

基于NOMA的移动边缘计算系统公平能效调度算法

胡 晗^{*①②} 鲍 楠^① 凌 章^② 沈 乐^②

^①(南京邮电大学物联网学院 南京 210003)

^②(江苏省无线通信重点实验室 南京 210003)

摘 要: 将移动边缘计算技术(MEC)与非正交多址技术(NOMA)结合,同时考虑公平性,该文研究了采用NOMA上行部分卸载的MEC系统公平能效问题。首先将基于公平函数的用户速率与功耗比值定义为公平能效函数,随后提出了两种公平能效调度准则下的能效调度算法,即最大化最小速率准则下DK-SCA算法及最大化系统能效准则下DK-SCALE算法,通过算法实现分别得到两种公平能效调度准则下用户最佳本地CPU处理频率及最佳传输功率。最后通过仿真表明,与基准方案相比,所提基于NOMA的部分卸载方案能够有效地将本地计算和基于NOMA的边缘卸载结合,达到最佳的公平能效性能。

关键词: 边缘计算; 计算卸载; NOMA; 能效; 公平

中图分类号: TN929.5

文献标识码: A

文章编号: 1009-5896(2021)12-3563-08

DOI: [10.11999/JEIT200898](https://doi.org/10.11999/JEIT200898)

Fair Energy Efficiency Scheduling in NOMA-Based Mobile Edge Computing

HU Han^{①②} BAO Nan^① LING Zhang^② SHEN Le^②

^①(College of Internet of Things, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

^②(Jiangsu Key Laboratory of Wireless Communications, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

Abstract: Combing Mobile Edge Computing (MEC) and Non-Orthogonal Multiple Access (NOMA) technologies while considering fairness, this paper studies the fair energy efficiency of the MEC system using NOMA partial offloading. First, the ratio of user rate to power consumption based on the fair function is defined as the fair energy efficiency function. Then, two energy efficiency scheduling algorithms under the fair energy efficiency scheduling criteria are proposed, namely the DK-SCA algorithm under the maximum-minimum rate criterion and the DK-SCALE algorithm under the maximum system energy efficiency criterion. The optimal CPU-frequency cycle and optimal transmit power under these two fair energy efficiency scheduling criteria are obtained, respectively. Finally, simulations show that compared with the benchmark schemes, the proposed NOMA-based partial offloading scheme can effectively combine local computing with edge offloading based on NOMA, which can achieve the best fair energy efficiency performance.

Key words: Edge computing; Computing offloading; Non-Orthogonal Multiple Access (NOMA); Energy efficiency; Fair

1 引言

移动通信技术的飞速发展以及移动终端的日益

普及,使得网络服务和新型应用不断涌现。然而,移动终端由于自身有限的能量及计算处理能力,难以针对高计算强度、时延敏感性的业务(例如,增强现实,虚拟现实)进行处理^[1]。针对上述问题,业界提出了移动边缘计算解决方案^[2]。通过在移动网络边缘部署较强的计算和存储资源,用户可以将全部或部分计算任务卸载到移动边缘计算节点,能够有效解决移动设备在资源存储、计算性能及能效等方面的不足,为用户提供低时延、低能耗的网络服务方案。

计算卸载是移动边缘计算研究中的一个重要研究领域^[3,4]。其主要包含卸载决策、资源分配两个

收稿日期: 2020-10-20; 改回日期: 2021-04-29; 网络出版: 2021-11-10

*通信作者: 胡晗 han_h@njupt.edu.cn

基金项目: 国家自然科学基金(61871446, 61801244), 江苏省科技厅自然科学基金项目(BK20191378), 江苏省高等学校自然科学基金面上项目(18KJB510034)

Foundation Items: The National Natural Science Foundation of China (61871446, 61801244), The National Science Foundation Program of Jiangsu Province (BK20191378), The National Science Research Project of Jiangsu Higher Education Institutions (18KJB510034)

方面。根据系统的优化目标,目前计算卸载的研究主要集中在以降低时延为目标^[5]、以降低能量消耗为目标^[6]以及权衡能耗和时延为目标^[7]的3种类型。文献^[5]研究了多MEC服务场景下的时延最小化问题,提出了一种无需使用业务模型的资源调度方案,和传统算法相比,能够较好地降低通信时延及计算时延的加权和。文献^[6]研究了IoT网络中的能效卸载问题,针对网络环境及任务到达的随机性,提出了一种基于Lyapunov的动态卸载方案,假设用户之间采用TDMA接入方式,通过优化各个用户的卸载时间,使得系统能耗最小的同时保证了平均任务队列长度的稳定性。文献^[7]研究了多MEC服务器场景下的联合任务卸载和资源分配问题,考虑用户采用OFDMA接入方式,采用一种启发式算法解决了优化问题,上述文献均是采用传统的正交多址接入(Orthogonal Multiple Access, OMA)技术。

非正交多址接入(Non Orthogonal Multiple Access, NOMA)作为一种能够有效解决未来网络频谱资源稀缺的问题,成为当今无线通信中的研究热点技术。与OMA技术相比,NOMA系统可以在同一个信道中同时服务多个用户,可以非常大幅度地提高频谱的使用效率,近年来基于NOMA技术的移动边缘计算系统的卸载决策及资源分配问题受到了广泛的关注^[8-11]。文献^[9]研究了上行NOMA移动边缘计算系统的资源分配问题,考虑相同群组内用户采用NOMA方式,不同群组之间用户采用TDMA方式,将用户的任务计算时延及能耗的线性组合最小化作为优化目标,提出了一种低复杂度的迭代算法,优化了不同群组之间的时间分配隙数,各个群组下每个用户卸载数据量,MEC的计算CPU频率分配以及用户发射功率。文献^[10]研究了最大化功率和能耗受限情况下,NOMA-MEC系统中的时延最小化问题。结果表明,NOMA系统与TDMA系统相比,可获得较低的时延。文献^[11]研究了MEC系统中的能耗最小化问题。考虑用户计算任务上传和计算结果下载都采用NOMA接入的方式,优化了用户传输功率、传输时间分配及任务分割因子,结果表明,与传统的正交接入相比,NOMA能够有效降低MEC系统中的能耗。

上述绝大多数文献^[9-11]考虑最小化能耗或者最小化时延,同时在研究NOMA-MEC系统卸载决策和资源分配时,没有考虑用户之间的公平性,这可能会导致系统资源分配的不均衡,例如,当网络中某些用户处于信道较差状态时,它们的卸载速率相对较低,为了提升系统总体传输效率,系统将会对

信道状态好的用户分配更多的传输资源,这样将难以保证信道较差状态下的用户的卸载速率,造成不公平的状况。

本文考虑了MEC系统中的部分卸载模型^[4],将任务分成本地计算和卸载计算两部分;并在卸载计算中采用NOMA提升系统频谱效率。将用户速率与功耗比值定义为公平能效函数,分析了两种公平能效调度准则,即最大化最小速率准则及最大化系统能效准则下的网络资源优化问题。由于问题的分式结构和速率约束的非凸性,提出了基于Dinkelbach的连续凸近似算法(Dinkelbach-based Successive Convex Approximation, DK-SCA)及基于Dinkelbach的低复杂度连续凸近似算法(Dinkelbach-based Successive Convex Approximation for Low Complexity, DK-SCALE),对这两种准则下的优化问题进行求解,分别得到用户本地最佳CPU处理频率及最佳传输功率。仿真结果表明,与基于NOMA的完全卸载方案及基于FDMA的部分卸载方案相比,本文所提出的基于NOMA的部分卸载方案,能够在系统处理速率及功耗之间取得一个较好的折中,进而有效提升网络的公平能效性能。

2 系统模型

如图1所示,本文考虑 N 个用户设备和1个MEC服务器组成的移动边缘计算模型。考虑每个用户设备采用部分卸载模型^[4],即将自身的计算任务分成两部分,一部分本地计算,另一部分卸载到MEC服务器上计算。在上行卸载过程中, N 个用户均采用NOMA技术以提高频谱效率。用户的本地计算和任务卸载是同时进行的,由于MEC服务器能量充足,本文仅关注用户在任务卸载及本地计算过程中消耗的能效问题。接下来将会给出本地计算和任务卸载的细节信息。

2.1 数据卸载

假设用户与边缘计算服务器之间采用NOMA

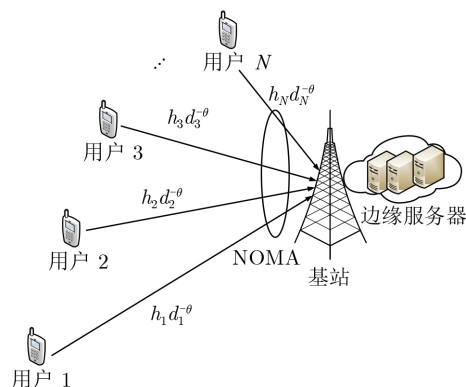


图1 系统模型

进行数据传输。NOMA根据各用户端信道增益的大小，在发送端进行功率控制，在接收端通过使用连续干扰消除(Successive Interference Cancellation, SIC)技术，将各用户间的干扰依次消除进而进行解码。由于各用户同时使用相同的频谱资源，因此用户可用带宽可以得到提升。假设信道带宽为 W ，信道无线电传播包括路径损耗和瑞利衰落，在接收端应用SIC时，对用户信道增益依次排序， $h_n d_n^{-\theta} \geq h_{n-1} d_{n-1}^{-\theta} \geq \dots \geq h_1 d_1^{-\theta}$ ，接收端可以先将增益较小的信号作为干扰，解出增益较大的信号，随后将其移除，顺次解码余下信号。因此，第 n 个用户的任务卸载速率可以表示为

$$r_n^{\text{off}} = W \log_2 \left(1 + \frac{p_n^u h_n d_n^{-\theta}}{\sum_{i=1}^{n-1} p_i^u h_i d_i^{-\theta} + \sigma^2} \right) \quad (1)$$

其中， d_n 为第 n 个用户与基站之间的距离， θ 是路径损耗指数； h_n 为第 n 个用户与基站之间的瑞利衰落； σ^2 为噪声功率。在卸载模式下，用户设备所消耗的功耗可以表示为

$$p_n^{\text{off}} = \zeta p_n^u + p_r \quad (2)$$

其中， ζ 为放大器系数， p_n^u 表示信息在卸载时的传输功耗， p_r 表示信号处理所消耗的恒定电路功耗。

2.2 本地计算

在本地计算模式中，假设用户设备具有一定的计算能力，根据卸载决策的选择情况，用户设备可以将任务的一部分用于本地计算，从而提高效率。本地计算将采用动态电压缩放(Dynamic Voltage Scaling, DVS)模型，使用DVS技术，用户可以自适应地调整其计算速度，以减少能耗或缩短计算时间。假设用户 n 的处理器处理1 bit数据需要的CPU处理周期为 γ_n (CPU cycle/bit)，定义用户 n 的CPU处理器频率为 f_n (CPU cycle/s)，则用户本地计算每秒可处理的比特数，即本地计算速度 r_n^{local} (bit/s)表示为^[4]

$$r_n^{\text{local}} = \frac{f_n}{\gamma_n} \quad (3)$$

根据DVS技术，本地计算的功耗为处理器速度的函数，可以表示为

$$p_n^{\text{local}} = \varepsilon f_n^3 \quad (4)$$

其中， ε 是处理器芯片的有效电容系数。

2.3 能量效率

能量效率是通信设计中的一个重要指标，特别是在边缘计算用户能源受限的情况下，以能量效率

作为系统目标在保障用户服务质量的同时也保障了系统运行的稳定性和长久性。与此同时，由于无线通信链路的不确定性以及各用户任务负载量的动态变化，用户公平性也至关重要，特别是NOMA通信中各用户采用功率控制进行协调传输，在不同的应用场景下会有不同的公平性需求和限制。因此，为了考虑各个用户之间计算速率的公平性，本文采用系统公平能效作为性能指标，将其定义为用户计算速率公平效用函数^[12-15]与用户总消耗功率的比值，具体公式为

$$\eta_\alpha = \frac{U_\alpha \left(\sum_{n \in N} R_n \right)}{\sum_{n \in N} P_n} \quad (5)$$

其中， $R_n = r_n^{\text{local}} + r_n^{\text{off}}$ ，表示第 n 个用户设备的本地和上传的总计算速率； $P_n = p_n^{\text{local}} + p_n^{\text{off}}$ ，表示第 n 个用户设备的本地和上传卸载的总功率消耗。 $U_\alpha \left(\sum_{n \in N} R_n \right)$ 为速率公平效用函数^[12-15]，其定义为

$$U_\alpha \left(\sum_{n \in N} R_n \right) = \begin{cases} \frac{\left(\sum_{n \in N} R_n \right)^{1-\alpha}}{1-\alpha}, & \alpha \geq 0, \alpha \neq 1 \\ \ln \left(\sum_{n \in N} R_n \right), & \alpha = 1 \end{cases} \quad (6)$$

其中， α 是公平指数。在实际应用中，根据不同的场景需求，需要选择不同的 α 值以保障用户的公平性^[12-15]。本文考虑两种特殊场景，一种是当 $\alpha = \infty$ 时，式(6)效用函数转换为所有用户中最小速率，系统旨在保障最小速率用户的公平性，系统能效即为最小速率用户的速度与系统所消耗功率的比值。另一种是当 $\alpha = 0$ 时，式(6)效用函数转换为用户和速率，此时系统旨在保障用户的总体性能，系统能效为所有用户速率之和与所有用户消耗功率的比值。具体论证过程详见文献^[13]中第3部分的引理3，该结论已广泛应用于文献^[12,14,15]。接下来，我们将分别针对这两种场景，通过优化用户计算处理频率以及发射功率，以最大化最小速率用户的系统能效($\alpha = \infty$)及最大系统总体能效($\alpha = 0$)这两种情况。

3 问题形成

本文旨在通过优化用户传输功率和本地计算频率来最大化系统的公平性能效，进而形成系统资源分配优化问题为

$$P_0: \max_{f_n, p_n^u} \eta_\alpha \quad (7)$$

$$\text{s.t. C1: } P_n^u \geq 0, \forall n \in N \quad (8)$$

$$\text{C2: } R_n \geq R_n^{\text{min}}, \forall n \in N \quad (9)$$

$$\text{C3: } P_n \leq P_n^{\text{th}}, \forall n \in N \quad (10)$$

$$C4: f_n^{\min} \leq f_n \leq f_n^{\max}, \forall n \in N \quad (10)$$

其中, C1为系统任务上传时的传输功率约束, 即传输功率不能为负值。在C2中是系统为保障服务质量要满足的最小速率约束 R_n^{\min} 。C3中的 P_n^{th} 是用户 n 可允许的最大可用功率上限。C4中 f_n^{\min} 和 f_n^{\max} 分别为用户设备处理器的最小和最大计算能力。在下文中, 根据 α 的不同取值, 本文将对系统的资源分配采用相应的算法进行优化, 以达到系统的最大能效。

4 算法设计

4.1 最大化最小速率用户系统能效($\alpha = \infty$)

当 $\alpha = \infty$, P_0 问题中速率效用函数可以写成:
 $U_\alpha \left(\sum_{n \in N} R_n \right) = \min_{n \in N} R_n$, 在这种情况下, 系统能效为最小速率用户的速度与系统整体消耗功率的比值, 即为

$$\eta_\infty = \frac{\min_{n \in N} R_n}{\sum_{n \in N} P_n} \quad (11)$$

因此, 原始优化问题 P_0 将转化为

$$P_{1.1}: \max_{f_n, p_n^u} \frac{\min_{n \in N} B \log_2 \left(1 + \frac{p_n^u h_n d_n^{-\theta}}{\sum_{i=1}^{n-1} p_i^u h_i d_i^{-\theta} + \sigma^2} \right) + \frac{f_n}{\gamma_n}}{\sum_{n \in N} (\zeta P_n^u + p_r + \varepsilon f_n^3)}$$

s.t. 式(7)–式(10)

为了解决 $\max - \min$ 问题, 我们引入新变量 $Z \leq R_n$, 作为 R_n 的下界, 因此添加约束变量

$$Z \leq B \log_2 \left(1 + \frac{p_n^u h_n d_n^{-\theta}}{\sum_{i=1}^{n-1} p_i^u h_i d_i^{-\theta} + \sigma^2} \right) + \frac{f_n}{\gamma_n}, \text{ 又由于}$$

系统的处理速度应满足大于等于 R_n^{\min} 。因此 $Z \geq R_n^{\min}$ 。则 $P_{1.1}$ 可以转化为

$$P_{1.2}: \max_{f_n, p_n^u} \frac{Z}{\sum_{n \in N} (\zeta P_n^u + p_r + \varepsilon f_n^3)} \quad (12)$$

s.t. 式(7), 式(9), 式(10)

$$Z \geq R_n^{\min}$$

$$Z \leq W \log_2 \left(1 + \frac{p_n^u h_n d_n^{-\theta}}{\sum_{i=1}^{n-1} p_i^u h_i d_i^{-\theta} + \sigma^2} \right) + \frac{f_n}{\gamma_n} \quad (13)$$

由于约束式(13)的不等式右边非凸, 引入辅助

变量 $\frac{Z}{B} - \frac{f_n}{\gamma_n B} \leq \exp(z_n^u)$ ^[12]并且将传输功率进行转换 $p_n^u = \exp(x_n^u)$, 则其可转换为两个等价约束

$$\frac{Z}{W} - \frac{f_n}{\gamma_n W} \leq \exp(z_n^u) \quad (14)$$

$$\log \left(2^{\exp(z_n^u)} - 1 \right) + \log \left(\sum_{i=1}^{n-1} \exp(x_i^u - x_n^u) \frac{h_i d_i^{-\theta}}{h_n d_n^{-\theta}} + \exp(-x_n^u) \frac{\sigma^2}{h_n d_n^{-\theta}} \right) \leq 0 \quad (15)$$

其中, 式(15)左侧第1项是一个凸函数, 第2项是 \log -sum-exp形式, 也是一个凸函数, 因此其为凸约束, 式(14)是一个指数函数大于线性函数, 因此其非凸, 为解决此问题, 本文采用SCA方法将其转换为凸问题形式, SCA的思想是通过引入辅助变量将原表达式的非凸部分采用1阶泰勒公式近似, 并连续不断更新辅助变量直到收敛至要求精度, 进而给出次优解。根据SCA, 则有

$$Z \leq \exp(\bar{z}_n^{k,u}) + \exp(\bar{z}_n^{k,u}) (z_n^u - \bar{z}_n^{k,u}) + \frac{f_n}{\gamma_n} \quad (16)$$

其中, $\bar{z}_n^{k,u}$ 是 z_n^u 在 k 次迭代时的近似值, 通过SCA多次迭代求解近似问题, 可以将 $P_{1.2}$ 转化为

$$P_{1.3}: \max_{Z, f_n, x_n^u, z_n^u} \frac{Z}{\sum_{n \in N} (\zeta \exp(x_n^u) + P_r + \varepsilon f_n^3)} \quad (17)$$

s.t. 式(7), 式(9), 式(10), 式(12), 式(15), 式(16)

$P_{1.3}$ 所有约束条件均为凸约束, 而其目标函数的分式结构为一个线性函数比凸函数的形式, 这种形式满足Dinkelbach方案要求的单调性条件, 因此, 可以采用Dinkelbach方案将其转换成等价的形式, 该方案通过设定一个初始值, 并经过多次迭代更新的方法不断逼近最优解^[16]。根据交替迭代算法的思想, 原始的目标函数可以转换为

$$Z - \eta_\infty^* \sum_{n \in N} (\zeta \exp(x_n^u) + P_r + \varepsilon f_n^3) \quad (17)$$

其中, η_∞^* 为最小速率能效的最优值, 通过预设一个暂态值 $\eta_\infty^i = \frac{Z^i}{\sum_{n \in N} (\zeta \exp(x_n^{u,i}) + P_r + \varepsilon f_n^{i3})}$ 不断迭代更新, 系统最终将收敛到最优能效值。因此, 在第 i 次迭代中, 目标函数可以转化为

$$Z - \eta_\infty^i \sum_{n \in N} (\zeta \exp(x_n^u) + P_r + \varepsilon f_n^3) \quad (18)$$

则问题可以转换为

$$P_{1.4}: \max_{Z, f_n, x_n^u, z_n^u} Z - \eta_\infty^i \sum_{n \in N} (\zeta \exp(x_n^u) + P_r + \varepsilon f_n^3) \quad (19)$$

s.t. 式(7), 式(9), 式(10), 式(12), 式(15), 式(16)

可以看出 $P_{1.4}$ 是关于 z, x_n^u, z_n^u 和 f_n 的凸问题, 可以用相应的凸优化工具进行求解, 通过求解 $P_{1.4}$, 不断迭代更新 z_n^u 与 $z_n^{k,u}$ 的值, 最终可以得到系统的最佳能效, 以下给出表1所示的求解最优值的算法。

表1 DK-SCA迭代算法

步骤1: 初始化本地计算速度 $f_n^{(0)}$ 和 $x_n^{u(0)}$, $Z^0, \eta_\infty^0 = 0$, 设置停止阈值 e , 迭代次数 I ;
步骤2: for $i=1: I$;
利用SCA迭代求解 $P_{1.4}$, 得到结果 $\{f_n^i, x_n^{u,i}, \eta_\infty^i, z_n^{u,i}(k)\}$, 更新能效暂态值 $\eta_\infty^i = \frac{Z^i}{\sum_{n \in N} (\zeta \exp(x_n^{u,i}) + P_r + \varepsilon f_n^{i3})} \quad (19)$
步骤3: if $\ \eta_\infty^i - \eta_\infty^{i-1}\ \leq e$ 获得最佳能效 $\eta_\infty^* = \eta_\infty^i$; break ;
步骤4: 输出最佳能效 η_∞^* 。

4.2 最大系统总体能效($\alpha = 0$)

当 $\alpha = 0$ 时, P_0 中速率效用函数为 $U_\alpha(\sum_{n \in N} R_n) = \sum_{n \in N} R_n$, 在这种情况下, 我们旨在最大化系统总体能效, 即为

$$\eta_0 = \max_{f_n, p_n^u} \frac{\sum_{n \in N} R_n}{\sum_{n \in N} P_n} \quad (20)$$

因此, 原始 P_0 可以转化为

$$P_{2.1}: \max_{f_n, p_n^u} \frac{\sum_{n \in N} \left(W \log_2 \left(1 + \frac{p_n^u h_n d_n^{-\theta}}{\sum_{i=1}^{n-1} p_i^u h_i d_i^{-\theta} + \sigma^2} \right) + \frac{f_n}{\gamma_n} \right)}{\sum_{n \in N} (\zeta P_n^u + p_r + \varepsilon f_n^3)}$$

s.t. 式(7)–式(10)

$$\text{在 } P_{2.1} \text{ 中, 同样由于 } \log_2 \left(1 + \frac{p_n^u h_n d_n^{-\theta}}{\sum_{i=1}^{n-1} p_i^u h_i d_i^{-\theta} + \sigma^2} \right)$$

非凸, 无法求得最优解, 此处采用SCALE方法将问题进行转换^[12], 然后通过求解转换后的问题来获得次优解, SCALE方法通过设置两个暂态变量 a_n 和 b_n , 通过交替更新暂态变量与目标变量来不断逼近最优值的, 具体方式如下:

首先, 我们进行变量代换 $p_n^u = \exp(x_n^u)$, 根据SCALE方法, 有

$$a_n \log_2 \left(\frac{\exp(x_n^u) h_n d_n^{-\theta}}{\sum_{i=1}^{n-1} \exp(x_i^u) h_i d_i^{-\theta} + \sigma^2} \right) + b_n \leq \log_2 \left(1 + \frac{\exp(x_n^u) h_n d_n^{-\theta}}{\sum_{i=1}^{n-1} \exp(x_i^u) h_i d_i^{-\theta} + \sigma^2} \right) \quad (21)$$

$$\text{令 } c_n = \frac{\exp(x_n^u) h_n d_n^{-\theta}}{\sum_{i=1}^{n-1} \exp(x_i^u) h_i d_i^{-\theta} + \sigma^2} \quad (22)$$

则式(22)在 a_n, b_n 满足式(23)和式(24)时取等号

$$a_n = \frac{c_n}{1 + c_n} \quad (23)$$

$$b_n = \log_2(1 + c_n) - \frac{c_n}{1 + c_n} a_n \log_2(c_n) \quad (24)$$

根据SCALE算法, 我们可以将 $P_{2.1}$ 中非凸部分转换为

$$Z_n(x_n^u, a_n, b_n) = a_n \ln 2 \left[(x_n^u + \ln(h_n d_n^{-\theta})) - \ln \left(\sum_{i=1}^{n-1} \exp(x_i^u) h_i d_i^{-\theta} + \sigma^2 \right) \right] + b_n \quad (25)$$

则 $P_{2.1}$ 可以通过交替迭代 a_n, b_n 和 x_n^u 进行求解, 问题进而转化为

$$P_{2.2}: \max_{f_n, x_n^u} \frac{\sum_{n \in N} \left(W Z_n(x_n^u, a_n, b_n) + \frac{f_n}{\gamma_n} \right)}{\sum_{n \in N} (\zeta \exp(x_n^u) + p_r + \varepsilon f_n^3)}$$

s.t. 式(7), 式(10),

$$W Z_n(x_n^u, a_n, b_n) + \frac{f_n}{\gamma_n} \geq R_n^{\min}, \forall n \in N \quad (26)$$

$$\zeta \exp(x_n^u) + p_r + \varepsilon f_n^3 \leq P_n^{\text{th}}, \forall n \in N \quad (27)$$

$P_{2.2}$ 同样是一个凹函数比凸函数的分式问题, 采用Dinkelbach方案将其转换成等价的形式, 原始目标函数可以转换为

$$\sum_{n \in N} \left(W Z_n(x_n^u, a_n, b_n) + \frac{f_n}{\gamma_n} \right) - \eta_0^* \sum_{n \in N} (\zeta \exp(x_n^u) + p_r + \varepsilon f_n^3) \quad (28)$$

其中, η_0^* 为能效的最优值, 通过不断迭代更新, 系统最终将得到其近似解。因此, 问题可转化为

$$P_{2.3}: \max_{f_n, x_n^u} \sum_{n \in N} \left(W Z_n(x_n^u, a_n, b_n) + \frac{f_n}{\gamma_n} \right) - \eta_0^i \sum_{n \in N} (\zeta \exp(x_n^u) + p_r + \varepsilon f_n^3)$$

s.t. 式(7), 式(10), 式(26), 式(27)

P_{2.3}的目标函数是最大化凹函数, 约束条件均为凸约束, 因此转化为一个标准的凸优化问题。可以采用凸优化工具进行求解, 求解过程如表2所示。

表 2 DK-SCALE迭代算法

步骤1:	取 $\zeta > 0$, $R_n^{\min} > 0$, $P_n^{\text{th}} > 0$, η_0^i ; 初始化P _{2.3} , $Z(0)$, $f_n(0)$, $p_n^u(0)$, $a_n(0)$, $b_n(0)$ 迭代次数I;
步骤2:	for $i = 1: I$ 利用SCALE方法交替迭代求解P _{2.3} , 得到近似能效: $\eta_0^i = \frac{\sum_{n \in N} \left(W Z_n(x_n^u, a_n, b_n) + \frac{f_n}{\gamma_n} \right)}{\sum_{n \in N} (\zeta \exp(x_n^u) + p_r + \varepsilon f_n^3)} \quad (29)$
步骤3:	if $\ \eta_0^i - \eta_0^{i-1}\ \leq e$ 获得最佳能效 $\eta_0^* = \eta_0^i$; break;
步骤4:	输出 (f_n^*, p_n^{u*}) 和最佳能效 η_0^* 。

5 性能评估

在本节中, 我们在MATLAB2018a平台下进行了实验仿真, 通过仿真结果来分析验证所提方法的性能。仿真参数设置参考文献[17]。假设用户服从 $[0, 1000] \times [0, 1000]$ 区域上的泊松分布, 瑞利衰落因子 h_n 服从均值为1的指数分布, 路径衰落指数 $\theta = 4$ 。系统带宽为 $W = 3 \times 10^6$ Hz, 干扰 $\sigma^2 = 10^{-7}$, 用户总数为 $N = 3$, 每个用户的任务计算强度为 $\gamma_n = 10^3$ cycle/bit。计算能量效率系数 $\varepsilon = 10^{-28}$, 放大器系数 $\zeta = 2$ 。每个用户设备的最大和最小计算能力均等设置, 即 $f_n^{\max} = 10^9$ Hz, $f_n^{\min} = 10^6$ Hz 时, 电路功率 p_r 为5 dBm。仿真结果中, 本文所提出的两种公平性准则下基于NOMA部分卸载方案均标记为NOMA+partial offloading。另外, 本文采用两种基准对比方案: 基准方案1为基于FDMA部分计算卸载方案, 在图中标记为FDMA+partial offloading; 基准方案2为基于NOMA的完全任务卸载方案, 在图中标记为NOMA+complete offloading。

图2—图4分别给出了当 $\alpha = \infty$ 时, 最大化最小速率准则下3种方案下的系统能效、处理速率和功耗的对比情况。系统能效与系统速率下限约束的关系如图2所示, 可以看出, 本文所提出的方案能取得比两个参考方案更好的能效性能。值得注意的是, 3种方案在速率下限比较小的时候, 均保持高

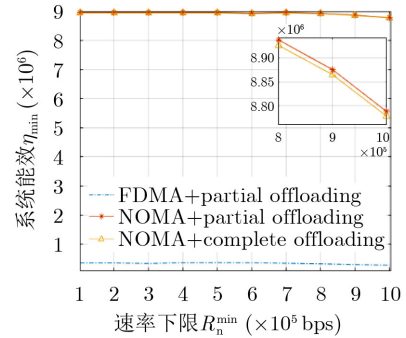


图 2 最大化最小速率准则下3种方案系统能效对比分析

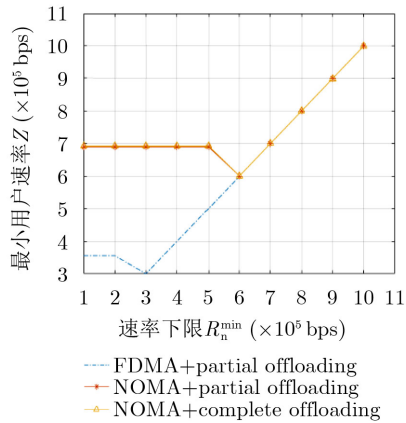


图 3 最大化最小速率准则下3种方案下系统速率对比分析

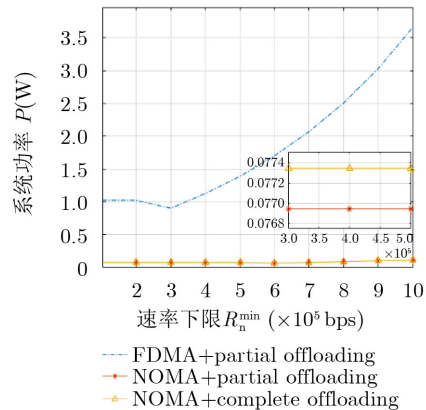


图 4 最大化最小速率准则下3种方案的系统功耗对比分析

水平, 但是随着速率下限逐渐增加, 各方案均出现了下降的趋势。这是因为在要求速率较小时, 基于能效的方案可以将处理速度维持在最佳能效处, 而当系统要求的速率下限逐渐增大时, 最佳能效处的速率无法达到要求, 于是各方案不得不增加任务处理的速率, 从而导致功耗的升高, 能效的下降。需要说明的是, 由于NOMA带来的卸载增益, 两种基于NOMA的方案取得的能效均远高于传统的FDMA方案。相比较NOMA完全卸载的方案, 本文提出的方案可以在本地处理和卸载间进行动态调整,

因而能够取得更好的能效表现。综上所述，本文所提出的方案，可以帮助系统提升公平性能效。

最大化最小速率准则下用户最小速率与系统速率下限的关系如图3所示，可以看出，各方案的系统最小速率在速度下限较小时保持在一个稳定的值，随着速率下限增大到一定程度，各方案的速率随速率下限要求进行线性增长。这与图2中观察到的结果相同，进一步说明了能效方案可以在速率和功耗间进行折中，同样，基于NOMA方案性能优于FDMA的方案。

最大化最小速率准则下系统功率与系统速率下限的关系如图4所示，其中基于FDMA部分卸载的方案功率远高于两种基于NOMA的方案，结合图2和图3的曲线趋势，可以看出，在速率下限小于 6×10^5 时，即使FDMA的方案消耗了更多的功率，其所能达到的速率远低于NOMA方案的速率，而当3种方案取得同样速率的时候，FDMA消耗的功率最高，能效最差。本文提出的方案所消耗的功耗是最低的，而取得的速率和其他两种方案相同，因此能效最好。综上，本文提出的方案能够取得最大化最小速率公平性场景中的最优能效。

图5—图7分别给出了当 $\alpha = 0$ 时，最大化系统能效准则下3种方案的对比情况。系统能效和速率下限的关系图如图5所示。从图中可以看出，随着速率下限的增加，基于FDMA部分卸载方案和本文提出的方案的能效均逐渐下降，而基于NOMA完全卸载方案维持在一个相对较稳定的范围，这是因为两种结合本地计算的方案会被速率下限迫使提升处理速率，进而导致功耗的增加，能效的降低。而基于NOMA完全卸载方案，由于维持在一个更高的速率，因此能效相对稳定，但是其能效是3种方案最差的。通过对比3种方案可以看出，本文所提方案能够取得最佳的能效。

图6给出最大化系统总体能效准则下3种方案总体速率和速率下限的对比关系。从图中可以看出，基于NOMA完全卸载方案的总体速率远高于其他两种方案，而基于NOMA部分卸载方案的方案次之，基于FDMA部分卸载方案的速率最小，这是因为有本地计算辅助的方案可以弥补纯卸载方案中信道差异带来的影响，所以系统能够更有效地进行功耗和处理速率间均衡分配，可以在满足最小速率下限的同时，将系统总体速率和能耗维持在一个更好的折中。同时，虽然基于NOMA部分卸载方案总体速率低于基于NOMA的完全卸载方案，但是其能效远高于其他方案，说明所提出的方案在保障速率要求的同时，能够更合理地进行功耗分配。

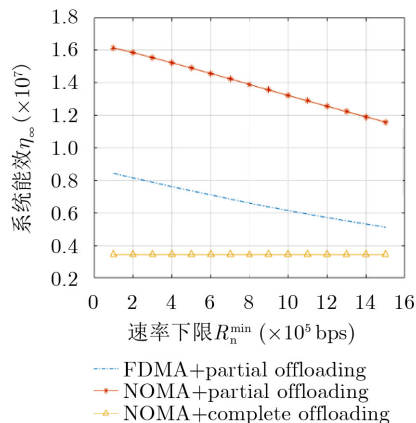


图5 最大化系统能效准则下3种方案系统能效对比分析

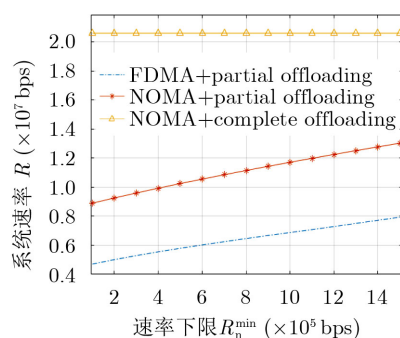


图6 最大化系统能效准则下3种方案的系统速率对比分析

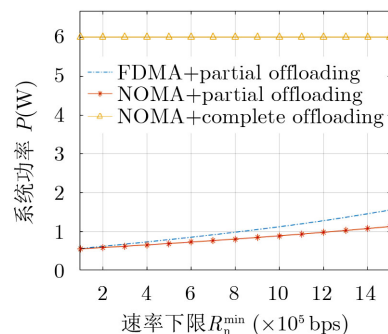


图7 最大化系统能效准则下3种方案的系统功耗对比分析

最大化系统能效准则下系统功耗与最小速率的关系如图7所示，与图5、图6趋势相同，本文所提出的方案功耗最低，并且相比其他两种方案，本文的方案仅通过消耗极低功率就能取得较高的数据处理速率，满足最小速率要求和服务质量的同时，还保障了最佳的能效，因而能够取得最佳的速率与功耗的折中。

6 结束语

本文从用户公平能效的角度，针对采用NOMA多址技术的上行部分卸载MEC系统方案，提出了两种公平性能效的调度算法：最大化最小速率准则下的DK-SCA算法，及最大化系统总体能效准则下

DK-SCALE算法。通过仿真表明, 本文所提出方案能够有效地将本地计算和基于NOMA的边缘卸载结合, 达到最佳的公平能效表现。

参考文献

- [1] BARBERA M V, KOSTA S, MEI A, *et al.* To offload or not to offload? The bandwidth and energy costs of mobile cloud computing[C]. Proceedings - IEEE INFOCOM, Turin, Italy, 2013: 1285–1293. doi: [10.1109/INFOCOM.2013.6566921](https://doi.org/10.1109/INFOCOM.2013.6566921).
 - [2] TAN L T, HU R Q, and HANZO L. Twin-timescale artificial intelligence aided mobility-aware edge caching and computing in vehicular networks[J]. *IEEE Transactions on Vehicular Technology*, 2019, 68(4): 3086–3099. doi: [10.1109/TVT.2019.2893898](https://doi.org/10.1109/TVT.2019.2893898).
 - [3] TAN L T and HU R Q. Mobility-aware edge caching and computing in vehicle networks: A deep reinforcement learning[J]. *IEEE Transactions on Vehicular Technology*, 2018, 67(11): 10190–10203. doi: [10.1109/TVT.2018.2867191](https://doi.org/10.1109/TVT.2018.2867191).
 - [4] HU Han, SONG Weiwei, WANG Qun, *et al.* Mobility-aware offloading and resource allocation in MEC-enabled IoT networks[C]. Proceedings of the 16th International Conference on Mobility, Sensing and Networking (MSN), Tokyo, Japan, 2020: 554–560. doi: [10.1109/MSN50589.2020.00092](https://doi.org/10.1109/MSN50589.2020.00092).
 - [5] ZHANG Yuan, DU Peng, WANG Jiang, *et al.* Resource scheduling for delay minimization in multi-server cellular edge computing systems[J]. *IEEE Access*, 2019, 7(99): 86265–86273. doi: [10.1109/ACCESS.2019.2924032](https://doi.org/10.1109/ACCESS.2019.2924032).
 - [6] CHEN Ying, ZHANG Ning, ZHANG Yongchao, *et al.* Energy efficient dynamic offloading in mobile edge computing for internet of things[J]. *IEEE Transactions on Cloud Computing*, 2021, 9(3): 1050–1060. doi: [10.1109/TCC.2019.2898657](https://doi.org/10.1109/TCC.2019.2898657).
 - [7] TRAN T X and 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. doi: [10.1109/TVT.2018.2881191](https://doi.org/10.1109/TVT.2018.2881191).
 - [8] SUN Haijian, ZHOU Fuhui, and HU R Q. Joint offloading and computation energy efficiency maximization in a mobile edge computing system[J]. *IEEE Transactions on Vehicular Technology*, 2019, 68(3): 3052–3056. doi: [10.1109/TVT.2019.2893094](https://doi.org/10.1109/TVT.2019.2893094).
 - [9] YANG Zhaohui, PAN Cunhua, HOU Jiancao, *et al.* Efficient resource allocation for mobile-edge computing networks with NOMA: Completion time and energy minimization[J]. *IEEE Transactions on Communications*, 2019, 67(11): 7771–7784. doi: [10.1109/TCOMM.2019.2935717](https://doi.org/10.1109/TCOMM.2019.2935717).
 - [10] ZENG Ming, NGUYEN N P, DOBRE O A, *et al.* Delay minimization for NOMA-assisted MEC under power and energy constraints[J]. *IEEE Wireless Communications Letters*, 2019, 8(6): 1657–1661. doi: [10.1109/LWC.2019.2934453](https://doi.org/10.1109/LWC.2019.2934453).
 - [11] PAN Yijin, CHEN Ming, YANG Zhaohui, *et al.* Energy-efficient NOMA-based mobile edge computing offloading[J]. *IEEE Communications Letters*, 2019, 23(2): 310–313. doi: [10.1109/LCOMM.2018.2882846](https://doi.org/10.1109/LCOMM.2018.2882846).
 - [12] PHAM Q V and HWANG W J. Fairness-aware spectral and energy efficiency in spectrum-sharing wireless networks[J]. *IEEE Transactions on Vehicular Technology*, 2017, 66(11): 10207–10219. doi: [10.1109/TVT.2017.2717926](https://doi.org/10.1109/TVT.2017.2717926).
 - [13] MO J and WALRAND J. Fair end-to-end window-based congestion control[J]. *IEEE/ACM Transactions on Networking*, 2000, 8(5): 556–567. doi: [10.1109/90.879343](https://doi.org/10.1109/90.879343).
 - [14] GUO Chongtao, ZHANG Yan, SHENG Min, *et al.* Alpha-fair power allocation in spectrum-sharing networks[J]. *IEEE Transactions on Vehicular Technology*, 2016, 65(5): 3771–3777. doi: [10.1109/TVT.2015.2436068](https://doi.org/10.1109/TVT.2015.2436068).
 - [15] SCHWARZ S, MEHLFUHRER C, and RUPP M. Throughput maximizing multiuser scheduling with adjustable fairness[C]. IEEE International Conference on Communications (ICC), Kyoto, Japan, 2011: 1–5. doi: [10.1109/icc.2011.5963489](https://doi.org/10.1109/icc.2011.5963489).
 - [16] YU Ye, BU Xiangyuan, YANG Kai, *et al.* Green large-scale fog computing resource allocation using joint benders decomposition, dinkelbach Algorithm, ADMM, and branch-and-bound[J]. *IEEE Internet of Things Journal*, 2019, 6(3): 4106–4117. doi: [10.1109/JIOT.2018.2875587](https://doi.org/10.1109/JIOT.2018.2875587).
 - [17] WEI Zhiqiang, ZHAO Lou, GUO Jiajia, *et al.* Multi-beam NOMA for hybrid mmWave systems[J]. *IEEE Transactions on Communications*, 2019, 67(2): 1705–1719. doi: [10.1109/TCOMM.2018.2879930](https://doi.org/10.1109/TCOMM.2018.2879930).
- 胡 晗: 女, 1985年生, 副教授, 研究方向为无线通信网络资源管理及优化等。
 鲍 楠: 女, 1985年生, 讲师, 研究方向为异构网络资源优化及干扰抑制等。
 凌 章: 男, 1993年生, 硕士, 研究方向为边缘计算及动态资源分配等。
 沈 乐: 男, 1997年生, 硕士生, 研究方向为边缘计算及动态资源分配等。

责任编辑: 陈 倩