

D2D 协作边缘缓存系统中基于传输时延的缓存策略

蔡艳, 吴凡, 朱洪波

(南京邮电大学江苏省无线通信重点实验室, 江苏 南京 210003)

摘要: 为了满足 5G 系统低时延高可靠的需求, 针对单缓存终端直传 (D2D) 协作边缘缓存系统, 提出了一种基于传输时延的缓存策略。运用随机几何理论, 将请求用户和空闲用户的动态分布建模为相互独立的齐次泊松点过程, 综合考虑内容流行度、用户位置信息、设备传输功率以及干扰, 推导出用户的平均传输时延与缓存概率分布的关系式。以平均传输时延为目标函数建立优化问题, 提出了一个低复杂度的迭代算法, 得到平均传输时延最优的缓存策略。仿真结果表明, 该缓存策略在传输时延方面优于常见的几种缓存策略。

关键词: 移动边缘网络; 终端直传协作缓存; 内容流行度; 缓存概率分布; 传输时延

中图分类号: TN92

文献标识码: A

DOI: 10.11959/j.issn.1000-436x.2021042

Caching strategy based on transmission delay for D2D cooperative edge caching system

CAI Yan, WU Fan, ZHU Hongbo

Key Laboratory of Wireless Communications of Jiangsu Province, Nanjing University of Posts and Telecommunications, Nanjing 210003, China

Abstract: To meet the requirements of 5G system with low delay and high reliability, for a single-caching D2D cooperative edge caching system, an optimization of the caching strategy based on transmission delay was proposed. The dynamic distributions of requesting users and idle users were modeled as the independent homogeneous poisson point process (HPPP) utilizing the stochastic geometry theory. Comprehensively considering the content popularity, user location information, device transmission power and interference, the relationship between the average transmission delay of user and the cache probability distribution was derived. Taking the average transmission delay as the objective function, the optimization problem was established, and a low complexity iterative algorithm was proposed to obtain the cache strategy with sub-optimal average transmission delay. Simulation results demonstrate that the proposed optimization cache strategy is superior to several common cache strategies in terms of transmission delay.

Keywords: mobile edge network, D2D cooperative caching, content popularity, distribution of caching probability, transmission delay

1 引言

随着无线通信技术的发展, 移动终端的数量不断增加, 人们对无线数据的需求也不断增长。为了满足这一需求, 5G 制定了将现有网络容量提高 1 000 倍的目标^[1]。现有的蜂窝网络属于中心式

架构, 在处理庞杂的数据时受限于回程链路的压力, 会出现网络拥塞, 无法满足“低时延高可靠”的要求, 因此移动边缘缓存、D2D(device-to-device)通信等能够缓解蜂窝网络压力的边缘通信技术成为研究热点。

文献[2]指出, 用户对视频内容的请求具有重复

收稿日期: 2020-07-20 ; 修回日期: 2020-10-21

基金项目: 国家自然科学基金资助项目 (No.61671251, No.61871446); 江苏省重点研发计划基金资助项目 (No.BE2020084-5)

Foundation Items: The National Natural Science Foundation of China (No.61671251, No.61871446), Jiangsu Provincial Key Research and Development Program (No.BE2020084-5)

性、可预测性的特点, 热门内容会在一定时间内被大量用户重复请求, 而用户的请求行为也可通过历史记录进行预估。根据这一特点, 边缘缓存系统可以在通信低谷时段于边缘节点处提前缓存热门内容, 并在通信高峰时段由边缘节点传给请求用户, 从而缓解回程链路压力, 避免网络拥塞^[3-4]。与传统的边缘缓存系统不同, D2D 协作边缘缓存系统充分利用了闲置的用户设备存储空间, 将热门内容分布式缓存, 用户之间可通过 D2D 协作完成内容的传输。相较于缓存在小基站 (BS, base station) 中, D2D 协作边缘缓存系统能够有效减轻基站负载, 降低传输时延^[5-6]。但是, 有限的缓存空间、用户喜好的差异性和终端设备的移动性为缓存策略的制定增加了复杂度, 如何缓存才能更有效地提高系统性能成为 D2D 协作边缘缓存系统的重点研究问题^[7-13]。

文献[7]运用随机几何理论, 将请求用户和空闲用户的分布建模为相互独立的齐次泊松点过程 (HPPP, homogeneous Poisson point process), 以缓存命中率为性能指标, 通过最大化平均缓存命中率优化缓存内容部署, 并提出最优对偶搜索算法解决优化问题。文献[8]针对最大化缓存命中率的优化问题, 提出一种截断式 Zipf 分布缓存策略, 通过低复杂度算法求解 2 个缓存参数, 即截断门限和 Zipf 指数, 联合优化缓存分布。文献[9]将缓存命中率和信道衰落相结合, 推导出缓存内容的覆盖率与缓存策略的关系式, 以覆盖率为目标函数优化缓存部署。文献[10]在用户建模服从 HPPP 分布的基础上, 分析了由其他 D2D 设备带来的干扰, 在满足接收信干噪比和 D2D 最远通信距离的双重约束下, 推导出成功传输概率的闭式表达式。文献[11]考虑到不同用户喜好之间的相似性, 提出一种基于 k-means 均值学习的协作算法, 对用户进行聚类, 再采用基于聚类的贪婪算法得到缓存策略。文献[12]针对不同用户缓存能力的差异, 根据缓存空间的大小将用户分别按相互独立的 HPPP 分布建模, 推导出缓存命中率, 最后采用基于坐标梯度的联合缓存布设算法优化缓存方案。文献[13]则以最大化蜂窝网络流量卸载为目标, 将缓存策略优化问题转化为背包问题, 通过贪婪算法求解。另有研究考虑了系统的能效问题, 文献[14]将用户分簇后联合考虑簇内用户之间的拓扑结构和用户设备剩余电量, 基于最小化传输能耗优化缓存策略。

以上对 D2D 协作边缘缓存系统的缓存策略研究中, 性能指标主要考虑缓存命中率^[7-8,11-12]、成功传输

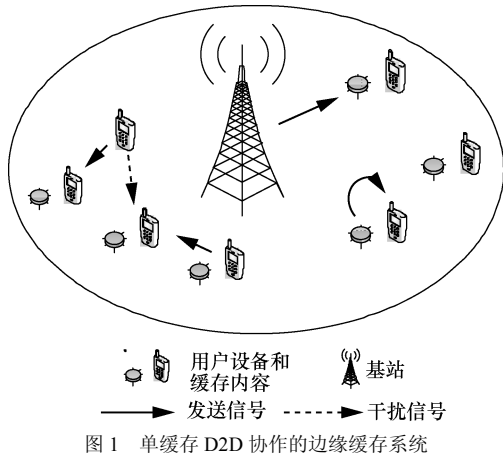
概率^[9-10,13]和传输能耗^[14], 没有考虑用户服务质量 (QoS, quality of service)。通信速率和传输时延是用户 QoS 的重要性能指标, 也是衡量系统性能的重要因素, 目前从通信速率、传输时延角度来优化缓存策略的研究较少。文献[15]在考虑用户偏好的基础上给出基于缓存命中率最大化的缓存策略, 通过比较传输时延证明该策略能够有效提升用户 QoS。文献[16]参考内容流行度, 为了在减少传输时延的同时增加本地命中率, 将缓存内容放置问题转化为 0-1 背包问题, 通过低复杂度的启发式算法优化缓存策略。但上述研究都只是间接地考虑缓存策略对传输时延的影响, 没有探究缓存策略与传输时延间的直接关系。

本文在 D2D 协作边缘缓存系统的缓存策略优化方案中考虑用户 QoS, 首先, 提出了一种基于平均传输时延最小化的缓存策略, 将缓存概率分布问题建模成平均传输时延最小化问题, 运用随机几何理论, 将请求用户和空闲用户的动态分布建模为相互独立的 HPPP, 再综合考虑内容流行度、用户位置信息、设备传输功率以及干扰, 推导出用户的平均传输时延与缓存概率分布的关系式; 然后, 以平均传输时延为目标函数建立优化问题; 最后, 提出一个低复杂度的迭代算法, 得到平均传输时延最优的缓存策略。通过仿真与其他考虑时延传输的缓存策略进行比较, 验证了所提缓存策略可有效降低传输时延, 提升通信服务质量。

2 系统模型

2.1 系统模型与缓存模型

考虑如图 1 所示的单缓存 D2D 协作的边缘缓存系统。该系统为单基站多用户小区, 基站位于小区中心, 用户位置服从密度为 λ_u 的 HPPP 分布, 记作 Φ_u 。用户设备可工作在蜂窝通信和 D2D 通信 2 种模式, D2D 通信与蜂窝通信之间采用 Overlay 方式共享资源, 相互间无干扰。考虑到 D2D 通信为带内 D2D, 用户设备不能同时进行蜂窝通信和 D2D 通信, 且 D2D 通信为半双工模式, 因此只有空闲的用户设备才能作为 D2D 发送端。当请求用户通过 D2D 链路接收内容时, 会受到作为 D2D 发送端的其他设备的同频干扰。本文假设同一时刻空闲用户占小区所有用户比例为 α , 则空闲用户服从密度为 $\alpha\lambda_u$ 的 HPPP 分布, 记作 Φ_r ^[10]。在满足最远 D2D 通信距离为 r_c 的约束下, D2D 发送设备采用自适应的发送功率, 使接收功率保持为 S , 以降低设备能耗。



基站收集小区内用户的分布信息和热门内容的流程度, 基于时延最小化来制定缓存策略 P 。本文假设云服务器端共有 K 个热门内容, 按流行度从高到低排序, 第 i 个内容可以表示为 $F_i, i=1,2,3,\dots,K$ 。每个内容数据大小为 B bit。热门内容的流行度可用请求概率表示, 将热门内容的请求概率记为 $\mathbf{Q}=[q_1, q_2, q_3, \dots, q_K]$, q_i 表示用户请求内容 F_i 的概率。文献[17]指出, 用户对内容的请求概率可以用 Zipf 分布近似描述, 将内容按其热度从高到低排序, 则用户请求内容 F_i 的概率为

$$q_i = \frac{i^{-\gamma}}{\sum_{j=1}^K j^{-\gamma}}, i=1,2,3,\dots,K \quad (1)$$

其中, γ 是 Zipf 流行度指数, γ 越大, 用户请求越集中在最热门的内容上。

基站在制定缓存策略后, 将热门内容在业务空闲时段通过蜂窝网络主动缓存在用户设备上, 以缓解业务高峰时段的基站压力。由于下一高峰时段的用户位置和请求行为都是随机值, 概率缓存模式相较于确定性缓存模式, 能带来更多的性能提升^[18], 因此本文采用概率缓存模式。假设用户设备的缓存空间只能缓存一个内容, 则缓存策略 $\mathbf{P}=[p_1, p_2, p_3, \dots, p_K]$, p_i 表示用户缓存内容 F_i 的概率, 且满足

$$\sum_{i=1}^K p_i = 1, 0 \leq p_i \leq 1 \quad (2)$$

相应地, 缓存有内容 F_i 的空闲用户服从密度为 $p_i \alpha \lambda_u$ 的 HPPP 分布, 记为 Φ_i^c 。

2.2 内容获取模式

在业务高峰时段, 当用户请求热门内容时, 设备首先检索本地缓存, 若缓存命中, 则直接从本地

缓存中获取目标内容, 此时的通信状态为本地命中; 若本地未命中, 则将该请求信息发送至基站, 由基站调度通信链路的建立。基站在收到用户设备发来的请求后, 根据各个用户的位置信息和缓存信息, 在以请求用户为圆心、以 r_c 为半径的范围内检索其他空闲设备的缓存空间。如果存在至少一个空闲设备缓存了目标内容, 则在所有命中的设备中, 选择距离最近的设备建立 D2D 链路, 此时通信状态为 D2D 通信; 否则用户只能通过蜂窝网由云端服务器下载内容, 此时通信状态为蜂窝通信。

3 系统性能分析

3.1 卸载概率分析

由 2.2 节可知, 基于 D2D 协作的边缘缓存系统中用户的通信状态有 3 种: 本地命中、D2D 通信以及蜂窝通信。制定缓存策略阶段, 用户在下一高峰时段的通信状态是一个随机量。定义用户处于 D2D 通信状态的概率为卸载概率, 即业务被卸载到边缘通过 D2D 完成的概率。

卸载概率与缓存策略 \mathbf{P} 相关。如果所有用户都以概率 p_i 缓存内容 F_i , 则最终缓存了内容 F_i 的空闲用户呈 Φ_i^c 分布, 密度 $\lambda_i^c = p_i \alpha \lambda_u$ 。因此, 在本地未命中, 且用户设备向基站发送对内容 F_i 请求信息的条件下, 其卸载概率记为 P_i^O ^[19]。

$$P_i^O = 1 - e^{-p_i \alpha \lambda_u \pi r_c^2} \quad (3)$$

用户请求热门内容的概率即内容流行度 $\mathbf{Q}=[q_1, q_2, q_3, \dots, q_K]$ 服从 Zipf 分布。用户只有本地未缓存目标内容时, 才会向基站发出请求。用 t_i 表示用户设备向基站发送对内容 F_i 请求信息的概率, 即

$$t_i = q_i(1 - p_i) \quad (4)$$

由全概率公式可知, 用户的卸载概率 P_O 为

$$P_O = \sum_{i=1}^K t_i P_i^O = \sum_{i=1}^K \frac{q_i}{\sum_{j=1}^K q_j} (1 - p_i) (1 - e^{-p_i \alpha \lambda_u \pi r_c^2}) \quad (5)$$

3.2 平均 D2D 通信速率分析

缓存策略只影响用户的 D2D 通信速率, 不影响蜂窝通信速率, 因此可以将蜂窝通信速率设为固定值。在缓存策略制定阶段, 基站可以根据已知的用户分布密度 λ_u 和最远 D2D 通信距离 r_c , 分别计算出请求不同内容的用户的平均 D2D 通信速率

$\bar{R}_i^{\text{D2D}}, i=1,2,3,\dots,K$ 。

如果用户在下一个高峰时段向基站发出对内容 F_i 的请求, 基站在收到请求后通过检索距离该用户 r_c 范围内的空闲用户, 建立 D2D 通信链路。用户设备除了接收已缓存内容 F_i 的空闲设备发来的信号外, 还收到了缓存其他内容的空闲设备发来的干扰信号。根据香农定理, 用户通过 D2D 传输内容 F_i 的平均速率为

$$\bar{R}_i^{\text{D2D}} = W \log \left(1 + \frac{S}{\sum_{j \neq i}^K I_j + N} \right) \quad (6)$$

其中, \bar{R}_i^{D2D} 为请求内容 F_i 的用户 D2D 通信速率的期望值, W 为信道带宽, S 为 D2D 接收功率, N 为白噪声功率, I_j 为来自自己缓存内容 F_i 的空闲设备的干扰功率。

由式(6)可知, 用户收到干扰的总功率与能够造成干扰的空闲设备数量相关, 而后者又与本文优化目标缓存策略相关。为简化处理, 假设每个设备产生的干扰功率平均值为 \bar{I} , 且该用户 r_c 范围外的干扰信号由于距离过远可以忽略不计, 则

$$\bar{R}_i^{\text{D2D}} = W \log \left(1 + \frac{S}{\sum_{j \neq i}^K p_j \alpha \lambda_u \pi r_c^2 \bar{I} + N} \right) = W \log \left(1 + \frac{S}{\alpha \lambda_u \pi r_c^2 \bar{I} (1 - p_i) + N} \right) \quad (7)$$

3.3 平均通信速率分析

当用户请求内容且本地命中时, 则直接从本地缓存中获取目标内容, 此时不考虑通信速率。当本地未命中时, 用户会向基站发出请求, 其可能存在的通信状态有 D2D 通信和蜂窝通信 2 种。定义平均通信速率为这 2 种通信状态下通信速率的条件期望。由式(5)可知, 用户请求内容 F_i 且被卸载的概率为 $t_i P_i^0 = q_i (1 - p_i) P_i^0$, 而用户向基站发出请求的概率为 $\sum_{j=1}^K q_j (1 - p_j)$ 。则在用户向基站发出请求的条件下, 请求内容 F_i 且通过 D2D 传输的条件概率为

$$P_i^{\text{D2D}} = \frac{q_i (1 - p_i) P_i^0}{\sum_{j=1}^K q_j (1 - p_j)} \quad (8)$$

同理, 在用户向基站发出请求的条件下, 请求

内容 F_i 且通过蜂窝传输的条件概率为

$$P_i^{\text{BS}} = \frac{q_i (1 - p_i) (1 - P_i^0)}{\sum_{j=1}^K q_j (1 - p_j)} \quad (9)$$

当用户向基站发出请求后, 通信状态有 D2D 通信和蜂窝通信 2 种, 因此, 请求用户的平均通信速率为

$$\bar{R} = \sum_{i=1}^K \bar{R}_i^{\text{D2D}} P_i^{\text{D2D}} + \bar{R}_{\text{BS}} P_i^{\text{BS}} \quad (10)$$

其中, \bar{R}_{BS} 为平均蜂窝通信速率。

3.4 平均传输时延分析

用户的 3 种通信状态分别对应以下 3 种传输时延: 本地命中传输时延 $\bar{T}_{\text{local}} = 0$, D2D 传输时延

$$\bar{T}_i^{\text{D2D}} = \frac{B}{\bar{R}_i^{\text{D2D}}}, \text{ 蜂窝传输时延 } \bar{T}_{\text{BS}} = \frac{B}{\bar{R}_{\text{BS}}}$$

对于任意用户以概率 q_i 请求内容 F_i 时, 本地命中的概率为 $q_i p_i$, D2D 通信概率为 $q_i (1 - p_i) P_i^0$, 蜂窝通信概率为 $q_i (1 - p_i) (1 - P_i^0)$ 。则用户的平均传输时延为

$$\bar{T} = \sum_{i=1}^K q_i \left\{ p_i \bar{T}_{\text{local}} + (1 - p_i) \left[P_i^0 \bar{T}_i^{\text{D2D}} + (1 - P_i^0) \bar{T}_{\text{BS}} \right] \right\} \quad (11)$$

令 $A = \alpha \lambda_u \pi r_c^2$, 则式(11)可以表示为

$$\bar{T} = \sum_{i=1}^K q_i (1 - p_i) \cdot \left[(1 - e^{-A p_i}) \frac{B}{W \log \left(1 + \frac{S}{A \bar{I} (1 - p_i) + N} \right)} + e^{-A p_i} \frac{B}{\bar{R}_{\text{BS}}} \right] \quad (12)$$

4 基于传输时延最小化的缓存策略

4.1 构建目标函数

基于第 3 节的分析, 本节提出基于传输时延最小化的缓存策略。最优缓存内容分布问题可以表示为

$$\min_{p_i} \bar{T} \quad (13)$$

$$\text{s.t. } \sum_{i=1}^K p_i = 1 \quad (14)$$

$$0 \leq p_i \leq 1 \quad (15)$$

式(14)和式(15)为优化问题式(13)的限制条件, 满足缓存概率非负且概率和为 1。

4.2 优化问题求解算法

式(13)为非凸优化问题。本文给出一种可以求解优化问题式(13)的次优迭代算法, 步骤如下。

步骤 1 输入用户分布密度 λ_u 、空闲用户比 α 、D2D 通信允许的最远距离 r_c 等系统参数; 输入信道带宽 W 、平均接受功率 S 、干扰功率 I 、噪声功率 N 、平均蜂窝通信速率 \bar{R}_{BS} 等通信参数; 输入热门内容数量 K 、热门内容大小 B 以及内容流行度 $\mathbf{Q}=[q_1, q_2, q_3, \dots, q_K]$ 。

步骤 2 初始化 $\mathbf{P}=[p_1, p_2, p_3, \dots, p_K]$, 令 $p_1=p_2=\dots=p_K=0$ 。

步骤 3 根据已知参数和 \mathbf{P} , 计算用户的平均传输时延 \bar{T} 关于缓存概率 p_i 的一阶偏导数 $\frac{\partial \bar{T}}{\partial p_i}, i=1, 2, \dots, K$ 。

$$\frac{\partial \bar{T}}{\partial p_i} = q_i(1-p_i) \left[\frac{ABe^{-Ap_i}}{W\text{lb}\left(1+\frac{S}{A\bar{I}(1-p_i)+N}\right)} - \frac{ABe^{-Ap_i}}{\bar{R}_{BS}} + \frac{AB\bar{I}S(e^{-Ap_i}-1)}{\ln 2W(A\bar{I}(1-p_i)+N)(A\bar{I}(1-p_i)+N+S)\left(\text{lb}\left(1+\frac{S}{A\bar{I}(1-p_i)+N}\right)\right)^2} \right] \quad (16)$$

$$q_i \left[\frac{Be^{-Ap_i}}{\bar{R}_{BS}} - \frac{B(e^{-Ap_i}-1)}{W\text{lb}\left(1+\frac{S}{A\bar{I}(1-p_i)+N}\right)} \right]$$

其中, $A = \alpha\lambda_u\pi r_c^2$ 。所得向量 $\left[\frac{\partial \bar{T}}{\partial p_1}, \frac{\partial \bar{T}}{\partial p_2}, \dots, \frac{\partial \bar{T}}{\partial p_K}\right]$ 为迭代方向。

步骤 4 比较 $\frac{\partial \bar{T}}{\partial p_i}, i=1, 2, \dots, K$ 的大小, 找出使 $\frac{\partial \bar{T}}{\partial p_i}$ 值最小的 $j, j = \arg \min_i \frac{\partial \bar{T}}{\partial p_i}$, 令 $p_j = p_j + d$, d 为迭代步长。

步骤 5 如果 $\sum_{i=1}^K p_i < 1$, 则重复步骤 3 和步骤 4; 如果 $\sum_{i=1}^K p_i = 1$, 则输出 \mathbf{P} 。

最终收敛精度受迭代步长影响, 与步长保持一致。算法循环的主体为步骤 3 和步骤 4。步骤 3 每次循环需要计算 K 个缓存概率 p_i 的一阶偏导, 时间复杂度为 $O(n)$; 步骤 4 找出 K 个偏导的最小值,

每次循环需比较 $K-1$ 次, 时间复杂度为 $O(n)$, 循环体的时间复杂度为 $O(n)$ 。共需要循环 $1/d$ 次, 因此算法的总时间复杂度为 $O(n^2)$ 。

5 仿真分析

为了验证本文所提策略的有效性, 本节对所提的优化缓存策略性能进行仿真验证, 并将其与其他考虑时延传输的缓存策略进行对比分析, 具体包括基于卸载概率优化的缓存策略^[15]、基于流行度的缓存策略^[16]和均匀随机的缓存策略。其中, 文献[15]提出的基于卸载概率优化的缓存策略使用户的卸载概率最大化, 尽可能多地让用户使用 D2D 通信, 缓解基站压力; 文献[16]提出的基于流行度的缓存策略中, 用户设备根据视频内容的流行度缓存, 用户请求热门内容的概率越大, 用户设备缓存该内容的概率越大; 均匀随机的缓存策略中, 中心基站不做优化, 所有内容的缓存概率保持一致, 这是为方便对比分析而采用的基本缓存策略。

假设小区覆盖半径为 500 m, 小区内用户服从密度为 0.02 的泊松点分布, 其他仿真参数如表 1 所示, 所有仿真结果均为 1 000 次蒙特卡罗实验的平均值。

表 1 仿真参数与取值

仿真参数	取值
迭代步长 d	10^{-4}
信道带宽 W/MHz	10
D2D 平均接受功率 S/dBm	-60
设备的平均干扰功率 \bar{I}/dBm	-80
噪声功率 N/dBm	-100
平均蜂窝通信速率 $\bar{R}_{BS}/(\text{Mbit}\cdot\text{s}^{-1})$	10
空闲用户占比 α	0.5
内容流行度指数 γ	1
内容大小 B/bit	8×10^5
D2D 通信允许的距离 r_c/m	20~50

图 2 给出了不同缓存策略条件下用户的平均传输时延 \bar{T} 随 D2D 通信允许的最远距离 r_c 的变化曲线。从图 2 中可以看出, 本文所提的基于时延优化的缓存策略在用户平均传输时延上的性能是最优的, 与理论推导吻合。随着 r_c 的增大, 4 种缓存策略的平均传输时延都是先减后增。其原因如下: 一方面, r_c 的增大使用户附近的检索范围增大, 请求用户采用 D2D 通信的概率提高, 而更多用户采用速率更快的 D2D 通信会导致用户平均传输时延的降低; 另一方面, r_c 的

增大也使用户设备的发送功率增加, D2D 链路与链路之间的干扰功率随之上升, 使 D2D 链路的信道质量降低, D2D 通信速率下降, 用户的平均传输时延增加。当 $r_c > 30\text{ m}$ 时, 干扰造成的影响明显大于 D2D 通信链路增加带来的增益, 使用户的平均时延增加。对比基于时延优化的缓存策略和基于卸载概率优化的缓存策略, 它们的性能在 r_c 较小时十分接近, 但随着 r_c 增大性能差距逐渐拉大, 当 r_c 从 30 m 增大到 50 m 时, 平均时延增幅分别为 15% 和 20%, 这是由于基于卸载概率优化的缓存策略的 D2D 用户数是最多的, 来自 D2D 用户数增加的收益较小, 更容易被干扰影响, 随着 r_c 增大其时延上升趋势更大。

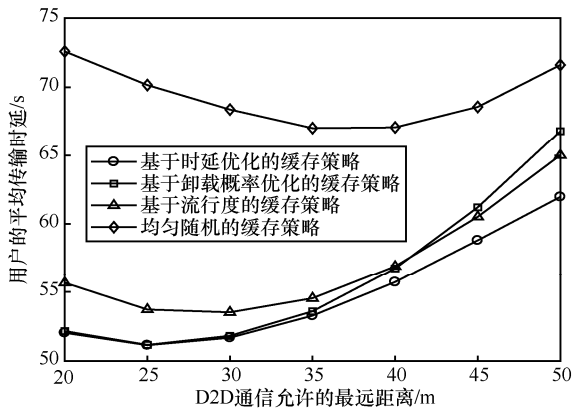


图 2 用户的平均传输时延随 D2D 通信允许的最远距离变化曲线

图 3 给出了不同缓存策略条件下用户的卸载概率 P_0 随 D2D 通信允许的最远距离 r_c 的变化曲线。从图 3 中可以看出, 除了本文所提的基于时延优化的缓存策略外, 其余缓存策略的用户卸载概率都随着 r_c 的增大而增大。而本文的缓存策略中, 当 r_c 小于 30 m 时, 用户卸载概率逐渐增大; 当 r_c 大于 30 m 时, 用户卸载概率逐渐减小, 用户卸载概率的先增后减是由于 r_c 的增大带来 D2D 干扰功率增大, 为了保证时延最低, 基站调整缓存概率分布, 使用户本地命中和蜂窝通信的概率都提高, 从而降低了卸载概率。

固定 D2D 通信允许的最远距离 $r_c=30\text{ m}$, 通过调节用户的分布密度 λ_u , 分析不同的缓存策略在用户密度不同的场景下的通信性能, 结果如图 4 和图 5 所示。图 4 与图 2 类似, 4 种缓存策略的平均传输时延的变化趋势都是先减后增。当 $\lambda_u < 0.01$ 时, 用户稀少, 距离请求用户 r_c 范围内的空闲用户较少, 能够建立 D2D 通信链路的概率较低, 大量用户只能选择蜂窝通信, 导致平均传输时延较高; 随着用户密度增大, 请求用户附近的空闲用户变多, 用户

的卸载概率提高, 部分用户通过 D2D 高速传输, 平均传输时延降低。但用户数的进一步增加导致干扰源变多, D2D 信道质量下降, D2D 通信速率降低, 平均传输时延反而增加。基于卸载概率优化的缓存策略在用户密度较大的场景下依然最大化卸载概率, 受 D2D 干扰的影响较大, 平均传输时延的上升趋势也较大; 与之相比, 本文所提的基于时延优化的缓存策略则综合考虑了卸载概率和干扰, 在 D2D 干扰较大时减少卸载概率, 增加本地命中概率, 保证了用户的平均传输时延最低。

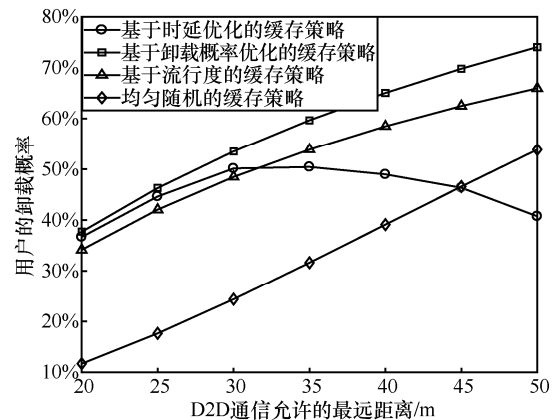


图 3 用户的卸载概率随 D2D 通信允许的最远距离变化曲线

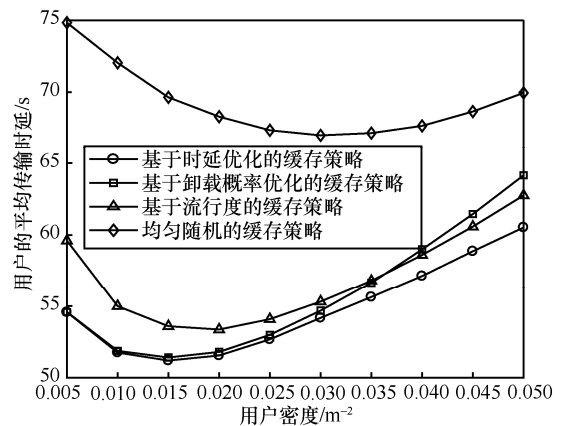


图 4 用户的平均传输时延随用户密度变化曲线

6 结束语

针对单缓存的 D2D 协作边缘缓存系统, 本文提出了一种基于平均传输时延最小化的缓存策略。基站根据内容流行度、用户位置信息、设备传输功率, 计算用户在下一个通信高峰段的平均通信速率, 并推导了用户的平均传输时延与缓存概率分布的关系式。在此基础上, 考虑缓存概率分布满足的限制条件, 建立基于平均传输时延最小化的缓存策略优化模型。针

对此非凸优化问题, 本文提出了一种次优迭代算法, 得到基于时延优化的缓存策略。仿真结果表明, 所提缓存策略通过控制卸载概率, 减少 D2D 之间的干扰, 能够有效降低平均传输时延。下一步的研究将考虑多缓存的系统模型和用户偏好对请求概率的影响, 以及进一步优化迭代算法的性能。

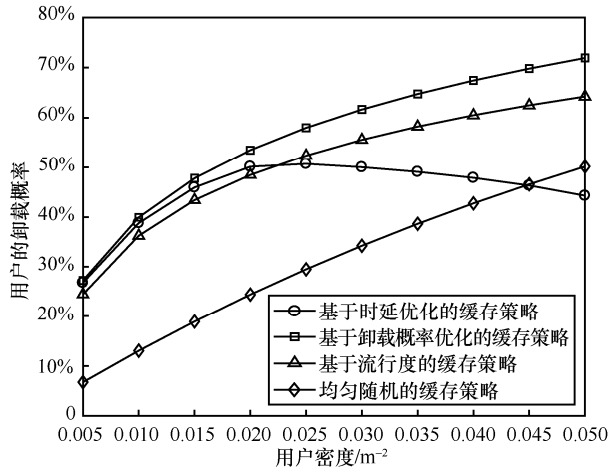


图 5 用户的卸载概率随用户密度变化曲线

参考文献:

- [1] ANDREWS J G, BUZZI S, CHOI W, et al. What will 5G be? [J]. IEEE Journal on Selected Areas in Communications, 2014, 32(6): 1065-1082.
- [2] DILLEY J, MAGGS B, PARIKH J, et al. Globally distributed content delivery[J]. IEEE Internet Computing, 2002, 6(5): 50-58.
- [3] BORST S C, GUPTA V, WALID A. Distributed caching algorithms for content distribution networks[C]//2010 Proceedings IEEE INFOCOM. Piscataway: IEEE Press, 2010: 1-9.
- [4] 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.
- [5] KOBO H I, ABU-MAHFOUZ A M, HANCKE G P. A survey on software-defined wireless sensor networks: challenges and design requirements[J]. IEEE Access, 2017, 5(2):1872-1899.
- [6] KAMILARIS A, PITSILLIDES A. Mobile phone computing and the Internet of things: a survey[J]. IEEE Internet of Things Journal, 2016, 3(6): 885-898.
- [7] BASAR E. Index modulation techniques for 5G wireless networks[J]. IEEE Communications Magazine, 2016, 54(7): 168-175.
- [8] 谷莎莎. 基于 D2D 通信的网络边缘协作缓存策略研究[D]. 武汉: 华中科技大学, 2018.
- [9] GU S S. Research on collaborative caching strategies at network edge based on D2D communications[D]. Wuhan: Huazhong University of Science and Technology, 2018.
- [9] DATTA T, ESHWARIAH H S, CHOCKALINGAM A. Generalized space and frequency index modulation[J]. IEEE Transactions on Vehicular Technology, 2016, 65(7): 4911-4924.
- [10] 龙彦汕, 富勤学, 郭继斌, 等. D2D 缓存网络中基于随机几何的成功卸载概率分析[J]. 计算机应用, 2019, 39(8): 2346-2353.
- LONG Y S, FU Q X, GUO J B, et al. Successful offloading probability analysis in device-to-device caching network based on stochastic geometry[J]. Journal of Computer Applications, 2019, 39(8): 2346-2353.
- [11] 牛煜霞. D2D 辅助蜂窝网络中的协作缓存策略[D]. 北京: 北京邮电大学, 2019.
- NIU Y X. Cooperative caching strategy in D2D assisted cellular networks[D]. Beijing: Beijing University of Posts and Telecommunications, 2019.
- [12] 龙彦汕, 吴丹, 蔡跃明, 等. 多缓存容量场景下的 D2D 内容缓存布设优化方案[J]. 计算机应用, 2018, 38(5): 1453-1457, 1462.
- LONG Y S, WU D, CAI Y M, et al. Cache placement optimization scheme in D2D networks with heterogeneous cache capacity[J]. Journal of Computer Applications, 2018, 38(5): 1453-1457, 1462.
- [13] 虞新颖, 谭冲, 刘洪, 等. 一种基于合作缓存的 D2D 通信缓存策略[J]. 计算机与数字工程, 2019, 47(9): 2242-2256.
- YU X Y, TAN C, LIU H, et al. A caching strategy in D2D communication based on cooperative caching[J]. Computer and Digital Engineering, 2019, 47(9): 2242-2256.
- [14] 左玉玲. 蜂窝与 D2D 网络中基于传输能耗的缓存及分发策略研究[D]. 北京: 北京邮电大学, 2019.
- ZUO Y L. Transmission energy cost optimization based caching and delivery scheme in cellular and D2D networks[D]. Beijing: Beijing University of Posts and Telecommunications, 2019.
- [15] CHEN B, YANG C. Caching policy for cache-enabled D2D communications by learning user preference[J]. IEEE Transactions on Communications, 2018, 66(12): 6586-6601.
- [16] ZHAO X, YUAN P, LI H, et al. Collaborative edge caching in context-aware device-to-device networks[J]. IEEE Transactions on Vehicular Technology, 2018, 67(10): 9583-9596.
- [17] BRESLAU L, CAO P, FAN L, et al. Web caching and Zipf-like distributions: evidence and implications[C]//IEEE INFOCOM Conference on Computer Communications. Piscataway: IEEE Press, 1999: 126-134.
- [18] LIN S, CHENG D, ZHAO G, et al. Energy-efficient wireless caching in device-to-device cooperative networks[C]//2017 IEEE 85th Vehicular Technology Conference. Piscataway: IEEE Press, 2017: 1-5.
- [19] CHEN Y, DING M, LI J, et al. Probabilistic small-cell caching: performance analysis and optimization[J]. IEEE Transactions on Vehicular Technology, 2017, 66(5): 4341-4354.

[作者简介]



蔡艳 (1974-), 女, 四川南充人, 博士, 南京邮电大学副教授、硕士生导师, 主要研究方向为认知无线电、D2D 通信、物联网与无线通信等。

吴凡 (1995-), 男, 江苏南京人, 南京邮电大学硕士生, 主要研究方向为边缘计算、D2D 通信等。

朱洪波 (1956-), 男, 江苏扬州人, 博士, 南京邮电大学教授、博士生导师, 主要研究方向为物联网与无线通信、宽带移动通信、无线通信与电磁兼容等。