

面向数据压缩的NOMA-MEC系统能耗最小化研究

施丽琴 刘璇* 卢光跃

(西安邮电大学陕西省信息通信网络及安全重点实验室 西安 710121)

摘要: 该文研究基于数据压缩的非正交多址-移动边缘计算(NOMA-MEC)系统中系统能耗最小化问题。考虑到部分压缩与卸载方案和基站端计算能力有限等条件,通过联合优化各用户的任务压缩和卸载比例、发射功率以及任务压缩时间等变量,建立一个系统能耗最小化优化问题。为了求解该问题,首先推导出各用户最佳发射功率的闭式表达式。接着利用连续凸逼近(SCA)方法对原问题的非凸约束进行近似,然后提出一个基于SCA的高效迭代算法来求解原问题,从而得到该系统的最佳资源分配方案。最后借助于计算机仿真对所提出方案的性能优势进行验证,仿真结果表明相比于其他基准方案,该文所提方案能有效降低系统能耗。

关键词: 部分压缩; 资源分配; 能耗最小化; NOMA

中图分类号: TN926

文献标识码: A

文章编号: 1009-5896(2024)07-2888-10

DOI: 10.11999/JEIT231033

Research on Energy Consumption Minimization for a Data Compression Based NOMA-MEC System

SHI Liqin LIU Xuan LU Guangyue

(Shaanxi Key Laboratory of Information Communication Network and Security, Xi'an University of Posts and Telecommunications Xi'an 710121, China)

Abstract: The system energy consumption minimization problem is studied for a data compression based Non-Orthogonal Multiple Access-Mobile Edge Computing (NOMA-MEC) system. Considering the partial compression and offloading schemes and the limited computation capacity at the base station, a system energy consumption minimization optimization problem is formulated by jointly optimizing the users' data compression and offloading ratios, transmit power, data compression time, etc. In order to solve this problem, closed-form expression of each user's optimal transmit power is firstly derived. Then the Successive Convex Approximation (SCA) method is used to approximate the non-convex constraints of the formulated problem, and An SCA based efficient iterative algorithm is proposed to solve the formulated problem, obtaining the optimal resource allocation scheme of the system. Finally, the simulation results verify the advantages of the proposed scheme via computer simulations and show that compared with other benchmark schemes, the proposed scheme can effectively reduce the system energy consumption.

Key words: Partial compression; Resource allocation; Energy consumption minimization; NOMA

1 引言

为了实现智能化服务,越来越多的移动设备连入网络,并需实时处理繁多且复杂的数据以支持多样化的移动业务。由于体积及制造成本的限制,移动设备所配置芯片的计算能力十分有限,从而难以在给定时间内完成高密度的数据计算并返回计算结

果。另外,移动设备进行高密度数据运算将消耗自身大量的能量,这加剧了移动设备能量短缺问题。移动边缘计算(Mobile Edge Computing, MEC)被认为是解决上述问题的一个有效手段^[1-5],其核心思想是移动设备将其部分或全部需要计算的任务数据卸载至离其较近的且具有更高计算能力的设备,如基站、边缘服务器等,进行计算,再将计算结果返回至移动设备,从而降低了任务计算时延与移动设备的计算能耗^[3]。

截止目前,MEC已获得了学术界与工业界的广泛研究^[4,5]。在这些研究中,如何合理分配任务计算与卸载所涉及资源,设计出适当的计算卸载

收稿日期: 2023-09-19; 改回日期: 2024-03-15; 网络出版: 2024-04-02

*通信作者: 刘璇 liuxuan202309@126.com

基金项目: 国家自然科学基金(62301421)

Foundation Item: The National Natural Science Foundation of China (62301421)

策略以提高任务卸载效率是实现MEC的关键。针对MEC网络,文献[4]通过优化用户的计算频率等来最小化系统时延。文献[5]设计出一种具有瞬时负荷计费机制的聚合博弈策略来最小化系统时延。在上述的研究中,用户均采用正交多址接入(Orthogonal Multiple Access, OMA)技术来进行任务卸载。相比于OMA技术,非正交多址接入(Non-Orthogonal Multiple-Access, NOMA)技术允许多个用户共享频谱和时间资源,从而取得更高的频谱效率与能量效率^[6,7]。因此,学者们将NOMA技术用于MEC网络中的用户任务卸载,并依次以系统时延和系统能效为优化目标研究最佳计算卸载策略设计^[6,7]。具体来说,文献[6]考虑了部分卸载方案,通过优化任务卸载比例等使得系统时延最小化。文献[7]将射频无线能量传输技术和NOMA-MEC网络相结合,通过联合优化时隙划分、发射功率等来最大化系统能效。

然而,随着移动用户数量及其计算任务的不断增加,这使得如何在给定的时间内利用有限的资源完成任务卸载成为实现未来MEC的重要挑战。为了解决这一挑战,部分学者将数据压缩融入到现有的MEC网络中,构建基于数据压缩的MEC网络。通过数据压缩,需要卸载的任务大小将大大降低,从而降低任务卸载能耗。然而数据压缩和解压缩也需消耗大量的时间和设备能量。因此,如何权衡任务压缩、卸载与计算的能耗是基于数据压缩的MEC网络的研究重点。目前,仅文献[8]和文献[9]将数据压缩引入到MEC网络,并依次以系统时延和系统能耗为优化目标设计了最佳资源分配方案。具体地,文献[8]对需要卸载的任务在何处进行压缩和解压缩进行了讨论,并设计出3种模型。然后针对这3种模型通过联合优化数据压缩的压缩比等来最小化系统时延。文献[9]中用户采用部分压缩与卸载策略,通过联合优化各用户的任务压缩卸载比例等来最小化系统能耗。然而上述基于数据压缩的MEC网络的研究工作仍存在以下不足:

(1) 文献[8]和文献[9]中用户均采用OMA方式进行任务卸载。相比于OMA技术,NOMA技术能取得更高的谱效和能效,而目前仍未有研究将NOMA技术与基于数据压缩的MEC网络相结合,并研究该网络的资源分配方法。

(2) 文献[8]中所有用户均采用全部压缩和卸载来处理计算任务,而全部压缩及卸载不能充分利用各用户自身的计算能力,进而无法通过权衡本地计算与任务压缩、卸载、解压缩、MEC计算等来以进一步降低整个系统的能耗。

(3) 文献[9]中假设基站上MEC服务器的计算能力是有限的,因此忽略了基站端任务解压缩和任务计算的时延。而在实际应用中,基站虽然拥有比用户更强大的计算能力,但面对庞大且复杂的任务,其任务解压缩和计算仍需要消耗大量的时间和能量,这对于整个系统来说均是不可忽视的。

受此启发,本文首次将数据压缩与NOMA-MEC网络相结合,构建了基于数据压缩的NOMA-MEC网络,并通过权衡该网络中任务压缩、卸载与计算资源,设计面向系统能耗最小的资源分配方案。具体而言,考虑到部分压缩与卸载方案及基站有限计算能力等条件,本文通过联合优化用户的任务压缩和卸载时间、发射功率、任务的压缩及卸载比例等变量,建立了一个系统能耗最小化优化问题。由于该优化问题是高度非凸的且难以求解,所以本文首先根据凸优化理论推导得到了各用户最佳发射功率的闭式表达式,然后利用连续凸逼近(Successive Convex Approximation, SCA)方法对原问题进行转换,并提出一个基于SCA的低复杂度迭代算法进行求解,从而得到本文所提的资源分配方案。最后借助于计算机仿真对本文所提方案的性能优势进行验证,仿真结果表明相比于其他基准方案,所提方案能有效降低系统能耗。

2 系统模型

如图1所示,本文考虑一个多用户NOMA-MEC系统,其中一个具备MEC功能的基站为 K 个边缘用户提供计算服务。假设所有的设备均装配了单根天线且工作在半双工模式。假设每个用户所需计算的任务均可被任意分割^[3,7],因而考虑了更加灵活的部分卸载方案,在该方案的指导下,每个用户均将其任务分成两个部分:一部分用于本地计算;另一部分则被卸载至基站计算。在任务卸载阶段,本文采用NOMA技术来指导 K 个用户同时同频卸载任务。另外,为了进一步降低任务卸载能耗,本文

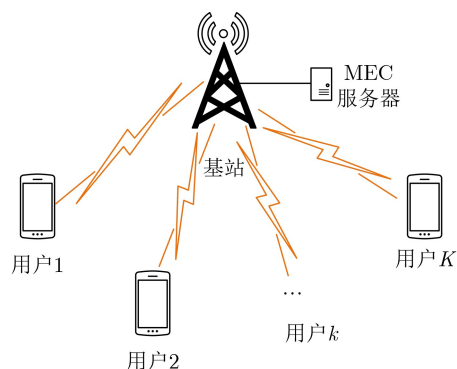


图1 系统模型

引入了部分压缩机制。具体而言,每个用户将其所需卸载的任务进一步分成两个部分:一部分直接卸载至基站;另一部分先被压缩至更少的比特数再卸载至基站。当基站接收到任务时,其先进行任务解压缩,然后再进行任务计算与结果下传。由于每个用户只有在任务压缩结束后才能进行任务卸载,且任务压缩与计算共用用户中央处理器(Central Processing Unit, CPU),所以用户在进行任务压缩时,无法进行本地计算。另外,在基站端,基站只有将所有压缩过的任务解压后,才能进行任务计算后将结果下传。本文用 $\mathcal{K} = \{1, 2, \dots, K\}$ 表示用户集。令 $h_k, k \in \mathcal{K}$ 表示第 k 个用户到基站的信道系数。假设所有信道均服从准静态衰落^[7],且可通过现有的信道估计算法计算得到所有信道增益。

令 T 表示给定时隙的时隙长度。根据部分压缩与卸载方案, T 可被分为两个子时隙,分别是任务压缩子时隙及任务卸载与本地计算子时隙,其分配情况如图2(a)所示。令 t_c 表示任务压缩子时隙的时隙长度。在该子时隙中,各个用户根据自身需求压缩一部分需要卸载的任务。令 $T - t_c$ 表示任务卸载与本地计算子时隙,该时隙又可进一步分为以下两种情况。

(1) 当用户进行任务卸载时,任务卸载与本地计算子时隙分布如图2(b)所示。具体而言,该子时隙可进一步被划分成4个阶段,依次是时长为 t_o 的任务卸载阶段、时长为 t_d 的任务解压缩阶段、时长为 t_e 的基站端任务计算阶段及计算结果下传阶段。具体地,所有用户采用NOMA技术在时间 t_o 内将任务卸载到基站。当基站接收到所有任务后,先经过 t_d 时间的解压缩,然后再进行 t_e 时间的任务计算。

考虑到任务计算结果的大小远小于各用户卸载任务的大小,本文假设计算结果下传时间非常小且可忽略不计^[5,6]。

(2) 当用户进行本地计算时,任务卸载与本地计算子时隙如图2(c)所示。令 $t_{l,k}$ 表示第 k 个用户本地计算的时间,各用户可在整个任务卸载与本地计算子时隙的任意时刻进行本地计算,则有 $0 \leq t_{l,k} \leq T - t_c$ 。

2.1 任务压缩

令 D_k 和 α_k 分别表示第 k 个用户需要计算的总任务量和任务卸载比例,则第 k 个用户进行卸载的总任务量为 $D_k \alpha_k$,而剩余的任务将进行本地计算。令 β_k 表示部分压缩机制下第 k 个用户任务的压缩比例。在该压缩比例引入后,卸载的总任务 $D_k \alpha_k$ 将进一步被分成两个部分:一部分是 $D_k \alpha_k \beta_k$,该部分任务通过无损压缩技术^[9]被压缩后再被卸载;另一部分是 $D_k \alpha_k (1 - \beta_k)$,该部分任务直接被卸载至基站。令 $f_{c,k}$ 和 ε_1 分别表示第 k 个用户进行任务压缩时的压缩频率及其CPU的能耗系数。令 C_{lc} 表示压缩1 bit任务时用户CPU需要的转数。根据文献^[8,9],第 k 个用户在任务压缩子时隙 t_c 内压缩时间和能耗需满足

$$t_c \geq D_k \alpha_k \beta_k C_{lc} f_{c,k}^{-1} \quad (1)$$

$$e_{c,k} = \varepsilon_1 f_{c,k}^3 D_k \alpha_k \beta_k C_{lc} / f_{c,k} = \varepsilon_1 f_{c,k}^2 D_k \alpha_k \beta_k C_{lc} \quad (2)$$

2.2 本地计算

令 C_{lc} 和 $f_{l,k}$ 依次表示第 k 个用户计算1 bit任务时CPU需要的转数和计算频率。根据文献^[4,5],第 k 个用户本地计算的时间及相应能耗可以被计算为

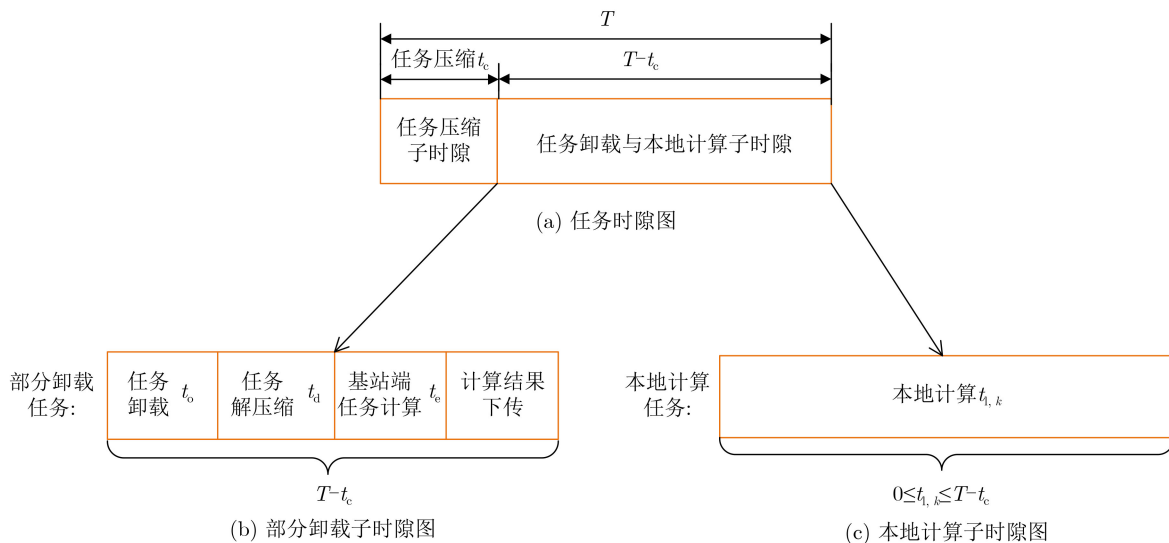


图2 系统时隙图

$$t_{l,k} = D_k (1 - \alpha_k) C_{le} f_{l,k}^{-1} \quad (3)$$

$$e_{l,k} = \varepsilon_l f_{l,k}^3 t_{l,k} = \varepsilon_l f_{l,k}^2 D_k (1 - \alpha_k) C_{le} \quad (4)$$

2.3 任务卸载

在任务卸载阶段，所有用户均采用NOMA技术来卸载任务。本文假设各用户之间使用同步算法^[10,11]，以严格同步保障多用户同时NOMA上传。根据NOMA技术的解码原则^[3,7]，基站将优先解码出强信号信息，避免强信号对弱信号解码的干扰。具体而言，假设所有用户与基站之间的信道增益服从 $h_1 \leq h_2 \leq \dots \leq h_K$ ，则基站将先解码第 K 个用户的卸载任务，再将解码出的任务从接收到的信号中减去，接着再解码第 $K-1$ 个用户的卸载任务，直到所有用户的卸载任务均被成功解码。令 B 表示系统信道带宽， p_k 表示第 k 个用户的发射功率， σ^2 表示信道噪声功率，则第 k 个用户在任务卸载阶段可卸载的任务比特数应被计算为 $R_k^u = t_o B \log_2 \left[1 + p_k h_k / \left(\sum_{i=1}^{k-1} p_i h_i + \sigma^2 \right) \right]$ 。令 δ_k 表示第 k 个用户任务压缩的压缩比，则第 k 个用户需要卸载的总任务比特数为 $D_k \alpha_k \beta_k \delta_k + D_k \alpha_k (1 - \beta_k)$ 。因此，为了在时间 t_o 内完成任务卸载，第 k 个用户可卸载的任务比特数应满足

$$R_k^u \geq D_k \alpha_k \beta_k \delta_k + D_k \alpha_k (1 - \beta_k) \quad (5)$$

令 $p_{c,k}$ 表示第 k 个用户在任务卸载阶段的固定电路损耗功率，则第 k 个用户的任务卸载能耗可被计算为

$$e_{o,k} = t_o (p_k + p_{c,k}) \quad (6)$$

2.4 任务解压缩

令 ε_m ， f_{md} 和 C_{md} 分别表示与基站连接的MEC服务器CPU的芯片能耗系数、解压缩频率和解压缩1 bit任务需要的转数。参考文献^[9]，基站上的MEC服务器利用无损压缩技术可准确地恢复所有卸载任务。根据文献^[8]，基站解压缩时间和相应的能耗可被计算为

$$t_d = \sum_{k=1}^K D_k \alpha_k \beta_k C_{md} f_{md}^{-1} \quad (7)$$

$$e_d = \varepsilon_m f_{md}^3 t_d = \sum_{k=1}^K \varepsilon_m f_{md}^2 D_k \alpha_k \beta_k C_{md} \quad (8)$$

2.5 基站端任务计算

当基站完成任务解压缩后，基站所需计算的总任务量为 $\sum_{k=1}^K D_k \alpha_k$ 。令 f_{me} 和 C_{me} 分别表示MEC服务器的CPU计算频率和计算1 bit任务所需的转

数，则基站进行任务计算所消耗的时间与能量可依次被计算为

$$t_e = \sum_{k=1}^K D_k \alpha_k C_{me} f_{me}^{-1} \quad (9)$$

$$e_e = \varepsilon_m f_{me}^3 t_e = \sum_{k=1}^K \varepsilon_m f_{me}^2 D_k \alpha_k C_{me} \quad (10)$$

3 最小化系统能耗

本节将通过联合优化用户任务压缩与卸载时间、用户压缩和卸载比例及发射功率来最小化系统能耗。具体地，系统能耗由所有用户任务压缩、本地计算与任务卸载能耗以及基站任务解压缩与计算能耗组成。根据式(2)、式(4)、式(6)、式(8)和式(10)，系统能耗可被写为

$$\begin{aligned} E &= \sum_{k=1}^K \varepsilon_l \frac{D_k \alpha_k \beta_k C_{le}}{f_{c,k}} f_{c,k}^3 + \sum_{k=1}^K \varepsilon_l t_{l,k} f_{l,k}^3 \\ &\quad + \sum_{k=1}^K t_o (p_k + p_{c,k}) + \varepsilon_m t_d f_{md}^3 + \varepsilon_m t_e f_{me}^3 \\ &= \sum_{k=1}^K \left(\varepsilon_l f_{c,k}^2 D_k \alpha_k \beta_k C_{le} + \varepsilon_l f_{l,k}^2 D_k (1 - \alpha_k) \right. \\ &\quad \cdot C_{le} + t_o (p_k + p_{c,k}) + \varepsilon_m f_{md}^2 D_k \alpha_k \beta_k C_{md} \\ &\quad \left. + \varepsilon_m f_{me}^2 D_k \alpha_k C_{me} \right) \end{aligned} \quad (11)$$

根据式(11)，系统能耗最小化优化问题可被建立为

$$\min_{\alpha, \beta, p, t_c, t_o} E \quad (12a)$$

$$\text{s.t. C1: } 0 \leq t_c + D_k (1 - \alpha_k) C_{le} f_{l,k}^{-1} \leq T, k \in \mathcal{K} \quad (12b)$$

$$\begin{aligned} \text{C2: } 0 \leq t_c + t_o + \sum_{k=1}^K D_k \alpha_k \beta_k C_{md} f_{md}^{-1} \\ + \sum_{k=1}^K D_k \alpha_k C_{me} f_{me}^{-1} \leq T \end{aligned} \quad (12c)$$

$$\text{C3: } t_c \geq D_k \alpha_k \beta_k C_{le} f_{c,k}^{-1}, k \in \mathcal{K} \quad (12d)$$

$$\text{C4: } R_k^u \geq D_k \alpha_k \beta_k \delta_k + D_k \alpha_k (1 - \beta_k), k \in \mathcal{K} \quad (12e)$$

$$\text{C5: } 0 \leq p_k \leq P, k \in \mathcal{K} \quad (12f)$$

$$\text{C6: } 0 \leq \alpha_k \leq 1, k \in \mathcal{K} \quad (12g)$$

$$\text{C7: } 0 \leq \beta_k \leq 1, k \in \mathcal{K} \quad (12h)$$

其中 $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_K]$ ， $\beta = [\beta_1, \beta_2, \dots, \beta_K]$ ， $p =$

$[p_1, p_2, \dots, p_K]$, P 表示用户的最大发射功率。在问题(12)中, 约束C1保障了每个用户都能在时间 T 内完成任务压缩和本地计算。C2约束了各用户任务压缩、卸载和基站端的任务解压缩、计算都必须在规定时间内完成。C3约束每个用户必须在预设的任务压缩时隙 t_c 内完成自身的任务压缩。C4保障了每个用户需要卸载的任务量都能在 t_o 内完成卸载。C5要求每个用户的发射功率不得超过 P 。C6和C7分别约束每个用户的任务压缩和卸载比例大小不得小于0, 也不得大于1。

通过观察, 可以发现上述所建立的优化问题是一个典型的非凸优化问题, 从而难以用现有的凸优化方法进行求解。而造成该问题非凸主要归因于以下两点:

(1)优化目标及约束C2和C3中均存在优化变量 α_k 和 β_k 之间的耦合, 从而导致了非凸的目标函数及非凸约束C2和C3;

(2)约束C4不但包含变量 α_k 与 β_k 之间的耦合, 还存在由NOMA共道干扰引起的高度非凸函数 R_k^u , 因此, C4也是一个高度非凸的约束。

为了求解该非凸问题, 本文先用凸优化理论推导出第 k 个用户最优发射功率 p_k^* 的闭式表达式。具体而言, 根据约束C4和C5首先推导出 p_k 的取值范围。然后, 通过分析目标函数与优化变量 p_k 之间的关系, 容易发现目标函数随着 p_k 的增大而增大, 因此为了最小化系统能耗, p_k^* 应在 p_k 取值范围的下界取得, 则第 k 个用户最优发射功率可被计算为

$$p_k^* = \begin{cases} \sigma^2 h_k^{-1} (\exp([D_k \delta_k \alpha_k \beta_k + D_k \alpha_k (1 - \beta_k)] \ln 2 / B t_o) - 1), & k = 1 \\ h_k^{-1} (\exp([D_k \delta_k \alpha_k \beta_k + D_k \alpha_k (1 - \beta_k)] \ln 2 / B t_o) - 1) \left(\sum_{i=1}^{k-1} p_i^* h_i + \sigma^2 \right), & 1 < k \leq K \end{cases} \quad (13)$$

令 $Q_k = \sum_{i=1}^k p_i^* h_i + \sigma^2 = \sigma^2 \exp\left(\left[\sum_{i=1}^k D_i \delta_i \alpha_i \beta_i + D_i \alpha_i (1 - \beta_i)\right] \ln 2 / B t_o\right)$, 且 $Q_0 = 0$ 。将 Q_k 代入式(13)中, 可以得到第 k 个用户最优发射功率为

$$p_k^* = (Q_k - Q_{k-1}) h_k^{-1} = \sigma^2 h_k^{-1} \left(\exp\left(\left[\sum_{i=1}^k D_i \delta_i \alpha_i \beta_i + D_i \alpha_i (1 - \beta_i)\right] \ln 2 / B t_o\right) - \exp\left(\left[\sum_{i=1}^{k-1} D_i \delta_i \alpha_i \beta_i + D_i \alpha_i (1 - \beta_i)\right] \ln 2 / B t_o\right) \right) \quad (14)$$

将 $p_k = p_k^*, k \in \mathcal{K}$ 代回问题(12), 问题(12)可被重写为

$$\min_{\alpha, \beta, t_o, t_c} E' \quad (15a)$$

$$\text{s.t. C1} - \text{C3} \quad (15b)$$

$$\bar{\text{C5}}: 0 \leq \sigma^2 h_k^{-1} \left(\exp\left(\left[\sum_{i=1}^k D_i \delta_i \alpha_i \beta_i + D_i \alpha_i (1 - \beta_i)\right] \ln 2 / B t_o\right) - \exp\left(\left[\sum_{i=1}^{k-1} D_i \delta_i \alpha_i \beta_i + D_i \alpha_i (1 - \beta_i)\right] \ln 2 / B t_o\right) \right) \leq P, k \in \mathcal{K} \quad (15c)$$

$$\text{C6, C7} \quad (15d)$$

其中优化目标可被重写为

$$E' = t_o \sum_{k=1}^K \left[\sigma^2 h_k^{-1} \left(\exp\left(\left[\sum_{i=1}^k D_i \delta_i \alpha_i \beta_i + D_i \alpha_i (1 - \beta_i)\right] \ln 2 / B t_o\right) - \exp\left(\left[\sum_{i=1}^{k-1} D_i \delta_i \alpha_i \beta_i + D_i \alpha_i (1 - \beta_i)\right] \ln 2 / B t_o\right) \right) + p_{c,k} \right] + \sum_{k=1}^K (\varepsilon_1 f_{c,k}^2 D_k x_k C_{lc} + \varepsilon_1 f_{l,k}^2 D_k (1 - \alpha_k) C_{le} + \varepsilon_m f_{md}^2 D_k x_k C_{md} + \varepsilon_m f_{me}^2 D_k \alpha_k C_{me}) \quad (16)$$

对于转化后的优化问题(15), 其仍是高度非凸的。为了处理变量 α_k 和 $\beta_k, k \in \mathcal{K}$ 之间的耦合关系, 本文引入了辅助变量 $x_k = \alpha_k \beta_k, k \in \mathcal{K}$ 并将其代入到优化问题(15)中。接着优化问题(15)可以重写为

$$\min_{\alpha, \mathbf{x}, t_o, t_c} E'' \quad (17a)$$

$$\text{s.t. C1} \quad (17b)$$

$$\bar{\text{C2}} : 0 \leq t_c + t_o + \sum_{k=1}^K D_k x_k C_{\text{md}} f_{\text{md}}^{-1} + \sum_{k=1}^K D_k \alpha_k C_{\text{me}} f_{\text{me}}^{-1} \leq T \quad (17c)$$

$$\bar{\text{C3}} : t_c \geq D_k x_k C_{\text{lc}} f_{\text{c},k}^{-1}, k \in \mathcal{K} \quad (17d)$$

$$\bar{\text{C5}} : 0 \leq t_o \left(\exp \left(\left[\sum_{i=1}^k D_i \delta_i \alpha_i \beta_i + D_i \alpha_i (1 - \beta_i) \right] \ln 2 / B t_o \right) - \exp \left(\left[\sum_{i=1}^{k-1} D_i \delta_i \alpha_i \beta_i + D_i \alpha_i (1 - \beta_i) \right] \ln 2 / B t_o \right) \right) \leq P t_o \gamma_k^{-1}, k \in \mathcal{K} \quad (17e)$$

$$\bar{\text{C6}} : 0 \leq x_k \leq \alpha_k \leq 1, k \in \mathcal{K} \quad (17f)$$

其中该问题的优化目标可被重写为

$$\begin{aligned} E'' = & t_o \sum_{k=1}^K \left[(\gamma_k - \gamma_{k+1}) \exp \left(\left[\sum_{i=1}^k D_i \delta_i \alpha_i \beta_i + D_i \alpha_i (1 - \beta_i) \right] \ln 2 / B t_o \right) + p_{\text{c},k} \right] - t_o \gamma_1 \\ & + \sum_{k=1}^K (\varepsilon_1 f_{\text{c},k}^2 D_k x_k C_{\text{lc}} + \varepsilon_1 f_{\text{l},k}^2 D_k (1 - \alpha_k) C_{\text{le}} + \varepsilon_{\text{m}} f_{\text{md}}^2 D_k x_k C_{\text{md}} + \varepsilon_{\text{m}} f_{\text{me}}^2 D_k \alpha_k C_{\text{me}}) \end{aligned} \quad (18)$$

且 $\gamma_k = \sigma^2 h_k^{-1}$ 及 $\gamma_{K+1} = 0$ 。

对于优化目标(18)来说, 对于优化目标(18)来说, 通过观察可以发现目标函数中 $\sum_{k=1}^K (\varepsilon_1 f_{\text{c},k}^2 D_k x_k C_{\text{lc}} + \varepsilon_1 f_{\text{l},k}^2 D_k (1 - \alpha_k) C_{\text{le}} + \varepsilon_{\text{m}} f_{\text{md}}^2 D_k x_k C_{\text{md}} + \varepsilon_{\text{m}} f_{\text{me}}^2 D_k \alpha_k C_{\text{me}} + t_o p_{\text{c},k}) - t_o \gamma_1$ 与各个优化变量之间均呈线性关系, 其关于各优化变量的2阶导均恒等于0(即非负), 因此该函数是凸函数。对于 $t_o \sum_{k=1}^K (\gamma_k - \gamma_{k+1}) \exp(\sum_{i=1}^k [D_i \delta_i x_i + D_i \alpha_i - D_i x_i] \ln 2 / B t_o)$ 而言, 令 $z_k = \sum_{i=1}^k [D_i \delta_i x_i + D_i \alpha_i - D_i x_i]$ 和 $y_k = t_o (\gamma_k - \gamma_{k+1}) \exp(\sum_{i=1}^k [D_i \delta_i x_i + D_i \alpha_i - D_i x_i] \ln 2 / B t_o) = t_o (\gamma_k - \gamma_{k+1}) \exp(z_k \ln 2 / B t_o)$, 则可以发现函数 $t_o \sum_{k=1}^K (\gamma_k - \gamma_{k+1}) \exp(\sum_{i=1}^k [D_i \delta_i x_i + D_i \alpha_i - D_i x_i] \ln 2 / B t_o)$ 的凹凸性取决于函数 y_k 的凹凸性。根据 $h_1 \leq h_2 \leq \dots \leq h_K$ 可以计算得到 $\gamma_k - \gamma_{k+1} \geq 0$ 。由于函数 y_k 是函数 $f_k = (\gamma_k - \gamma_{k+1}) \exp(z_k \ln 2 / B)$ 的透视函数, 所以 y_k 的凹凸性与函数 f_k 的凹凸性一致。根据函数 f_k 的表达式可以计算出 f_k 的2阶导 $f_k''(z_k) = (\gamma_k - \gamma_{k+1}) (\ln 2 / B)^2 \exp(z_k \ln 2 / B) \geq 0$, 因此函数 f_k 为凸函数。相应地, 函数 y_k 也是一个凸函数。综上所述, 优化目标(18)是一个凸函数。且约束 C1, $\bar{\text{C2}}$, $\bar{\text{C3}}$, $\bar{\text{C6}}$ 为线性约束为凸约束。对于约束 $\bar{\text{C5}}$, 当 $k = 1$ 时, 该约束可被写为 $t_o \exp(z_1 \ln 2 / B t_o) - t_o \leq P / \gamma_1 t_o$ 。由于函数 $t_o \exp(z_1 \ln 2 / B t_o)$

是一个凸函数, 因此 $k = 1$ 时的约束 $\bar{\text{C5}}$ 是凸约束。当 $k \geq 2$ 时, 约束 $\bar{\text{C5}}$ 可被写为 $0 \leq t_o (\exp(z_k \ln 2 / B t_o) - \exp(z_{k-1} \ln 2 / B t_o)) \leq P / \gamma_k t_o$ 。由于函数 $t_o (\exp(z_k \ln 2 / B t_o) - \exp(z_{k-1} \ln 2 / B t_o))$ 是一个凸差函数。因此, $k \geq 2$ 时的约束 $\bar{\text{C5}}$ 是非凸约束。根据以上分析, 问题(17)仍是一个非凸优化问题, 且该问题的求解难点在于如何处理非凸约束 $\bar{\text{C5}}$ 。

为了求解该非凸问题, 本文采用了SCA方法, 其主要思想是利用1阶泰勒展开得到的函数下界来近似导致原问题非凸的函数从而得到一个凸优化问题, 再利用凸优化方法, 如CVX^[12], 对该问题的最优解进行求解, 并通过不断迭代直到所得最优解满足收敛条件, 进而输出达到收敛时所得的最优解。

具体地, 令函数 $g_{k-1}(z_{k-1}, t_o) = t_o \exp(z_{k-1} \ln 2 / t_o B)$, $k \in \{2, 3, \dots, K\}$ 表示 $k \geq 2$ 时约束 $\bar{\text{C5}}$ 的非凸部分, 然后对该函数在给定初始值 (z_{k-1}^0, t_o^0) 处进行1阶泰勒展开, 可以得到

$$\begin{aligned} g_{k-1}(z_{k-1}, t_o) \geq & t_o \exp(z_{k-1} \ln 2 / t_o B) + B^{-1} (\ln 2) \\ & \cdot \exp(z_{k-1}^0 \ln 2 / t_o^0 B) (z_{k-1} - z_{k-1}^0) \\ & + (1 - z_{k-1}^0 \ln 2 / t_o^0 B) \\ & \cdot \exp(z_{k-1}^0 \ln 2 / t_o^0 B) (t_o - t_o^0) \end{aligned} \quad (19)$$

将式(19)代入到问题(21)的约束 $\bar{\text{C5}}$ 中, 则在 $k \in \{2, 3, \dots, K\}$ 时, $\bar{\text{C5}}$ 可近似为

$$\begin{aligned}
& 0 \leq t_o \exp(z_k \ln 2 / B t_o) - t_o^0 \exp(z_{k-1}^0 \ln 2 / B t_o^0) \\
& - B^{-1} (\ln 2) \exp(z_{k-1}^0 \ln 2 / B t_o^0) \\
& \cdot (z_{k-1} - z_{k-1}^0) - (1 - z_{k-1}^0 \ln 2 / t_o^0 B) \\
& \cdot \exp(z_{k-1}^0 \ln 2 / B t_o^0) (t_o - t_o^0) \leq P t_o \gamma_k^{-1}, \\
& k \in \{2, 3, \dots, K\}
\end{aligned} \quad (20)$$

相应地, 在给定初始值 $(\{\alpha_k^0\}, \{\beta_k^0\}, t_o^0)$ 的基础上, 首先计算得出 $x_k^0 = \alpha_k^0 \beta_k^0, k \in \mathcal{K}$, 进而得到 $z_k^0 = \sum_{i=1}^k [D_i \delta_i x_i^0 + D_i \alpha_i^0 - D_i x_i^0], k \in \mathcal{K}$ 。然后在给定 $(\{z_k^0\}, t_o^0)$ 的情况下, 优化问题(21)可被转化为

$$\min_{\alpha, x, t_o, t_c} E'' \quad (21a)$$

$$\text{s.t. } \bar{C}1, \bar{C}2, \bar{C}3, \bar{C}6 \quad (21b)$$

$$\begin{aligned}
& \bar{C}5: 0 \leq t_o \exp(z_1 \ln 2, B t_o) - t_o \leq P t_o \gamma_1^{-1}, \\
& k = 1 \\
& 0 \leq t_o \exp(z_k \ln 2, B t_o) \\
& - t_o^0 \exp(z_k^0 \ln 2 / B t_o^0) \\
& - \frac{\ln 2}{B} \exp(z_k^0 \ln 2, B t_o^0) (z_k - z_{k-1}^0) \\
& - (1 - z_{k-1}^0 \ln 2, t_o^0 B) \\
& \cdot \exp(z_{k-1}^0 \ln 2, B t_o^0) (t_o - t_o^0) \\
& \leq P t_o \gamma_k^{-1}, k \in \{2, 3, \dots, K\}
\end{aligned} \quad (21c)$$

很显然问题(21)是一个典型的凸优化问题, 其可以直接用CVX等工具进行求解。基于以上分析, 本文设计出一种基于SCA的迭代算法来解决优化问题(21), 并得到其最优解。具体步骤如算法1所示。

在此算法中, 本文先给定一组初始值 $(\{\alpha_k^0\},$

$\{\beta_k^0\}, t_o^0, t_c^0)$, 然后根据此初始值计算得出中间变量 $\{x_k^0\}$ 、初始最小系统能耗 E^0 和中间变量 $\{z_k^0\}$ 。在第 n 次迭代中, 在给定 $(\{z_k^0\}, t_o^0)$ 的情况下求解问题(21)并得到其最优解, 接着根据所得到的最优解计算得出此时的最小系统能耗 E^n 。判断此时是否满足循环终止条件, 如若满足, 则此时的最优解和最小系统能耗即为问题(17)的最优解和最小系统能耗。否则令 $\alpha_k^n = \alpha_k^n; x_k^n = x_k^n, \forall k; t_o^n = t_o^n$ 并更新循环次数直至循环中止条件满足。

下面给出所提算法1的复杂度分析。假设所提算法达到收敛时其算法迭代次数为 N_u 。根据参考文献[13,14], 可以得到算法1的计算复杂度为 $O(N_u(2K+2)^3)$ 。其中, 每次迭代时优化问题(21)中的优化变量个数为 $2K+2$, 函数 $O(\cdot)$ 表示渐进上界, 代表算法的最坏复杂度。

4 仿真结果与分析

在本节中, 本文利用计算机仿真来验证本文所提方案的优越性。如无特殊说明, 本文的基本系统参数设置如下。参考文献[3,7,9], 令用户总个数 K 为2, 系统带宽 B 为3 MHz, 各用户最大发射功率 P 为23 dBm, 参数 γ_1 和 γ_2 分别为1/110和1/180, 各用户的任务量 D_1 和 D_2 均为1 Mbits。根据文献[8,9], 设定各用户任务压缩比 δ_1 和 δ_2 均为0.1, 各用户的计算频率 $f_{1,1}$ 和 $f_{1,2}$ 均为 5×10^6 Hz, 压缩频率 $f_{c,1}$ 和 $f_{c,2}$ 均为 5×10^7 Hz, 基站端计算频率 f_{me} 和解压缩频率 f_d 分别为10 GHz和5 MHz, 各用户计算所需周期数 C_{le} 和压缩所需周期数 C_{lc} 分别为1000 cycles/bit 和120 cycle/bit, 基站端计算所需周期数 C_{me} 和解压

算法 1 基于SCA的迭代算法

输入: 给定初始值 $(\{\alpha_k^0\}, \{\beta_k^0\}, t_o^0, t_c^0)$; 设置迭代次数 $n = 1$ 、最大收敛次数 N 和收敛精度 ε ; 设定循环中止标志 $\text{Flag} = 0$;

输出: 最优解 $(\{\alpha_k^*\}, \{\beta_k^*\}, t_c^*, \{p_k^*\}, t_o^*)$; 最小系统能耗 E^* 。

(1) 根据 $(\{\alpha_k^0\}, \{\beta_k^0\}, t_o^0, t_c^0)$ 计算得出中间变量 x_k^0 , 进而计算出初始系统能耗为 E^0 ;

(2) Repeat

(3) 根据初始值计算得出 z_k^0 ;

(4) 在给定 $(\{z_k^0\}, t_o^0)$ 的情况下, 利用CVX工具求解问题(21), 并得到其最优解, 即: $(\{\alpha_k^n\}, \{x_k^n\}, t_c^n, t_o^n)$;

(5) 根据上述所得到的最优解计算得出此时的最小系统能耗为 E^n ;

(6) if $|E^n - E^{n-1}| < \varepsilon$;

(7) 此时最优解即为问题(18)的最优解, 即: $\alpha_k^* = \alpha_k^n; \beta_k^* = x_k^n / \alpha_k^n, \forall k; t_o^* = t_o^n; t_c^* = t_c^n$;

$p_k^* = \sigma^2 / h_k (\exp(z_k^0 \ln 2 / B t_o) - \exp(z_{k-1}^0 \ln 2 / B t_o))$; 系统最小能耗为 $E^* = E^n$;

(8) 输出问题(17)的最优解和系统最小能耗, 即: $(\{\alpha_k^*\}, \{\beta_k^*\}, \{p_k^*\}, t_o^*, t_c^*)$; E^* ; 设置 $\text{Flag} = 1$;

(9) else

(10) 设置 $\alpha_k^0 = \alpha_k^n; x_k^0 = x_k^n, \forall k; t_o^0 = t_o^n; n = n + 1$;

(11) end

(12) until $\text{Flag} = 1$ or $n = N$ 。

缩所需周期数 C_{md} 依次为100 cycle/bit和10 cycle/bit, 各用户和MEC服务器的CPU芯片能耗系数分别为 10^{-26} 和 10^{-28} 。

图3展现了在不同MEC计算频率下, 本文所提算法的收敛性。通过观察可以发现本文所提算法仅经过3次迭代, 就能达到收敛状态, 这显示了本文所提算法的快速收敛性与高效的计算效率。另外, 通过比较, 还可以发现随着MEC计算频率的不断增大, 系统能耗也随之增加, 这是因为系统能耗与MEC计算频率之间呈现单调递增关系。

为了充分展现本文所提方案的性能优势, 本文设置五种不同的基准方案, 各方案具体介绍如下:

(1)无数据压缩下的部分卸载方案^[3]: 在该方案下各用户根据任务卸载比例将其部分任务通过NOMA技术卸载至基站进行计算, 而剩余部分则进行本地计算。

(2)基于全压缩的部分卸载方案: 此方案下的各用户根据任务卸载比例将计算任务一分为二, 其中一部分任务先全部压缩至最小比特数再通过NOMA技术卸载至基站进行计算而另一部分则直接进行本地计算。

(3)基于时分多址接入(Time Division Multiple Access, TDMA)的部分压缩与卸载方案: 此方案下各用户根据任务卸载比例将其计算任务一分为二, 一部分任务直接进行本地计算, 而另一部分任务再根据任务压缩比例分成两个部分, 第1部分任务先进行数据压缩, 再将压缩后的任务与第2部分任务通过TDMA方式卸载到基站上进行计算。

(4)基于部分压缩的完全卸载方案: 此方案下各用户将其需要计算的任务分成两个部分, 一部分任务先进行数据压缩, 再将压缩后的任务与剩余部分任务通过NOMA技术全部卸载至基站进行计算。

(5)完全本地计算方案: 该方案中每个用户的所有任务都进行本地计算。

除基于TDMA的部分压缩与卸载方案外, 其

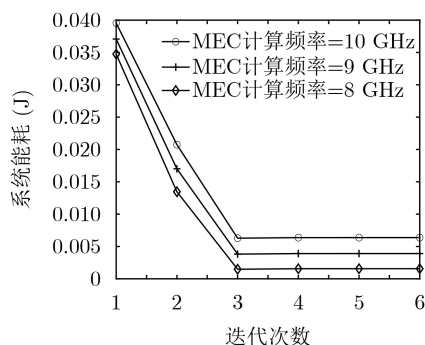


图3 本文所提算法的收敛性

余方案都可通过本文所提算法得到。具体而言, 无数据压缩下的部分卸载方案和基于全压缩的部分卸载方案可通过分别设置用户任务压缩比例 $\beta_k = 0$ 和 $\beta_k = 1, k \in \mathcal{K}$ 得到; 基于部分压缩的完全卸载方案和完全本地计算方案可以通过分别设置任务卸载比例 $\alpha_k = 1$ 和 $\alpha_k = 0, k \in \mathcal{K}$ 得到。对于基于TDMA的部分压缩与卸载方案, 需要重新建立系统模型及相应的优化问题, 再通过对该优化问题进行转化与求解, 从而得到该方案。

图4描绘了不同方案下系统能耗与各用户计算任务量的对比情况, 其中令 $D_1 = D_2 = D_c$ 。如图所示, 随着用户任务量 D_c 的增加, 系统能耗也随之增大。这是因为系统能耗与 D_c 之间呈单调递增关系。另外, 与无数据压缩下的部分卸载方案、基于全压缩的部分卸载方案、基于部分压缩的完全卸载方案及完全本地计算方案比较, 容易发现本文所提方案能够取得最低的系统能耗, 且随着 D_c 的增加, 本文所提方案的性能优势越明显。这是因为本文所提方案采用了更加灵活的部分压缩和卸载机制, 从而能更好地利用可用资源实现系统能耗最小化。与基于TDMA的部分压缩与卸载方案不同, 本文所提方案采用NOMA技术来卸载各用户的计算任务, 而NOMA技术允许多个用户共享频谱和时间资源, 从而能取得更高的频谱效率和能量效率, 因而本文所提方案的系统能耗低于基于TDMA的部分压缩和卸载方案的系统能耗。

图5展示了不同方案下系统能耗随着基站端计算频率 f_{me} 的变化情况。如图所示, 无数据压缩下的部分卸载方案、基于全压缩的部分卸载方案、基于部分压缩的完全卸载方案的系统能耗均随着 f_{me} 的增加而增加, 而完全本地计算方案下的系统能耗

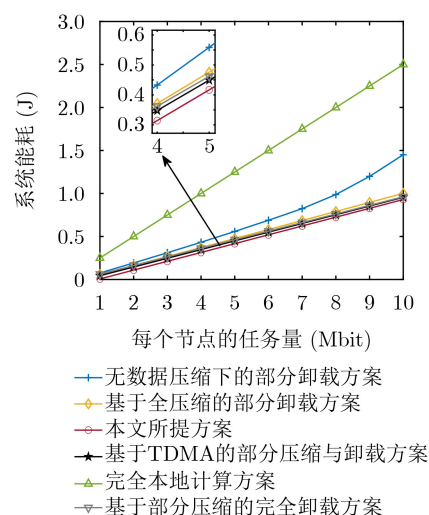


图4 不同方案下系统能耗与各用户计算任务量的对比情况

不随 f_{me} 的变化而变化。这是因为根据式(10)，基站端计算所消耗的能量随着 f_{me} 的增加而增加，而完全本地计算方案下的系统能耗与 f_{me} 无关。通过比较，可以发现无压缩的部分卸载方案和基于全压缩的部分卸载方案下的系统能耗都远远大于本文所提方案的系统能耗。这是因为本文采用了部分压缩机制，其可以根据当前的信道状况更加灵活地调节需要压缩的任务量，从而取得更小的系统能耗。与基于部分压缩的完全卸载方案和完全本地计算方案相比，本文所提方案可以灵活调度本地计算资源和基站计算资源，从而取得更低的系统能耗。相比于基于TDMA的部分压缩与卸载方案，本文所提方案使用了NOMA技术，其具有更高的频谱效率和能量效率，因而本文所提方案的系统能耗始终低于基于TDMA的部分压缩与卸载方案的系统能耗。

图6描绘了不同用户接入数量下系统能耗与各用户任务量的对比情况，其中令 $D_1 = D_2 = D_3 = D_4 = D_c$ 。如图所示，随着接入用户数量的增加，系统的能耗也随之增加，且这种变化随着 D_c 的增加也愈发明显。这是因为用户数越多，总体需要计算的任务量也随之增加，根据式(18)可知，系统能

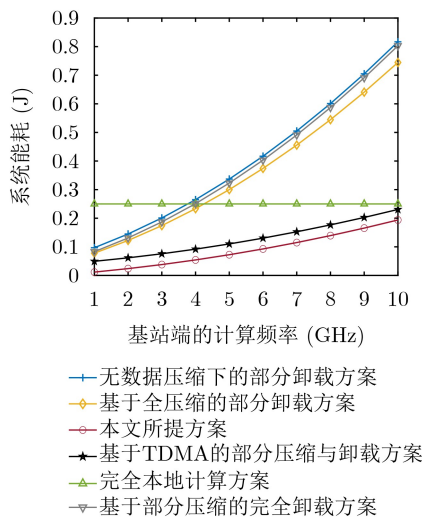


图5 不同方案下系统能耗随基站端计算频率的变化情况

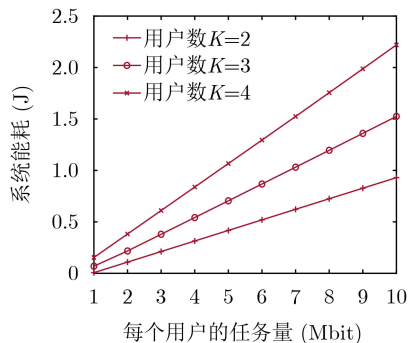


图6 不同用户接入数量下系统能耗随用户任务量的变化情况

耗与计算任务比特数之间呈单调递增关系，所以系统能耗将随着用户数的增大而增大。

5 结束语

本文建立了一个全新的基于数据压缩的NOMA-MEC网络，并研究了该网络中系统能耗最小化问题。具体而言，在考虑基站有限的计算能力和各用户所采用的部分压缩与卸载机制情况下，本文通过联合优化用户发射功率和压缩、卸载比例等变量，建立一个系统能耗最小化的非凸优化问题。为了求解该非凸问题，本文先根据凸优化理论推导得到各用户最佳功率表达式，再利用SCA方法对原问题进行转化并设计出一个基于SCA的迭代算法求解该问题。最后借助于计算机仿真对本文所提方案的性能优势进行验证。仿真结果表明，本文所提方案较其他基准方案在节能方面有明显优势。

参考文献

- [1] SONG Zhiyuan, MA Ruijiang, and XIE Yong. A collaborative task offloading strategy for mobile edge computing in internet of vehicles[C]. 2021 IEEE 5th Advanced Information Technology, Electronic and Automation Control Conference (IAEAC), Chongqing, China, 2021: 1379–1384. doi: [10.1109/IAEAC50856.2021.9390817](https://doi.org/10.1109/IAEAC50856.2021.9390817).
- [2] SATRIA D, PARK D, and JO M. Recovery for overloaded mobile edge computing[J]. *Future Generation Computer Systems*, 2017, 70: 138–147. doi: [10.1016/j.future.2016.06.024](https://doi.org/10.1016/j.future.2016.06.024).
- [3] PAN Yijin, CHEN Ming, YANG Zhaohui, et al. Energy-efficient NOMA-based mobile edge computing offloading[J]. *IEEE Communications Letters*, 2018, 23(2): 310–313. doi: [10.1109/LCOMM.2018.2882846](https://doi.org/10.1109/LCOMM.2018.2882846).
- [4] ZHANG Siqi, YI Na, and MA Yi. Correlation-based device energy-efficient dynamic multi-task offloading for mobile edge computing[C]. 2021 IEEE 93rd Vehicular Technology Conference (VTC2021-Spring), Helsinki, Finland, 2021: 1–5. doi: [10.1109/VTC2021-Spring51267.2021.9448864](https://doi.org/10.1109/VTC2021-Spring51267.2021.9448864).
- [5] GAO Mingjin, SHEN Rujing, LI Jun, et al. Computation offloading with instantaneous load billing for mobile edge computing[J]. *IEEE Transactions on Services Computing*, 2022, 15(3): 1473–1485. doi: [10.1109/TSC.2020.2996764](https://doi.org/10.1109/TSC.2020.2996764).
- [6] HE Tianmi, WANG Dawei, ZHOU Fuhui, et al. Delay-aware offloading for cooperative NOMA-based near-and-far MEC networks[C]. 2021 IEEE/CIC International Conference on Communications in China (ICCC), Xiamen, China, 2021: 978–983. doi: [10.1109/ICCC52777.2021.9580414](https://doi.org/10.1109/ICCC52777.2021.9580414).

- [7] SHI Liqin, YE Yinghui, CHU Xiaoli, *et al.* Computation energy efficiency maximization for a NOMA-based WPT-MEC network[J]. *IEEE Internet of Things Journal*, 2021, 8(13): 10731–10744. doi: [10.1109/JIOT.2020.3048937](https://doi.org/10.1109/JIOT.2020.3048937).
- [8] REN Jinke, RUAN Yangjun, and YU Guanding. Data transmission in mobile edge networks: Whether and where to compress?[J]. *IEEE Communications Letters*, 2019, 23(3): 490–493. doi: [10.1109/LCOMM.2019.2894415](https://doi.org/10.1109/LCOMM.2019.2894415).
- [9] XU Ding, LI Qun, and ZHU Hongbo. Energy-saving computation offloading by joint data compression and resource allocation for mobile-edge computing[J]. *IEEE Communications Letters*, 2019, 23(4): 704–707. doi: [10.1109/LCOMM.2019.2897630](https://doi.org/10.1109/LCOMM.2019.2897630).
- [10] MHEICH Z and DUPRAZ E. Short length non-binary rate-adaptive LDPC codes for Slepian-Wolf source coding[C]. 2018 IEEE Wireless Communications and Networking Conference (WCNC), Barcelona, Spain, 2018: 1–5. doi: [10.1109/WCNC.2018.8377291](https://doi.org/10.1109/WCNC.2018.8377291).
- [11] LIU Chenglin. Predictor-based synchronization algorithms for multiple harmonic oscillators with communication delay[C]. 2015 IEEE International Conference on Information and Automation, Lijiang, China, 2015: 1003–1008. doi: [10.1109/ICInfA.2015.7279433](https://doi.org/10.1109/ICInfA.2015.7279433).
- [12] GRANT M and BOYD S. CVX: Matlab software for disciplined convex programming[EB/OL]. <http://cvxr.com/cvx>, 2023.
- [13] BOYD S and VANDENBERGHE L. *Convex Optimization*[M]. Cambridge: Cambridge University Press, 2009: 104–112.
- [14] LUO Weiran, SHEN Yanyan, YANG Bo, *et al.* Joint 3-D trajectory and resource optimization in multi-UAV-enabled IoT networks with wireless power transfer[J]. *IEEE Internet of Things Journal*, 2021, 8(10): 7833–7848. doi: [10.1109/JIOT.2020.3041303](https://doi.org/10.1109/JIOT.2020.3041303).
- 施丽琴：女，副教授，研究方向为无线供能通信、移动边缘计算。
刘璇：女，硕士生，研究方向为移动边缘计算。
卢光跃：男，教授，研究方向为宽带无线通信。
- 责任编辑：马秀强