

Cell-Free 大规模 MIMO 系统中基于传输时延的缓存策略研究

王蕊^{1,2}, 申敏¹, 何云¹, 刘香燕¹

(1. 重庆邮电大学通信与信息工程学院, 重庆 400065; 2. 玉溪师范学院物理与电子工程学院, 云南 玉溪 653100)

摘要: 为了满足未来移动互联网中用户的超低时延和超高可靠需求, 将无线缓存技术与无小区大规模多输入多输出系统相结合, 设计了基于 AP 间协作缓存及区域流行度评估的缓存模型。推导出涉及 AP 分簇、协作缓存以及区域流行度的传输时延表达式, 并将内容放置问题表述为最小化总内容传输时延问题。通过对优化问题的 NP-hard 和拟阵约束下次模单调性的证明, 提出了基于贪婪算法的缓存部署优化策略。仿真结果表明, 所提策略可有效降低系统内容传输时延和提升缓存命中率。

关键词: Cell-Free 大规模 MIMO 系统; 缓存部署; 协作缓存; 内容传输时延

中图分类号: TN92

文献标识码: A

DOI: 10.11959/j.issn.1000-436x.2021234

Research on caching strategy based on transmission delay in Cell-Free massive MIMO systems

WANG Rui^{1,2}, SHEN Min¹, HE Yun¹, LIU Xiangyan¹

1. School of Communication and Information Engineering, Chongqing University of Posts and Telecommunications, Chongqing 400065, China

2. School of Physics and Electronic Engineering, Yuxi Normal University, Yuxi 653100, China

Abstract: To meet the ultra-low latency and ultra-high reliability requirements of users in the future mobile Internet, the wireless caching technology was combined with Cell-Free massive MIMO systems. The caching model was designed based on AP cooperative caching and regional popularity evaluation. The transmission delay expression involving AP clustering, cooperative caching, and regional popularity was derived, and the content placement problem was expressed as total content transmission delay minimization. Through the demonstration of the NP-hard and submodular monotony of the optimization problem, the greedy algorithm-based optimization strategy was proposed. Simulation results show that the proposed strategy can effectively reduce the content transmission delay and improve the cache hit rate.

Keywords: Cell-Free massive MIMO system, caching placement, cooperative caching, content transmission delay

1 引言

随着全球移动互联网的发展, 移动通信流量呈爆发式增长。然而, 现有的移动通信系统架构都以蜂窝小区为基础, 随着网络的密集化, 小区间干扰愈发严重, 这将严重影响小区边缘用户的信号质量

和系统服务质量的一致性, 难以满足自动驾驶、远程医疗等低时延高可靠业务的需求。Cell-Free 大规模多输入多输出 (MIMO, multiple-input multiple-output) 系统对传统蜂窝小区架构进行变革, 在解决这些挑战方面具有巨大潜力^[1]。在 Cell-Free 大规模 MIMO 系统中, 所有接入点 (AP, access

收稿日期: 2021-07-19; 修回日期: 2021-10-06

通信作者: 申敏, shenmin@cqupt.edu.cn

基金项目: 国家科技重大专项基金资助项目 (No.2018ZX03001026-002); 云南省地方本科高校 (部分) 基础研究联合专项基金资助项目 (No.2018FH001-120); 重庆邮电大学博士研究生创新人才基金资助项目 (No.BYJS202102)

Foundation Items: The National Science and Technology Major Project of China (No.2018ZX03001026-002), The Joint Special Fund Project for Basic Research of Local Undergraduate Universities (part of) in Yunnan Province (No.2018FH001-120), Chongqing University of Posts and Telecommunications Ph.D. Innovative Talents Project (No.BYJS202102)

point) 通过前传网络与中央处理单元 (CPU, central processing unit) 连接, 在相同的时频资源^[2]上通过时分双工 (TDD, time division duplex) 技术为所有用户提供服务, 具有服务质量均衡、覆盖范围广等优点, 消除了传统蜂窝网络的小区边界, 可显著提高系统频谱效率。然而, 由于 AP 数量众多, AP 与 CPU 之间传输的大量应用数据加重了前传链路的负荷, 容易导致前传链路拥塞, 一旦前传链路的容量受限, 将大大影响 Cell-Free 大规模 MIMO 系统的传输时延和可扩展性。这对于未来移动通信网络中时延敏感型应用来说是不可接受的。

为了解决上述问题, 研究者提出了发送量化信号^[3-4]、接入点选择^[5]等方式来降低前传链路压力。其中, Bashar 等^[3-4]提出发送定量信号的方法来减轻前传链路的负担。然而, 上述工作假设所有 AP 同时为所有用户服务。这样的框架在实践中是不现实和不可扩展的。为了实现可扩展性和降低前传压力, Ngo 等^[5]提出了一种基于最大尺度衰落系数的 AP 选择方案, 该方案中每个用户只由附近的一些 AP 为其提供服务, 由于考虑的因素较少, 该方案中的 AP 选择算法不能保证系统性能的优越性。因此, 找到一种可兼顾系统时延性能和前传流量压力的方案成为亟待解决的问题。

无线缓存技术将内容 (视频、网页等) 存储在靠近无线网络边缘的存储设备上供将来使用, 可以有效减少前传/回传链路的流量压力和时延。此外, 安装内存的成本要低于提升前传/回传容量的成本^[6-7]。基于上述优点, 无线缓存技术在无线网络中受到了广泛的关注和研究。文献[8]研究了宏蜂窝网络场景下以平均下载时延为目标的缓存部署问题, 通过松弛整数约束, 将混合整数非线性规划问题转化为凸优化问题, 并采用连续凸逼近算法进行求解。文献[9]在缓存容量受限情况下, 提出了集中式缓存部署方案来最小化所有用户的平均下载时延。文献[10]基于内容流行度和用户位置等信息, 推导出用户平均时延表达式, 采用低复杂度的迭代算法优化缓存方案。文献[11]设计了一个针对无线网络的移动感知缓存框架, 在考虑用户移动性的基础上, 采用编码和未编码的内容放置策略来最大化缓存命中率。为了进一步提高缓存命中率和有效利用有限的存储资源, 文献[12]在用户簇为中心的小蜂窝网络中提出了一种组合的编码/未编码缓存策略, 分析了位于簇中心用户的内容成功交

付概率。在组合缓存策略中, 对每个基站的存储空间进行分区, 分别存储最流行的内容和较不流行的内容。文献[13]以最小化内容传输时延为优化目标, 提出了超密集蜂窝网络中基于协作多点 (CoMP, coordinated multi-point) 传输技术的协作缓存方案, 通过对问题 NP-hard 特性分析, 采用遗传算法进行求解。

以上研究将缓存与多种无线蜂窝网络结合使系统时延等性能得以提升, 但均是针对蜂窝网络架构下的缓存部署策略, 未考虑 Cell-Free 大规模 MIMO 系统中去蜂窝化、大量 AP 部署、AP 分簇和以用户为中心的特性。针对如何有效利用 Cell-Free 大规模 MIMO 系统多 AP 部署特性来增加缓存内容的多样性以及如何在 AP 分簇场景下准确地评估文件流行度, 本文提出了基于 AP 间协作缓存及区域流行度评估的缓存模型, 通过 AP 间缓存内容的协作分发提高缓存内容部署多样性, 以服务用户的 AP 簇为单位计算流行度来提高流行度评估的准确性; 推导出考虑 AP 分簇、协作缓存及区域流行度的内容传输时延表达式, 且以传输时延为优化目标建立缓存优化问题, 证明了优化问题的 NP-hard 和拟阵约束下的次模单调特性, 以此提出一种基于贪婪算法的低复杂度缓存部署策略。通过仿真与其他以时延为优化目标的缓存策略^[9,12-14]进行了比较, 验证了所提策略可始终保持较低的传输时延和较高的缓存命中率, 从而降低前传链路压力, 有效提升通信服务质量。

2 系统模型

本节将介绍支持缓存的 Cell-Free 大规模 MIMO 系统, 并对系统的无线内容传输和缓存进行分析和建模。

2.1 通信模型

缓存辅助的 Cell-Free 大规模 MIMO 系统架构如图 1 所示。系统由 L 个具有 N 根天线的 AP 和 K 个具有单天线的用户设备 (UE, user equipment) 组成, 系统工作在 TDD 模式, 根据信道互易性, 仅需进行上行信道估计。AP 通过前传链路连接至 CPU, CPU 通过回传链路连接到核心网络。

为了反映实际信道特性, 本文考虑由视距和多径分量组成的莱斯 (Rician) 衰落信道, 第 l 个 AP 到第 k 个 UE 的信道可表示为 $\mathbf{h}_{lk} \in \mathbb{C}^{N \times 1}$, 且满足

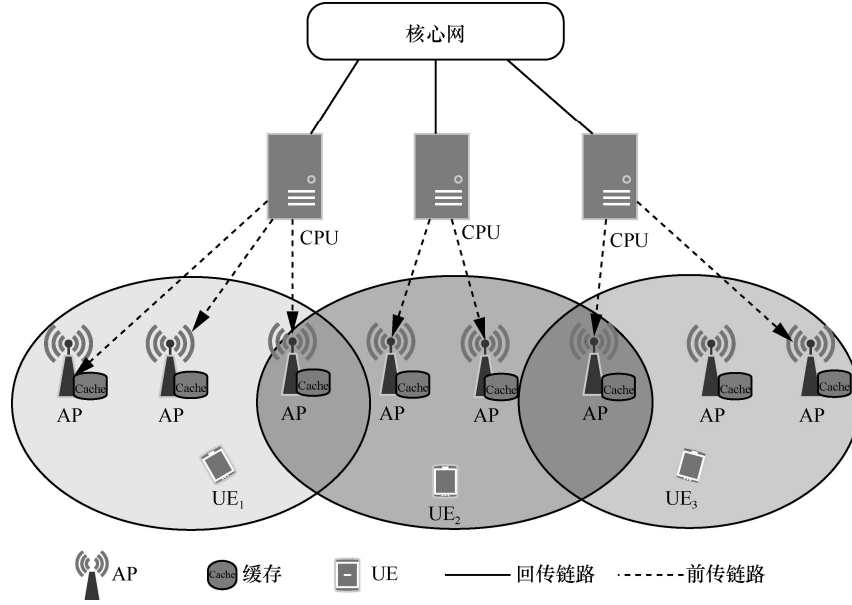


图 1 缓存辅助的 Cell-Free 大规模 MIMO 系统架构

$$\mathbf{h}_{lk} = \sqrt{\beta_{lk}} \left(\sqrt{\frac{k_{lk}}{k_{lk}+1}} \bar{\mathbf{g}}_{lk} + \sqrt{\frac{1}{k_{lk}+1}} \mathbf{A}_{lk}^{\frac{1}{2}} \tilde{\mathbf{g}}_{lk} \right) \quad (1)$$

其中, β_{lk} 表示大尺度衰落系数; 括号中的部分为小尺度衰落系数, $\bar{\mathbf{g}}_{lk}$ 表示直视 (LoS, line-of-sight) 路径分量, $\tilde{\mathbf{g}}_{lk} \sim \mathcal{CN}(0, \mathbf{I}_N)$ 表示非直视 (NLoS, non-LoS) 路径分量, k_{lk} 表示 Rician 因子, \mathbf{A}_{lk} 表示信道相关矩阵。对 UE 移动可能引起 LoS 分量^[15]的相移效应, 本文假设可以准确地跟踪。因此, 信道参数定义为^[16-17]

$$\bar{\mathbf{h}}_{lk} = \sqrt{\frac{\beta_{lk} k_{lk}}{k_{lk}+1}} \bar{\mathbf{g}}_{lk}, \quad \forall l, k \quad (2)$$

$$\mathbf{R}_{lk} = \frac{\beta_{lk}}{k_{lk}+1} \mathbf{A}_{lk}, \quad \forall l, k \quad (3)$$

由以上分析可知, $\mathbf{h}_{lk} \sim \mathcal{CN}(\bar{\mathbf{h}}_{lk}, \mathbf{R}_{lk})$, 其中 $\bar{\mathbf{h}}_{lk}$ 表示 LoS 分量, \mathbf{R}_{lk} 表示 NLoS 分量的空间相关特性, $\mathcal{CN}(\cdot, \cdot)$ 表示圆对称复高斯分布。

假设 τ_c 表示相干时间长度, τ_p 表示上行训练阶段的持续时间长度, $\tau_c - \tau_p$ 表示下行数据发送阶段的持续时间长度。

1) 上行信道估计。在上行链路训练阶段, 所有 UE 同时向所有 AP 发送相互正交或相同的导频序列, 本文将分配给每个 UE 的导频序列记为 $\boldsymbol{\varphi}_k \in \mathbb{C}^{1 \times \tau_p}$, 且满足 $\|\boldsymbol{\varphi}_k\|^2 = 1$, 假设导频序列的长度

小于 UE 的数量, 即存在导频污染, 并定义 \mathcal{P}_k 为与 UE k 使用相同导频的 UE 集合。由此可知, 第 l 个 AP 接收到的导频信号为

$$\mathbf{y}_{lk} = \sum_{k_1 \in \mathcal{P}_k} \mathbf{h}_{lk_1} + \frac{N_{p,l} \boldsymbol{\varphi}_k^H}{\sqrt{\tau_p P_p}} \quad (4)$$

其中, p_p 为上行导频序列的归一化发送信噪比, $N_{p,l}$ 为第 l 个 AP 的接收噪声矩阵, $(\cdot)^H$ 为共轭转置运算。本文利用最小均方误差 (MMSE, minimum mean square error) 估计器^[16], 可获得信道估计为

$$\hat{\mathbf{h}}_{lk} = \bar{\mathbf{h}}_{lk} + \mathbf{R}_{lk} \boldsymbol{\Phi}_{lk} (\mathbf{y}_{lk} - \bar{\mathbf{y}}_{lk}) \quad (5)$$

其中, $\bar{\mathbf{y}}_{lk} = \sum_{k_1 \in \mathcal{P}_k} \bar{\mathbf{h}}_{lk_1}$, $\boldsymbol{\Phi}_{lk} = \left(\sum_{k_1 \in \mathcal{P}_k} \mathbf{R}_{lk_1} + \frac{1}{\tau_p P_p} \mathbf{I}_N \right)^{-1}$ 。

由式(5)可知, $\hat{\mathbf{h}}_{lk} \sim \mathcal{CN}(\bar{\mathbf{h}}_{lk}, \boldsymbol{\Omega}_{lk})$, $\boldsymbol{\Omega}_{lk} = \mathbf{R}_{lk} \boldsymbol{\Phi}_{lk} \mathbf{R}_{lk}$ 。

2) 下行数据传输。本文使用共轭波束成形和文献[18]中的 AP 选择方案进行下行信号传输。第 K 个 UE 接收到的信号为

$$\begin{aligned} \mathbf{y}_k &= \sum_{l=1}^L \mathbf{h}_{lk}^T \sqrt{p_d} \sum_{k=1}^K \zeta_{lk} \boldsymbol{\psi}_{lk} s_k + \mathbf{w}_k = \\ & \sqrt{p_d} \mathbf{v}_{kk} s_k + \sqrt{p_d} \sum_{k_1 \neq k}^K \mathbf{v}_{kk_1} s_{k_1} + \mathbf{w}_k \end{aligned} \quad (6)$$

其中, p_d 为最大归一化下行传输功率, $\zeta_{lk} = \rho_{lk} \alpha_{lk}$, ρ_{lk} 为功率控制系数, α_{lk} 为天线选择系数, $\boldsymbol{\psi}_{lk} = \hat{\mathbf{h}}_{lk}^*$

为波束成形预编码向量, $\mathbf{v}_{kk_1} = \sum_{l=1}^L \varsigma_{lk_1} \mathbf{h}_{lk}^T \boldsymbol{\psi}_{lk_1}$, $(\cdot)^*$ 和 $(\cdot)^T$ 分别为复共轭和转置运算。第 k 个 UE 的下行可达速率为

$$R_k = B \left(1 - \frac{\tau_p}{\tau_c} \right) \text{lb} \left(1 + \frac{|\mathbb{E}\{\mathbf{v}_{kk}\}|^2}{\text{Var}(\mathbf{v}_{kk}) + \sum_{k_1 \neq k}^K \mathbb{E}\left\{|\mathbf{v}_{kk_1}|^2\right\} + \frac{1}{P_d}} \right) \quad (7)$$

其中, 各部分的定义和计算如下

$$|\mathbb{E}\{\mathbf{v}_{kk}\}|^2 = \sum_{l=1}^L \sum_{l_1=1}^L \varsigma_{lk} \varsigma_{l_1k} A_{ll_1kk} \quad (8)$$

$$\text{Var}(\mathbf{v}_{kk}) + \sum_{k_1 \neq k}^K \mathbb{E}\left\{|\mathbf{v}_{kk_1}|^2\right\} = \sum_{l=1}^L \sum_{k_1=1}^K \varsigma_{lk_1}^2 B_{llk_1} + \sum_{l=1}^L \sum_{l_1=1}^L \sum_{k_1 \neq k}^K \varsigma_{lk_1} \varsigma_{l_1k_1} A_{ll_1kk_1} \quad (9)$$

$$A_{ll_1kk_1} = \Re\left\{C_{llk_1} C_{l_1kk_1}^H\right\} \quad (10)$$

$$B_{llk_1} = \bar{\mathbf{h}}_{lk}^H \boldsymbol{\Omega}_{lk_1} \bar{\mathbf{h}}_{lk} + \bar{\mathbf{h}}_{l_1k_1}^H \mathbf{R}_{lk} \bar{\mathbf{h}}_{l_1k_1} + \text{trace}(\mathbf{R}_{lk} \boldsymbol{\Omega}_{lk_1}) \quad (11)$$

$$C_{llk_1} = \text{trace}(\boldsymbol{\Omega}_{lk_1}) \delta_{kk_1} + \bar{\mathbf{h}}_{mk}^H \bar{\mathbf{h}}_{mk_1} \quad (12)$$

其中, $\mathbb{E}\{\cdot\}$ 和 $\text{Var}\{\cdot\}$ 表示求期望和方差值的运算。如果 $k = k_1$, 则 $\delta_{kk_1} = 1$; 否则, $\delta_{kk_1} = 0$ 。

2.2 缓存模型

本文令 M 表示文件库的大小, $\mathcal{F} = \{f_1, \dots, f_M\}$ 表示文件库中文件的索引。为了便于分析, 假设所有文件具有相同的长度, 大小为 S_{f_m} bit。事实上, 在传输过程中, 不同大小的文件总能被分割成相同大小的文件块。另外, 每个文件都有一个内容流行度, 由 P_{f_m} 表示, 其分布遵循广义 Zipf 函数, 即

$$P_{f_m} = \left(\text{rank}(f_m)^\gamma \sum_{n=1}^F n^{-\gamma} \right)^{-1}, \quad \forall f_m \in \mathcal{F} \quad (13)$$

其中, $0 \leq \gamma \leq 1$ 为 Zipf 分布的偏态因子。如果 $\gamma = 0$, 则文件流行度分布均匀, 即所有文件具有相同的流行度; 如果 $\gamma > 0$, 则文件流行度遵循经典的 Zipf 定律, 意味着 \mathcal{F} 中的文件具有不均匀的流行度。然而, 考虑到不同区域的用户都有自己个性化的内容兴趣和偏好, 可能与基于大量用户统计的全局流行度不一致。为了更准确地评估文件流行度, 本文考虑了一个区域流行度模型, 不同 AP 簇中的

文件具有不同的区域流行度分布。令 $P_{M_k, f_m} = \sum_{l \in M_k} P_{l, f_m}$ 表示文件 f_m 的区域流行度, 即表示用户 k 在 AP 簇 M_k 上对文件 f_m 的偏好, 且满足 $\sum_{l \in \mathcal{L}} \sum_{f_m \in \mathcal{F}} P_{l, f_m} = \sum_{f_m \in \mathcal{F}} P_{f_m} = 1$ 。其中, M_k 表示服务于第 k 个 UE 的 AP 簇且每个簇的 AP 数相同; P_{l, f_m} 表示用户在第 l 个 AP 上对文件 f_m 的偏好, 可基于机器学习技术^[19-20]进行预测, 并假定在内容放置阶段是已知的。

此外, 本文将文件请求过程建模为马尔可夫调制速率过程^[21]。也就是说, 对于与 AP 簇相关的 UE k , 其请求数遵循平均速率为 $\lambda_{M_k, k}$ 的泊松过程, 且不同活跃度的用户具有不同的业务请求平均速率。然后, 根据归一化区域流行度 \bar{P}_{M_k, f_m} , 得到 UE k 对文件 f_m 的平均请求到达率 $\lambda_{M_k, k}^{f_m}$ 。因此可得

$$\bar{P}_{M_k, f_m} = \frac{P_{M_k, f_m}}{\sum_{f_m \in \mathcal{F}} P_{M_k, f_m}}, \quad \forall k, f_m \quad (14)$$

$$\lambda_{M_k, k}^{f_m} = \lambda_{M_k, k} \bar{P}_{M_k, f_m}, \quad \forall k, f_m \quad (15)$$

本文提出了基于协作的缓存部署策略。设 $\Theta = \{\eta_{ml}^j \in \{0, 1\} : m = 1, \dots, M, l = 1, \dots, L, j = 0, 1, \dots, L\}$ 表示协作缓存参数, $\eta_{ml}^j = 1$ 表示 AP l 可以从 AP j 获取文件 f_m , $\eta_{ml}^j = 0$ 表示不能获取。 $j = l$ 表示 AP 从本地缓存中获取所需的内容, $j = 0$ 表示 UE 必须通过 CPU 从核心网络中获取所需的内容。本文假设 CPU 侧不具备缓存功能, 仅处理 AP 与核心网之间的传输内容。下一步将对 CPU 侧增加缓存功能的情况进行研究。

2.3 时延模型

在 Cell-Free 大规模 MIMO 系统中, UE 获取所需内容的传输时延取决于文件在 AP 的缓存状态以及协作缓存参数。内容传输时延由无线接入时延、前传传输时延和回传传输时延组成。从 AP 簇到 UE k 之间的无线接入时延 $\mathcal{D}_{k, f_m}^{\text{tr}}$ 可表示为

$$\mathcal{D}_{k, f_m}^{\text{tr}} = \frac{S_{f_m}}{R_k} = \frac{S_{f_m}}{B \left(1 - \frac{\tau_p}{\tau_c} \right) \text{lb} \left(1 + \frac{|\mathbb{E}\{\mathbf{v}_{kk}\}|^2}{\text{Var}(\mathbf{v}_{kk}) + \sum_{k_1 \neq k}^K \mathbb{E}\left\{|\mathbf{v}_{kk_1}|^2\right\} + \frac{1}{P_d}} \right)} \quad (16)$$

令 r^{FH} 表示从 CPU 到 AP 的平均传输速率, 则传输文件 f_m 的平均传输时延 $\mathcal{D}_{\text{th}} = S_{f_m} / r^{\text{FH}}$ 。同样, CPU 从核心网获取文件 f_m 的平均传输时延为 $\mathcal{D}_{\text{bh}} = S_{f_m} / r^{\text{BH}}$, 其中 r^{BH} 为核心网到 CPU 的平均传输速率。

3 基于传输时延的缓存优化

3.1 问题描述

基于系统缓存策略, UE k 的传输时延由无线接入时延 $\mathcal{D}_{k,f_m}^{\text{r}}$ 、前传时延 \mathcal{D}_{th} 和回传时延 \mathcal{D}_{bh} 三部分组成。因此, UE k 接收内容 f_m 的传输时延可以表示为

$$\mathcal{D}_{k,f_m} = \lambda_{M_k,k}^{f_m} \left(\mathcal{D}_{k,f_m}^{\text{r}} + \frac{\sum_{l=1}^L \sum_{j=0}^L \alpha_{lk} (1 - v_{ml}) \mathcal{D}_{ml}^j \eta_{ml}^j}{|M_k|} \right) \quad (17)$$

其中, $\lambda_{M_k,k}^{f_m}$ 表示 UE k 对内容 f_m 的请求数; $v = \{v_{ml} \in \{0,1\} : m=1,\dots,M, l=1,\dots,L\}$ 表示缓存部署决策变量, $v_{ml}=1$ 表示内容缓存在 AP l 中, $v_{ml}=0$ 则表示未缓存; \mathcal{D}_{ml}^j 表示 AP l 为 UE 提供服务时从缓存 j 检索内容 f_m 的时延, 可以通过下述方式计算: 如果 $j \neq 0$, 则 $\mathcal{D}_{ml}^j = 2\mathcal{D}_{\text{th}}$; 如果 $j=0$, 则 $\mathcal{D}_{ml}^j = \mathcal{D}_{\text{th}} + \mathcal{D}_{\text{bh}}$ 。

基于以上分析, 本文提出了以用户总传输时延最小化为目标的优化问题, 表述如下

$$\begin{aligned} & \min_v \sum_{k=1}^K \sum_{f_m \in \mathcal{F}} \mathcal{D}_{k,f_m}(v) \\ & \text{s.t. C1: } \sum_{f_m \in \mathcal{F}} S_{f_m} v_{ml} \leq C_l, \quad \forall l \\ & \quad \text{C2: } v_{ml} \in \{0,1\}, \quad \forall l, \forall f_m \end{aligned} \quad (18)$$

第一组约束条件确保 AP 缓存的文件之和不超过 AP 的存储容量, 第二组约束条件表明缓存部署变量是具有离散特性的 0-1 变量。

3.2 NP-Hard 证明

为证明式(18)的整数规划问题属于 NP-hard 问题, 则需要将已知的 NP-complete 问题规约为所提问题相应的判定问题的一种特例。加权集合覆盖问题是一种经典的 NP-complete 问题, 其定义为给定集合 $E = \{e_l : l=1,\dots,L\}$ 和集合 $S = \{s_j : j=1,\dots,J\}$, 集合 S 的每个元素 s_j 是 E 的一个子集且有一个权值 $\phi_j \geq 0$ 。加权集合覆盖问题的目标是找到集合 S

的一组子集, 使这组子集的并集等于集合 E 并且使总权值最小。

引理 1 整数规划问题式(18)的判定问题是 NP-complete 问题。

证明 为证明所提问题的判定问题是 NP-complete 问题, 将加权集合覆盖问题归约为总传输时延最小化问题, 可作如下改写。1) 将问题中的内容库的大小设为 1, 即 $M=1$; 2) 将加权集覆盖问题中集合 E 的每个元素 e_l 映射为 AP l 请求所需文件; 3) 将每个子集 S_j 映射为 AP j 存储了文件, 可以为所包含 AP 提供所需的文件, 其中 $e_l \in s_j$ 表示 AP l 可以从 AP j 的缓存获取内容; 4) 对文件进行传输的时延视为每个子集 S_j 的权值, 可由式(17)计算。本文的总传输时延最小化优化问题的目标可表示为选择一组子集 $\{s_1, \dots, s_j\} \subseteq S$, 满足 $s_1 \cup s_2 \cup \dots \cup s_j = E$, 使子集 $\{s_1, \dots, s_j\}$ 的总传输时延最小。证毕。

3.3 问题优化

本节将证明优化问题式(18)表述为拟阵约束的次模函数的最小化问题。通过引入优化问题的拟阵和次模特征, 可采用低复杂度贪婪算法求解该问题。首先证明问题式(18)的限制条件是一个划分拟阵。

1) 拟阵证明

拟阵的定义如下。拟阵 \mathcal{M} 是一个元组 $\mathcal{M} = (S, \mathcal{I})$, 其中 S 是一个有限的基集, $\mathcal{I} \subseteq 2^S$ 是一个独立集合簇且满足以下条件。

① \mathcal{I} 是非空集合且 $\emptyset \in \mathcal{I}$ 。

② 如果集合 $Y \in \mathcal{I}$ 且 $X \subseteq Y$, 则 $X \in \mathcal{I}$ 。

③ 如果 $X, Y \in \mathcal{I}$, $|X| < |Y|$, 那么一定存在一个元素 $y \in Y \setminus X$, 使 $X \cup \{y\} \in \mathcal{I}$ 。

划分拟阵的定义如下。将有限基集 S 划分为 n 类并表示为 $S = S_1 \cup S_2 \dots \cup S_n$, 且对于 $\forall 1 \leq i \leq j \leq n$, 存在 $S_i \cap S_j = \emptyset$ 。此时定义的独立集合簇 \mathcal{I} 可以表示为

$$\mathcal{I} = \{X \subseteq S : |X \cap S_i| \leq \varpi_i, \forall i = 1, \dots, n\} \quad (19)$$

其中, $\varpi_1, \dots, \varpi_n$ 是固定的参数, 则 $\mathcal{M} = (S, \mathcal{I})$ 就是一个划分拟阵。

引理 2 整数规划问题式(18)中的约束条件可以写成一个划分拟阵。

证明 基于上述拟阵的定义, 将有限的基集定

义如下

$$S = \{s_1^1, s_2^1, \dots, s_M^1, \dots, s_1^l, \dots, s_m^l, \dots, s_M^l, \dots, s_1^L, \dots, s_M^L\} \quad (20)$$

其中, s_m^l 表示文件与 AP 之间的关系。定义缓存部署集合 $\nu \subseteq S$, 当 $\nu_{ml}=1$ 时元素 s_m^l 存在于集合 ν 中。基于上述定义, 缓存部署策略中第 l 列的非零元素对应于 $\nu \cap S_l$ 中的元素。因此, AP 缓存容量的约束可以表示为 $\nu \in \mathcal{I}$, 其中,

$$\mathcal{I} = \{\nu \subseteq S : |\nu \cap S_l| \leq C_l, \forall l=1, \dots, L\} \quad (21)$$

将式(21)与划分拟阵的定义式(19)对比可以发现, 优化问题式(18)约束下的缓存部署可写成一个划分拟阵, 其中 $n=L$, $\varpi_i=C_i, \forall i=1, \dots, L$ 。

因此, 划分拟阵可以表示为 $\mathcal{M}=(S, \mathcal{I})$ 。证毕。

2) 次模函数证明

次模函数的定义如下。有限集合 S 和定义在其幂 2^S 的一个实函数 f , 记 $f_W(i) = f(W+i) - f(W)$ 为边际效用值, 如果对于 S 的任意 2 个子集 W 和 Z , 且 $Z \supset W, i \in S \setminus Z$ 有如下关系

$$f_W(i) \geq f_Z(i) \quad (22)$$

则函数 f 为一个次模函数。

引理 3 式(18)中的目标函数是单调非递增的次模函数。

单调性证明 由于在已有缓存 ν 的基础上增加一个新的文件, 不会增加系统的传输时延。因此, 目标函数是关于 ν 的非递增函数。

次模性证明 对于目标函数, 如果某个用户请求某个文件的传输时延是次模函数, 那么, 总传输时延也是次模函数。接下来讨论一个用户请求一个文件的情况。根据次模性的定义, 需要对如下性质进行证明。如果向任何缓存添加新内容, 则边际效用值会随着放置集 $\nu = \{\nu_{ml} \in \{0, 1\} : m=1, \dots, M, l=1, \dots, L\}$ 的增加而减少。基于以上分析, 本文用 $f_\nu(\nu_{mj}) = \mathcal{D}_{k, f_m}(\nu) - \mathcal{D}_{k, f_m}(\nu \cup \nu_{mj})$ 表示给缓存策略 ν 添加一个新内容 ν_{mj} 的边际效用值, 用其定义将新内容 f_m 添加至缓存 l 后传输时延的减少值。假设给定 2 个缓存部署策略 ν 和 $\bar{\nu}$, 且 $\nu \subset \bar{\nu}$ 。则需从以下 3 种情况考虑边际效用值的变化。

情况 1 2 种缓存部署策略均有 AP 部署文件 f_m , 如果再向未部署 f_m 的其中一个 AP 添加内容 f_m , 则 2 种策略的边际效用值相等, 且等于 $2\mathcal{D}_{th}/|M_k|$ 。

情况 2 2 种缓存部署策略均未部署文件 f_m , 如果向其中一个 AP 添加内容 f_m , 则 2 种策略的边际效用值依然相等, 且等于 $\mathcal{D}_{bh} - \mathcal{D}_{th} + 2\mathcal{D}_{th}/|M_k|$ 。

情况 3 策略 $\bar{\nu}$ 已有 AP 部署文件 f_m , 策略 ν 没有, 则策略 $\bar{\nu}$ 的边际效用值 $f_{\bar{\nu}}(\nu_{mj}) = 2\mathcal{D}_{th}/|M_k|$, 策略 ν 的边际效用值 $f_\nu(\nu_{mj}) = \mathcal{D}_{bh} - \mathcal{D}_{th} + 2\mathcal{D}_{th}/|M_k|$ 。由于 $\mathcal{D}_{bh} > \mathcal{D}_{th}$, 因此 $f_\nu(\nu_{mj}) > f_{\bar{\nu}}(\nu_{mj})$ 。综合以上 3 种情况可知, $f_\nu(\nu_{mj}) \geq f_{\bar{\nu}}(\nu_{mj})$ 。因此, 式(18)中的目标函数是单调非递增的次模函数。

证毕。

3) 优化策略

在优化问题拟阵和次模单调性证明的基础上, 可利用贪婪算法求解问题^[22-23]。算法 1 给出了缓存资源部署的贪婪策略。该策略首先将缓存部署策略 ν 设置为空集, 然后循环计算每个文件在 AP 中部署时系统的边际效用 $f_\nu(\nu_{mj})$, 选择效用最大的部署 ν_{mj} 添加至集合 ν 中, 重复此操作, 直到缓存空间存满或边际效用值小于或等于 0 时停止循环。

算法 1 基于贪婪策略的缓存部署算法

输入 用户业务请求到达率 $\{\lambda_{M_k, k}^f\}$, 频谱效率 $\{SE_k\}$, 天线选择系数 $\{\alpha_k\}$, 内容大小 $\{S_{f_m}\}$ 和 AP 缓存容量 $\{C_l\}$

输出 缓存部署策略

- 1) 初始化 $i=0, \nu^i = \emptyset, f_{\nu^i} = \mathcal{D}_{total}(\emptyset)$,
- $C^i = \sum_{f_m \in \mathcal{F}} S_{f_m} \nu_{ml}^i = 0$
- 2) while $f_{\nu^i} > 0$ or $C^i \leq \sum_{l \in L} C_l$ do
- 3) 求出使边际效用值最大的缓存部署 $(m^*, l^*) \leftarrow \operatorname{argmax} f_\nu(\nu_{mj})$
- 4) 令 $\nu_{m^* l^*}^i = 1$
- 5) 令 $i = i + 1$
- 6) 更新缓存占用容量 $C^i = C^i + S_{f_m}$
- 7) 更新边际效用 $f_{\nu^i} = f_{\nu^i}(\chi_{m^* l^*})$
- 8) 更新缓存部署 $\nu^i = \nu^i \cup \chi_{m^* l^*}$
- 9) end while

算法复杂度分析。算法 1 中, 如果所有 AP 具有相同的缓存大小 $C_l = C$, 则平均迭代次数为 LC 。每次迭代需要计算不超过 LF 次的边际效用值。每次计算边际效用值需要 $O(K)$ 时间。因此, 本文所提贪婪算法的运行时间为 $O(CFL^2K)$ 。

4 仿真结果与分析

本节提供了数值和仿真结果，从传输时延和缓存命中率两方面对缓存辅助的 Cell-Free 大规模 MIMO 系统的性能进行分析。假设 50 个配置 4 根天线的 AP 和 5 个单天线 UE 均匀独立分布在 $1\ 000\text{ m} \times 1\ 000\text{ m}$ 的正方形区域内。LoS 分量 $\bar{\mathbf{g}}_{lk}$ 被建模为

$$\bar{\mathbf{g}}_{lk} = [1, e^{-j2\pi d \sin(\theta_{lk})}, \dots, e^{-j2\pi d(N-1)\sin(\theta_{lk})}]^T \quad (23)$$

其中， $d=0.5$ 表示天线间距系数， $\theta_{lk} \sim \mathcal{U}[-\pi, \pi]$ 表示第 l 个 AP 与第 k 个 UE 之间的到达角。此外，构造 NLoS 分量的相关矩阵为 $[\mathbf{A}_{lk}]_{n_1 n_2} = r_{lk}^{|n_1 - n_2|}$ ，其中 $0 < r_{lk} \leq 1$ 是第 l 个 AP 和第 k 个 UE 的天线相关系数^[24]。

基于文献[25]中的 3GPP 信道模型和文献[16]中的建议，第 l 个 AP 和第 k 个 UE 之间具有 LoS 分量的可能性主要取决于它们之间的距离 d_{lk} 。假设所有 AP 和 UE 对的距离 $d_{lk} \geq 20\text{ m}$ ，定义具有 LoS 分量的概率如下

$$\text{Possibility(LoS)} = \begin{cases} 1 - \frac{d_{lk}}{300}, & 0 < d_{lk} < 300\text{ m} \\ 0, & d_{lk} \geq 300\text{ m} \end{cases} \quad (24)$$

根据式(24)，可以计算出第 l 个 AP 和第 k 个 UE 之间的 Rician 因子^[8,14]为

$$k_{lk} = \begin{cases} 10^{1.3-0.003d_{lk}}, & \text{LoS 分量存在} \\ 0, & \text{LoS 分量不存在} \end{cases} \quad (25)$$

对于路径损失模型，采用 COST 321 Walfisch Ikegami 模型，AP 高度为 12.5 m，UE 高度为 1.5 m。根据式(24)，给出第 l 个 AP 与第 k 个 UE 之间对应的大规模信道衰落系数（以 dB 为单位）如下^[16,24]

$$\beta_{l,k} = \begin{cases} -30.18 - 26\lg\left(\frac{d_{lk}}{1\text{m}}\right) + F_{lk}, & k_{lk} \neq 0 \\ -34.53 - 38\lg\left(\frac{d_{lk}}{1\text{m}}\right) + F_{lk}, & k_{lk} = 0 \end{cases} \quad (26)$$

其中， $F_{lk} = \sqrt{\delta}a_l + \sqrt{1-\delta}b_k$ 表示相关阴影衰落系数， $a_l \sim \mathcal{N}(0, \sigma_{sf}^2)$ ， $b_k \sim \mathcal{N}(0, \sigma_{sf}^2)$ 。仿真中使用的其他系统参数如表 1 所示^[2,13]。

表 1 仿真中使用的其他系统参数

参数	含义
带宽/MHz	20
噪声功率 N_0 /dBm	-90
p_p, p_d /mW	100/ N_0 , 200/ N_0
σ_{sf}, δ	8, 0.5
τ_p, τ_c	10, 200
文件大小 S_{fm} /MB	5
传输速率 $r^{\text{FH}}, r^{\text{BH}}$ / (Mbit·s ⁻¹)	20, 5

本文重点分析缓存策略的归一化内容传输时延和缓存命中率。归一化内容传输时延定义为

$$\text{归一化内容传输时延} = \frac{\mathcal{D}_{\text{total}}(\text{采用缓存策略})}{\mathcal{D}_{\text{total}}(\text{未采用缓存策略})} \quad (27)$$

其中， $\mathcal{D}_{\text{total}}(\bullet)$ 表示在某一缓存策略下所有用户的内容请求总传输时延。缓存命中率定义为在 AP 簇的本地缓存中找到请求内容的概率。

本文将与以下现有缓存策略进行比较。

1) 基于最大流行度的微 (FemtoMPC, femto-caching with most popular content) 缓存策略。该策略是通过修改微缓存策略^[9]得到。在该策略中，UE 附近的 AP 会根据本地流行度分布来缓存最流行的内容，直到它们的缓存空间被填满。同时，该策略用于小蜂窝网络，不考虑 AP 间的协作传输和协作缓存，UE 只能由网络中锚定的一个 AP 提供服务。

2) 基于最大流行度与最大内容多样性的联合缓存策略，简称为 MPC&LCD (joint content placement with most popular content and largest content diversity) 缓存策略。该策略源于基于流行度的组合缓存策略^[12]，其中，AP 的缓存空间被划分为存储最流行的内容和不太流行的内容。该策略用于 Cell-Free 大规模 MIMO 系统。

3) 基于遗传算法 (GA, genetic algorithm) 缓存策略。该策略通过修改文献[13]中基于遗传算法的缓存部署策略得出。同时，该策略用于 Cell-Free 大规模 MIMO 系统，且考虑 AP 之间协作传输和协作缓存。

4) 随机缓存策略。该策略从系统中随机选择 AP 作为缓存点来缓存内容^[14]，该策略用于 Cell-Free 大规模 MIMO 系统。

图 2 给出了不同归一化缓存大小下，5 种缓存策略在缓存命中率和内容传输时延方面的性能。其

中, 归一化缓存定义为 AP 处的存储空间总量与系统存储空间总量的比值, 设置为 5%~35%, 文件数 $M=300$ 。从图 2(a)可以看出, 所有策略的缓存命中率都随着归一化缓存的增加而增大。本文所提缓存策略在提高缓存命中率方面始终具有优势, 其性能优于其他策略, 所提缓存策略的平均缓存命中率约为 0.9, 比 GA 策略高约 7%, 比随机策略高约 30%, 比 FemtoMPC 和 MPC&LCD 高近 2.5 倍。从图 2(b)可以看出, 本文所提缓存策略传输时延性能优于 FemtoMPC 策略, 因为 FemtoMPC 策略采用蜂窝网络的传输方式, 每个 UE 只由一个选中的 AP 提供服务, 而本文所提缓存策略利用 Cell-Free 大规模 MIMO 系统协作传输特性, 可降低用户间干扰和提升数据传输速率。GA 缓存策略的传输时延性能优于 MPC&LCD 协作策略, 原因在于 GA 缓存策略有更多的机会缓存不同的内容, 从存储的角度提高了

缓存命中率并降低了传输时延。

图 3 比较了 5 种缓存策略在 Zipf 分布的偏态因子 γ 变化下的缓存命中率和传输时延性能, 偏态因子在 0.2~1.7 变化, 文件数 $M=300$ 。由图 3(a)可以看出, 本文所提缓存策略的命中率比 GA 缓存策略提高了约 4%, 且具有更低的计算复杂度; 同时, 本文所提缓存策略的命中率比 FemtoMPC 策略、MPC&LCD 策略和随机缓存策略提高了约 2 倍。由于本文所提缓存策略不仅基于文件流行度, 还基于用户活跃度及文件请求量, 因此所提缓存策略和 GA 缓存策略的命中率并未严格随着偏态因子的增加而增大。由图 3(b)可以看出, 本文所提缓存策略、GA 缓存策略、FemtoMPC 策略和 MPC&LCD 策略的内容传输时延随着偏态因子 γ 的增加而急剧下降, 因为不均匀的流行度分布可以更好地体现缓存策略的优势。随机缓存策略由于其随机部署特性,

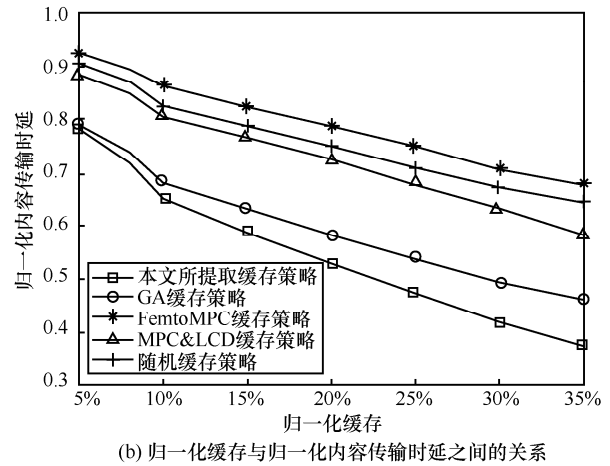
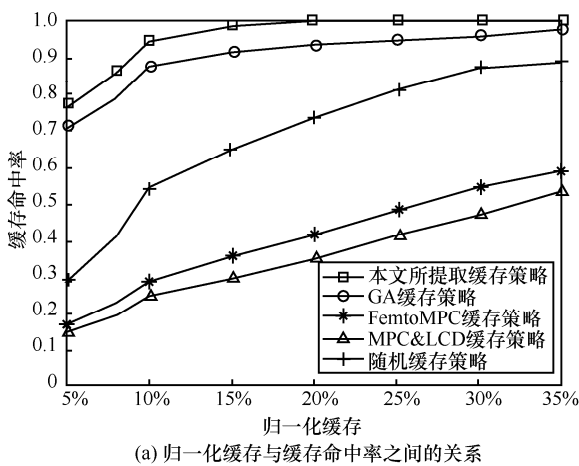


图 2 不同归一化缓存大小下, 5 种缓存策略在缓存命中率和内容传输时延方面的性能

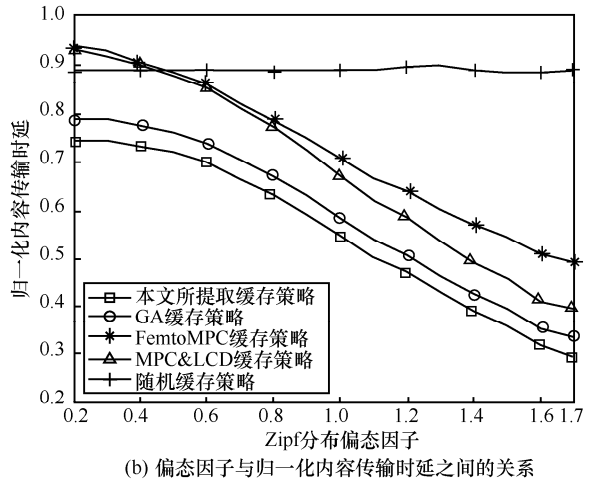
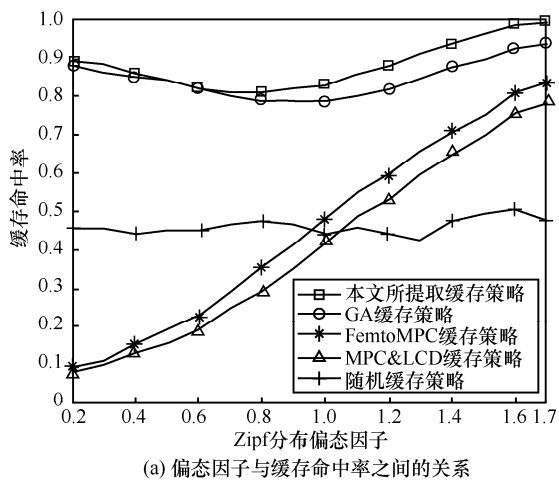


图 3 5 种缓存策略在 Zipf 分布的偏态因子 γ 变化下的缓存命中率和传输时延性能

其缓存部署未考虑文件流行度，因此未随着偏态因子的增加而降低时延。本文所提缓存策略在传输时延性能上比 GA 缓存策略降低约 6%，比 MPC&LCD 策略降低约 19%，比 FemtoMPC 策略降低约 23%，比随机缓存策略降低约 37%。改进的原因是本文所提缓存策略进行了协作传输及协作缓存，提高数据传输速率及内容部署多样性、灵活性。此外，本文所提缓存策略可跟踪本地流行度及用户需求的变化，从而能进一步降低传输时延。

图 4 评估了本文所提缓存策略与其他策略在增加内容数量情况下的性能。仿真中的内容总数设置为 100~1 000。从图 4(a)可知，本文所提缓存策略与 GA 缓存策略、FemtoMPC 策略、MPC&LCD 策略和随机缓存策略相比，分别提高了约 35%、60%、70%和 35%的缓存命中率。同时，随着内容数量的增加，GA 缓存策略、FemtoMPC 策略、MPC&LCD 策略和随机缓存策略的缓存命中率都会下降。这是因为当整个网络的内容越多时，用户请求的分散程度就越大，导致热门内容的受欢迎程度被稀释。同时，由于 AP 的缓存容量有限，无法存储更多流行的内容，从而导致缓存命中率的降低。而本文所提缓存策略为了最小化用户传输时延，通过更准确地跟踪区域文件流行度和用户活跃度，可保持较高的缓存命中率和容纳更多的内容文件。

由图 4(b)可以观察到，所有缓存策略的归一化传输时延都随着内容数量的增加而增加。传输时延增加的原因在于随着文件数量的增加和缓存容量的限制，AP 簇中会有更多的 AP 无法缓存用户所需的内容。从图 3(b)中可以看出，当系统中的内容数量不大（为 100~400）时，所提缓存策略与 GA 缓存策略、MPC&LCD、FemtoMPC 和随机缓存策略相比，可分别减少约 11%、34%、24%和 46%的系统归一化内容传输时延。随着内容数量的增加，所提策略的性能依然优于其他缓存策略。对比本文所提缓存策略与 GA 缓存策略可发现，随着文件规模的增大，所提缓存策略与 GA 缓存策略传输时延性能差异变大，原因在于 GA 虽然是自适应的全局搜索算法，但具有过早收敛、进化后期物种多样性降低的缺点，因此，随着文件规模变大，常常会陷入局部最优情况。结果表明，本文所提缓存策略在不同规模的缓存网络下都能保持良好的性能，表明了该策略的稳定性。

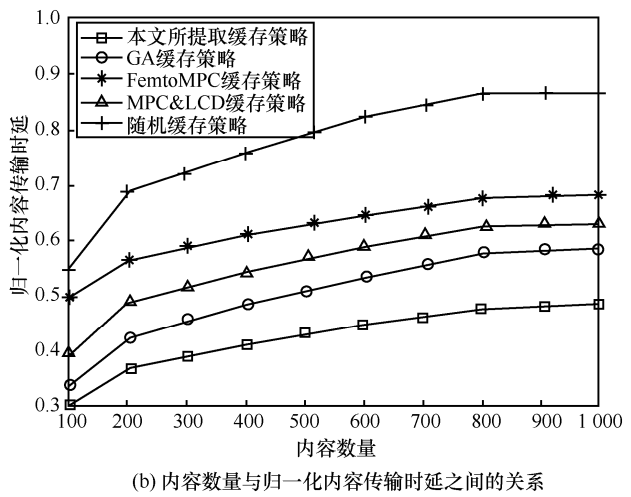
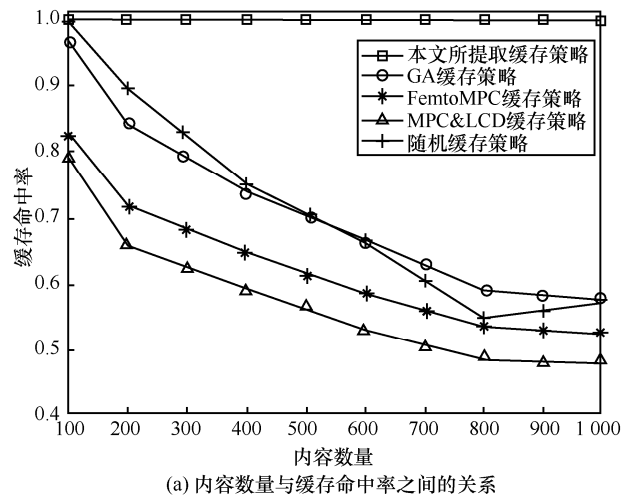


图 4 本文所提缓存策略与其他策略在增加内容数量情况下的性能

5 结束语

本文研究了莱斯衰落信道下缓存辅助的 Cell-Free 大规模 MIMO 系统的传输时延和缓存命中率。首先基于 AP 间协作缓存及区域流行度评估进行缓存建模，推导出考虑 AP 分簇、协作缓存及区域流行度的传输时延表达式。为了在降低前传链路压力的同时降低系统的传输时延，提出了以最小化传输时延为目标的缓存部署优化问题。接着，通过对优化问题 NP-hard 及拟阵约束下次模单调性的证明，提出了低复杂度的贪婪缓存部署策略。仿真结果表明，与缓存辅助的小蜂窝网络相比，缓存辅助的 Cell-Free 大规模 MIMO 系统能更有效地降低用户的传输时延。同时，与 Cell-Free 架构下的 MPC&LCD 组合、遗传算法等方案相比，本文所提缓存策略在命中率及传输时延方面均有优势，从而提高了通信服务质量。

参考文献:

- [1] RAJATHEVA N, ATZENI I, BJORNSSON E, et al. White paper on broadband connectivity in 6G[J]. arXiv Preprint, arXiv:2004.14247, 2020.
- [2] NGO H Q, ASHIKHMIN A, YANG H, et al. Cell-Free massive MIMO versus small cells[J]. IEEE Transactions on Wireless Communications, 2017, 16(3): 1834-1850.
- [3] BASHAR M, CUMANAN K, BURR A G, et al. Energy efficiency of the Cell-Free massive MIMO uplink with optimal uniform quantization[J]. IEEE Transactions on Green Communications and Networking, 2019, 3(4): 971-987.
- [4] BASHAR M, NGO H Q, BURR A G, et al. On the performance of backhaul constrained Cell-Free massive MIMO with linear receivers[C]// Proceedings of 2018 52nd Asilomar Conference on Signals, Systems, and Computers. Piscataway: IEEE Press, 2018: 624-628.
- [5] NGO H Q, TRAN L N, DUONG T Q, et al. On the total energy efficiency of Cell-Free massive MIMO[J]. IEEE Transactions on Green Communications and Networking, 2018, 2(1): 25-39.
- [6] 黄永明, 郑冲, 张征明, 等. 大规模无线通信网络移动边缘计算和缓存研究[J]. 通信学报, 2021, 42(4): 44-61.
- HUANG Y M, ZHENG C, ZHANG Z M, et al. Research on mobile edge computing and caching in massive wireless communication network [J]. Journal on Communications, 2021, 42(4): 44-61.
- [7] 任佳智, 田辉, 范绍帅, 等. 基于用户偏好预测的无人机部署和缓存策略[J]. 通信学报, 2020, 41(6): 1-13.
- REN J Z, TIAN H, FAN S S, et al. UAV deployment and caching scheme based on user preference prediction[J]. Journal on Communications, 2020, 41(6): 1-13.
- [8] PENG X, SHEN J C, ZHANG J, et al. Backhaul-aware caching placement for wireless networks[C]//Proceedings of 2015 IEEE Global Communications Conference (GLOBECOM). Piscataway: IEEE Press, 2015: 1-6.
- [9] SHANMUGAM K, GOLREZAEI N, DIMAKIS A G, et al. Femto-Caching: wireless content delivery through distributed caching helpers[J]. IEEE Transactions on Information Theory, 2013, 59(12): 8402-8413.
- [10] 蔡艳, 吴凡, 朱洪波. D2D 协作边缘缓存系统中基于传输时延的缓存策略[J]. 通信学报, 2021, 42(3): 183-189.
- CAI Y, WU F, ZHU H B. Caching strategy based on transmission delay for D2D cooperative edge caching system [J]. Journal on Communications, 2021, 42(3): 183-189.
- [11] WANG R, PENG X, ZHANG J, et al. Mobility-aware caching for content-centric wireless networks: modeling and methodology[J]. IEEE Communications Magazine, 2016, 54(8): 77-83.
- [12] CHEN Z, LEE J, QUEK T Q S, et al. Cooperative caching and transmission design in cluster-centric small cell networks[J]. IEEE Transactions on Wireless Communications, 2017, 16(5): 3401-3415.
- [13] LIN P, SONG Q Y, JAMALIPOUR A. Multidimensional cooperative caching in CoMP-integrated ultra-dense cellular networks[J]. IEEE Transactions on Wireless Communications, 2020, 19(3): 1977-1989.
- [14] LI Y, GURSOY M C, VELIPASALAR S. A delay-aware caching algorithm for wireless D2D caching networks[C]//Proceedings of 2017 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS). Piscataway: IEEE Press, 2017: 456-461.
- [15] ÖZDOĞAN Ö, BJÖRNSSON E, ZHANG J Y. Performance of Cell-Free massive MIMO with Rician fading and phase shifts[J]. IEEE Transactions on Wireless Communications, 2019, 18(11): 5299-5315.
- [16] ÖZDOĞAN Ö, BJÖRNSSON E, LARSSON E G. Massive MIMO with spatially correlated Rician fading channels[J]. IEEE Transactions on Communications, 2019, 67(5): 3234-3250.
- [17] BOUKHEDIMI I, KAMMOUN A, ALOUINI M S. LMMSE receivers in uplink massive MIMO systems with correlated Rician fading[J]. IEEE Transactions on Communications, 2019, 67(1): 230-243.
- [18] WANG R, SHEN M, HE Y, et al. Performance of Cell-Free massive MIMO with joint user clustering and access point selection[J]. IEEE Access, 2021, 9: 40860-40870.
- [19] MÜLLER S, ATAN O, VAN DER SCHAAR M, et al. Context-aware proactive content caching with service differentiation in wireless networks[J]. IEEE Transactions on Wireless Communications, 2017, 16(2): 1024-1036.
- [20] ZHANG S, YAO L N, SUN A X, et al. Deep learning based recommender system[J]. ACM Computing Surveys, 2019, 52(1): 1-38.
- [21] CAROFIGLIO G, GALLO M, MUSCARIELLO L. On the performance of bandwidth and storage sharing in information-centric networks[J]. Computer Networks, 2013, 57(17): 3743-3758.
- [22] XIA W C, QUEK T Q S, ZHANG J, et al. Resource allocation by submodular optimization in programmable hierarchical C-RAN[C]// Proceedings of 2018 IEEE/CIC International Conference on Communications in China (ICCC). Piscataway: IEEE Press, 2018: 558-562.
- [23] CALINESCU G, CHEKURI C, PÁL M, et al. Maximizing a monotone submodular function subject to a matroid constraint[J]. SIAM Journal on Computing, 2011, 40(6): 1740-1766.
- [24] TARICCO G, RIEGLER E. On the ergodic capacity of correlated Rician fading MIMO channels with interference[J]. IEEE Transactions on Information Theory, 2011, 57(7): 4123-4137.
- [25] 3rd Generation Partnership Project. Technical specification group radio access network; spatial channel model for multiple input multiple output (MIMO) simulations: 3GPP TR 25.996 V14.0.0[S]. 2017.

[作者简介]



王蕊 (1988-), 女, 云南玉溪人, 重庆邮电大学博士生, 主要研究方向为 Cell-Free 大规模 MIMO 系统的资源分配、动态协作和预编码。

申敏 (1963-), 女, 贵州湄潭人, 重庆邮电大学教授、博士生导师, 主要研究方向为通信核心芯片、协议与系统应用技术。

何云 (1979-), 女, 湖北武汉人, 重庆邮电大学博士生, 主要研究方向为协作通信、大规模 MIMO 系统能效优化。

刘香燕 (1992-), 女, 四川资阳人, 重庆邮电大学博士生, 主要研究方向为移动边缘计算、资源分配和 D2D 通信。