

doi: 10.14132/j.cnki.1673-5439.2025.05.005

# RIS 辅助的 NOMA-MEC 系统时延最小化问题研究

徐宇欣, 王 鸿

(南京邮电大学 通信与信息工程学院, 江苏 南京 210003)

**摘要:** 移动边缘计算(Mobile Edge Computing, MEC)与智能超表面(Reconfigurable Intelligent Surface, RIS)是未来无线通信领域中两种极具发展前景的技术,分别致力于提升通信系统的计算能力与优化信道传输性能。考虑一种多用户非正交多址接入(Non-Orthogonal Multiple Access, NOMA)-MEC 网络的任务时延最小化问题,并通过部署 RIS 改善通信系统性能,其中多个用户可以通过同一频段同时卸载任务。由于优化问题非凸,且需联合优化 RIS 相移、卸载功率、波束成型和分配因子,很难直接求解建立的优化问题。为了进一步求解优化问题,利用半正定松弛(Semidefinite Relaxation, SDR)算法优化 RIS 的相移,采用逐次凸逼近(Successive Convex Approximation, SCA)算法求解功率和分配因子,同时采用最小均方误差(Minimum Mean Square Error, MMSE)准则设计均衡器,并进一步利用交替优化使其时延最小。

**关键词:** 智能超表面;非正交多址;移动边缘计算;时延

中图分类号:TN929.5 文献标志码:A 文章编号:1673-5439(2025)05-0042-08

## Minimized delay of the RIS-assisted NOMA-MEC system

XU Yuxin, WANG Hong

(School of Communications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China)

**Abstract:** Mobile edge computing (MEC) and reconfigurable intelligent surface (RIS) are two highly promising key technologies in future wireless communications, dedicated to enhancing the computational capabilities of communication systems and optimizing channel transmission performance, respectively. To address the problem of minimizing the task delay in a multi-user non-orthogonal multiple access (NOMA)-MEC network, RIS is deployed to improve the communication performance, enabling multiple users to offload tasks simultaneously through the same frequency band. We jointly optimize the RIS phase shift, offloading power, beamforming, and distribution factors, and design an equalizer at the receiving end. Since the problem is non-convex, the semidefinite relaxation (SDR) algorithm is used to optimize the phase shift of RIS, the successive convex approximation (SCA) algorithm is used for the power and allocation factors, the equalizer is designed by the minimum mean square error (MMSE) criterion, and the delay is minimized by the alternating optimization.

**Keywords:** reconfigurable intelligent surface (RIS); non-orthogonal multiple access (NOMA); mobile edge computing (MEC); delay

收稿日期:2025-04-09;修回日期:2025-06-09 本刊网址:<http://nyzr.njupt.edu.cn>

基金项目:国家自然科学基金(62171237)和江苏省自然科学基金(BK20231285)资助项目

作者简介:徐宇欣,男,硕士研究生;王鸿(通信作者),男,博士,副教授, [wanghong@njupt.edu.cn](mailto:wanghong@njupt.edu.cn)

引用本文:徐宇欣,王鸿.RIS 辅助的 NOMA-MEC 系统时延最小化问题研究[J].南京邮电大学学报(自然科学版),2025,45(5):42-49.

**Citation:** XU Yuxin, WANG Hong. Minimized delay of the RIS-assisted NOMA-MEC system[J]. Journal of Nanjing University of Posts and Telecommunications(Natural Science Edition), 2025, 45(5): 42-49.

可重构智能超表面(Reconfigurable Intelligent Surface, RIS)技术可以智能地调整无线传播环境,被视为实现第六代移动通信的关键技术之一。通过部署大量反射元件改善信号的传输条件,可以大幅提高信号的频谱/能量效率<sup>[1]</sup>。相较于中继器,无源RIS反射信号无须依赖有源器件,能满足全双工工作模式,同时不引入任何噪声,因这些特点被学术界应用于各种通信场景。此外,RIS还能与其他通信技术很好地融合,进一步提升通信性能,例如RIS辅助的非正交多址接入技术(Non-Orthogonal Multiple Access, NOMA)<sup>[2-3]</sup>,RIS辅助的速率拆分多址技术(Rate Splitting Multiple Access, RSMA)<sup>[4]</sup>。同时,随着物联网的快速发展,移动边缘计算(Mobile Edge Computing, MEC)逐渐成为很有前途的技术,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具有更高计算能力的设备进行计算,如基站、边缘服务器等,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<sup>[5]</sup>。

对于MEC系统,由于部分任务无法在本地完成,需要传输到边缘节点进行处理,此时传输的时延是比较重要的指标,文献[6]给出了MEC时延最基本的3种模型,即局部压缩、边缘云压缩和部分压缩卸载,并推导时延的相关表达式。在文献[7]中,RIS可以提高MEC系统中卸载链路的信道增益,因为RIS可以减轻传播损耗,提高频谱和能量效率。具体而言,每个被动反射元件都可以对入射信号施加相移,从而协同改善传播环境。文献[8]考虑两个用户场景下RIS辅助的MEC系统,用户将数据传输分为NOMA和时分多址(Time Division Multiple Access, TDMA)两部分,研究两个用户时延总和的最小化问题。文献[9]将RIS与MEC结合,以边缘计算能力和RIS相移为约束条件,构建单天线基站、多用户模型的时延最小化问题,仿真结果表明,RIS辅助MEC系统显著优于无RIS辅助的传统MEC系统。文献[10]考虑了异构网络场景,研究RIS-MEC辅助的异构网络时延最小化问题,联合优化缓存大小、任务卸载和资源分配得到问题最优解。文献[11]研究RIS辅助MEC卸载场景下的全网延迟优化问题,通过图论将其转化为最短路径问题,实现无源波束成形增益和反射损耗之间的折衷。

与此同时,NOMA技术允许多个用户复用相同的时间频率资源进行数据传输,能够大幅提高系统的频谱利用效率,不少学者将NOMA与MEC结合,

NOMA-MEC网络有利于实现大规模连接和分布式计算,从而支持海量用户完成任务卸载,有潜力为更复杂的应用场景提供高性能服务。文献[12]研究了两用户情况下NOMA-MEC系统的时延,利用Dinkelbach法和Newton法求解分式规划问题。文献[13]研究了NOMA-MEC网络的时延问题,利用分层结构求解最优的卸载系数。文献[14]联合优化任务卸载因子和发送功率,最小化多用户NOMA-MEC网络中的延迟,利用二分搜索算法求出最优解。然而,文献[12]并没有考虑多用户下NOMA-MEC网络的时延,实际上,多用户场景在通信系统中更为常见。文献[13]没有考虑发送功率的影响,然而发送功率是决定时延的一个重要因素。文献[14]只考虑了NOMA与MEC结合,在实际通信场景中,由于障碍物遮挡或传播环境恶劣等因素,用户与基站之间可能存在非视距或信道条件恶劣的情况。为了解决这一问题,RIS被引入无线通信系统,在多用户的场景下,同时引入RIS和NOMA两种技术的研究还较为鲜见,因此,本文将在多用户场景下,研究RIS-NOMA-MEC网络的时延性能优化问题。主要贡献如下:

(1)将RIS与NOMA两种技术结合进MEC网络,并在接收端设计一个均衡器,该方案可以同时优化RIS相移,发送功率,分配系数。(2)针对方案中的非凸问题,采用高效的交替优化(Alternating Optimization, AO)算法,该算法利用MMSE准则优化均衡器,采用SDR算法优化相移,利用SCA算法优化功率与分配系数。

## 1 系统模型

### 1.1 通信模型

如图1所示,建立一个RIS辅助的上行多用户NOMA-MEC系统,基站配备了一个MEC服务器。用户与基站之间存在直连链路,用户与基站之间部署一个拥有 $N$ 个反射元件的RIS来改善信道,基站处部署 $M$ 根天线,用户处部署单个天线,系统共有 $K$

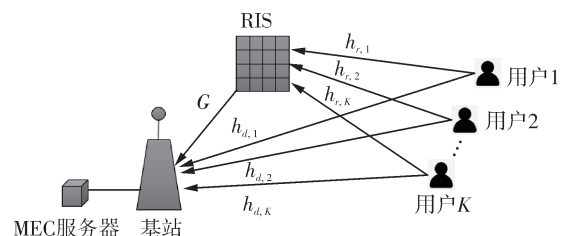


图1 RIS辅助上行多用户NOMA-MEC系统

个用户,多个用户可以通过同一频段同时进行任务卸载,用户标记为  $k \in \{1, 2, \dots, K\}$ 。假设基站到用户  $k$  的距离为  $D_{d,k}$ ,基站到 RIS 的距离为  $D_{B,r}$ ,RIS 到用户  $k$  的距离为  $D_{r,k}$ 。

假设信道状态信息(Channel State Information, CSI)已知,信道模型包含路径损耗和小尺度衰落两部分,基站到用户的直连链路建模为瑞利衰落,则用户  $k$  到基站的信道为

$$\mathbf{h}_{d,k} = D_{d,k}^{-\alpha_{d,k}/2} \mathbf{q}_{d,k} \quad (1)$$

其中,  $\mathbf{h}_{d,k} \in \mathbb{C}^{M \times 1}$ ,  $\mathbf{q}_{d,k}$  为小尺度衰落,每个元素服从均值为0、方差为1的复高斯分布,即  $\mathbf{q}_{d,k} \sim \text{CN}(0, 1)$ 。 $\alpha_{d,k}$  为用户  $k$  到基站的路径损耗指数。

由于 RIS 的位置一般是设计好的,因此基站到 RIS 与用户到 RIS 之间都存在视距链路,故信道采用莱斯衰落。基站到 RIS 的信道矩阵和用户到 RIS 的信道矩阵分别为

$$\mathbf{G} = D_{B,r}^{-\alpha_{B,r}/2} \left( \sqrt{\frac{\beta_{Br}}{1 + \beta_{Br}}} \mathbf{G}^{\text{LoS}} + \sqrt{\frac{1}{1 + \beta_{Br}}} \mathbf{G}^{\text{NLoS}} \right) \quad (2)$$

$$\mathbf{h}_{r,k} = D_{r,k}^{-\alpha_{r,k}/2} \left( \sqrt{\frac{\beta_{rk}}{1 + \beta_{rk}}} \mathbf{h}^{\text{LoS}} + \sqrt{\frac{1}{1 + \beta_{rk}}} \mathbf{h}^{\text{NLoS}} \right) \quad (3)$$

其中,  $\mathbf{G} \in \mathbb{C}^{M \times N}$ ,  $\mathbf{h}_{r,k} \in \mathbb{C}^{N \times 1}$ ,  $\beta_{Br}$  和  $\beta_{rk}$  为莱斯因子,  $\alpha_{B,r}$  和  $\alpha_{r,k}$  分别为基站到 RIS 的路径损耗指数和 RIS 到用户  $k$  的路径损耗指数,  $\mathbf{G}^{\text{LoS}}$  和  $\mathbf{h}^{\text{LoS}}$  代表视距分量,  $\mathbf{G}^{\text{NLoS}}$  和  $\mathbf{h}^{\text{NLoS}}$  代表非视距分量。

定义  $P_k$  为用户  $k$  的卸载发送功率,  $x_k$  表示用户  $k$  的发送数据,基站处的接收信号可以表示为

$$\mathbf{y} = \sum_{k=1}^K \sqrt{P_k} (\mathbf{G}\mathbf{\Theta}\mathbf{h}_{r,k} + \mathbf{h}_{d,k})x_k + \mathbf{n} \quad (4)$$

其中,  $\mathbf{\Theta} = \text{diag}(\alpha_1 e^{j\theta_1}, \dots, \alpha_N e^{j\theta_N})$  表示 RIS 的反射矩阵,其为对角阵,  $\alpha_n \in [0, 1]$  表示第  $n$  个 RIS 反射元件的幅度,  $\theta_n \in [0, 2\pi]$  表示第  $n$  个元件的相移。假设 RIS 的幅度和相移相互独立,并且可以实现连续的相移调控,即满足  $\alpha_n = 1$  和  $\theta_n \in [0, 2\pi]$ 。 $\mathbf{n} \sim \text{CN}(0, \sigma^2)$  表示在基站处的加性高斯白噪声。

在基站处为了解调各个用户的信号,设计一个均衡器  $\mathbf{w}_k \in \mathbb{C}^{M \times 1}$ ,其表示用户  $k$  的均衡器,用户  $k$  在基站处的解码信号可以表示为

$$\hat{y}_k = \mathbf{w}_k^H \left( \sum_{k=1}^K \sqrt{P_k} (\mathbf{G}\mathbf{\Theta}\mathbf{h}_{r,k} + \mathbf{h}_{d,k})x_k + \mathbf{n} \right) \quad (5)$$

不失一般性地,假设各用户的功率增益排序为  $\|\mathbf{G}\mathbf{\Theta}\mathbf{h}_{r,1} + \mathbf{h}_{d,1}\|^2 \leq \|\mathbf{G}\mathbf{\Theta}\mathbf{h}_{r,2} + \mathbf{h}_{d,2}\|^2 \leq \dots \leq \|\mathbf{G}\mathbf{\Theta}\mathbf{h}_{r,K} + \mathbf{h}_{d,K}\|^2$ ,在上行 NOMA 系统中,  $K$  个用户的信号叠加在一起,为了减轻用户间的干扰,基站使用串行干

扰消除(Successive Interference Cancellation, SIC)技术先将信道条件较好的用户信号解调,然后从叠加信号中删除该用户的信号,直到完成对所有用户信号的解调。因此,用户  $k$  处的信干噪比可以表示为

$$\gamma_k = \frac{P_k |\mathbf{w}_k^H (\mathbf{G}\mathbf{\Theta}\mathbf{h}_{r,k} + \mathbf{h}_{d,k})|^2}{\sum_{j=1}^{k-1} P_j |\mathbf{w}_j^H (\mathbf{G}\mathbf{\Theta}\mathbf{h}_{r,j} + \mathbf{h}_{d,j})|^2 + \sigma^2 \|\mathbf{w}_k\|^2} \quad (6)$$

相应地,用户  $k$  的速率表示为

$$R_k = B \log_2(1 + \gamma_k) \quad (7)$$

其中  $B$  为带宽。

## 1.2 计算模型

本文采用部分卸载方案,即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。定义  $\mu_k$  表示用户  $k$  的任务卸载因子,  $(1 - \mu_k)$  则表示用户  $k$  在本地计算的占比,其中  $\mu_k \in [0, 1]$ 。

### 1.2.1 边缘计算

$L_k$  表示数据长度,因此用户  $k$  发送到 MEC 服务器的任务卸载时间表示为

$$T_k^{\text{off}} = \frac{\mu_k L_k}{R_k} \quad (8)$$

同时,用户  $k$  发送数据到 MEC 服务器的能耗可以表示为

$$E_k^{\text{off}} = T_k^{\text{off}} p_k \quad (9)$$

### 1.2.2 本地计算

由前文可知,有  $(1 - \mu_k)$  部分的数据在本地计算,假设  $f_k^{\text{loc}}$  表示用户  $k$  的 CPU 频率(cycles per second),  $C_k$  表示用户  $k$  处理 1 比特数据需要的 CPU 数(cycles per bit),用户  $k$  处的本地计算时间可以表示为

$$T_k^{\text{loc}} = \frac{(1 - \mu_k) L_k C_k}{f_k^{\text{loc}}} \quad (10)$$

MEC 服务器的 CPU 内核功耗为  $\kappa_k (f_k^{\text{loc}})^3$ ,而  $\kappa_k$  表示在用户  $k$  处每个 CPU 周期的有效电容系数,因此能耗可以表示为

$$E_k^{\text{loc}} = \kappa_k (1 - \mu_k) L_k C_k (f_k^{\text{loc}})^2 \quad (11)$$

## 2 问题建模与解决方案

考虑这样一个场景,由于基站处的 MEC 服务器计算能力强大,与卸载时间相比,基站处的计算时间可以忽略,同时基站到 MEC 服务器的传输时间很小,也可以忽略。从基站到用户的下行时间不考虑,因为基站计算完的结果通常较小,同时基站处

的传输能力比用户的传输能力强很多,因此相较于上行传输时间,下行的传输时间可以忽略,并且下行主要从基站角度考虑,而本文主要考虑用户角度。因此本文的目标是通过优化任务分配、传输功率、RIS相移以及波束成形,来最小化用户之间的传输时延。因此,优化问题可以表示为(P0)

$$(P0) \min_{w, \theta, \mu, p} \max \left\{ \frac{(1 - \mu_k) L_k C_k}{f_k^{\text{loc}}}, T_k^{\text{off}} \quad \forall k \right\} \quad (12)$$

$$\text{s. t. } 0 \leq \mu_k \leq 1 \quad \forall k \quad (12a)$$

$$0 \leq P_k \leq P_{\max} \quad \forall k \quad (12b)$$

$$E_k^{\text{off}} + E_k^{\text{loc}} \leq E_{\max} \quad \forall k \quad (12c)$$

$$\| \mathbf{G}\mathbf{O}\mathbf{h}_{r,1} + \mathbf{h}_{d,1} \|^2 \leq \| \mathbf{G}\mathbf{O}\mathbf{h}_{r,2} + \mathbf{h}_{d,2} \|^2 \leq \dots \leq \| \mathbf{G}\mathbf{O}\mathbf{h}_{r,K} + \mathbf{h}_{d,K} \|^2 \quad (12d)$$

$$\| w_k \|^2 = 1 \quad \forall k \quad (12e)$$

$$0 \leq \theta \leq 2\pi \quad (12f)$$

其中,  $w = [w_1, w_2, \dots, w_K]$ ,  $\mu = [\mu_1, \mu_2, \dots, \mu_K]$ ,  $P = [P_1, P_2, \dots, P_K]$ 。式(12a)表示任务分配因子约束,式(12b)表示传输功率约束,  $P_{\max}$  表示用户的最大传输功率,式(12c)表示用户的总能耗约束,  $E_{\max}$  表示用户的最大能耗,式(12d)表示基站处的解码顺序,式(12e)表示均衡器的归一化功率要求,式(12f)表示RIS相移约束。

由于不同变量存在耦合的情况,因此问题(P0)是个非凸问题,难以直接求解,为了解决该问题,本文采用交替优化方法,利用SDR和SCA方法联合优化问题(P0),进一步将其转化为凸问题进行求解。为了便于分析问题(P0),根据文献[14],引入两个引理。

**引理1** 对于不同的用户,边缘计算的时延最小问题表示为

$$\min_{\{\mu_k, \mu_{k'}, P_k, P_{k'}\}} \max \{ T_k^{\text{off}}, T_{k'}^{\text{off}} \quad \forall k \neq k' \} \quad (13)$$

当  $T_k^{\text{off}} = T_{k'}^{\text{off}}, \forall k \neq k'$  时,  $\{\mu_k, p_k\}$  的最优解才会出现。

**证明** 不同用户卸载传输时间最小化问题写作:  $\min_{\mu, P} \max \{ T_1^{\text{off}}, T_2^{\text{off}}, \dots, T_K^{\text{off}} \}$  记最小时延为  $T^*$ , 最优解为  $\{\mu^*, P^*\}$ , 这个最优解只有当  $T_k^{\text{off}} = T_{k'}^{\text{off}}, \forall k \neq k'$  时才会获得,证明采用反证法。

假设用户  $U_{k'}$  先进行解调,最优解在  $T_k^{\text{off}} > T_{k'}^{\text{off}}, k < k'$  时取得。因此  $T^*$  对应的最优解为  $\{\mu_k^*, \mu_{k'}^*, P_k^*, P_{k'}^*\}$ 。如果增加  $P_k^*$  到  $\hat{P}_k$ , 此时  $T_k^{\text{off}}$  会减小。由于  $P_{k'}^*$  保持不变,因此  $T_{k'}^{\text{off}}$  会增加。一定会存在一组解  $\{\hat{P}_k, P_{k'}^*\}$  满足  $T_k^{\text{off}}(P_k^*, P_{k'}^*) > \hat{T}_k^{\text{off}}(\hat{P}_k, P_{k'}^*) =$

$\hat{T}_{k'}^{\text{off}}(\hat{P}_k, P_{k'}^*) > T_{k'}^{\text{off}}(P_k^*, P_{k'}^*)$ , 这进一步减小时延,因此  $\hat{T}_k^{\text{off}}(\hat{P}_k, P_{k'}^*) = \hat{T}_{k'}^{\text{off}}(\hat{P}_k, P_{k'}^*)$  是最优值,这与假设  $\{\mu_k^*, \mu_{k'}^*, P_k^*, P_{k'}^*\}$  是最优解矛盾。

假设最优解在  $T_k^{\text{off}} < T_{k'}^{\text{off}}, k < k'$  时取得,因此  $T^*$  对应的最优解为  $\{\mu_k^*, \mu_{k'}^*, P_k^*, P_{k'}^*\}$ 。如果减小  $P_k^*$  到  $\hat{P}_k$ , 此时  $T_k^{\text{off}}$  会增加。由于  $P_{k'}^*$  保持不变,因此  $T_{k'}^{\text{off}}$  会减小。一定会存在一组解  $\{\hat{P}_k, P_{k'}^*\}$  满足  $T_k^{\text{off}}(P_k^*, P_{k'}^*) < \hat{T}_k^{\text{off}}(\hat{P}_k, P_{k'}^*) = \hat{T}_{k'}^{\text{off}}(\hat{P}_k, P_{k'}^*) < T_{k'}^{\text{off}}(P_k^*, P_{k'}^*)$ , 这进一步减小时延,因此  $\hat{T}_k^{\text{off}}(\hat{P}_k, P_{k'}^*) = \hat{T}_{k'}^{\text{off}}(\hat{P}_k, P_{k'}^*)$  是最优值,这与假设  $\{\mu_k^*, \mu_{k'}^*, P_k^*, P_{k'}^*\}$  是最优解矛盾。

因此,最优解只有当  $T_k^{\text{off}} = T_{k'}^{\text{off}}, \forall k \neq k'$  时才会获得。(证毕)

由于采用NOMA方案,其中不同用户在同一频段同时进行任务卸载,在引理1的基础上,引入引理2,进一步简化问题。

**引理2** 如果所有用户到MEC服务器的卸载时延相同,即

$$T = T_1^{\text{off}} = T_2^{\text{off}} = \dots = T_K^{\text{off}} \quad (14)$$

那么,式(14)可以化简成

$$\hat{T}_k^{\text{off}} = \frac{\sum_{i=1}^k \mu_i L_i}{B \log_2 \left( 1 + \frac{\sum_{i=1}^k P_i |w_i^H (\mathbf{G}\mathbf{O}\mathbf{h}_{r,i} + \mathbf{h}_{d,i})|^2}{\sigma^2} \right)} \quad \forall k \quad (15)$$

证明:因为  $\frac{a_1}{b_1} = \frac{a_2}{b_2} = \dots = \frac{a_K}{b_K}$  可以写成  $\frac{a_1}{b_1} = \frac{e_2 a_2}{e_2 b_2} = \dots = \frac{e_K a_K}{e_K b_K}$ , 其中  $e_k = \frac{a_k}{a_1} = \frac{b_k}{b_1}$ 。故  $\frac{a_1}{b_1} = \frac{a_1(1 + e_2 + \dots + e_K)}{b_1(1 + e_2 + \dots + e_K)} = \frac{a_1 + \dots + a_K}{b_1 + \dots + b_K}$ , 因此  $T =$

$$\frac{\mu_1 L_1 + \dots + \mu_K L_K}{R_1 + \dots + R_K} = \frac{\sum_{i=1}^k \mu_i L_i}{\sum_{i=1}^k R_i}$$

$$B \log_2 \left( 1 + \frac{|z_1|^2 P_1}{\sigma^2} \right) + B \log_2 \left( 1 + \frac{|z_2|^2 P_2}{|z_1|^2 P_1 + \sigma^2} \right) + \dots +$$

$$B \log_2 \left( 1 + \frac{|z_k|^2 P_k}{\sum_{i=1}^{k-1} |z_i|^2 P_i + \sigma^2} \right) = B \log_2 \left( 1 + \frac{\sum_{i=1}^k |z_i|^2 P_i}{\sigma^2} \right)$$

其中,  $z_i = |w_i^H (\mathbf{G}\mathbf{O}\mathbf{h}_{r,i} + \mathbf{h}_{d,i})|^2$ 。因此  $\hat{T}_k^{\text{off}} =$

$$\frac{\sum_{i=1}^k \mu_i L_i}{B \log_2 \left( 1 + \frac{\sum_{i=1}^k P_i |w_i^H (\mathbf{G}\mathbf{O}\mathbf{h}_{r,i} + \mathbf{h}_{d,i})|^2}{\sigma^2} \right)} \quad \forall k$$

(证毕)。

基于引理1和引理2并根据文献[14],问题(P0)可以重写为问题(P1)。

$$(P1) \min_{w, \theta, \mu, P} \max \left\{ \frac{(1 - \mu_k)L_k C_k}{f_k^{\text{loc}}}, \hat{T}_k^{\text{off}} \quad \forall k \right\} \quad (16)$$

$$\text{s. t. } P_k \hat{T}_k^{\text{off}} + E_k^{\text{loc}} \leq E_{\max} \quad \forall k \quad (16a)$$

$$\text{式(12a, 12b, 12d-12f)} \quad (16b)$$

## 2.1 RIS相移优化

将问题(P1)分成3个子问题,对于子问题(P2),给定其余变量,优化RIS相移。由于问题(P1)的目标函数难以直接求解,因此引入松弛变量 $Y$ 便于后续处理。

$$(P2) \min_{Y, \theta} Y \quad (17)$$

$$\text{s. t. } (1 - \mu_k)L_k C_k f_k^{\text{loc}^2} \leq E_{\max} - P_k Y \quad \forall k \quad (17a)$$

$$\frac{\sum_{i=1}^k \mu_i L_i}{B \log_2 \left( 1 + \frac{\sum_{i=1}^k P_i |w_i^H (\mathbf{G}\Theta \mathbf{h}_{r,i} + \mathbf{h}_{d,i})|^2}{\sigma^2} \right)} \leq Y \quad \forall k \quad (17b)$$

$$0 \leq \theta \leq 2\pi \quad (17c)$$

$$\begin{aligned} \|\mathbf{G}\Theta \mathbf{h}_{r,1} + \mathbf{h}_{d,1}\|^2 &\leq \|\mathbf{G}\Theta \mathbf{h}_{r,2} + \mathbf{h}_{d,2}\|^2 \leq \\ \dots &\leq \|\mathbf{G}\Theta \mathbf{h}_{r,K} + \mathbf{h}_{d,K}\|^2 \end{aligned} \quad (17d)$$

由于式(17b)中变量是耦合的,难以直接处理,为了便于后续求解,进行如下变换。令 $\mathbf{Q}_k = \mathbf{G} \text{diag}(\mathbf{h}_{r,k})$ ,  $\mathbf{u} = [e^{-j\theta_1}, e^{-j\theta_2}, \dots, e^{-j\theta_K}]^H$ ,  $\tilde{\mathbf{u}} = [\mathbf{u}, 1]^H$ ,  $\mathbf{U} = \tilde{\mathbf{u}}\tilde{\mathbf{u}}^H$ ,可以进一步得到

$$\left| w_k^H (\mathbf{G}\Theta \mathbf{h}_{r,k} + \mathbf{h}_{d,k}) \right|^2 = \text{tr}(\mathbf{U}\Phi_k) \quad (18)$$

$$\|\mathbf{G}\Theta \mathbf{h}_{r,k} + \mathbf{h}_{d,k}\|^2 = \text{tr}(\mathbf{U}\varphi_k) \quad (19)$$

其中,  $\Phi_k = \begin{bmatrix} \mathbf{Q}_k^H \mathbf{w}_k \mathbf{w}_k^H \mathbf{Q}_k & \mathbf{Q}_k^H \mathbf{w}_k \mathbf{w}_k^H \mathbf{h}_{d,k} \\ \mathbf{h}_{d,k}^H \mathbf{w}_k \mathbf{w}_k^H \mathbf{Q}_k & \mathbf{h}_{d,k}^H \mathbf{w}_k \mathbf{w}_k^H \mathbf{h}_{d,k} \end{bmatrix}$  以及  $\varphi_k =$

$$\begin{bmatrix} \mathbf{Q}_k^H \mathbf{Q}_k & \mathbf{Q}_k^H \mathbf{h}_{d,k} \\ \mathbf{h}_{d,k}^H \mathbf{Q}_k & \mathbf{h}_{d,k}^H \mathbf{h}_{d,k} \end{bmatrix}.$$

因此子问题(P2)可以进一步改写为问题(P3),表示为

$$(P3) \min_{Y, U} Y \quad (20)$$

$$\text{s. t. } (1 - \mu_k)L_k C_k f_k^{\text{loc}^2} \leq E_{\max} - P_k Y \quad \forall k \quad (20a)$$

$$\sum_{i=1}^k \text{tr}(\mathbf{U}\Phi_i) P_i \geq \sigma^2 \left( 2^{\frac{\sum_{i=1}^k \mu_i L_i}{BY}} - 1 \right) \quad \forall k \quad (20b)$$

$$\text{tr}(\mathbf{U}\varphi_1) \leq \text{tr}(\mathbf{U}\varphi_2) \leq \dots \leq \text{tr}(\mathbf{U}\varphi_K) \quad (20c)$$

$$\mathbf{U} \geq 0 \quad (20d)$$

$$\mathbf{U}_{m,m} = 1 \quad (20e)$$

$$\text{rank}(\mathbf{U}) = 1 \quad (20f)$$

此时,在问题(P3)中,除了式(20f)秩一约束外,其余都是凸的,经过处理后便可利用CVX进行求解。对于秩一约束,采用SDR将其松弛,并对获得的结果进行高斯随机化或特征值分解,从而得到满足要求的最优解。

## 2.2 均衡器优化

给定其余优化变量,优化均衡器 $w$ 。为了最大化各个用户的信干噪比,采用MMSE准则设计均衡器 $w$ ,基站通过均衡器 $w_k \in \mathbb{C}^{M \times 1}$ 恢复用户 $k$ 的信号 $x_k$ ,因此用户 $k$ 的检测信号可以表示为 $\hat{x}_k = w_k^H y_k$ ,因此MSE矩阵表示为

$$\mathbf{E}_k = E[(\hat{x}_k - x_k)(\hat{x}_k - x_k)^H] \quad (21)$$

由于 $y_k = \sum_{i=1}^K \sqrt{P_i} (\mathbf{G}\Theta \mathbf{h}_{r,i} + \mathbf{h}_{d,i}) x_i + n$ ,记 $\mathbf{h}_i = \mathbf{h}_{d,i} + \mathbf{G}\Theta \mathbf{h}_{r,i}$ ,将其代入式(20)展开可得到

$$\begin{aligned} \mathbf{E}_k &= w_k^H \left( \sum_{i=1}^k \mathbf{h}_i \mathbf{h}_i^H P_i + \sigma^2 \mathbf{I}_M \right) w_k - \sqrt{P_k} w_k^H \mathbf{h}_k - \\ &\quad \sqrt{P_k} \mathbf{h}_k^H w_k + 1 \end{aligned} \quad (22)$$

利用 $\partial \mathbf{E}_k / \partial w_k = 0$ ,得到最优均衡器表达式为

$\left( \sum_{i=1}^k \mathbf{h}_i \mathbf{h}_i^H P_i + \sigma^2 \mathbf{I}_M \right)^{-1} \mathbf{h}_k \sqrt{P_k}$ ,对其归一化后,得到最优解为

$$w_k^* = \frac{\left( \sum_{i=1}^k \mathbf{h}_i \mathbf{h}_i^H P_i + \sigma^2 \mathbf{I}_M \right)^{-1} \mathbf{h}_k \sqrt{P_k}}{\left\| \left( \sum_{i=1}^k \mathbf{h}_i \mathbf{h}_i^H P_i + \sigma^2 \mathbf{I}_M \right)^{-1} \mathbf{h}_k \sqrt{P_k} \right\|} \quad (23)$$

## 2.3 分配因子和发送功率优化

对于子问题(P4),给定其余变量,优化分配因子 $\mu$ 和发送功率 $P$ 。问题可以简化为

$$(P4) \min_{t, P, \mu} t \quad (24)$$

$$\text{s. t. } \frac{(1 - \mu_k)L_k C_k}{f_k^{\text{loc}}} \leq t \quad \forall k \quad (24a)$$

$$\frac{\sum_{i=1}^k \mu_i L_i}{B \log_2 \left( 1 + \frac{\sum_{i=1}^k P_i |w_i^H (\mathbf{G}\Theta \mathbf{h}_{r,i} + \mathbf{h}_{d,i})|^2}{\sigma^2} \right)} \leq t \quad \forall k \quad (24b)$$

$$\kappa_k (1 - \mu_k) L_k C_k f_k^{\text{loc}^2} \leq E_{\max} - P_k t \quad \forall k \quad (24c)$$

$$0 \leq \mu_k \leq 1, \quad \forall k \quad (24d)$$

$$0 \leq P_k \leq P_{\max}, \quad \forall k \quad (24e)$$

子问题P3中约束式(24b)、式(24c)存在变量耦合,是非凸约束,为了将其变为凸问题,首先处理约束条件式(24b)。引入松弛变量 $x$ ,令

$$\frac{\sum_{i=1}^k P_i \left| \mathbf{w}_i^H (\mathbf{G}\Theta \mathbf{h}_{r,i} + \mathbf{h}_{d,i}) \right|^2}{\sigma^2} \geq x^2 \quad (25)$$

此时式(25)是凸约束,将其回代入约束条件式(24b),可得

$$x^2 \geq 2 \frac{\sum_{i=1}^k \mu_i L_i}{Bt} - 1 \quad (26)$$

此时,式(26)依然是非凸约束,再引入松弛变量 $y$ ,并令

$$y \geq \frac{\sum_{i=1}^k \mu_i L_i}{Bt} \quad (27)$$

将其回代式(26),并进行一阶泰勒展开,可得

$$\begin{cases} x^2 \geq 2y - 1 \\ 2x^{(n)}x - [x^{(n)}]^2 - 2y + 1 \geq 0 \end{cases} \quad (28)$$

其中, $x^{(n)}$ 代表第 $n$ 次迭代值,此时非凸约束式(24b)已经转化为式(27)和式(28),并且式(28)是凸约束,接着只需进一步处理式(27)。

对于式(27),引入松弛变量 $z$ ,并回代得

$$z^2 \geq \sum_{i=1}^k \mu_i L_i \quad (29)$$

$$Bty \geq z^2 \quad (30)$$

对于式(29)进行一阶泰勒展开得式(31),对于式(30)根据LMI约束转化为半正定矩阵的形式,得式(32),表达式为

$$2z^{(n)}z - [z^{(n)}]^2 - \sum_{i=1}^k \mu_i L_i \geq 0 \quad (31)$$

$$\begin{bmatrix} Bt & z \\ z & y \end{bmatrix} \succeq 0 \quad (32)$$

至此非凸约束式(24b)转变为凸约束式(25, 28, 31, 32),接着处理非凸约束式(24c),引入松弛变量 $m$ ,可以得到

$$P_k t \leq m \quad (33)$$

$$\kappa_k (1 - \mu_k) L_k C_k f_k^{\text{loc}^2} \leq E_{\max} - m \quad (34)$$

其中,式(33)依然是非凸约束,注意到 $P_k$ 和 $t$ 都是正数,利用基本不等式进行放缩,得到

$$\left( m_k^{(n)} t \right)^2 + \left( \frac{P_k}{m_k^{(n)}} \right)^2 \leq 2m \quad (35)$$

其中, $m_k^{(n)}$ 代表第 $n$ 次的迭代值,并取 $m_k^{(n)} = \sqrt{P_k^{(n-1)}/t^{(n-1)}}$ 。从而,非凸问题(P4)转化为问题(P5)

$$(P5) \quad \min_{t, P, \mu, x, y, z, m} t \quad (36)$$

$$\text{s. t.} \quad \frac{(1 - \mu_k) L_k C_k}{f_k^{\text{loc}}} \leq t, \forall k \quad (36a)$$

$$\frac{\sum_{i=1}^k P_i \left| \mathbf{w}_i^H (\mathbf{G}\Theta \mathbf{h}_{r,i} + \mathbf{h}_{d,i}) \right|^2}{\sigma^2} \geq x^2 \quad (36b)$$

$$2x^{(n)}x - [x^{(n)}]^2 - 2y + 1 \geq 0 \quad (36c)$$

$$2z^{(n)}z - [z^{(n)}]^2 - \sum_{i=1}^k \mu_i L_i \geq 0 \quad (36d)$$

$$\begin{bmatrix} Bt & z \\ z & y \end{bmatrix} \succeq 0 \quad (36e)$$

$$\kappa_k (1 - \mu_k) L_k C_k f_k^{\text{loc}^2} \leq E_{\max} - m, \forall k \quad (36f)$$

$$\left( m_k^{(n)} t \right)^2 + \left( \frac{P_k}{m_k^{(n)}} \right)^2 \leq 2m \quad (36g)$$

$$m_k^{(n)} = \sqrt{\frac{P_k^{(n-1)}}{t^{(n-1)}}}, \forall k \quad (36h)$$

此时子问题(P5)是凸问题,可以利用CVX进行求解。

### 3 仿真结果

在本节中,给出仿真结果和参数来评估其性能。考虑3个用户的情景,具体参数如表1所示。

表1 各参数取值

| 参数  | 值                   |
|---|---------------------|
| 基站天线数   | 2                   |
| 基站到RIS距离/m  | 60                  |
| 用户到RIS距离/m  | [50, 30, 20]        |
| 用户到基站距离/m   | [100, 80, 70]       |
| 莱斯因子  | 4                   |
| 路径损耗指数 $[\alpha_{d,k}, \alpha_{B,r}, \alpha_{r,k}]$   | [3.8, 2.3, 2.3]     |
| 噪声功率 $\sigma^2$ /dBm                                  | -120                |
| 数据长度 $l_k$ /bit                                       | $1 \times 10^6$     |
| 最大总能耗 $E_{\max}$ /J                                   | 4                   |
| 硬件参数 $\kappa_k$                                       | $1 \times 10^{-27}$ |
| 最大功率 $p_{\max}$ /W                                    | 0.01                |
| 信道带宽 $B$ /Hz  | $1 \times 10^5$     |
| CPU频率 $f_k^{\text{loc}}$ /(cycles $\cdot$ s $^{-1}$ ) | $1 \times 10^8$     |
| 每比特所需CPU周期 $C_k$ /(cycles $\cdot$ bit $^{-1}$ )       | 400                 |

本文将通过数值分析评估所提方案的性能。为了进行对比,还引入了4种基准方案,第一种是RIS-OMA(Orthogonal Multiple Access, OMA)与MEC网络下的时延,即使用正交多址技术卸载不同用户的任务数据;第二种是RIS-NOMA与MEC网络下时延,即使用非正交多址技术,但RIS相移不进行优化

而是随机取值;第三种是将莱斯因子设为0,即考虑瑞利信道下RIS-NOMA的时延;第四种是在不引入RIS的情况下NOMA-MEC的时延。

在图2中,绘制了不同RIS数下各方案的时延结果。仿真结果表明,随着RIS反射单元数目的增加,各方案都呈现出时延下降的趋势,这是因为在通信的传输受阻甚至中断的情况下,RIS通过创建用户设备和MEC服务器之间额外的反射链路,调节其反射相位矩阵后,可以提高信号的传输速率,从而降低用户的卸载延迟。然而,与RIS-OMA方案相比,RIS-NOMA的时延结果明显优于基准方案,这是因为NOMA可以实现频谱资源的复用,当天线数小于用户数时,用户数越多NOMA的性能往往会更好。与RIS-NOMA瑞利信道条件相比,由于缺少视距分量,与莱斯信道相比,信道条件比较差,因此时延会变高。与随机相位方案相比,由于RIS相位进行过优化,RIS-NOMA方案的时延明显更小,进一步表明优化RIS相位是很有必要的。

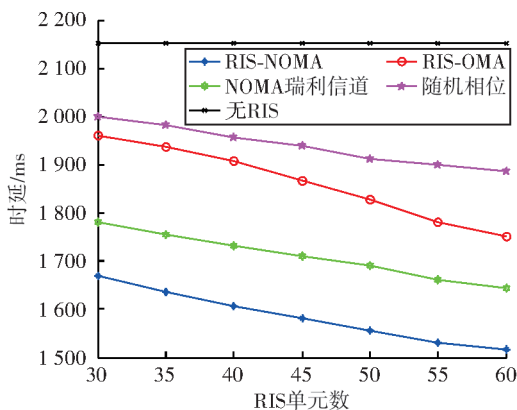


图2 RIS单元数对时延的影响

在图3中,当RIS单元数为35时,绘制了不同最大卸载功率下各方案的时延结果。仿真结果表明,RIS-NOMA的方案优于RIS-OMA,这是因为NOMA能够实现的卸载速率大于OMA,因此时延往往更低。同时,优化后的RIS相位能够显著改善信道条件,因此RIS-NOMA方案明显优于随机相位方案。相较于莱斯信道,由于瑞利信道缺失视距分量,因此瑞利信道方案时延会增加。如果不引入RIS,就无法提供额外的反射链路,因此无RIS的时延也会增加。此外,随着卸载功率增大,5种方案的时延都会减小。

图4给出不同RIS单元数和最大卸载功率情况下提出方法的收敛性能。随着交替优化的次数增

加,时延逐渐趋向于稳定值,在3到4次交替优化后,优化结果实现收敛。同时,不难看出,当卸载功率固定时,增大RIS单元数可以有效地减少时延,这是因为RIS能够改善信道条件;当RIS单元数固定时,增大卸载功率也能够降低时延。

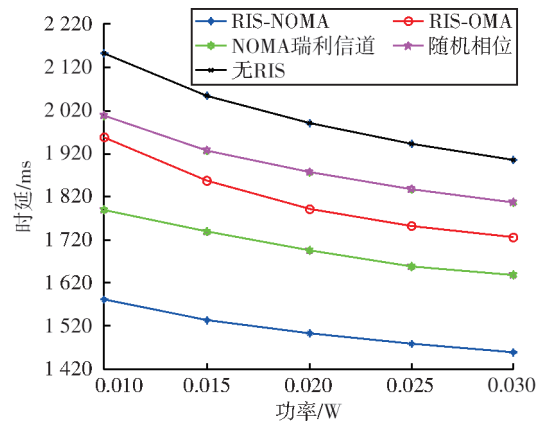


图3 卸载功率对时延的影响

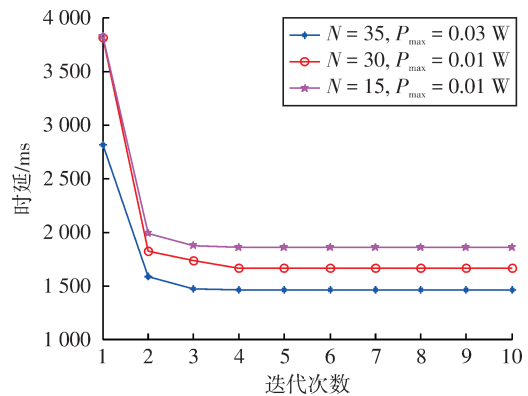


图4 算法收敛性

## 4 结束语

本文研究了多用户下上行RIS-NOMA-MEC网络中时延优化问题,通过优化RIS相移、卸载功率、波束成型和分配因子,使得不同用户间的传输时延最小。对于非凸问题形式,利用SDR算法优化相移,同时采用SCA算法优化算法功率和分配因子,采用MMSE准则设计均衡器,对3个子问题进行交替优化得到优化结果。仿真结果表明,RIS-NOMA结合MMSE均衡器的时延明显小于其余两种基准方案。

### 参考文献:

- [1] ELMOSSALLAMY M A, ZHANG H L, SONG L Y, et al. Reconfigurable intelligent surfaces for wireless communications: principles, challenges, and opportuni-

- ties [J]. *IEEE Transactions on Cognitive Communications and Networking*, 2020, 6(3): 990–1002.
- [2] ZHENG B X, WU Q Q, ZHANG R. Intelligent reflecting surface-assisted multiple access with user pairing: NOMA or OMA? [J]. *IEEE Communications Letters*, 2020, 24(4): 753–757.
- [3] HOU T W, LIU Y W, SONG Z Y, et al. Reconfigurable intelligent surface aided NOMA networks [J]. *IEEE Journal on Selected Areas in Communications*, 2020, 38(11): 2575–2588.
- [4] ZHANG W L, XIONG K, ZHANG R C, et al. SEE maximization in RIS-aided network with RSMA: a PPO-SCF method [J]. *IEEE Wireless Communications Letters*, 2024, 13(12): 3315–3319.
- [5] 施丽琴, 刘璇, 卢光跃. 面向数据压缩的NOMA-MEC系统能耗最小化研究[J]. *电子与信息学报*, 2024, 46(7): 2888–2897.  
SHI Liqin, LIU Xuan, LU Guangyue. Research on energy consumption minimization for a data compression based NOMA-MEC system [J]. *Journal of Electronics & Information Technology*, 2024, 46(7): 2888–2897. (in Chinese)
- [6] REN J K, YU G D, CAI Y L, et al. Latency optimization for resource allocation in mobile-edge computation offloading [J]. *IEEE Transactions on Wireless Communications*, 2018, 17(8): 5506–5519.
- [7] WU Q Q, ZHANG S W, ZHENG B X, et al. Intelligent reflecting surface-aided wireless communications: a tutorial [J]. *IEEE Transactions on Communications*, 2021, 69(5): 3313–3351.
- [8] ZHOU F S, YOU C S, ZHANG R. Delay-optimal scheduling for IRS-aided mobile edge computing [J]. *IEEE Wireless Communications Letters*, 2021, 10(4): 740–744.
- [9] BAI T, PAN C H, DENG Y S, et al. Latency minimization for intelligent reflecting surface aided mobile edge computing [J]. *IEEE Journal on Selected Areas in Communications*, 2020, 38(11): 2666–2682.
- [10] WANG Y Y, NIU J P, CHEN G J, et al. RIS-aided latency-efficient MEC HetNet with wireless backhaul [J]. *IEEE Transactions on Vehicular Technology*, 2024, 73(6): 8705–8719.
- [11] HUANG A, QU L, KHABBAZ M J. Latency-aware computation offloading in multi-RIS-assisted edge networks [J]. *IEEE Open Journal of the Communications Society*, 2024, 5: 1204–1221.
- [12] DING Z G, NG D W K, SCHOBER R, et al. Delay minimization for NOMA-MEC offloading [J]. *IEEE Signal Processing Letters*, 2018, 25(12): 1875–1879.
- [13] WU Y, QIAN L P, NI K J, et al. Delay-minimization nonorthogonal multiple access enabled multi-user mobile edge computation offloading [J]. *IEEE Journal of Selected Topics in Signal Processing*, 2019, 13(3): 392–407.
- [14] FANG F, XU Y Q, DING Z G, et al. Optimal resource allocation for delay minimization in NOMA-MEC networks [J]. *IEEE Transactions on Communications*, 2020, 68(12): 7867–7881.

(责任编辑:李小溪)