占用同一子信道的小小区数量增加,导致严重的小区间干扰,进而增加了卸载过程的能耗。

图6为总能耗与任务输入数据大小的关系。可以看出,总能耗随着输入数据量的增加而增加。此外,随着任务计算密度 $\rho_{u,m}$ 的增大,总能耗增

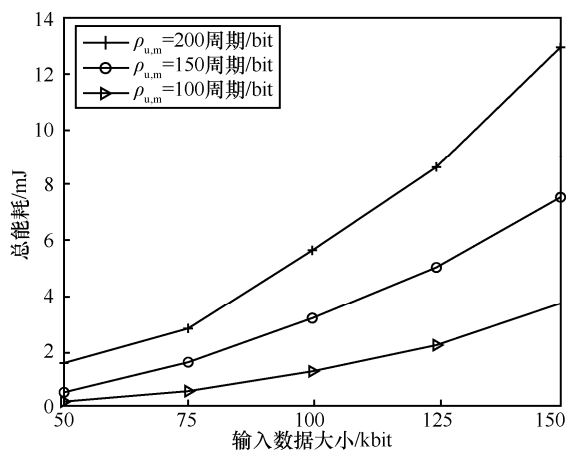


图6 总能耗与输入数据大小关系

长率也随之提高。这是因为较大的任务计算负荷会导致本地计算的能耗大幅上升,与此同时更多用户选择将任务卸载到MEC服务器执行,卸载用户比例与输入数据大小的关系如图7所示。另一

方面,从图 7 可以看出,随着用户任务输入数据量的增加,系统中卸载用户的比例也在增加。同样,因为任务数据量增加会加大大地计算的能耗,进而促使更多用户选择将任务卸载到边缘服务器上执行。这进一步证明了执行复杂任务时边缘计算在节约用户能耗方面的优势。

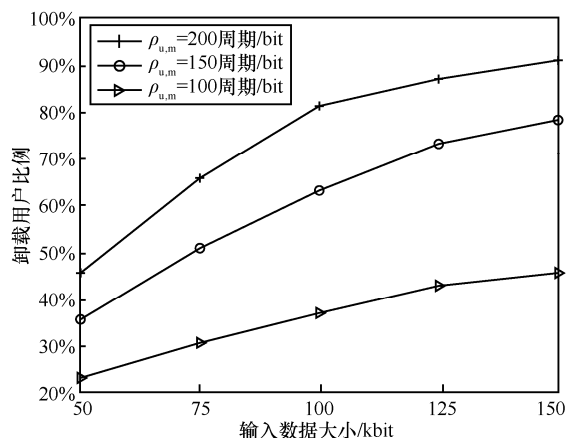


图 7 卸载用户比例与输入数据大小的关系

4 结束语

本文研究了基于 NOMA 的超密集 MEC 网络中以能耗最小化为目标的任务卸载和资源分配问题。为了在保证每个任务的 QoS 的要求下,最小化所有用户的能耗,本文联合优化了卸载决策、发射功率、计算资源和子信道资源分配。由于该优化问题是一个混合整数非线性问题,通常无法在多项式时间内得到最优解,因此,本文将原问题解耦为卸载决策和资源分配两个子问题,并提出了一种高效迭代优化方案获得原问题的近似解。仿真结果表明,与其他基准方案相比,所提方案能够实现更低的系统能耗,并且以较低的复杂度获得接近穷举方案的结果。在未来,可将目前的工作扩展到协同计算场景,允许用户的任务可以由不同的 MEC 服务器协同执行。此外,笔者还将考虑 MEC 服务器的缓存能力,通过预先缓存多个用户所需的数据库或程序节约系统的通信和计算资源。

参考文献:

- [1] MACH P, BECVAR Z. Mobile edge computing: a survey on architecture and computation offloading[J]. IEEE Communications Surveys & Tutorials, 2017, 19(3): 1628-1656.
- [2] 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.
- [3] ZHAO J H, LI Q P, GONG Y, et al. Computation offloading and resource allocation for cloud assisted mobile edge computing in vehicular networks[J]. IEEE Transactions on Vehicular Technology, 2019, 68(8): 7944-7956.
- [4] DING Z G, FAN P Z, POOR H V. Impact of non-orthogonal multiple access on the offloading of mobile edge computing[J]. IEEE Transactions on Communications, 2019, 67(1): 375-390.
- [5] FENG S M, ZHANG R, XU W, et al. Multiple access design for ultra-dense VLC networks: orthogonal vs non-orthogonal[J]. IEEE Transactions on Communications, 2019, 67(3): 2218-2232.
- [6] LI J, CHU S F, SHU F, et al. Contract-based small-cell caching for data disseminations in ultra-dense cellular networks[J]. IEEE Transactions on Mobile Computing, 2019, 18(5): 1042-1053.
- [7] ABU MAHADY I, BEDEER E, IKKI S, et al. Sum-rate maximization of NOMA systems under imperfect successive interference cancellation[J]. IEEE Communications Letters, 2019, 23(3): 474-477.
- [8] ZHANG Q, LUO K, WANG W, et al. Joint C-OMA and C-NOMA wireless backhaul scheduling in heterogeneous ultra dense networks[J]. IEEE Transactions on Wireless Communications, 2020, 19(2): 874-887.
- [9] QIN Z J, YUE X W, LIU Y W, et al. User association and resource allocation in unified NOMA enabled heterogeneous ultra dense networks[J]. IEEE Communications Magazine, 2018, 56(6): 86-92.
- [10] WU Y, NI K J, ZHANG C, et al. NOMA-assisted multi-access mobile edge computing: a joint optimization of computation offloading and time allocation[J]. IEEE Transactions on Vehicular Technology, 2018, 67(12): 12244-12258.
- [11] PAN Y J, CHEN M, YANG Z H, et al. Energy-efficient NOMA-based mobile edge computing offloading[J]. IEEE Communications Letters, 2019, 23(2): 310-313.
- [12] SENG S M, LUO C Q, LI X, et al. User matching on blockchain for computation offloading in ultra-dense wireless networks[J]. IEEE Transactions on Network Science and Engineering, 2021, 8(2): 1167-1177.
- [13] LI L X, CHENG Q Q, TANG X, et al. Resource allocation for NOMA-MEC systems in ultra-dense networks: a learning aided mean-field game approach[J]. IEEE Transactions on Wireless Communications, 2021, 20(3): 1487-1500.



- [14] NOURI N, ABOUEI J, JASEEMUDDIN M, et al. Joint access and resource allocation in ultradense mmWave NOMA networks with mobile edge computing[J]. IEEE Internet of Things Journal, 2020, 7(2): 1531-1547.
- [15] SENG S M, LI X, JI H, et al. Joint access selection and heterogeneous resources allocation in UDNs with MEC based on non-orthogonal multiple access[C]//Proceedings of 2018 IEEE International Conference on Communications Workshops. Piscataway: IEEE Press, 2018: 1-6.
- [16] LI Y, XU G C, GE J Q, et al. Jointly optimizing helpers selection and resource allocation in D2D mobile edge computing[C]//Proceedings of 2020 IEEE Wireless Communications and Networking Conference. Piscataway: IEEE Press, 2020: 1-6.
- [17] HAN S J, XU X D, FANG S S, et al. Energy efficient secure computation offloading in NOMA-based mMTC networks for IoT[J]. IEEE Internet of Things Journal, 2019, 6(3): 5674-5690.
- [18] 吴柳青, 朱晓荣. 基于边-端协同的任务卸载资源分配联合优化算法[J]. 电信科学, 2020, 36(3): 42-52.
- WU L Q, ZHU X R. Joint optimization algorithm for task offloading resource allocation based on edge-end collaboration[J]. Telecommunications Science, 2020, 36(3): 42-52.
- [19] BANDO K. Many-to-one matching markets with externalities among firms[J]. Journal of Mathematical Economics, 2012, 48(1): 14-20.
- [20] 崔静静. 非正交多址接入系统功率分配优化设计与性能分析[D]. 成都: 西南交通大学, 2018.
- CUI J J. Power allocation optimization design and performance analysis of non-orthogonal multiple access system[D]. Chengdu: Southwest Jiaotong University, 2018.
- [21] BODINE-BARON E, LEE C, CHONG A, et al. Peer effects and stability in matching markets[C]//Proceedings of International Symposium on Algorithmic Game Theory. [S.l.:s.n.], 2011.
- [22] LI S L, TAO Y Z, QIN X Q, et al. Energy-aware mobile edge computation offloading for IoT over heterogenous networks[J]. IEEE Access, 2019(7): 13092-13105.
- [23] PAPANDRIOPOULOS J, EVANS J S. SCALE: a low-complexity distributed protocol for spectrum balancing in multiuser DSL networks[J]. IEEE Transactions on Information Theory, 2009, 55(8): 3711-3724.
- [24] ZHAO J J, LIU Y W, CHAI K K, et al. Spectrum allocation and power control for non-orthogonal multiple access in HetNets[J]. IEEE Transactions on Wireless Communications, 2017, 16(9): 5825-5837.
- [25] WU W, ZHOU F H, HU R Q, et al. Energy-efficient resource allocation for secure NOMA-enabled mobile edge computing networks[J]. IEEE Transactions on Communications, 2020, 68(1): 493-505.
- [26] TANG L R, HU H L. Computation offloading and resource allocation for the Internet of Things in energy-constrained MEC-enabled HetNets[J]. IEEE Access, 2020(8): 47509-47521.
- [27] DI B Y, SONGL Y, LI Y H. Sub-channel assignment, power allocation, and user scheduling for non-orthogonal multiple access networks[J]. IEEE Transactions on Wireless Communications, 2016, 15(11): 7686-7698.

[作者简介]



绳韵(1998-), 女, 华北电力大学电气与电子工程学院硕士生, 主要研究方向为移动边缘计算网络中的资源管理。



许晨(1988-), 女, 博士, 北京邮电大学人工智能学院副研究员、博士生导师, 主要研究方向为智能网络优化、协作通信与计算、未来移动通信系统。



郑光远(1996-), 男, 华南理工大学电子与信息学院博士生, 主要研究方向为移动边缘计算网络中的资源分配、能量管理和干扰管理。