

基于多天线NOMA的移动边缘计算网络公平 计算卸载策略研究

曾 胜¹, 黄小红¹, 李丹丹¹, 于维军²

(1. 北京邮电大学计算机学院(国家示范性软件学院), 北京 100876; 2. 火炬高技术产业开发区党政办公室, 山东威海 264209)

摘要: 为了解决移动边缘计算网络中不断增加的用户计算需求与有限的计算和通信资源之间的矛盾, 以及在处理用户的计算任务时公平性难以得到保证等问题, 本文针对基于多天线NOMA(Non-Orthogonal Multiple Access)的移动边缘计算网络, 提出了一种保证用户公平性的计算卸载策略. 通过联合优化用户的分簇、NOMA簇的解码顺序、用户的发送功率、用户的CPU处理频率和传输时间, 建立了一个系统公平计算效率最大化问题. 考虑到优化问题的非凸性, 将其分解为两个阶段进行求解. 在第一个阶段中, 设计了一种有效的启发式算法来实现用户的分簇, 并根据分簇的结果来确定NOMA簇的解码顺序; 在第二个阶段中, 提出了一种基于Dinkelbach方法和连续凸逼近(Successive Convex Approximation, SCA)方法的迭代算法来获得用户的发送功率、用户的CPU处理频率和传输时间. 仿真结果表明了所提出的方案具有较好的收敛性, 且与几个基线方案进行对比, 能够实现更加优越的系统性能.

关键词: 移动边缘计算; 多天线; 非正交多址接入; 公平性; 计算卸载

基金项目: 教育部-中国移动科研基金项目(No.MCM20160304)

中图分类号: TN929.5

文献标识码: A

文章编号: 0372-2112(2023)09-2457-12

电子学报 URL: <http://www.ejournal.org.cn>

DOI: 10.12263/DZXB.20211306

Fair Computation Offloading for a Multi-Antenna NOMA Aided Mobile Edge Computing Network

ZENG Sheng¹, HUANG Xiao-hong¹, LI Dan-dan¹, YU Wei-jun²

(1. School of Computer Science(National Pilot Software Engineering School),

Beijing University of Posts and Telecommunications, Beijing 100876, China;

2. Office of Party and Government Affairs, Torch Hi-tech Industrial Development Zone, Weihai, Shandong 264209, China)

Abstract: In order to solve the contradiction between the increasing computing demands of users and the limited computing and communication resources in mobile edge computing networks, and the difficulty to guarantee the fairness in handling users' computing tasks, this paper proposes a fair computation offloading policy for a mobile edge computing network based on multi-antenna NOMA (Non-Orthogonal Multiple Access). By jointly optimizing the user clustering, the decoding order of NOMA cluster, the transmission power, the CPU processing frequency, and the transmission time, an optimization problem is formulated to maximize the fair computation efficiency of the system. Considering the non-convexity of the optimization problem, it is decomposed into two solving stages. In the first stage, an effective heuristic algorithm is designed to realize user clustering, and determine the decoding order of NOMA cluster according to the clustering results. In the second stage, based on the Dinkelbach method and the SCA (Successive Convex Approximation), an iterative algorithm is proposed to obtain the transmission power, the CPU processing frequency, and the transmission time. Simulation results show that the proposed scheme has good convergence, and it can achieve the superior system performance compared with several baseline schemes.

Key words: mobile edge computing; multi-antenna; NOMA; fairness; computation offloading

Foundation Item(s): Research Fund of Ministry of Education-China Mobile (No.MCM20160304)

1 引言

移动边缘计算(Mobile Edge Computing, MEC)被认为是一种能够促进新一代移动通信网络发展的、很有前景的技术^[1]. 通过将计算资源下沉到无线网络边缘侧,资源受限的移动设备(Mobile Device, MD)可以将其繁重的任务卸载到边缘侧的MEC服务器中进行处理,可以有效地应对一些计算密集型和时延敏感型应用. 计算卸载问题是MEC中一个基本的问题,它的主要目的是去找到一个最优的计算和通信资源的分配策略来卸载MD的计算任务,以最大化系统效用. 然而,考虑到日益紧张的通信资源,以及MEC中有限的计算资源,如何设计一种高效的计算卸载策略变得至关重要.

在过去的研究中,大量的工作主要采用了正交多址接入(Orthogonal Multiple Access, OMA)的方式,分别从系统的计算速率、时延、能耗以及它们的组合角度对MEC中的计算卸载问题展开了研究^[2-5]. 近年来,非正交多址接入(Non-Orthogonal Multiple Access, NOMA)被认为是5G/B5G网络中非常有潜力的接入技术^[6]. 不同于传统OMA方案像“切蛋糕”似的资源分配方式,NOMA利用多用户信道增益的差异,在功率域内对用户信号进行区分. 因此,多个用户的信号能够在同一时频资源块上进行传输,可以显著提高系统的频谱效率. 受此启发,结合NOMA和MEC被期待是一种能够进一步提高MEC系统计算卸载性能的有效途径,在学术界也受到了广泛的关注^[7-15].

文献[7]针对基于NOMA的无线供能MEC系统,提出了一种系统总的计算速率最大化的计算卸载策略,其中,联合优化了用户的计算模式、发送功率、CPU处理频率和传输时间. 文献[8]分别在用户的最大发送功率限制下和最大能量限制下建立了两种优化问题. 通过优化用户的发送功率,以最小化NOMA-MEC系统中所有用户的任务完成时间. 文献[9]研究了一种包含多个MEC服务器场景的计算卸载策略,以最小化所有用户任务完成时间和边缘服务器计算资源使用成本的加权和,其中,用户利用NOMA将其计算任务同时发送给多个MEC服务器进行处理. 文献[7-9]都侧重于系统整体性能的优化,忽略了由于用户地理位置的差异造成任务执行之间的不公平性,从而会导致系统的资源向信道条件好的用户倾斜. 文献[10,11]考虑了用户的公平性,分别以最大化由基于 α 公平效用函数的计算速率和用户总能耗所组成的系统效用、最大化基于 α 公平效用函数的系统能量效率为优化目标,提出了两种公平计算卸载策略. 上述文献[7-11]仅仅考虑了单个NOMA簇的场景. 由于接收端解码复杂度和误差传播的存在,往往单个NOMA簇能够容纳的用户数有限. 随着系统用户数的增加,研究多个NOMA簇的场景是很

有必要的,但这也给计算卸载策略的设计带来了新的挑战:(1)如何去匹配用户与NOMA簇;(2)如何去消除NOMA簇间的干扰.

目前,已有一些文献研究了多个NOMA簇场景下的计算卸载策略. 文献[12]考虑了在多个NOMA簇之间采用时分的方式来消除干扰,在用户分簇已知的前提下,通过优化任务卸载量、传输时间、发送功率和MEC服务器计算资源分配,研究了一种计算卸载策略来最小化用户的时延和能耗的加权和. 文献[13-15]采用了频分的方式来消除簇间干扰. 文献[13]设计了一种用户能耗最小化的策略,采用了一种高效的启发式算法来实现用户分簇、计算和通信资源块的分配. 文献[14]通过优化用户分簇、发送功率和计算资源分配建立了一个用户卸载效率最大化问题. 不同于文献[12-14]中的单MEC服务器场景,文献[15]针对多个MEC服务器场景展开了研究,所有用户被分成了多个相互交叠的簇,用户可以将其计算任务同时卸载到多个MEC服务器中进行处理,每个MEC服务器也可以处理多个用户的计算任务,通过优化用户分簇和发送功率研究了一个系统长期平均效用最大化问题. 文献[12-15]均未考虑用户的公平性.

基于以上研究,本文进一步研究了多个NOMA簇场景下的公平计算卸载策略. 与现有的方案相比,主要不同之处在于:(1)本文建立了一个基于上行多天线NOMA的部分卸载模型,其中,考虑到无线网络中可用时频资源的稀缺性,本文假设所有的用户共享相同的时频资源,利用多天线技术^[16,17],在基站(Base Station, BS)端设计迫零波束赋形矩阵,采用空分的方式来部分地消除簇间干扰,以均衡系统性能和设计复杂度;(2)利用BS端的全局网络信息优化了多个NOMA簇的解码顺序,进一步提升了系统的性能;(3)在优化目标中考虑了多种指标,包括用户的计算速率、能耗和公平性,通过联合优化用户的分簇、NOMA簇的解码顺序、用户的发送功率、用户的CPU处理频率和传输时间,建立了一个系统公平计算效率最大化问题,其中,系统公平计算效率被定义为用户基于 α 公平效用函数的计算速率与用户的总能耗的比值. 尽管本文所提出的方案弥补了现有文献的一些不足,然而也面临着新的挑战:(1)非凸的 α 公平效用函数与分式优化目标相互耦合,导致优化目标更加复杂;(2)优化变量中同时存在2种离散变量和3种连续变量,并且变量之间紧密耦合,这导致问题的求解难度增加. 本文在后续将围绕解决上述挑战而展开.

2 系统模型

考虑一个基于多天线NOMA的多用户MEC网络,

如图 1 所示. 网络包含了一个 BS 和 K 个用户, BS 有 M 根天线, 每个用户有 1 根天线. BS 与 MEC 服务器相连, 位于小区中心, 为用户提供计算服务, 所有的用户分布在小区的覆盖范围内. 为了更好地应用多天线 NOMA 技术, K 个用户被分成 M 个簇. 这里, 本文假设 $K \geq 2M$, 用户分簇将在第 4.1 小节中进行讨论. 令 U_m 表示簇 m ($m \in \{1, 2, \dots, M\}$) 中用户的数量, 满足 $\sum_{m=1}^M U_m = K$. 考虑到簇间用户数量的公平性^[18], 每个簇中能够容纳的用户数量为 $\lfloor \frac{K}{M} \rfloor$ 或者 $\lfloor \frac{K}{M} \rfloor - 1$. 此外, 给出了本文所采用的主要参数及其含义, 如表 1 所示.

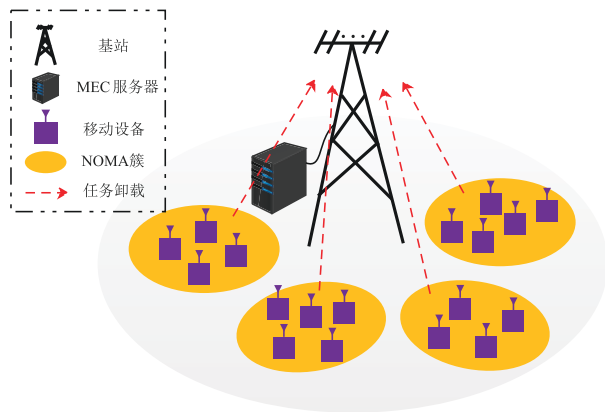


图 1 系统模型

表 1 主要参数及其含义

参数	含义
K	用户的数量
M	BS 的天线数/NOMA 簇的个数
U_m	簇 m 中用户的数量
B	系统带宽
$P_{m,u}$	用户 $[m,u]$ 的发送功率
T	时间块的长度
t_1	用户的传输时间
$S_{m,u}^e$	用户 $[m,u]$ 可以卸载的比特数
$E_{m,u}^e$	用户 $[m,u]$ 卸载任务时的能耗
L_{mec}	MEC 服务器的 CPU 处理频率
$S_{m,u}^1$	用户 $[m,u]$ 本地计算的比特数
$E_{m,u}^1$	用户 $[m,u]$ 本地计算的能耗
$C_{m,u}$	处理用户 $[m,u]$ 的 1 比特计算任务所需要的 CPU 周期数
$L_{m,u}$	用户 $[m,u]$ 的 CPU 处理频率
$\omega_{m,u}$	用户 $[m,u]$ 的有效转换电容
$S_{m,u}$	用户 $[m,u]$ 可完成的总比特数
$E_{m,u}$	用户 $[m,u]$ 的总能耗
η_{CE}^α	系统公平计算效率

2.1 上行多天线 NOMA

令用户 $[m, u]$ ($m \in \{1, 2, \dots, M\}, u \in \{1, 2, \dots, U_m\}$)

表示簇 m 中的第 u 个用户. 对于上行传输, BS 接收到的信号可以表示为:

$$\mathbf{y}_{\text{total}} = \sum_{m=1}^M \sum_{u=1}^{U_m} \mathbf{h}_{m,u} p_{m,u} s_{m,u} + \mathbf{n} \quad (1)$$

其中, $s_{m,u}$ 为用户 $[m, u]$ 要发送给 BS 的信号, 满足期望 $E[\|s_{m,u}\|^2] = 1$; $p_{m,u}$ 为用户的发送功率; $\mathbf{n} \in \mathbb{C}^{M \times 1}$ 为 BS 端的加性高斯白噪声 (Additive White Gaussian Noise, AWGN) 向量; $\mathbf{h}_{m,u} \in \mathbb{C}^{M \times 1}$ 是从用户 $[m, u]$ 到 BS 的复合信道向量, 模型如下:

$$\mathbf{h}_{m,u} = \sqrt{\rho(d_{m,u})} \mathbf{g}_{m,u} \quad (2)$$

其中, $\rho(d_{m,u})$ 为大尺度信道衰落增益, 它是用户 $[m, u]$ 与 BS 之间距离 $d_{m,u}$ 的函数, $\mathbf{g}_{m,u} \in \mathbb{C}^{M \times 1}$ 为小尺度信道衰落向量.

利用波束赋形矩阵 $\mathbf{W} \in \mathbb{C}^{M \times M}$ 后, 可以得到 BS 端接收到的信号为:

$$\begin{aligned} \mathbf{y} &= \mathbf{W}^H \mathbf{y}_{\text{total}} = \mathbf{W}^H \left(\sum_{m=1}^M \sum_{u=1}^{U_m} \mathbf{h}_{m,u} p_{m,u} s_{m,u} + \mathbf{n} \right) \\ &= \begin{bmatrix} \mathbf{w}_1^H \\ \mathbf{w}_2^H \\ \vdots \\ \mathbf{w}_M^H \end{bmatrix} \left(\sum_{m=1}^M \sum_{u=1}^{U_m} \mathbf{h}_{m,u} p_{m,u} s_{m,u} + \mathbf{n} \right) \end{aligned} \quad (3)$$

其中, \mathbf{w}_m 为波束赋形矩阵 \mathbf{W} 的第 m 列, 也是簇 m 的波束赋形向量, 满足 $\|\mathbf{w}_m\|^2 = 1$.

由式 (3) 可以看出, 在 BS 端当解码用户 $[m, u]$ 的信号时会受到两种类型的干扰: 簇间干扰 $\mathbf{w}_m^H \sum_{j=1, j \neq m}^M \sum_{u=1}^{U_j} \mathbf{h}_{j,u} p_{j,u} s_{j,u}$ 和簇内干扰 $\mathbf{w}_m^H \sum_{i=1, i \neq u}^{U_m} \mathbf{h}_{m,i} p_{m,i} s_{m,i}$. 为了降低簇间干扰和设计复杂度, 本文采用了迫零波束赋形方案, 它能够很好地均衡系统性能和设计复杂度^[19]. 考虑到 BS 的天线数和分簇的个数相等, BS 无法通过设计波束赋形矩阵来完全消除簇间干扰. 因此, 本文在设计迫零波束赋形矩阵时仅仅考虑了每个簇中信道增益最高的用户^[16,17]. 不失一般性, 簇 m 中用户的信道增益排序为: $\|\mathbf{h}_{m,1}\|^2 \geq \|\mathbf{h}_{m,2}\|^2 \geq \dots \geq \|\mathbf{h}_{m,U_m}\|^2$, 迫零波束赋形矩阵可以表示为:

$$\mathbf{Q} = (\mathbf{H}^\dagger)^H \quad (4)$$

其中, $\mathbf{H} = [\mathbf{h}_{1,1}, \mathbf{h}_{2,1}, \dots, \mathbf{h}_{M,1}]$, $\mathbf{H}^\dagger = (\mathbf{H}^H \mathbf{H})^{-1} \mathbf{H}^H$. 对 \mathbf{Q} 中的每一列归一化, 可以得到波束赋形矩阵 \mathbf{W} , 也就是 $\mathbf{w}_m = \frac{\mathbf{q}_m}{\|\mathbf{q}_m\|}$, 其中, \mathbf{q}_m 为 \mathbf{Q} 的第 m 列. 此时, 通过利用 \mathbf{W} , 每个簇中第 1 个用户将不会对其它簇的用户产生干扰, 其余用户之间仍然存在簇间干扰.

对于簇内干扰,本文采用了NOMA技术.针对每一个NOMA簇,BS通过执行串行干扰消除(Successive Interference Cancellation, SIC),可以按一定的顺序解码出簇内用户的信号,其中,先解码的用户信号将被移除,不会对后解码的用户造成干扰.为了减小系统开销,本文基于簇内用户的信道增益排序,对接收到的叠加信号采用了一个固定的降序解码方式.具体来说,在簇 m 中,解码第 i ($i \in \{1, 2, \dots, U_m\}$)个用户的信号会受到来自

$$R_{m,u} = B \log_2 \left(1 + \frac{p_{m,u} \|\mathbf{w}_m^H \mathbf{h}_{m,u}\|^2}{\sum_{i=u+1}^{U_m} p_{m,i} \|\mathbf{w}_m^H \mathbf{h}_{m,i}\|^2 + \sum_{j=m+1}^M \sum_{k=2}^{U_j} p_{j,k} \|\mathbf{w}_m^H \mathbf{h}_{j,k}\|^2 + \|\mathbf{n}\|^2} \right) \quad (5)$$

其中, B 为系统带宽.此外,当 $u=U_m$ 时,

$$\sum_{i=u+1}^{U_m} p_{m,i} \|\mathbf{w}_m^H \mathbf{h}_{m,i}\|^2 = 0; \quad \text{当 } m=M \text{ 时},$$

$$\sum_{j=m+1}^M \sum_{k=2}^{U_j} p_{j,k} \|\mathbf{w}_m^H \mathbf{h}_{j,k}\|^2 = 0.$$

2.2 MEC 服务

假设在持续时间为 T 的时间块内每个用户都有一个要处理的计算任务,并使用了按位独立的卸载模型^[20].因此,本文考虑了一个部分计算卸载方案.如图2所示,该方案涉及到了两种基本的计算模式:任务卸载模式和本地计算模式.

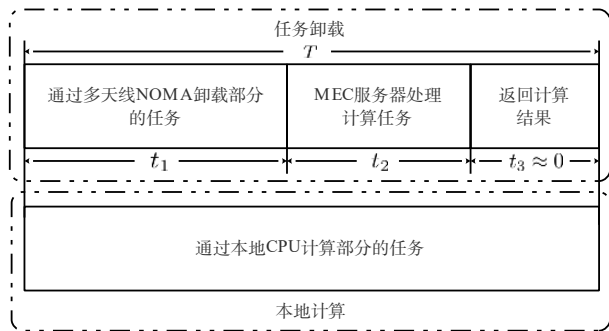


图2 计算模式

(1)任务卸载模式:主要包含三个阶段,令 t_1, t_2, t_3 分别表示这三个阶段的持续时间.在第一个阶段,通过应用多天线NOMA,所有用户能够在相同的时频资源块上卸载部分的任务到MEC服务器.在第二个阶段,MEC服务器处理用户卸载的计算任务.在第三个阶段,MEC服务器将计算结果返回给用户.考虑到计算结果通常远小于原始计算任务的大小、BS的发送功率通常高于用户的发送功率,本文假设第三个阶段忽略不计.在这个模式下,用户 $[m, u]$ 可以部分地卸载 $S_{m,u}^c$ 比特的数据到MEC服务器.根据前面的讨论, $S_{m,u}^c$ 可表示为:

$$S_{m,u}^c = R_{m,u} t_1 \quad (6)$$

第 $\{i+1, i+2, \dots, U_m\}$ 个用户信号的干扰.对于第 U_m 个用户而言,解码它的信号将不会受到其它用户信号的干扰.此外,考虑到BS有网络的全局信息,假设 M 个簇是按先后顺序逐个解码各个簇的用户信号,先解码的NOMA簇中的用户将不会对其它簇产生干扰.为了便于表述,不失一般性,假设先解码簇号小的NOMA簇,具体的解码顺序方案将在第4.1小节中进行讨论.因此,可以计算出用户 $[m, u]$ 的可实现传输速率为:

对应的能耗可以表示为:

$$E_{m,u}^c = p_{m,u} t_1 \quad (7)$$

此外,针对不同用户卸载的计算任务,MEC服务器采用了负载均衡^[21,22].具体来说,根据用户卸载的计算任务的大小,MEC服务器成比例的将其计算资源分配给不同的用户.因此,不同用户的计算任务在MEC服务器中的处理时间相等,即:

$$t_2 = \frac{\sum_{m=1}^M \sum_{u=1}^{U_m} S_{m,u}^c C_{m,u}}{L_{mec}} \quad (8)$$

其中, $C_{m,u}$ 为处理用户 $[m, u]$ 的1比特计算任务所需要的CPU周期数, L_{mec} (CPU周期数/秒)为MEC服务器的CPU处理频率.

(2)本地计算模式:在给定的时间块内,用户 $[m, u]$ 可以通过它的本地CPU来完成部分的计算任务.定义 $L_{m,u}$ 为用户 $[m, u]$ 的CPU处理频率.由此可得,通过本地计算模式,用户 $[m, u]$ 可以计算的总比特数为:

$$S_{m,u}^l = \frac{L_{m,u} T}{C_{m,u}} \quad (9)$$

对应的能耗可以表示为:

$$E_{m,u}^l = \omega_{m,u} (L_{m,u})^3 T \quad (10)$$

其中, $\omega_{m,u}$ 为用户 $[m, u]$ 的有效转换电容^[1].

结合上述两种计算模式,用户 $[m, u]$ 完成的总比特数可以表示为:

$$S_{m,u} = S_{m,u}^c + S_{m,u}^l = R_{m,u} t_1 + \frac{L_{m,u} T}{C_{m,u}} \quad (11)$$

对应的总能耗可以表示为:

$$E_{m,u} = E_{m,u}^c + E_{m,u}^l = p_{m,u} t_1 + \omega_{m,u} (L_{m,u})^3 T \quad (12)$$

2.3 系统效用

本文同时考虑了用户的计算速率(在时间块 T 内用户可卸载的总比特数^[7])、能耗和公平性,将系统公平计算效率作为系统效用^[11],它被定义为用户基于 α 公平

效用函数的计算速率与用户的总能耗的比值,也就是:

$$\eta_{CE}^\alpha = \frac{\sum_{m=1}^M \sum_{u=1}^{U_m} U_\alpha(S_{m,u})}{\sum_{m=1}^M \sum_{u=1}^{U_m} E_{m,u}} \quad (13)$$

其中,可以根据不同的场景,通过调整参数 $\alpha (\alpha \geq 0)$ 来控制公平性程度. $U_\alpha(S_{m,u})$ 为用户 $[m, u]$ 基于 α 公平效用函数的计算速率,定义为:

$$U_\alpha(S_{m,u}) = \begin{cases} \ln(S_{m,u}), & \alpha = 1 \\ \frac{1}{1-\alpha} S_{m,u}^{1-\alpha}, & \alpha \geq 0, \alpha \neq 1 \end{cases} \quad (14)$$

其中,当 $\alpha = 0$ 时, $\sum_{m=1}^M \sum_{u=1}^{U_m} U_\alpha(S_{m,u})$ 为用户总的计算速率,此时在用户之间没有考虑公平性;当 $\alpha = 1$ 时, $\sum_{m=1}^M \sum_{u=1}^{U_m} U_\alpha(S_{m,u})$ 为用户计算速率的自然对数之和,此时在用户之间存在比例公平性;当 $\alpha \rightarrow +\infty$, $\sum_{m=1}^M \sum_{u=1}^{U_m} U_\alpha(S_{m,u})$ 为用户之间可实现的最小的计算速率,此时 $\sum_{m=1}^M \sum_{u=1}^{U_m} U_\alpha(S_{m,u}) = \min_{[m,u]} S_{m,u}$, 在用户之间存在绝对的公平性. 随着参数 α 的增大,用户之间的公平性程度逐渐增大. 在本文中仅仅考虑了当 $0 \leq \alpha \leq 1$ 时的场景^[23].

3 问题描述

本文通过优化用户的分簇、NOMA 簇的解码顺序、用户的发送功率、用户的 CPU 处理频率和传输时间,建立了一个系统公平计算效率最大化问题. 其中,针对用户的分簇和 NOMA 簇的解码顺序,本文在下一节中提出了一种高效的启发式算法. 为了便于表述,在本节给出的优化问题 P1 中移除了它们,如下所示:

$$\begin{aligned} \text{P1: } & \max_{p_{m,u}, L_{m,u}, t_1} \eta_{CE}^\alpha \\ \text{s.t. C1: } & t_1 + \frac{\sum_{m=1}^M \sum_{u=1}^{U_m} S_{m,u}^e C_{m,u}}{L_{mcc}} \leq T \\ & \text{C2: } S_{m,u} \geq S_{m,u}^{\min}, \forall m, u \\ & \text{C3: } E_{m,u} \leq E_{m,u}^{\max}, \forall m, u \\ & \text{C4: } 0 \leq p_{m,u} \leq p_{m,u}^{\max}, \forall m, u \\ & \text{C5: } 0 \leq L_{m,u} \leq L_{m,u}^{\max}, \forall m, u \\ & \text{C6: } 0 \leq t_1 \leq T \end{aligned} \quad (15)$$

约束条件 C1 为用户计算任务的时延约束;约束条件 C2 保证了每个用户基本的计算需求;约束条件 C3 确保了每个用户的能耗不超过它的上界;约束条件 C4~C6 为优化变量的边界约束,其中, $p_{m,u}^{\max}$ 和 $L_{m,u}^{\max}$

分别为用户 $[m, u]$ 的最大发送功率和最大 CPU 处理频率.

4 问题求解

显然,所建立的优化问题是一个混合整数非线性规划问题,包含离散变量(用户的分簇和 NOMA 簇的解码顺序)和连续变量(用户的发送功率、CPU 处理频率和传输时间). 由于问题的非凸性,很难找到一个有效的方法来获取最优解. 为了求解这个问题,本文采用了两阶段的优化方式. 在第一个阶段中,提出了一种有效的启发式算法来优化用户的分簇,进而根据分簇的结果来确定 NOMA 簇的解码顺序. 移除它们后,在第二个阶段中,利用 Dinkelbach 方法和连续凸逼近 (Successive Convex Approximation, SCA) 方法将问题转换成凸优化问题,然后迭代求解该凸优化问题来优化连续变量.

4.1 优化用户的分簇和 NOMA 簇的解码顺序

对于上行 NOMA, BS 通过执行 SIC, 信道增益高的用户的信号能够被首先解码,然后从接收到的叠加信号中移除,直到在叠加信号中解码出所有用户的信号. 因此,信道增益高的用户不会对信道增益低的用户产生干扰. 有着最高信道增益的用户能够以它的最大功率发射信号,以提高其所在 NOMA 簇的总速率,从而可以提升 NOMA 簇的计算速率或者降低 NOMA 簇的能耗. 此外,从文献[24]中可以发现,在同一个 NOMA 簇中,不同用户之间信道增益的差异对于实现簇内用户间干扰的最小化是十分重要的. 基于上述分析,本文提出了一种有效的启发式算法来实现用户的分簇. 首先将 K 个用户按其信道增益从大到小进行排序;然后根据排序结果依次给用户分配索引,即:最高信道增益的用户的索引为 1,最低信道增益的用户的索引为 K ;最后依次将索引为 1~ K 的用户分别分配到簇 1~ M 中,也就是,将索引为 1~ M 的用户分别分配给簇 1~ M ,将索引为 $M+1$ ~ $2M$ 的用户分别分配给簇 1~ M ,以此类推,直到所有用户都完成分簇. 为了便于理解,本文也给出了一个例子来展示整个分簇过程,如图 3 所示,其中, $K=8$, $M=3$. 在执行完上述步骤后, K 个用户将与 $[m, u]$ 一一对应.

接下来,根据分簇的结果来确定 NOMA 簇的解码顺序. 由 2.1 小节可知,每个簇中除了第 1 个用户外,剩余用户仍然会对其它簇中的用户造成干扰. 定义每个簇中用户造成的簇间干扰为 $\Phi_m = \sum_{j=1, j \neq m}^M \sum_{k=2}^{U_m} \|w_j^H h_{m,k}\|^2$. 为了尽可能地降低簇间干扰,根据 Φ_m 按降序排列的顺序,重新给每个簇分配簇号, Φ_m 越大,分配的簇号越小,也就是,本文选择先解码 Φ_m 最大的簇,随后解码 Φ_m 第

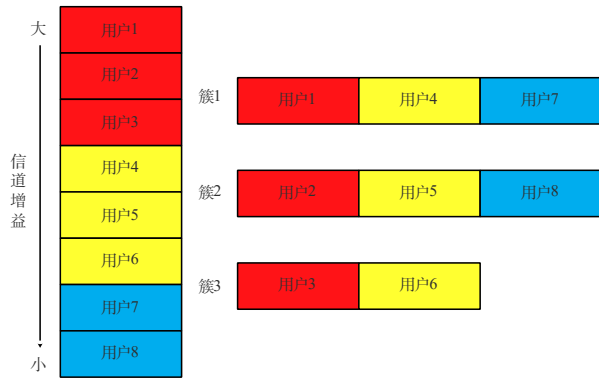


图3 用户分簇的示意图

二大的簇,以此类推,直到解码完所有的簇.

此外,也总结了上述方案的执行流程,如算法1所示.

算法1 用户的分簇和NOMA簇的解码顺序

用户的分簇

1. 根据用户的复合信道向量,计算用户的信道增益,并按信道增益从大到小进行排序;
2. 按信道增益的排序依次给用户分配索引,即:最高信道增益的用户的索引为1,最低信道增益的用户的索引为K;
3. 根据用户的索引进行分簇,具体如下:

For $k=1: \lfloor \frac{K}{M} \rfloor$

IF $Mk \leq K$

将索引编号为 $\{M(k-1)+1, M(k-1)+2, \dots, M(k-1)+M\}$

的用户分别分配到NOMA簇 $\{1, 2, \dots, M\}$;

ELSE

将索引编号为 $\{M(k-1)+1, M(k-1)+2, \dots, K\}$ 的用户分别分配到NOMA簇 $\{1, 2, \dots, K-M(k-1)\}$;

END IF

END FOR

NOMA簇的解码顺序

4. 根据用户的分簇结果计算每个簇中用户造成的簇间干扰 Φ_m ;
5. 将 Φ_m 按降序排列,按照此排列顺序重新给每个簇分配簇号, NOMA簇的解码顺序和此排列顺序一致.

4.2 优化用户的发送功率、CPU处理频率和传输时间

通过4.1小节,可以得到用户的分簇和NOMA簇的解码顺序. 移除它们后,本文所建立的优化问题能够被简化为P1. 考虑到问题P1是一个分式优化问题,受Dinkelbach方法^[25]的启发,可得引理1.

引理1 当优化问题P1取得最优解的时候,当且仅当下列等式成立,也就是:

$$\begin{aligned} Z(\lambda^*) &= \max_{p_{m,u}, L_{m,u}, t_1} \sum_{m=1}^M \sum_{u=1}^{U_m} U_\alpha(S_{m,u}(p_{m,u}, L_{m,u}, t_1)) \\ &\quad - \lambda^* \sum_{m=1}^M \sum_{u=1}^{U_m} E_{m,u}(p_{m,u}, L_{m,u}, t_1) \\ &= \sum_{m=1}^M \sum_{u=1}^{U_m} U_\alpha(S_{m,u}(p_{m,u}^*, L_{m,u}^*, t_1^*)) \\ &\quad - \lambda^* \sum_{m=1}^M \sum_{u=1}^{U_m} E_{m,u}(p_{m,u}^*, L_{m,u}^*, t_1^*) \\ &= 0 \end{aligned} \quad (16)$$

其中, λ^* 为问题P1中系统最大的公平计算效率, $(p_{m,u}^*, L_{m,u}^*, t_1^*)$ 为最优解. 详细证明可以参考文献[25]中3.2.1小节.

根据引理1,可以将分式优化问题P1转换为更易于处理的减法形式的优化问题P2.

$$\begin{aligned} \text{P2: } \max_{p_{m,u}, L_{m,u}, t_1} & \sum_{m=1}^M \sum_{u=1}^{U_m} U_\alpha(S_{m,u}) - \lambda \sum_{m=1}^M \sum_{u=1}^{U_m} E_{m,u} \quad (17) \\ \text{s.t. } & \text{C1} \sim \text{C6} \end{aligned}$$

通过迭代求解问题P2来得到P1的最优解,具体的算法执行流程如算法2所示.

算法2 基于Dinkelbach的迭代算法

输入:系统参数,初始化 $\lambda^{(1)}=0$,最大迭代次数 D ,迭代精度参数 τ

输出:系统的最优解 $p_{m,u}^*, L_{m,u}^*$ 和 t_1^*

1. FOR $d=1:D$

2. 给定 $\lambda=\lambda^{(d)}$,通过求解问题P2,得到其最优解 $p_{m,u}^{(d)*}, L_{m,u}^{(d)*}$ 和 $t_1^{(d)*}$;

3. $Z(\lambda^{(d)}) = \sum_{m=1}^M \sum_{u=1}^{U_m} U_\alpha(S_{m,u}(p_{m,u}^{(d)*}, L_{m,u}^{(d)*}, t_1^{(d)*})) - \lambda^{(d)} \sum_{m=1}^M \sum_{u=1}^{U_m} E_{m,u}(p_{m,u}^{(d)*}, L_{m,u}^{(d)*}, t_1^{(d)*})$

4. IF $Z(\lambda^{(d)}) \leq \tau$

5. 获得问题P1的最优解 $p_{m,u}^*, L_{m,u}^*, t_1^* = p_{m,u}^{(d)*}, L_{m,u}^{(d)*}, t_1^{(d)*}$;

6. BREAK;

7. ELSE

8. $\lambda^{(d+1)} = \frac{\sum_{m=1}^M \sum_{u=1}^{U_m} U_\alpha(S_{m,u}(p_{m,u}^{(d)*}, L_{m,u}^{(d)*}, t_1^{(d)*}))}{\sum_{m=1}^M \sum_{u=1}^{U_m} E_{m,u}(p_{m,u}^{(d)*}, L_{m,u}^{(d)*}, t_1^{(d)*})}$;

9. END IF

10. END FOR

从算法2可以看出,其难点在于在给定 λ 的情况下如何求解问题P2(步骤2). 考虑到P2中用户的发送功率 $p_{m,u}$ 和传输时间 t_1 相互耦合,引入辅助变量 $\xi_{m,u} = p_{m,u} t_1$ 对其进行解耦. 此时,问题P2可以等价转换为P3.

$$\begin{aligned}
 \text{P3: } & \max_{\zeta_{m,u}^*, L_{m,u}^*, t_1} \sum_{m=1}^M \sum_{u=1}^{U_m} U_{\alpha}(S'_{m,u}) - \lambda \sum_{m=1}^M \sum_{u=1}^{U_m} E'_{m,u} \\
 & \text{s.t. C1-1: } t_1 + \frac{\sum_{m=1}^M \sum_{u=1}^{U_m} S_{m,u}^{e'} C_{m,u}}{L_{\text{mec}}} \leq T \quad (18) \\
 & \text{C2-1: } S'_{m,u} \geq S_{m,u}^{\min}, \forall m, u \\
 & \text{C3-1: } E'_{m,u} \leq E_{m,u}^{\max}, \forall m, u \\
 & \text{C4-1: } 0 \leq \zeta_{m,u}^* \leq p_{m,u}^{\max} t_1, \forall m, u \\
 & \text{C5, C6}
 \end{aligned}$$

$$\text{其中, } S'_{m,u} = S_{m,u}^{e'} + S_{m,u}^1 = R'_{m,u} t_1 + \frac{L_{m,u} T}{C_{m,u}}, R'_{m,u} = B \log_2 \left(1 + \frac{\frac{\zeta_{m,u}}{t_1} \|\mathbf{w}_m^H \mathbf{h}_{m,u}\|^2}{\sum_{i=u+1}^{U_m} \frac{\zeta_{m,i}}{t_1} \|\mathbf{w}_m^H \mathbf{h}_{m,i}\|^2 + \sum_{j=m+1}^M \sum_{k=2}^{U_j} \frac{\zeta_{j,k}}{t_1} \|\mathbf{w}_m^H \mathbf{h}_{j,k}\|^2 + \|\mathbf{n}\|^2}} \right)$$

$E'_{m,u} = \zeta_{m,u} + \omega_{m,u} (L_{m,u})^3 T$. 本文根据 α 值的不同, 提出了三种迭代算法来求解问题 P3.

4.2.1 当 α 的取值为 0 时

在这种情况下, 问题 P3 变为

$$\begin{aligned}
 \text{P4: } & \max_{\zeta_{m,u}^*, L_{m,u}^*, t_1} \sum_{m=1}^M \sum_{u=1}^{U_m} S'_{m,u} - \lambda \sum_{m=1}^M \sum_{u=1}^{U_m} E'_{m,u} \quad (19) \\
 & \text{s.t. C1-1} \sim \text{C4-1, C5, C6}
 \end{aligned}$$

问题 P4 是一个 NP-hard 问题^[10,26]. 观察问题 P4 可以看出, 由于 $S_{m,u}^{e'}$ 的非凸性, 导致了 P4 中目标函数和约束条件 C1-1、C2-1 的非凸性. 为了求解这个问题, 将 $S_{m,u}^{e'}$ 写成两个表达式相减的形式:

$$S_{m,u}^{e'} = S_{m,u}^{e1} - S_{m,u}^{e2} \quad (20)$$

其中,

$$\begin{aligned}
 S_{m,u}^{e1} &= B t_1 \log_2 \left(\sum_{i=u}^{U_m} \frac{\zeta_{m,i}}{t_1} \|\mathbf{w}_m^H \mathbf{h}_{m,i}\|^2 + \sum_{j=m+1}^M \sum_{k=2}^{U_j} \frac{\zeta_{j,k}}{t_1} \|\mathbf{w}_m^H \mathbf{h}_{j,k}\|^2 + \|\mathbf{n}\|^2 \right), \\
 S_{m,u}^{e2} &= B t_1 \log_2 \left(\sum_{i=u+1}^{U_m} \frac{\zeta_{m,i}}{t_1} \|\mathbf{w}_m^H \mathbf{h}_{m,i}\|^2 + \sum_{j=m+1}^M \sum_{k=2}^{U_j} \frac{\zeta_{j,k}}{t_1} \|\mathbf{w}_m^H \mathbf{h}_{j,k}\|^2 + \|\mathbf{n}\|^2 \right)
 \end{aligned} \quad (21)$$

可以看出, $S_{m,u}^{e1}$ 和 $S_{m,u}^{e2}$ 分别为 $B \log_2 \left(\sum_{i=u}^{U_m} \zeta_{m,i} \|\mathbf{w}_m^H \mathbf{h}_{m,i}\|^2 + \sum_{j=m+1}^M \sum_{k=2}^{U_j} \zeta_{j,k} \|\mathbf{w}_m^H \mathbf{h}_{j,k}\|^2 + \|\mathbf{n}\|^2 \right)$

和 $B \log_2 \left(\sum_{i=u+1}^{U_m} \zeta_{m,i} \|\mathbf{w}_m^H \mathbf{h}_{m,i}\|^2 + \sum_{j=m+1}^M \sum_{k=2}^{U_j} \zeta_{j,k} \|\mathbf{w}_m^H \mathbf{h}_{j,k}\|^2 + \|\mathbf{n}\|^2 \right)$ 的透视函数. 由于透视运算的保凹/凸性, 以及这两个表达式都是凹的, 因此, $S_{m,u}^{e1}$ 和 $S_{m,u}^{e2}$ 也是凹的.

利用 SCA 的方法, 分别使用 $S_{m,u}^{e1}$ 和 $S_{m,u}^{e2}$ 的一阶泰勒展开式近似代替它们本身, 可以得到 $S_{m,u}^{e'}$ 的一个上界和一个下界, 也就是:

$$\begin{aligned}
 S_{m,u}^{e',\text{up}} &= S_{m,u}^{e1}(\mathbf{x}^{(d)}) + \nabla S_{m,u}^{e1}(\mathbf{x}^{(d)}) (\mathbf{x} - \mathbf{x}^{(d)}) - S_{m,u}^{e2}, \\
 S_{m,u}^{e',\text{low}} &= S_{m,u}^{e1} - \left(S_{m,u}^{e2}(\mathbf{x}^{(d)}) + \nabla S_{m,u}^{e2}(\mathbf{x}^{(d)}) (\mathbf{x} - \mathbf{x}^{(d)}) \right) \quad (22)
 \end{aligned}$$

其中, $\mathbf{x} = [\zeta_{1,1}^*, \zeta_{1,2}^*, \dots, \zeta_{M,U_M}^*, t_1]^T$, $\mathbf{x}^{(d)} = [\zeta_{1,1}^{(d)}, \zeta_{1,2}^{(d)}, \dots, \zeta_{M,U_M}^{(d)}, t_1^{(d)}]^T$ 为 \mathbf{x} 在第 d 次迭代时的局部点, ∇ 为偏导运算, $\nabla S_{m,u}^{e1}(\mathbf{x}^{(d)})$ 和 $\nabla S_{m,u}^{e2}(\mathbf{x}^{(d)})$ 为两个 $(K+1)$ 维行向量.

用 $S_{m,u}^{e',\text{up}}$ 替换约束条件 C1-1 中的 $S_{m,u}^{e'}$, 用 $S_{m,u}^{e',\text{low}}$ 替换目标函数和约束条件 C2-1 中的 $S_{m,u}^{e'}$, 可以重写问题 P4 为

$$\begin{aligned}
 \text{P5: } & \max_{\zeta_{m,u}^*, L_{m,u}^*, t_1} \sum_{m=1}^M \sum_{u=1}^{U_m} (S_{m,u}^{e',\text{low}} + S_{m,u}^1) - \lambda \sum_{m=1}^M \sum_{u=1}^{U_m} E'_{m,u} \\
 & \text{s.t. C1-2: } t_1 + \frac{\sum_{m=1}^M \sum_{u=1}^{U_m} S_{m,u}^{e',\text{up}} C_{m,u}}{L_{\text{mec}}} \leq T \quad (23) \\
 & \text{C2-2: } S_{m,u}^{e',\text{low}} + S_{m,u}^1 \geq S_{m,u}^{\min}, \forall m, u \\
 & \text{C3-1, C4-1, C5, C6}
 \end{aligned}$$

引理 2 问题 P5 是一个凸优化问题.

证明 由式 (22) 可以判断出 $S_{m,u}^{e',\text{up}}$ 是一个凸表达式, $S_{m,u}^{e',\text{low}}$ 是一个凹表达式. 此外, 考虑到表达式 $\frac{L_{m,u} T}{C_{m,u}}$ 关于 $L_{m,u}$ 是线性的、表达式 $(L_{m,u})^3$ 关于 $L_{m,u}$ 是凸的, 可以判断出问题 P5 的目标函数是凹的, 它的约束条件所形成的集合是凸集. 因此, 问题 P5 是一个凸优化问题. 证毕.

通过内点法可以有效地对问题 P5 进行求解^[27]. 通过迭代求解问题 P5 可以得到问题 P4 的解. 具体的算法执行流程如算法 3 所示.

4.2.2 当 α 的取值为 0 到 1 之间时

在这种情况下, 问题 P3 变为:

$$\begin{aligned}
 \text{P6: } & \max_{\zeta_{m,u}^*, L_{m,u}^*, t_1} \sum_{m=1}^M \sum_{u=1}^{U_m} \frac{1}{1-\alpha} (S'_{m,u})^{1-\alpha} - \lambda \sum_{m=1}^M \sum_{u=1}^{U_m} E'_{m,u} \quad (24) \\
 & \text{s.t. C1-1} \sim \text{C4-1, C5, C6}
 \end{aligned}$$

问题 P6 是一个 NP-hard 问题^[28,29]. 为了求解这个问题, 将 $S'_{m,u}$ 写成两个凹函数之差的形式:

算法 3 基于 SCA 的迭代算法 ($\alpha=0$)

输入: 系统参数, 初始化 $\zeta_{m,u}^{(1)}, L_{m,u}^{(1)}$ 和 $t_1^{(1)}$, 最大迭代次数 D_1 , 迭代精度参数 τ_1

输出: 系统的最优解 $\zeta_{m,u}^*, L_{m,u}^*$ 和 t_1^*

1. FOR $d_1 = 1 : D_1$
2. 通过内点法对问题 P5 进行求解, 可以获得其最优解 $\zeta_{m,u}^{(d_1)*}, L_{m,u}^{(d_1)*}, t_1^{(d_1)*}$ 和目标值 $\text{obj}^{(d_1)*}$;
3. IF $|\text{obj}^{(d_1)*} - \text{obj}^{(d_1-1)*}| \leq \tau_1$
4. 获得问题 P4 的最优解 $\zeta_{m,u}^*, L_{m,u}^*, t_1^* = \zeta_{m,u}^{(d_1)*}, L_{m,u}^{(d_1)*}, t_1^{(d_1)*}$;
5. BREAK;
6. ELSE
7. $\zeta_{m,u}^{(d_1+1)}, L_{m,u}^{(d_1+1)}, t_1^{(d_1+1)} = \zeta_{m,u}^{(d_1)*}, L_{m,u}^{(d_1)*}, t_1^{(d_1)*}$;
8. END IF
9. END FOR

$$S'_{m,u} = \left(S_{m,u}^{e1} + \frac{L_{m,u}T}{C_{m,u}} \right) - S_{m,u}^{e2} \quad (25)$$

其中, $S_{m,u}^{e1}$ 和 $S_{m,u}^{e2}$ 详见式(21).

因此, $\frac{1}{1-\alpha} (S'_{m,u})^{1-\alpha}$ 能够被表示为:

$$\frac{1}{1-\alpha} (S'_{m,u})^{1-\alpha} = \frac{1}{1-\alpha} \left(\left(S_{m,u}^{e1} + \frac{L_{m,u}T}{C_{m,u}} \right) - S_{m,u}^{e2} \right)^{1-\alpha} = r_{m,u} - v_{m,u} \quad (26)$$

其中, $r_{m,u} = \frac{1}{1-\alpha} \left(\left(S_{m,u}^{e1} + \frac{L_{m,u}T}{C_{m,u}} \right) - S_{m,u}^{e2} \right)^{1-\alpha} + \varphi_{m,u} S_{m,u}^{e2}$ 和 $v_{m,u} = \varphi_{m,u} S_{m,u}^{e2}$ ($\varphi_{m,u} \geq Q_{m,u}^{-\alpha}, Q_{m,u} > 0$). 由于 $S_{m,u}^{e2}$ 是凹的, $\varphi_{m,u}$ 为正数, 因此, $v_{m,u}$ 是凹的. 此外, 根据文献[29], 可以发现 $r_{m,u}$ 能够被表示为一系列凹函数的逐点下确界, 因此, $r_{m,u}$ 也是凹的. 详细的推导过程可以参考文献[29]中第4小节.

问题 P6 进一步改写为

$$\text{P7: } \max_{\zeta_{m,u}, L_{m,u}, t_1} \sum_{m=1}^M \sum_{u=1}^{U_m} (r_{m,u} - v_{m,u}) - \lambda \sum_{m=1}^M \sum_{u=1}^{U_m} E'_{m,u} \quad (27)$$

s.t. C1-1 ~ C4-1, C5, C6

通过观察可以发现, 由于目标函数中表达式 $(r_{m,u} - v_{m,u})$ 和约束条件 C1-1、C2-1 中表达式 $S_{m,u}^{e'}$ 是非凸的, 问题 P7 仍然是非凸的. 为了能够有效地求解它, 本文采用了 SCA 的方法. 针对表达式 $(r_{m,u} - v_{m,u})$, 使用 $v_{m,u}$ 的一阶泰勒展开式来近似代替它本身, 目标函数可以转换成一个凹函数. 针对约束条件 C1-1 和 C2-1, 与 4.2.1 小节相同, 用 $S_{m,u}^{e',up}$ 替换约束条件 C1-1 中的 $S_{m,u}^{e'}$, 用 $S_{m,u}^{e',low}$ 替换约束条件 C2-1 中的 $S_{m,u}^{e'}$. 经过上述替换后, 问题 P7 能够转换成如下凸优化问题:

$$\text{P8: } \max_{\zeta_{m,u}, L_{m,u}, t_1} \sum_{m=1}^M \sum_{u=1}^{U_m} (r_{m,u} - v_{m,u}^{up}) - \lambda \sum_{m=1}^M \sum_{u=1}^{U_m} E'_{m,u} \quad (28)$$

s.t. C1-2, C2-2, C3-1, C4-1, C5, C6

其中, 类似于式(22), $v_{m,u}^{up}$ 为 $v_{m,u}$ 的一阶泰勒展开式.

采用内点法对问题 P8 进行求解. 通过迭代求解问题 P8 可以得到问题 P6 的解. 具体的算法执行流程类似于算法 3.

4.2.3 当 α 的取值为 1 时

在这种情况下, 问题 P3 变为

$$\text{P9: } \max_{\zeta_{m,u}, L_{m,u}, t_1} \sum_{m=1}^M \sum_{u=1}^{U_m} \ln(S'_{m,u}) - \lambda \sum_{m=1}^M \sum_{u=1}^{U_m} E'_{m,u} \quad (29)$$

s.t. C1-1 ~ C4-1, C5, C6

考虑到目标函数中 $\ln(S'_{m,u})$ 的非凸性, 引入辅助变量 $b_{m,u}$, 满足:

$$S'_{m,u} \geq e^{b_{m,u}} \geq S_{m,u}^{\min} \quad (30)$$

同时借助于式(25), 问题 P9 可以改写为

$$\text{P10: } \max_{\zeta_{m,u}, L_{m,u}, t_1, b_{m,u}} \sum_{m=1}^M \sum_{u=1}^{U_m} b_{m,u} - \lambda \sum_{m=1}^M \sum_{u=1}^{U_m} E'_{m,u}$$

s.t. C2-3: $b_{m,u} \geq \ln(S_{m,u}^{\min})$ (31)

C2-4: $e^{b_{m,u}} \leq S'_{m,u}$
C1-1, C3-1, C4-1, C5, C6

此时, 消除了目标函数的非凸性. 考虑到问题 P10 和 P4 的相似性, 可以采用 SCA 的方法对问题 P10 进行求解. 因此, 针对约束条件 C1-1 和 C2-4, 与 4.2.1 小节相同, 用 $S_{m,u}^{e',up}$ 替换约束条件 C1-1 中的 $S_{m,u}^{e'}$, 用 $S_{m,u}^{e',low}$ 替换约束条件 C2-4 中的 $S_{m,u}^{e'}$. 经过上述替换后, 问题 P10 能够转换成如下凸优化问题.

$$\text{P11: } \max_{\zeta_{m,u}, L_{m,u}, t_1, b_{m,u}} \sum_{m=1}^M \sum_{u=1}^{U_m} b_{m,u} - \lambda \sum_{m=1}^M \sum_{u=1}^{U_m} E'_{m,u} \quad (32)$$

s.t. C2-5: $e^{b_{m,u}} \leq S_{m,u}^{e',low} + S_{m,u}^1$
C1-2, C2-3, C3-1, C4-1, C5, C6

针对问题 P11, 采用内点法进行求解. 进一步通过迭代求解问题 P11 可以得到问题 P9 的解. 具体的算法执行流程类似于算法 3.

5 仿真分析

本节通过仿真评估了所提方案的性能, 并且与 4 种基线方案进行了对比. 基线方案如下所示.

(1) 基线方案 1: 不考虑 NOMA 簇的解码顺序, 此时, 用户 $[m, u]$ 的可实现传输速率为 $R_{m,u} =$

$$B \log_2 \left[1 + \frac{p_{m,u} \|\mathbf{w}_m^H \mathbf{h}_{m,u}\|^2}{\sum_{i=1}^{U_m} p_{m,i} \|\mathbf{w}_m^H \mathbf{h}_{m,i}\|^2 + \sum_{j=1, j \neq m}^M \sum_{k=2}^{U_j} p_{j,k} \|\mathbf{w}_m^H \mathbf{h}_{j,k}\|^2 + \|\mathbf{n}\|^2} \right],$$

其余部分与所提出的方案相同。

(2)基线方案 2:用户采用随机分簇的方法进行分簇,其余部分与所提出的方案相同。

(3)基线方案 3:在每个簇中,用户采用等自由度 OMA 方式进行接入^[18],此时,用户 $[m,u]$ 的可实现传输速

$$率为 R_{m,u} = \frac{B}{U_m} \log_2 \left(1 + \frac{p_{m,u} \| \mathbf{w}_m^H \mathbf{h}_{m,u} \|^2}{\sum_{j=1, j \neq m}^M \sum_{k=2}^{U_j} p_{j,k} \| \mathbf{w}_m^H \mathbf{h}_{j,k} \|^2 + \frac{\| \mathbf{n} \|^2}{U_m}} \right),$$

其余部分与所提出的方案相同;

(4)基线方案 4:不考虑任务卸载,用户仅仅通过本地计算模式处理计算任务。

在仿真中,用户均匀分布在距离 BS 50 m 到 500 m 的范围内,大尺度衰落采用了奥村城市传播模型^[30],小尺度衰落服从复高斯分布,其它仿真参数如表 2 所示。在没有特殊说明的情况下,表 2 中的参数将作为默认值。

表 2 仿真参数

参数	值
用户的数量 K	12
BS 的天线数 M	4
系统带宽 B	10 MHz
公平性参数 α	0.5
用户的最大发送功率 $p_{m,u}^{\max}$	23 dBm
用户的最大 CPU 处理频率 $L_{m,u}^{\max}$	1 GHz
每比特需要的 CPU 周期数 $C_{m,u}$	1 000
有效转换电容 $\omega_{m,u}$	10^{-28}
用户计算需求 $S_{m,u}^{\min}$	4×10^5 bits
用户最大可用能量 $E_{m,u}^{\max}$	0.5 J
MEC 服务器的 CPU 处理频率 L_{mec}	12 GHz
用户时延需求 T	1 s
噪声功率	5×10^{-15} W
迭代精度 ϵ/τ_1	0.001/0.01

所提出的方案在不同的用户数量 K 和不同的公平性参数 α 下的收敛性能如图 4 所示,其中纵坐标为方案收敛时的迭代次数。根据第 4 节,总的求解过程主要包含外层迭代和内层迭代,此处的迭代次数为方案收敛时所有内层迭代次数之和。从图 4 中可以看出,随着公平性参数 α 从 0 增加到 1,在不同的用户数量下,总的迭代次数的趋势均逐渐下降。这是由于随着公平性参数 α 的增加,优化问题的目标值将会变小,在一定的迭代精度下,收敛时所需的迭代次数也会降低。此外,也可以看出,在不同的用户数量 K 和不同的公平性参数 α 下,方案收敛时的迭代次数均小于 95。因此,所提出的方案具有较好的收敛性。

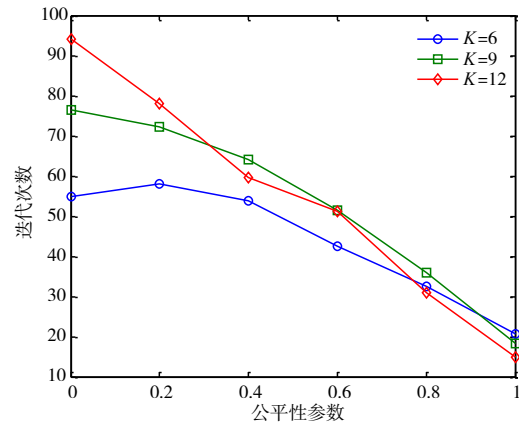


图 4 所提方案的收敛性能

系统效用随公平性参数 α 的变化曲线如图 5 所示。从图 5 中可以看出,对于提出的方案和基线方案,随着公平性参数 α 从 0 增加到 1,系统效用均逐渐降低,并逐渐趋于平缓。这是由于随着公平性参数 α 的增加,系统将会执行更加公平的资源分配。在此过程中,系统的总计算速率和总能耗会逐渐降低,从而会导致系统效用也逐渐降低,并最终趋于平缓以维持用户基本的计算需求。同时,也表明了在此过程中,用户基于 α 公平效用函数的计算速率的减少要快于其能耗的减少。从图 5 中还可以看出,在不同的公平性参数 α 下,所提方案的系统效用始终高于 4 种基线方案。这是由于所提出的方案充分利用了信道增益高的用户给 NOMA 簇带来速率提升的能力,以及提出的方案同时利用了基于 NOMA 的任务卸载模式和本地计算模式,能够更加高效地利用有限的系统资源。此外,这也凸显了考虑 NOMA 簇解码顺序的有效性。

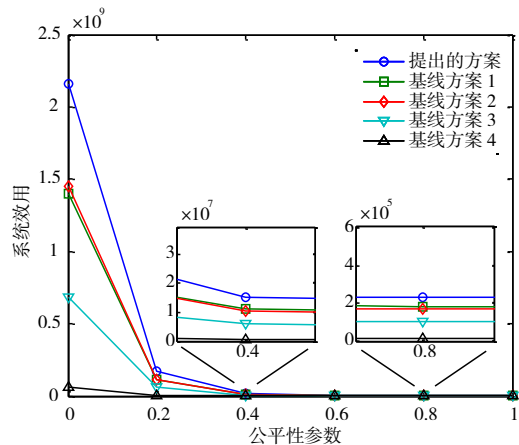


图 5 系统效用随公平性参数的变化曲线

系统效用随用户计算需求 $S_{m,u}^{\min}$ 的变化曲线如图 6 所示。从图 6 中可以看出,对于提出的方案和基线方案,随着用户计算需求 $S_{m,u}^{\min}$ 从 1×10^5 bits 增加到 7×10^5 bits,

系统的总效用均逐渐降低. 这是由于随着用户计算需求 $S_{m,u}^{\min}$ 的增加, 用户需要去增大它的发射功率和分配更多的计算资源来完成更多的计算任务, 以满足用户增加的计算需求, 这将会导致系统总能耗的升高和系统效用的降低. 这也表明了在此过程中, 用户能耗的增长要快于其基于 α 公平效用函数的计算速率的增长. 此外, 还可以发现, 提出的方案相比于 4 种基线方案始终能够取得更优的性能. 当用户计算需求 $S_{m,u}^{\min}$ 为 3×10^5 bits 时, 所提方案的系统效用相比于基线方案 1/2/3 分别提升了 22.52%、33.05%、82.12%, 且为基线方案 4 的 16 倍.

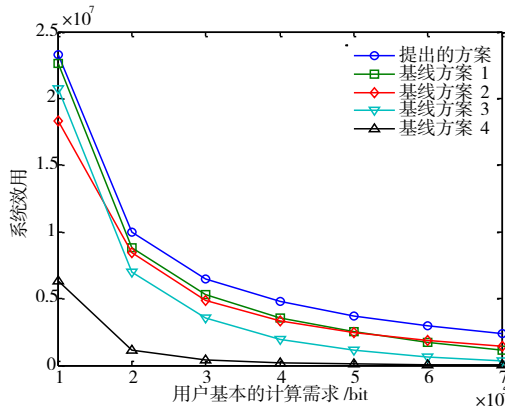


图6 系统效用随用户计算需求的变化曲线

系统效用随时间块长度 T 的变化曲线如图 7 所示. 从图 7 中可以看出, 对于提出的方案和基线方案, 随着时间块长度 T 从 0.6 s 增加到 1.2 s, 系统的总效用均逐渐升高. 这是由于随着时间块长度 T 的增加, 用户满足基本的计算需求 $S_{m,u}^{\min}$ 将会消耗更少的能量, 用户可以消耗更少的能量来完成更多的计算任务, 这会导致系统的总计算速率逐渐升高, 总能耗逐渐降低, 从而系统效用逐渐升高. 此外, 与 4 种基线方案进行对比, 可以看出, 提出的方案能够带来显著的性能提升. 当时间块长度 T 等于 1.1 s 时, 所提方案的系统效用相比于基线方案 1/2/3 分别提升了 28.92%、39.39%、116.45%, 且为基线方案 4 的 21 倍.

系统效用随 MEC 服务器 CPU 处理频率 L_{mcc} 的变化曲线如图 8 所示. 从图 8 中可以看出, 对于提出的方案和基线方案, 随着 MEC 服务器 CPU 处理频率 L_{mcc} 从 6 GHz 增加到 24 GHz, 提出的方案和基线方案 1/2/3 的系统效用均逐渐升高, 基线方案 4 保持不变. 这是由于随着 MEC 服务器 CPU 处理频率 L_{mcc} 的增加, 用户的计算任务在 MEC 服务器中的处理时间将会缩短, 在一定的时间块长度 T 下, 用户可用于卸载的时间将会增加, 此时, 用户可以消耗更少的能量来卸载更多的计算任务到 MEC 服务器中进行处理. 因此, 提出的方案和

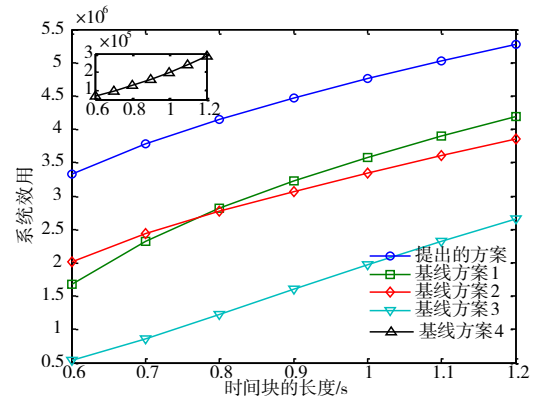


图7 系统效用随时间块长度的变化曲线

基线方案 1/2/3 的系统效用均逐渐升高. 由于基线方案 4 不考虑任务卸载, 其不受 L_{mcc} 的影响, 所以始终保持不变. 此外, 从图 8 中还可以看出, 所提方案始终优于 4 种基线方案. 当 MEC 服务器 CPU 处理频率 L_{mcc} 等于 24 GHz 时, 所提方案的系统效用相比于基线方案 1/2/3 分别提升了 30.43%、42.84%、130.56%, 且为基线方案 4 的 27.6 倍.

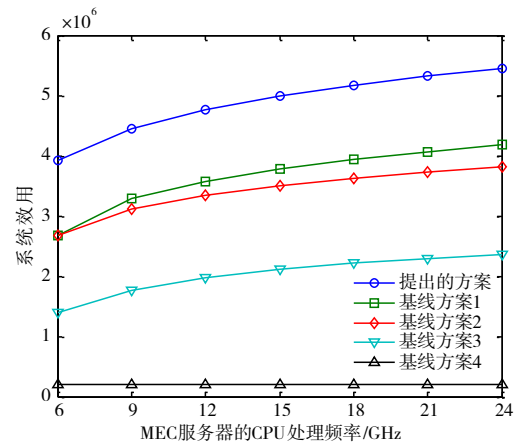


图8 系统效用随MEC服务器CPU处理频率的变化曲线

6 结束语

本文研究了多个 NOMA 簇场景下的公平计算卸载策略, 提出了一个基于上行多天线 NOMA 的部分卸载框架. 利用多天线技术, 设计了一种低复杂度的迫零波束赋形方案来部分地消除簇间干扰, 以均衡系统性能和设计复杂度. 在满足用户计算需求和时延等约束条件下, 建立了一个系统公平计算效率最大化问题. 考虑到问题的非凸性, 且同时包含连续变量和离散变量, 采用了一种基于两阶段的优化方式对问题进行求解, 分别对离散变量和连续变量进行优化. 仿真结果表明了所提出的方案具有较好的收敛性, 且在提升系统效用方面, 始终优于几种基线方案, 凸显了所提方案的有

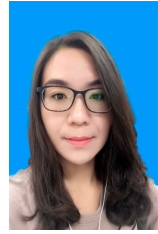
效性.

参考文献

- [1] HUANG X H, ZENG S, LI D D, et al. Fair computation efficiency scheduling in NOMA-aided mobile edge computing[J]. *IEEE Wireless Communications Letters*, 2020, 9(11): 1812-1816.
- [2] ZHOU F H, WU Y P, HU R Q, et al. Computation rate maximization in UAV-enabled wireless-powered mobile-edge computing systems[J]. *IEEE Journal on Selected Areas in Communications*, 2018, 36(9): 1927-1941.
- [3] HE W, ZHANG Y Z, HUANG Y H, et al. Latency minimization for full-duplex mobile-edge computing system[C]// 2019 IEEE International Conference on Communications (ICC). Shanghai: IEEE, 2019: 1-6.
- [4] ZHANG T K, XU Y, LOO J, et al. Joint computation and communication design for UAV-assisted mobile edge computing in IoT[J]. *IEEE Transactions on Industrial Informatics*, 2020, 16(8): 5505-5516.
- [5] YU Z, GONG Y M, GONG S M, et al. Joint task offloading and resource allocation in UAV-enabled mobile edge computing[J]. *IEEE Internet of Things Journal*, 2020, 7(4): 3147-3159.
- [6] 史清江, 洪明毅, 罗智泉. 面向 5G/B5G 通信的智能无线资源管理技术[J]. *中国科学基金*, 2020, 34(2): 142-149.
SHI Q J, HONG M Y, LUO Z Q. Intelligent radio resource management for 5G/B5G communications[J]. *Science Foundation of China*, 2020, 34(2): 142-149. (in Chinese)
- [7] ZENG M, DU R, FODOR V, et al. Computation rate maximization for wireless powered mobile edge computing with NOMA[C]// 2019 IEEE 20th International Symposium on "A World of Wireless, Mobile and Multimedia Networks"(WoWMoM). Washington: IEEE, 2019: 1-9.
- [8] ZENG M, 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.
- [9] QIAN L P, SHI B H, WU Y, et al. NOMA-enabled mobile edge computing for Internet of Things via joint communication and computation resource allocations[J]. *IEEE Internet of Things Journal*, 2020, 7(1): 718-733.
- [10] WANG Q, ZHOU F H. Fair resource allocation in an MEC-enabled ultra-dense IoT network with NOMA[C]// 2019 IEEE International Conference on Communications Workshops. Shanghai: IEEE, 2019: 1-6.
- [11] 胡晗, 鲍楠, 凌章, 等. 基于 NOMA 的移动边缘计算系统公平能效调度算法[J]. *电子与信息学报*, 2021, 43(12): 3563-3570.
- [12] HU H, BAO N, LING Z, et al. Fair energy efficiency scheduling in NOMA-based mobile edge computing[J]. *Journal of Electronics & Information Technology*, 2021, 43(12): 3563-3570. (in Chinese)
- [13] YANG Z H, PAN C H, HOU J C, 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.
- [14] KIANI A, ANSARI N. Edge computing aware NOMA for 5G networks[J]. *IEEE Internet of Things Journal*, 2018, 5(2): 1299-1306.
- [15] BAIDAS M W. Resource allocation for offloading-efficiency maximization in clustered NOMA-enabled mobile edge computing networks[J]. *Computer Networks*, 2021, 189: 107919.
- [16] LI Q, MA X, ZHOU A, et al. User-oriented edge node grouping in mobile edge computing[J]. *IEEE Transactions on Mobile Computing*, 2023, 22(6): 3691-3705.
- [17] KIM Y B, LIM S, KIM H, et al. Non-orthogonal multiple access in a downlink multiuser beamforming system[C]// MILCOM 2013 - 2013 IEEE Military Communications Conference. San Diego: IEEE, 2013: 1278-1283.
- [18] ZENG M, HAO W M, DOBRE O A, et al. Energy-efficient power allocation in uplink mmWave massive MIMO with NOMA[J]. *IEEE Transactions on Vehicular Technology*, 2019, 68(3): 3000-3004.
- [19] ZENG M, YADAV A, DOBRE O A, et al. Energy-efficient joint user-RB association and power allocation for uplink hybrid NOMA-OMA[J]. *IEEE Internet of Things Journal*, 2019, 6(3): 5119-5131.
- [20] KABIR M T, MASOUIROS C. A scalable energy vs. latency trade-off in full-duplex mobile edge computing systems[J]. *IEEE Transactions on Communications*, 2019, 67(8): 5848-5861.
- [21] SUN H J, ZHOU F H, 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.
- [22] 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.
- [23] 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.

- [23] CELIK A, TSAI M C, RADAYDEH R M, et al. Distributed user clustering and resource allocation for imperfect NOMA in heterogeneous networks[J]. IEEE Transactions on Communications, 2019, 67(10): 7211-7227.
- [24] ALI M S, TABASSUM H, HOSSAIN E. Dynamic user clustering and power allocation for uplink and downlink non-orthogonal multiple access (NOMA) systems[J]. IEEE Access, 2016, 4: 6325-6343.
- [25] ZAPPONE A, JORSWIECK E. Energy efficiency in wireless networks via fractional programming theory[J]. Foundations and Trends® in Communications and Information Theory, 2015, 11(3/4): 185-396.
- [26] LUO Z Q, ZHANG S Z. Dynamic spectrum management: Complexity and duality[J]. IEEE Journal of Selected Topics in Signal Processing, 2008, 2(1): 57-73.
- [27] BOYD S, VANDENBERGHE L. Convex Optimization [M]. Cambridge: Cambridge University Press, 2004.
- [28] PHAM Q V, 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.
- [29] GUO C T, ZHANG Y, SHENG M, et al. α -fair power allocation in spectrum-sharing networks[J]. IEEE Transactions on Vehicular Technology, 2016, 65(5): 3771-3777.
- [30] SONG Z Y, NI Q, NAVAIE K, et al. On the spectral-energy efficiency and rate fairness tradeoff in relay-aided cooperative OFDMA systems[J]. IEEE Transactions on Wireless Communications, 2016, 15(9): 6342-6355.



李丹丹 女, 1987年3月出生于河南省平顶山市. 现为北京邮电大学计算机学院(国家示范性软件学院)副教授. 主要研究方向为下一代网络技术、网络安全、密码学等.

E-mail: dandl@bupt.edu.cn



于维军 男, 1977年7月出生于山东省威海市. 现为威海市高新区党政办公室高级工程师. 主要研究方向为政务大数据的应用、网络安全等.

E-mail: whgqyj@163.com

作者简介



曾胜 男, 1994年10月出生于湖北省洪湖市. 现为北京邮电大学计算机学院(国家示范性软件学院)在读博士研究生. 主要研究方向为移动边缘计算、5G通信、深度强化学习等. 中国电子学会会员编号: E190065881M.

E-mail: zengsheng@bupt.edu.cn



黄小红(通讯作者) 女, 1978年1月出生于广东省佛山市. 现为北京邮电大学计算机学院(国家示范性软件学院)教授. 主要研究方向为移动边缘计算、下一代网络技术、网络安全等.

E-mail: huangxh@bupt.edu.cn