

文章编号:1007-5321(2019)03-0014-07

DOI:10.13190/j.jbupt.2018-256

# 无线 MEC 系统中队列状态感知的卸载和传输联合优化

滕颖蕾<sup>1</sup>, 刘薇<sup>1</sup>, 欧阳卫平<sup>2</sup>, 李鸷<sup>3</sup>, 宋梅<sup>1</sup>

(1. 北京邮电大学 电子工程学院, 北京 100876; 2. 华为技术有限公司, 深圳 518129;

3. 中国铁道科学研究院集团有限公司标准计量研究所, 北京 100081)

**摘要:** 针对多用户-多移动边缘计算服务器系统的动态计算任务卸载问题, 基于用户端和服务端的任务队列模型, 以系统的长期平均时延和长期平均功耗为优化目标, 求解最优的卸载策略及相应的上行预编码。通过李雅普诺夫优化方法将长期平均问题转化成单阶段目标优化问题, 考虑到卸载策略和预编码之间存在范数约束关系, 通过连续近似和半正定松弛, 可转化成典型的 DC 规划求预编码解问题。仿真结果表明, 所提方案比传统方法具有更低的时延和功耗。

**关键词:** 移动边缘计算; 预编码; 李雅普诺夫优化; 凸优化

中图分类号: TN929.53

文献标志码: A

## Queue-Aware Joint Optimization of Offloading and Transmission in Wireless Mobile Edge Computing Systems

TENG Ying-lei<sup>1</sup>, LIU Wei<sup>1</sup>, OUYANG Wei-ping<sup>2</sup>, LI Kun<sup>3</sup>, SONG Mei<sup>1</sup>

(1. School of Electronic Engineering, Beijing University of Posts and Telecommunications, Beijing 100876, China;

2. Huawei Technologies CO., LTD, Shenzhen 518129, China;

3. Institute of Standard Metrology, China Academy of Railway Sciences, Beijing 100081, China)

**Abstract:** Considering task queue model on both user and mobile edge computing (MEC) server side, a dynamic computing task offloading problem in multi-user-multi-MEC-server system is proposed. To find the optimal offloading and corresponding uplink precoding strategy, a long-term average overhead containing delay and power consumption of the whole system optimization problem is formulated. The original problem is transformed into a single-stage cost target optimization problem based on Lyapunov optimization method and further converted into a typical DC programming problem by utilizing successive approximation and semi-definite relaxation. Simulation shows that the proposed scheme characterized by optimizing precoding design can meet the lower delay and power consumption requirements.

**Key words:** mobile edge computing; precoding; Lyapunov optimization; convex optimization

随着第5代移动通信系统时代的到来, 移动设备和节点数量呈现爆炸式增长, 同时随着虚拟现实、

增强现实、车联网等新兴技术的发展, 移动业务在呈现多样化特征的同时对计算资源的需求也日益提

收稿日期: 2018-10-16

基金项目: 国家重点研发计划项目(2018YFB1201500); 国家自然科学基金项目(61771072); 北京市自然科学基金项目(L171011); 北京市重大专项项目(Z181100003118012); 中国铁路总公司科技研究开发计划项目(J2018G004)

作者简介: 滕颖蕾(1983—), 女, 副教授, 博士生导师, E-mail: lilytengt@bupt.edu.cn.

升. 然而, 由于边缘移动设备计算能力和能源存储能力受限, 无法执行计算密集型和超高时延要求的新兴业务. 为解决这一难题, 传统网络架构中, 任务被卸载到计算资源丰富的集中式云平台上处理, 从而降低移动设备的自身能耗, 延长设备的使用时间. 然而数据从边缘传输到云端以及计算结果从云端返回到边缘用户的过程会耗费大量时间, 任务执行的时延要求将很难满足. 此外, 大量边缘业务涌向云端会增加网络链路的传输压力和核心网的负担, 甚至造成网络瘫痪. 针对传统云网络架构中存在的问题, 移动边缘计算 (MEC, mobile edge computing) 应运而生<sup>[1]</sup>, MEC 是云计算网络架构的边缘化演进, 旨在将计算和存储资源推进到网络的边缘, 并进行分布式部署, 从而在满足时延要求的同时处理高计算要求的新兴业务.

MEC 系统中卸载策略和联合无线电源优化问题是当前 MEC 领域的研究热点之一. Wang 等<sup>[2]</sup>针对单个 MEC 服务器系统提出了用于计算卸载和干扰管理的集成框架; Xu 等<sup>[3]</sup>提出以用户为中心的移动性管理方案, 在保证通信能耗约束的前提下最大化单个用户的边缘计算性能; Han 等<sup>[4]</sup>基于随机几何网络建模, 研究链路和服务器稳定性约束的系统时延问题, 包括通信和排队计算时延, 并分析了在轻重负载、不同网络密度等情况下的系统性能. 综上, 目前国内外很多学者一致认为卸载时延是系统的关键指标, 但是多数卸载时延仅考虑通信和计算方面, 对由排队所造成的时延研究不足, Han 等<sup>[4]</sup>对排队时延的研究也仅是在下界的推导, 未对联合问题进行优化.

笔者提出多用户-多移动边缘计算服务器系统, 综合考虑信道状态和任务队列状态变化的马尔可夫特性, 以长期平均时延和长期平均功耗的联合优化为目标构建马尔可夫决策问题, 利用李雅普诺夫优化方法将系统的稳定性约束条件转化成确定性的约束条件, 通过半正定松弛处理, 将非凸的确定性优化问题转为凸函数优化, 用成熟的凸优化理论求解最优的预编码决策.

## 1 系统模型

多用户的移动边缘计算传输系统如图 1 所示. 单核移动设备有  $K$  个, 集合表示为  $\mathcal{K} = \{1, 2, 3, \dots, K\}$ ; MEC 服务器个数为  $M$ , 集合为  $\mathcal{M} = \{1, 2, \dots, M\}$ . 用户端是单天线的, MEC 服务器端的天线数目为  $N_s$ . 采取用户为中心的服务模式, 任意用户可由

任意若干个服务器提供服务. 系统带宽为  $W$ , 时间被等分成长度为  $\tau$  的离散时间间隔, 用集合  $\mathcal{T} = \{0, 1, 2, \dots\}$  表示,  $t$  表示第  $t$  个决策时间间隔.

时刻  $t$ , 用户端  $k$  上存在的任务队列记作  $Q_k(t)$ , 用户  $k$  卸载到 MEC 服务器端处理的任务队列记作  $T_k(t)$ . 当用户端产生计算任务时, 用户选择将计算任务卸载至 MEC 服务器端处理或者直接在本地处理, 任务卸载调度直接影响本地或 MEC 服务器端的任务队列状态. 若选择在本地处理, 将会产生本地功耗; 上传至 MEC 服务器端处理则会产生传输功耗与 MEC 服务器端计算功耗.

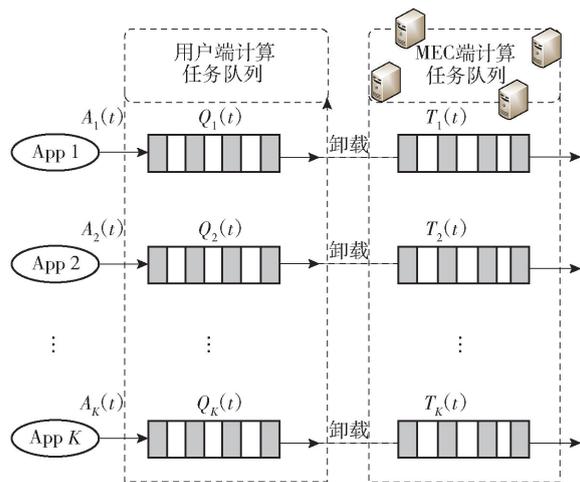


图 1 多用户-多移动边缘计算服务器系统模型

### 1.1 信道模型

令  $\mathbf{H}_k(t) \in \mathbb{C}^{N_r \times 1}$  和  $\mathbf{F}_k(t) \in \mathbb{C}^{1 \times N_t}$  分别表示在第  $t$  个时刻用户  $k$  到 MEC 服务器端的信道编码和预编码. 假定在每个决策间隔  $\tau$  内  $\mathbf{H}_k(t)$  保持不变, 其中每一个元素均服从期望为 0、方差为 1 的标准正态分布,  $\mathbf{H}_k(t)$  与  $k$  相互独立. 若给定信道状态信息  $\mathbf{H} = \{\mathbf{H}_k(t) : \forall k \in \mathcal{K}\}$ ,  $\mathbf{F} = \{\mathbf{F}_k(t) : \forall k \in \mathcal{K}\}$ , 则用户  $k$  将任务上传到 MEC 服务器端的信干噪比为

$$\theta_k = \frac{|\mathbf{H}_k(t) \mathbf{F}_k(t)|^2}{1 + \sum_{i=1, i \neq k}^K |\mathbf{H}_k(t) \mathbf{F}_i(t)|^2},$$

相应的传输速率  $R_k$  的表达式为

$$R_k(\mathbf{H}(t), \mathbf{F}(t)) = \alpha_k(t) \text{Wlb} \left( 1 + \frac{|\mathbf{H}_k(t) \mathbf{F}_k(t)|^2}{1 + \sum_{i=1, i \neq k}^K |\mathbf{H}_k(t) \mathbf{F}_i(t)|^2} \right)$$

其中:  $\boldsymbol{\alpha}(t) = \{\alpha_1(t), \alpha_2(t), \dots, \alpha_K(t)\}$  为分配的带宽向量, 满足  $\sum_{k=1}^K \alpha_k(t) \leq 1$ . 假设带宽均匀分配, 对

于每个用户  $k$ ,  $\alpha_k(t) = 1/K$ .

## 1.2 用户端任务队列模型

假设  $\mathbf{A}(t) = \{A_k(t) : k \in \mathcal{K}\}$  表示时刻  $t$  到达  $K$  个用户的比特数量,  $A_k(t)$  表示在第  $t$  个时刻到达用户  $k$  的计算任务量, 并且在第  $t+1$  个时刻立即被计算处理. 在不同的决策时间间隔,  $A_k(t)$  独立且服从泊松分布,  $A_k(t) \in [A_{k,\min}, A_{k,\max}]$ ,  $E[A_k(t)] = \lambda_k$ ,  $k \in \mathcal{K}$ . 令  $\{\lambda_k : \forall k \in \{1, 2, \dots, K\}\}$  属于系统的稳定域, 以保证系统的稳定性.

$\mathbf{Q}(t) = \{Q_k(t) : \forall k \in \mathcal{K}\}$  为  $t$  时刻所有用户端计算任务的队列状态信息.  $Q_k(t) \in [0, +\infty)$  表示  $t$  时刻用户  $k$  上的任务队列状态信息. 假设所有的移动设备独立运作, 并且执行细粒度任务(比特数), 则用户  $k$  上计算任务的动态队列表示为

$$Q_k(t+1) = [Q_k(t) - D_{l,k}(t) - D_{r,k}(t)]^+ + A_k(t), t \in \mathcal{T}$$

其中:  $[y]^+ = \max\{y, 0\}$ ,  $D_{l,k}(t)$  为用户  $k$  在第  $t$  个时刻分配到本地处理的计算任务量, 它取决于本地 CPU 此刻的计算能力;  $D_{r,k}(t)$  为  $t$  时刻用户  $k$  卸载到 MEC 服务器端的任务量, 与传输速率  $R_k$  有关. 假设卸载策略为  $S_k(t)$ ,  $S_k(t) = 1$  代表用户  $k$  选择将计算任务卸载到 MEC 服务器端执行,  $S_k(t) = 0$  表示任务留在本地处理的策略.  $D_{l,k}(t)$  和  $D_{r,k}(t)$  与对应的卸载策略相关, 即  $D_{l,k}(t) = (1 - S_k(t)) D_{\text{normal}}^l$ . 其中:  $D_{\text{normal}}^l$  为单位时间内在本地处理的计算任务量,  $D_{\text{normal}}^l = f_k/X_l$ ,  $X_l$  为衡量用户端 CPU 运行速度的指标, 即处理每比特任务量所需要的 CPU 周期数,  $f_k$  为用户端计算频率, 且在相当长的时间内恒定不变.  $D_{r,k}(t) = S_k(t) R_k(\mathbf{H}(t), \mathbf{F}(t)) \tau$ .

## 1.3 MEC 服务器端任务队列模型

令  $\mathbf{T}(t) \triangleq [T_1(t), T_2(t), \dots, T_K(t)]$  表示各用户卸载到 MEC 服务器端的计算任务队列状态信息.  $T_k(t)$  为  $t$  时刻用户  $k$  卸载到 MEC 服务器端的计算任务队列, 动态队列方程表示为

$$T_k(t+1) = [T_k(t) - D_{s,k}(t)]^+ + D_{r,k}(t)$$

其中:  $D_{s,k}(t)$  表示在第  $t$  个时隙 MEC 服务器端处理的来自用户  $k$  的计算任务量,  $D_{s,k}(t) = f_{C,m}/X_{\text{ser}}$ ,  $X_{\text{ser}}$  为衡量 MEC 服务器端 CPU 运行速度的指标,  $f_{C,m}$  为 MEC 服务器端计算频率, 在相当长的时间内是恒定值. 从 MEC 服务器端队列表达式可以看出, 只有没有处理完的任务量才会被存储在 MEC 服务器端任务队列中, 调节预编码也可以控制 MEC 服务器端的

任务队列长度. 对所有用户  $k$ , 令  $Q_k(0) = T_k(0) = 0$ .

根据信道状态、MEC 服务器端及用户端任务队列状态的马尔可夫转移特性, 通过最优化预编码决策, 用户可以选择将自己的计算任务上传到 MEC 服务器端或者直接在本地处理, 达到优化系统时延的目的.

## 1.4 本地和 MEC 服务器端的功耗模型

根据文献[5],  $t$  时刻, 用户  $k$  的本地功耗  $p_{l,k}(t)$  以及 MEC 服务器端的计算功耗  $p_{\text{ser},k}(t)$  分别为

$$\left. \begin{aligned} p_{l,k}(t) &= (1 - S_k(t)) \cdot \mathcal{K}_{\text{mob},k} f_k^3(t), \quad k \in \mathcal{K} \\ p_{\text{ser},k}(t) &= S_k(t) \sum_{m \in \mathcal{M}} \mathcal{K}_{\text{ser},m} f_{C,m}^3(t), \quad k \in \mathcal{K}, m \in \mathcal{M} \end{aligned} \right\} \quad (1)$$

其中:  $\mathcal{K}_{\text{mob},k}$  和  $\mathcal{K}_{\text{ser},m}$  分别为第  $k$  个移动设备和第  $m$  个 CPU 的有效开关电容, 这 2 个参数为常数, 仅与用户设备和服务器的硬件设备有关.

## 2 优化问题

### 2.1 系统的平均功耗

系统产生功耗包括 2 个方面: 一是传输功耗; 二是在本地以及服务器端处理计算任务产生的计算功耗. 系统的平均功耗表达式为

$$\bar{P}_{\Sigma}(t) \triangleq \lim_{T \rightarrow +\infty} \frac{1}{T} \sum_{t=0}^{T-1} E \left[ \sum_{k \in \mathcal{K}} (w_k (p_{\text{tx},k}(t) + p_{l,k}(t)) + w_M p_{\text{ser},k}(t)) \right] \quad (2)$$

其中:  $w_k$ 、 $w_M$  分别为用户端、MEC 服务器端的功耗权重系数,  $p_{\text{tx},k}(t)$  表示传输功耗. 卸载策略  $S_k(t)$  与上行预编码之间存在如下关系:

$$S_k(t) = \|\|\mathbf{F}_k(t)\|^2\|_{l_0} \quad (3)$$

其中:  $\|\cdot\|$  为 2-范数,  $\|\cdot\|_{l_0}$  为 0-范数. 给定信道状态信息  $\mathbf{H} = \{\mathbf{H}_k(t) : \forall k \in \mathcal{K}\}$ ,  $\mathbf{F} = \{\mathbf{F}_k(t) : \forall k \in \mathcal{K}\}$ , 若  $\mathbf{F}_k(t) = \mathbf{0}$ , 表示  $t$  时刻用户  $k$  选择将计算任务留在本地处理;  $\mathbf{F}_k(t) \neq \mathbf{0}$ , 则表示用户  $k$  将计算任务上传至 MEC 服务器端处理. 传输功耗  $p_{\text{tx},k}(t) = \|\|\mathbf{F}_k(t)\|^2\|_{l_0}$ , 则系统的平均功耗可以表示为

$$\bar{P}_{\Sigma}(t) \triangleq \lim_{T \rightarrow +\infty} \frac{1}{T} \sum_{t=0}^{T-1} E \left[ \sum_{k \in \mathcal{K}} (w_k (\|\|\mathbf{F}_k(t)\|^2\|_{l_0} + (1 - \|\|\mathbf{F}_k(t)\|^2\|_{l_0}) K_{\text{mob},k} f_k^3(t)) + w_M \|\|\mathbf{F}_k(t)\|^2\|_{l_0} \sum_{m \in \mathcal{M}} \mathcal{K}_{\text{ser},m} f_{C,m}^3(t)) \right]$$

### 2.2 系统的平均时延

在无限时域中, 时延是衡量系统性能的重要特

性之一. 通过用户端及 MEC 服务器端的任务队列长度之和来度量系统的时延特性, 可以得到系统的平均时延表达式为

$$\bar{q}_{\Sigma,k}(t) \triangleq \lim_{T \rightarrow +\infty} \frac{1}{T} \sum_{t=0}^{T-1} E \left[ \sum_{k \in \mathcal{K}} (Q_k(t) + T_k(t)) \right],$$

$$k \in \mathcal{K}$$

优化无限时域内系统的平均时延本质上是优化系统的任务队列动态演变趋势. 用户可以调节预编码使系统的队列长度向最优的方向演变, 这是典型的随机优化模型问题.

### 2.3 总的优化问题

基于系统成本模型分析<sup>[6-7]</sup>, 考虑平均功耗和时延的联合优化, 总的优化问题数学表达式为

$$\min_{X(t)} \bar{P}_{\Sigma}(t)$$

$$\text{s. t. } \lim_{T \rightarrow \infty} \frac{1}{T} \sum_{t=0}^T \left( \sum_{k \in \mathcal{K}} \bar{q}_{\Sigma,k}(t) \right) < \infty$$

$$\theta_k \geq \rho_k, \forall k \in \mathcal{K} \quad (4)$$

其中:  $\rho_k$  为信号质量阈值, 表示用户任务卸载时的传输质量要求.

整理得单阶段的功耗方程为

$$c(t) = \sum_{k \in \mathcal{K}} \left( w_k ( \| \mathbf{F}_k(t) \|^2 + \mathcal{N}_{\text{mob},k} f_k^3(t) ) + \left( w_M \sum_{m \in \mathcal{M}} \mathcal{N}_{\text{ser},m} f_{C,m}^3(t) - w_k \mathcal{N}_{\text{mob},k} f_k^3(t) \right) \times \| \mathbf{F}_k(t) \|^2 \| l_0 \right)$$

用连续方程近似单阶段功耗方程中的离散函数, 其中,

$$\| \mathbf{F}_k(t) \|^2 \| l_0 = \lim_{\alpha \rightarrow 0} \frac{\ln(1 + \| \mathbf{F}_k(t) \|^2 \alpha^{-1})}{\ln(1 + \alpha^{-1})}$$

对于充分小的  $\alpha$ , 有

$$\| \mathbf{F}_k(t) \|^2 \| l_0 \approx \frac{\ln(1 + \| \mathbf{F}_k(t) \|^2 \alpha^{-1})}{\ln(1 + \alpha^{-1})}$$

成立, 则单阶段功耗方程可以进一步重写为

$$c(t) \approx \sum_{k \in \mathcal{K}} \left( \left( w_M \sum_{m \in \mathcal{M}} \mathcal{N}_{\text{ser},m} f_{C,m}^3(t) - w_k \mathcal{N}_{\text{mob},k} f_k^3(t) \right) \frac{\ln(1 + \| \mathbf{F}_k(t) \|^2 \alpha^{-1})}{\ln(1 + \alpha^{-1})} + w_k ( \| \mathbf{F}_k(t) \|^2 + \mathcal{N}_{\text{mob},k} f_k^3(t) ) \right)$$

### 3 问题求解

利用李雅普诺夫优化方法<sup>[3]</sup>对优化问题中的队列稳定性条件做如下处理, 令

$$L(\boldsymbol{\Theta}(t)) = \frac{1}{2} \sum_{k \in \mathcal{K}} [Q_k^2(t) + T_k^2(t)]$$

其中  $\boldsymbol{\Theta}(t) \triangleq [Q(t), T(t)]$ . 李雅普诺夫加值函数为

$$\Delta(\boldsymbol{\Theta}(t)) = L(\boldsymbol{\Theta}(t+1)) - L(\boldsymbol{\Theta}(t))$$

李雅普诺夫加值惩罚函数为

$$\Delta_V(\boldsymbol{\Theta}(t)) = \Delta(\boldsymbol{\Theta}(t)) + VE[P_{\Sigma}(t) | \boldsymbol{\Theta}(t)]$$

其中  $V$  为李雅普诺夫控制参数.

对用户端的动态队列  $Q_k^2(t+1)$  做如下不等式处理, 其中,  $D_{\Sigma,k}(t) = D_{l,k}(t) + D_{r,k}(t)$ :

$$Q_k^2(t+1) = ([Q_k(t) - D_{\Sigma,k}(t)]^+)^2 + A_k^2(t) + 2A_k(t)[Q_k(t) - D_{\Sigma,k}(t)]^+ \leq (Q_k(t) - D_{\Sigma,k}(t))^2 + A_k^2(t) + 2A_k(t)Q_k(t) \leq Q_k^2(t) - 2Q_k(t)(D_{\Sigma,k}(t) - A_k(t)) + A_k^2(t) + D_{\Sigma,k}^2(t) \quad (5)$$

把不等式(5)右边的  $Q_k^2(t)$  移到不等式的左边, 整理得

$$\frac{1}{2} [Q_k^2(t+1) - Q_k^2(t)] \leq -Q_k(t)(D_{\Sigma,k}(t) - A_k(t)) + \frac{A_k^2(t) + D_{\Sigma,k}^2(t)}{2}$$

同理, 对 MEC 服务器端的动态队列  $T_k^2(t+1)$  进行处理:

$$T_k^2(t+1) \leq (T_k(t) - D_{s,k}(t))^2 + D_{r,k}^2(t) + 2D_{r,k}(t)T_k(t) \leq T_k^2(t) - 2T_k(t)(D_{s,k}(t) - D_{r,k}(t)) + D_{r,k}^2(t) + D_{s,k}^2(t) \quad (6)$$

可得

$$\frac{1}{2} [T_k^2(t+1) - T_k^2(t)] \leq -T_k(t)(D_{s,k}(t) - D_{r,k}(t)) + \frac{D_{r,k}^2(t) + D_{s,k}^2(t)}{2}$$

综合式(5)和式(6)可得李雅普诺夫加值函数:

$$\Delta(\boldsymbol{\Theta}(t)) \leq - \sum_{k \in \mathcal{K}} [Q_k(t)(D_{\Sigma,k}(t) - A_k(t)) + T_k(t)(D_{s,k}(t) - D_{r,k}(t))] + \frac{1}{2} \sum_{k \in \mathcal{K}} [A_k^2(t) + D_{\Sigma,k}^2(t) + D_{r,k}^2(t) + D_{s,k}^2(t)]$$

同时对不等式两边取数学期望, 可得  $\Delta_V(\boldsymbol{\Theta}(t))$  上界:

$$\Delta_V(\boldsymbol{\Theta}(t)) \leq$$

$$C - E \left[ \sum_{k \in \mathcal{K}} [Q_k(t)(D_{\Sigma,k}(t) - A_k(t))] \middle| \boldsymbol{\Theta}(t) \right] - E \left[ \sum_{k \in \mathcal{K}} [T_k(t)(D_{s,k}(t) - D_{r,k}(t))] \middle| \boldsymbol{\Theta}(t) \right] +$$

$$VE[P_{\Sigma}(t) | \Theta(t)] \quad (7)$$

其中:  $C$  为常数, 通过优化  $\Delta_V(\Theta(t))$  的上界 (不等式(7)的右半部分) 可以求得近似最优解. 则总的优化问题(4)可以转化成综合考虑队列状态和功耗的单阶段优化问题, 数学表达式为

$$\begin{aligned} \min_F \sum_{k \in \mathcal{K}} [ & (T_k(t) - Q_k(t))D_{r,k}(t) - \\ & Q_k(t)D_{l,k}(t) ] + VP_{\Sigma}(t) \\ \text{s. t } & \theta_k \geq \rho_k, \forall k \in \mathcal{K} \end{aligned} \quad (8)$$

其中:  $P_{\Sigma}(t) \triangleq \sum_{k \in \mathcal{K}} (w_k(p_{tx,k}(t) + p_{l,k}(t)) + w_M p_{ser,k}(t))$ . 结合式(1)~式(3), 优化问题整理为

$$\begin{aligned} \min_F \sum_{k \in \mathcal{K}} \left\{ V \left( w_M \sum_{m \in \mathcal{M}} \mathcal{K}_{ser,m}^3 f_{C,m}^3(t) - \right. \right. \\ \left. \left. w_k \mathcal{K}_{mob,k}^3 f_k^3(t) + \frac{Q_k(t)D_{normal}^l}{V} \right) \times \right. \\ \left. \frac{\ln(1 + \|F_k(t)\|^2 \alpha^{-1})}{\ln(1 + \alpha^{-1})} + \right. \\ \left. w_k (\|F_k(t)\|^2 + \mathcal{K}_{mob,k}^3 f_k^3(t)) \right) - \\ Q_k(t)D_{normal}^l + (T_k(t) - Q_k(t))\tau\alpha_k(t)W \times \\ \text{lb} \left( 1 + \frac{|\mathbf{H}_k(t)^H F_k(t)|^2}{1 + \sum_{i=1, i \neq k}^K |\mathbf{H}_k(t)^H F_i(t)|^2} \right) \Bigg\} \\ \text{s. t. } \sqrt{1 + \sum_{i=1, i \neq k}^K |\mathbf{H}_k^H F_i|^2} \leq \frac{1}{\sqrt{\rho_k}} \text{Re}(\mathbf{H}_k^H F_k), \\ \forall k \in \mathcal{K} \end{aligned} \quad (9)$$

接下来做半正定松弛处理<sup>[8]</sup>, 令  $\mathbf{W}_k = F_k F_k^H$ ,  $\mathbf{V}_k = \mathbf{H}_k \mathbf{H}_k^H$ , 如果去掉矩阵秩为 1 的约束, 优化方程(9)转化成

$$\begin{aligned} \min_W \sum_{k \in \mathcal{K}} \left\{ V \left( w_M \sum_{m \in \mathcal{M}} \mathcal{K}_{ser,m}^3 f_{C,m}^3(t) - \right. \right. \\ \left. \left. w_k \mathcal{K}_{mob,k}^3 f_k^3(t) + \frac{Q_k(t)D_{normal}^l}{V} \right) \times \right. \\ \left. \frac{\ln(1 + \alpha^{-1} \text{Tr}(\mathbf{W}_k))}{\ln(1 + \alpha^{-1})} + w_k (\text{Tr}(\mathbf{W}_k) + \mathcal{K}_{mob,k}^3 f_k^3(t)) \right) - \\ Q_k(t)D_{normal}^l + (T_k(t) - Q_k(t))\tau\alpha_k(t)W \times \\ \text{lb} \left( 1 + \frac{\text{Tr}(\mathbf{V}_k \mathbf{W}_k)}{1 + \sum_{i=1, i \neq k}^K \text{Tr}(\mathbf{V}_k \mathbf{W}_i)} \right) \Bigg\} \\ \text{s. t. } \sum_{i=1, i \neq k}^K \text{Tr}(\mathbf{V}_k \mathbf{W}_i) - \frac{1}{\rho_k} \text{Tr}(\mathbf{V}_k \mathbf{W}_k) \leq -1, \forall k \in \mathcal{K} \end{aligned} \quad (10)$$

其中:  $\text{Tr}(\mathbf{W}_k)$  为凸函数, 含有变量  $\mathbf{W}_k$  的  $\ln$  和  $\text{lb}$  的形式是凹函数. 即式(10)可以拆分成一个凸函数和

2 个凹函数之和, 约束条件是线性的凸函数. 综上转化分析, 最终优化问题是一个典型的 DC 规划问题, 可以用 CCP 算法<sup>[9]</sup> 求解最优预编码决策.

**算法: CCP 算法求解  $\{\mathbf{W}^*\}$**

**步骤 1** 初始化: 迭代次数  $n = 0$ ; 以式(10)中的凸函数部分为目标求解式(11), 作为初始点记作  $\{\mathbf{W}_k^{[0]}\}$ .

$$\begin{aligned} \min_W \sum_{k \in \mathcal{K}} \text{Tr}(\mathbf{W}_k) \\ \text{s. t. } \sum_{i=1, i \neq k}^K \text{Tr}(\mathbf{V}_k \mathbf{W}_i) - \frac{1}{\rho_k} \text{Tr}(\mathbf{V}_k \mathbf{W}_k) \leq -1, \forall k \in \mathcal{K} \end{aligned} \quad (11)$$

**步骤 2**  $n = n + 1$ .

**步骤 3** 更新  $\mathbf{W}$ : 基于  $\{\mathbf{W}_k^{[n]}\}$ , 求解式(10)的一阶泰勒近似问题, 得到最优解  $\{\mathbf{W}^*\}$ .

**步骤 4** 检查  $\{\mathbf{W}_k^{[n]}; \forall k \in \mathcal{K}\}$ , 如果不满足终止条件  $\|\mathbf{W}_k^{[n]} - \mathbf{W}_k^{[n+1]}\| < \varepsilon$ , 则返回步骤 3; 否则终止.

输出: 最优解  $\{\mathbf{W}^*\}$ .

## 4 仿真结果与分析

**Baseline 1** 忽略队列信息, 只考虑传输功耗最小化的上行预编码策略.

**Baseline 2** 考虑 Cheng 等<sup>[5]</sup> 基于最小化瞬时传输和计算功耗和的卸载策略, 同时采用迫零预编码算法.

将笔者设计的方案与其他 2 个典型模型 (Baseline1, 2) 进行性能对比, 仿真参数如表 1 所示.

表 1 仿真参数

参数	数值
用户数量 $K$	6
MEC 服务器数量 $J$	3
天线数量 $N_t$	6
时间间隔 $\tau/\text{ms}$	2
系统带宽 $W/\text{MHz}$	100
服务器的额定发射功率 $P/W$	100
MEC 服务器端计算频率 $f_{C,m}/\text{GHz}$	=4
用户端计算频率 $f_k/\text{GHz}$	2
李雅普诺夫控制参数 $V$	5 000
平滑参数 $\alpha$	0.001
功耗权重系数 $w_k, w_M$	1

图 2 所示为本文方案中系统总功耗、用户端和 MEC 服务器端任务队列长度与李雅普诺夫控制参

数的关系. 从该双纵轴图可以看出, 李雅普诺夫控制参数增加, 系统总功耗随之减少, 而队列囤积逐渐增加, 这是因为李雅普诺夫控制参数在优化中相当于时延和功耗的权重, 该参数越大, 即目标更着重优化系统功耗, 对时延的关注变小. 因此, 合理的  $V$  值设置有利于取得优化功耗和队列囤积的折中效果.

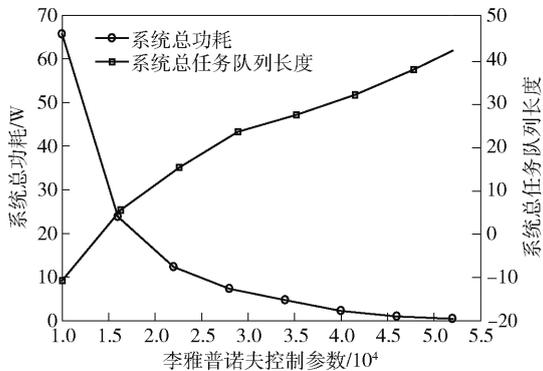


图 2 系统总功耗、任务队列长度与李雅普诺夫控制参数关系

图 3 所示为系统总功耗与用户数量的关系. 可见, 随着用户任务数量增加, 系统总功耗随之增加. 本文方案、Baseline 2 的系统功耗比 Baseline 1 低, 且在用户数量较多时更明显, 这是因为 Baseline 1 只考虑传输功耗优化, 而本文方案和 Baseline 2 同时优化传输和计算功耗. 本文方案的系统功耗比 Baseline 2 更低, 这是因为 Baseline 2 只关注单阶段瞬时目标优化, 忽略队列信息的马尔可夫特性, 而本文方案的优化目标是长期平均功耗. 可见, 笔者所提的预编码设计方案在优化功耗性能方面有较大提升.

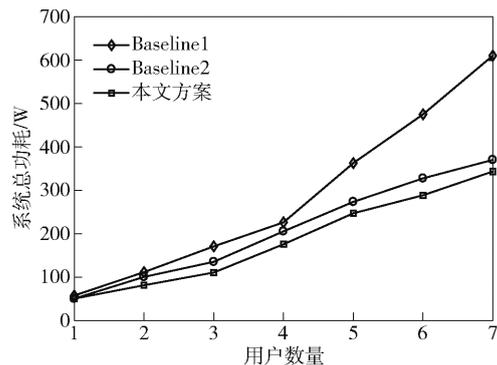


图 3 系统总功耗与用户数量关系

图 4 所示为系统总功耗随天线数量变化的关系. 可以看出, 随着天线数量增加, 系统总功耗减少. 这是因为天线数量增加, 用户的任务分配有更

多选择, MEC 服务器端能提供更多服务, 减少用户端的能耗负荷, 从而降低系统总功耗. 本文方案的系统总功耗保持最低, 原因与图 3 解释类似. 同时可以看出, 当天线数量由 1 增至 2 时, 系统总功耗大幅度下降, 之后随着天线数量增加, 系统总功耗下降趋势趋于平缓, 这是由于多天线增益趋于稳定.

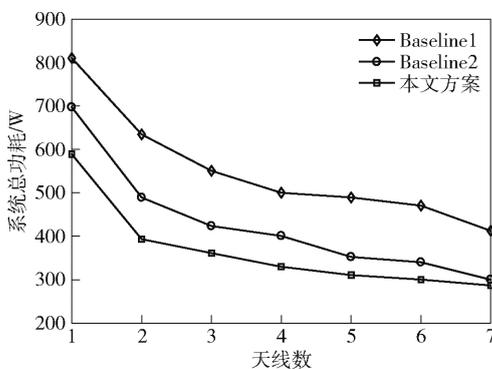


图 4 系统总功耗与天线数量关系

图 5 所示为系统时延和用户数量的关系. 随着用户数量的增加, 系统时延随之增加. 这是因为用户数量增加, 需要处理的任务量增多, 信道会逐渐拥塞, 所以系统时延随之增加. 本文方案的系统时延比 Baseline 2、Baseline 1 小, 这是因为本文方案同时考虑信道和队列状态的马尔可夫特性, 优化系统长期平均时延, 而 Baseline 1 只优化传输功耗, 没有考虑时延特性. Baseline 2 考虑时延特性, 但是只优化单阶段瞬时目标. 可见, 笔者所提预编码设计方案在时延性能优化上能满足更低的时延要求.

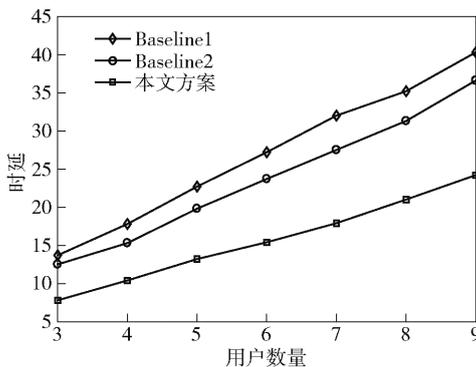


图 5 系统时延与用户数量关系

综合图 2 ~ 图 5 可得, 通过调整李雅普诺夫控制参数, 本文方案对长期平均功耗和长期平均时延联合优化, 可以达到一个折中的效果. 相比于其他 2 个方案, 本文方案综合考虑信道和队列状态的马尔可夫特性, 以长期平均性能为优化目标, 在时延和功

耗性能优化中,能达到更好的优化效果.

图6所示为系统功耗随功耗权重系数变化的关系.其中, $w_k + w_M = 1$ .由图可知,随着MEC服务器端功耗权重系数增加,MEC服务器端功耗持续减小,用户端功耗随之增加,同时系统加权总功耗随系数先增后减.这是因为优化目标函数中采用了最小加权功耗,随 $w_M$ 的增大,MEC服务器端的功耗逐渐成为主要优化目标,造成卸载减少,MEC服务器端功耗降低.因此,通过设置合适的功耗权重系数可以均衡系统对MEC服务器端与用户端的功耗.

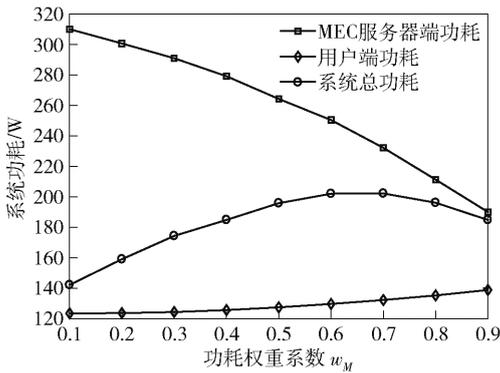


图6 系统功耗与MEC服务器端功耗权重系数关系

## 5 结束语

针对多用户-多移动边缘计算服务器系统进行优化,综合考虑时延、功耗等因素,以降低系统的平均开销为目标,利用李雅普诺夫优化方法对系统的综合性能进行优化.仿真结果表明,相比于其他2个方案,通过预编码设计优化的方案能满足更低的功耗和时延要求.

### 参考文献:

[1] Blanco B, Fajardo J O, Giannoulakis I, et al. Technology pillars in the architecture of future 5G mobile networks: NFV, MEC and SDN[J]. Computer Standards & Interfaces, 2017, 54(4): 216-228.

[2] Wang Chenmeng, Yu F R, Liang Chengchao, et al. Joint computation offloading and interference management in wireless cellular networks with mobile edge computing [J]. IEEE Transactions on Vehicular Technology, 2017, 66(8): 7432-7445.

[3] Xu Jie, Sun Yuxuan, Chen Lixing, et al. E2M2: energy efficient mobility management in dense small cells with mobile edge computing [C] // IEEE International Conference on Communications (ICC). New York: IEEE Press, 2017: 17065927.

[4] Han Kaifeng, Ko S W, Huang Kaibin. Spatial modeling and latency analysis for mobile edge computing in wireless networks [C] // IEEE International Conference on Communications (ICC). New York: IEEE Press, 2018: 17971408.

[5] Cheng Kang, Teng Yinglei, Sun Weiqi, et al. Energy-efficient joint offloading and wireless resource allocation strategy in multi-MEC server systems [C] // IEEE International Conference on Communications (ICC). New York: IEEE Press, 2018: 8422877.

[6] Wang Wei, Lau V K N, Peng Mugen. Delay-aware uplink Fronthaul allocation in cloud radio access networks [J]. IEEE Transactions on Wireless Communications, 2017, 16(7): 4275-4287.

[7] Li Jian, Wu Jingxian, Peng Mugen, et al. Queue-aware joint Remote radio head activation and beamforming for green cloud radio access networks [C] // IEEE Global Communications Conference. New York: IEEE Press, 2015: 15808266.

[8] Chen M H, Liang Ben, Dong Min. Joint offloading decision and resource allocation for multi-user multi-task mobile cloud [C] // IEEE International Conference on Communications (ICC). New York: IEEE Press, 2016: 16141619.

[9] Wu Xiaoxiao, So A, Pan Jiaxian, et al. Semidefinite relaxation and approximation analysis of a beamformed alamouti scheme for relay beamforming networks [J]. IEEE Transactions on Signal Processing, 2017, 65(9): 2428-2443.