Journal of Computer Applications
计算机应用,2018,38(5):1453 – 1457,1462
ISSN 1001-9081
CODEN JYIIDU
2018-05-10
http://www.joca.cn

文章编号:1001-9081(2018)05-1453-05
DOI:10.11772/j.issn.1001-9081.2017112710

多缓存容量场景下的D2D内容缓存布设优化方案

龙彦汕,吴　丹*,蔡跃明,王　萌,郭继斌

(中国人民解放军陆军工程大学 通信工程学院,南京 210007)

(*通信作者电子邮箱 wujing1958725@126.com)

摘　要:在终端直传(D2D)缓存网络中,用户有限且各异的缓存能力是制约缓存效率的一个关键参数,然而现有文献大多考虑不同用户具有相同的缓存能力,针对这一不足有必要进行用户具有不同缓存容量下的D2D内容缓存布设方案优化。首先,鉴于用户终端的移动性和随机分布特性,利用随机几何理论将网络中不同缓存容量的用户节点建模为相互独立的齐次泊松点过程;其次,考虑本地卸载和邻近D2D链路卸载两种内容卸载方式,推导得到网络缓存命中率;最后,将最大化缓存命中率作为优化目标函数,提出了基于坐标梯度的联合缓存布设(JCP)算法,从而得到多用户多缓存容量场景下的内容缓存布设方案。仿真结果表明,与现有的缓存布设方案相比,由JCP得到的缓存布设方案可以有效提高缓存命中率。

关键词:终端直传缓存网络;多缓存容量;随机几何;缓存命中率;缓存布设优化

中图分类号:TN929.5　　**文献标志码**:A

Cache placement optimization scheme in D2D networks with heterogeneous cache capacity

LONG Yanshan, WU Dan*, CAI Yueming, WANG Meng, GUO Jibin

(*College of Communications Engineering, Army Engineering University of PLA, Nanjing Jiangsu* 210007, *China*)

Abstract: Limited and heterogeneous cache capacity is one of the key parameters which affect the cache efficiency in Device-to-Device (D2D) caching networks. However, most of existing literatures assume all users have homogeneous cache capability. In this regard, cache placement optimization is necessary for practical scenarios with heterogeneous cache capacities. Considering the mobility and random distribution of terminal users, users with different cache capacities were modeled as mutually independent homogeneous Poisson point processes with stochastic geometry. Moreover, the average cache hit ratio was derived with considering both self-offloading and D2D-offloading cases. Finally, a Joint Cache Placement (JCP) algorithm based on coordinate gradient optimization was proposed to obtain the optimal cache placement scheme which can maximize the cache hit ratio. Simulation results show that the proposed JCP can achieve larger cache hit ratio than the existing cache placement schemes.

Key words: Device-to-Device (D2D) caching network; heterogeneous cache capacity; stochastic geometry; cache hit ratio; cache placement optimization

0　引言

终端直传(Device-to-Device, D2D)通信由于允许邻近的通信双方可以不经过基站直接通信,成为缓解蜂窝回程压力的关键技术之一[1]。更进一步地,借助于网络中大量分散的用户终端缓存资源,基于分布式缓存的D2D通信技术更利于减轻基站负载,确保用户的服务体验,使得内容分发具有高可扩展性和低成本优势[2-3]。但是,由于用户设备的缓存容量有限以及用户对于不同内容的偏好不同,缓存什么内容、由哪个位置的用户来缓存,即缓存布设方案,成为影响缓存效率的关键因素,也是想要获取缓存增益必须解决的问题。近年来,考虑到移动用户的移动性和空间随机分布的特点,越来越多的研究者开始利用随机几何理论来建模动态的网络拓扑结构以及在D2D缓存网络中内容的可获得性[4-9]。

文献[4]假设缓存用户和请求用户服从相互独立的齐次泊松点过程,将分布式缓存的布设优化问题建模为平均缓存命中率的最大化问题,提出了低复杂度的最优对偶搜索算法;文献[5]在此基础上进一步考虑内容传输过程中信道衰落的影响,对每个文件的覆盖概率进行分析,在假设每个节点只能缓存一个文件的条件下,以网络中文件总的覆盖概率为目标函数进行缓存方案优化;文献[6]利用随机几何理论得到D2D缓存网络中缓存辅助的网络吞吐量,并通过数值仿真获得最优缓存概率;文献[7-8]根据用户对内容的不同偏好将网络用户分组并得到最大化网络收益的缓存策略,但是两者依然假设所有用户缓存容量相同,而且后者只考虑一个文件的情况。在实际的蜂窝网络中,移动用户所携设备的缓存容

收稿日期:2017-11-16;**修回日期**:2017-12-25;**录用日期**:2017-12-27。

基金项目:国家自然科学基金资助项目(61671474);江苏省优秀青年基金资助项目(BK20170089)。

作者简介:龙彦汕(1989—),女,山东荣成人,博士研究生,主要研究方向:D2D缓存网络、随机几何、资源管理;　吴丹(1983—),女,四川成都人,副教授,博士,主要研究方向:D2D通信、博弈论、资源管理;　蔡跃明(1961—),男,福建永泰人,教授,博士,主要研究方向:协同无线网络、无线物理层安全、信号处理;　王萌(1983—),男,山东临沂人,讲师,硕士,主要研究方向:D2D通信;　郭继斌(1973—),男,河北唐山人,讲师,硕士,主要研究方向:协同通信、资源分配。

量大小是不同的，而每个终端缓存已使用情况的特殊性使得这一情况更加突出。上述文献都假设所有缓存节点具有相同的缓存容量的情况，由此获得的最优缓存布设方案并不能直接应用于缓存容量多样的情况。此外，虽然文献[9]在静态网络拓扑的假设下研究小蜂窝基站缓存容量不同以及文件大小不同的情况下最小化回程链路速率的编码缓存方案，但该方案并不适用于D2D缓存网络。如上文所述，在D2D缓存网络中，用户移动性导致的位置不确定性，使得网络拓扑结构是动态变化的；加之用户对不同内容的偏好程度不同以及具体请求内容的不确定性，使得在静态拓扑结构下得到的最优缓存布设方案只适用于某一时刻的网络需求。为了满足时变的动态网络需求，则需要频繁地更新最优缓存策略，势必带来较高的复杂度并且较难实现[11]。因此，如何设计在动态网络拓扑环境下的多用户多缓存容量的内容缓存布设方案仍亟待研究。

因此，针对D2D缓存网络中用户缓存能力异质的特点，本文利用随机几何理论建模不同缓存容量大小的用户位置分布，分析了多用户多缓存容量情况下的缓存命中率。在此基础上以最大化缓存命中率为目标设计了适用于任意多种缓存容量大小的缓存布设方案，提出了基于坐标梯度优化的联合缓存布设（Joint Cache Placement, JCP）算法。仿真结果表明，与现有的几种缓存布设方案相比，通过JCP得到的缓存布设方案可以获得更高的缓存命中率。

1 系统模型

考虑如图1所示的D2D缓存网络，网络中所有用户的位置服从密度为 λ 的齐次泊松点过程（Homogeneous Poisson Point Process, HPPP）Φ[4]。在现实情况中，移动用户手持终端的缓存容量大小不等，例如，常见的智能手机有32 GB、64 GB、128 GB等硬盘容量大小，常见的笔记本电脑有256 GB、512 GB、1 TB等硬盘容量大小。显然，终端缓存容量大小与用户位置之间相互独立，并且持有不同缓存容量终端的用户位置之间也相互独立。因此，假设网络中共有 K 种不同缓存容量的移动用户终端，终端缓存容量为 M_k 的用户个数占总用户数的比例为 $0 < \tau_k < 1$。根据HPPP的稀释定理，可知对于 $\forall k \in \{1,2,\cdots,K\}$，缓存容量为 M_k 的用户位置分布服从密度为 $\lambda_k = \tau_k \lambda$ 的齐次泊松点过程 Φ_k，称为第 k 类用户。显然，$\sum_{k=1}^{K} \tau_k = 1$，不同缓存容量的用户位置之间相互独立。考虑拥有 N 个文件的数据库，假设所有文件具有归一化大小，在内容请求方面，用户总是对当下最流行的多媒体内容感兴趣，即大部分的用户频繁地请求少部分的内容；而且一个内容越流行，被用户请求的概率越大。文献[10]研究了计算机网络中网页请求的分布情况，表明网页请求的分布可以用Zipf分布来近似；更进一步，在文献[11]中，作者通过对YouTube等视频点播平台上大量用户的请求数据，分析了视频文件的请求分布统计特性，结果表明，视频文件的请求分布也可以由Zipf分布来近似。因此，假设每个用户按照Zipf分布独立地从 N 个文件中请求文件，则流行度排名第 i 的文件被请求的概率可以表示为：

$$q_i = i^{-\nu} \Big/ \sum_{j=1}^{N} j^{-\nu} \tag{1}$$

其中，ν 为内容流行度参数，表征文件被请求的密集程度。ν 越大，大部分用户集中请求的流行文件数越少。显然 $q_1 > q_2 > \cdots > q_N$。与传统的服务器-客户模式不同，在D2D缓存网络中，终端用户既是缓存提供者又是文件请求者。因此，当请求用户发起文件请求，有3种方式获得所需文件：1）自卸载：当用户已经缓存该文件，可以立即从本地缓存获得该文件；2）D2D卸载：当本地未缓存所需文件时，由距离请求用户 R 范围内已缓存目标文件的缓存用户通过D2D通信发送该文件；3）蜂窝卸载：当前两者都无效时，该请求将转由其服务基站通过回程链路从核心网络获取。由于本文是以最大化D2D缓存网络的缓存命中率为目标来优化缓存布设方案，因此暂不考虑蜂窝卸载的情况。

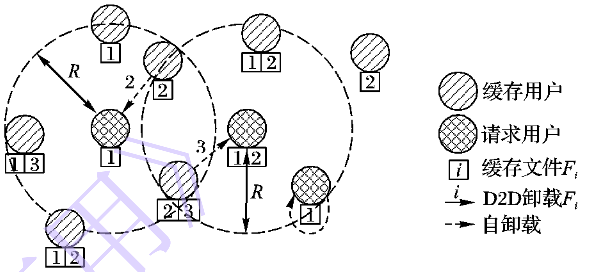

图1 多缓存容量下的D2D缓存网络示意图

Fig. 1 Schematic diagram of D2D caching network with heterogeneous cache capacities

由于用户位置的随机分布以及请求内容的不确定性，本文考虑基于概率的缓存布设方案[11]。假设第 k 类缓存用户的缓存方案为 $\boldsymbol{b}_k = [c_{k1}, c_{k2}, \cdots, c_{kN}]$，其中 $0 \leqslant c_{ki} \leqslant 1$ 表示缓存容量为 M_k 的用户缓存文件 F_i 的概率。所有用户的缓存方案表示为 $\boldsymbol{A} \triangleq [c_{ki}]_{K \times N}$。根据HPPP的稀释定理可知，缓存用户 Φ_k 中缓存文件 F_i 的用户服从密度为 $\lambda_{ki} = \lambda_k c_{ki}$ 的齐次泊松点过程 Φ_{ki}。此外，在半双工模式下，用户不能同时作为D2D通信的发送者和接收者，因此，假定某个时刻网络中发起D2D请求的用户比例为 $\beta \in [0,1]$，余下的用户成为潜在的D2D发送者。因此，Φ_{ki} 中潜在的D2D发送者服从密度为 $\lambda_{ki}^{t} = (1-\beta)\lambda_{ki} = (1-\beta)\lambda_k c_{ki}$ 的齐次泊松点过程 Φ_{ki}^{t}[6]。

2 缓存命中率分析

缓存命中率定义为任意请求用户可以在自己的缓存或者周边 R 范围内找到目标文件的概率，主要由自缓存命中率和D2D缓存命中率两部分构成[6]：

$$p_{\mathrm{hit}} = p_{\mathrm{hit}}^{\mathrm{self}} + p_{\mathrm{hit}}^{\mathrm{d2d}} \tag{2}$$

2.1 自缓存命中率分析

当第 k 类用户请求文件 F_i 时，如果该用户已经缓存文件 F_i，那么他可以立刻通过自卸载获得该文件。这一事件发生的概率为 $p_{\mathrm{hit}|k,i}^{\mathrm{self}} = c_{ki}$。进而，考虑请求用户缓存容量以及被请求文件的不同可能性，根据全概率公式可以得到自缓存命中率为：

$$p_{\mathrm{hit}}^{\mathrm{self}} = \sum_{k=1}^{K} \tau_k \sum_{i=1}^{N} q_i c_{ki} \tag{3}$$

2.2 D2D缓存命中率分析

当请求用户没有缓存自己需要的文件时，则向邻近的缓存用户发送请求。如果在请求用户半径为 R 的区域内存在缓存该文件的用户，D2D缓存命中。当第 k 类的用户向周边用

户请求文件 F_i 时,D2D 缓存命中的概率为:

$$p_{\mathrm{hit}\mid k,i}^{\mathrm{d2d}} = (1-c_{ki})(1-p_{\mathrm{fail},i}^{\mathrm{d2d}}) \tag{4}$$

其中,$p_{\mathrm{fail},i}^{\mathrm{d2d}}$ 为当用户请求文件 F_i 时,D2D 缓存命中失败的概率,即在请求者周边 R 距离内没有已缓存文件 F_i 的用户的概率。已知对于用户密度为 λ 的 HPPP 分布来说,在半径为 r 的区域中有 n 个用户的概率为 $F(n,r,\lambda) = [(\pi r^2\lambda)^n \mathrm{e}^{-\pi r^2\lambda}]/(n!)$,因此,对请求文件 F_i 的用户来说,通过 D2D 请求文件失败的概率为:

$$p_{\mathrm{fail},i}^{\mathrm{d2d}} = \prod_{k=1}^{K} F(0,R,\lambda_{ki}^{c}) = \exp\left(-\sum_{k=1}^{K} a_k c_{ki}\right) \tag{5}$$

其中,$a_k = \pi R^2(1-\beta)\lambda_k$ 表示请求用户半径 R 范围内的第 k 类缓存用户的平均个数。将式(5)代入式(4)中,并根据全概率公式,可得 D2D 缓存命中率为:

$$p_{\mathrm{hit}}^{\mathrm{d2d}} = \sum_{k=1}^{K}\tau_k\sum_{i=1}^{N} q_i(1-c_{ki})\left(1-\exp\left(-\sum_{k=1}^{K}a_k c_{ki}\right)\right) \tag{6}$$

将式(3)和(6)代入式(2)中,可以得到缓存命中率为:

$$\begin{aligned} p_{\mathrm{hit}} &= \sum_{j=1}^{K}\tau_j\sum_{i=1}^{N}q_i c_{ji} + \sum_{j=1}^{K}\tau_j\sum_{i=1}^{N}q_i(1-c_{ji})(1-p_{\mathrm{fail}\mid i}^{\mathrm{d2d}}) = \\ &\sum_{j=1}^{K}\tau_j\sum_{i=1}^{N}(q_i - q_i p_{\mathrm{fail}\mid i}^{\mathrm{d2d}} + q_i c_{ji} p_{\mathrm{fail}\mid i}^{\mathrm{d2d}}) = \\ &\sum_{j=1}^{K}\tau_j\left(1-\sum_{i=1}^{N}q_i(1-c_{ji})\exp\left(-\sum_{k=1}^{K}a_k c_{ki}\right)\right) = \\ &1+\sum_{j=1}^{K}\tau_j\sum_{i=1}^{N}q_i(c_{ji}-1)\exp\left(-\sum_{k=1}^{K}a_k c_{ki}\right) \end{aligned} \tag{7}$$

与只有一种缓存容量的情况[6]相比,多种缓存容量带来的影响主要体现在:1)请求用户的缓存容量是不相同的,从而导致其自卸载概率的不同,在式(7)中主要体现在对第 j 类请求用户出现概率 τ_j 的全概率求和;2)请求用户半径 R 内缓存目标文件的缓存用户的缓存容量大小不确定,在式(7)中主要体现在指数项中对 K 种情况的全概率求和。当 $K=1$ 时,式(7)可以退化成与文献[6]相吻合的结论,即:

$$\hat{p}_{\mathrm{hit}} = 1-\sum_{i=1}^{N}q_i(1-c_i)\mathrm{e}^{-\pi R^2(1-\beta)\lambda c_i} \tag{8}$$

当 $K=1$ 并且不考虑自卸载情况时,式(7)进一步退化成文献[4]的结论,即:

$$\tilde{p}_{\mathrm{hit}} = \sum_{i=1}^{N}q_i(1-\mathrm{e}^{-\pi R^2(1-\beta)\lambda c_i}) \tag{9}$$

3 最优缓存布设方案

如上所述,本文的目标是在考虑多缓存容量的情况下获得最大化缓存命中率 p_{hit} 的最优缓存布设方案 $\boldsymbol{A}$。

上述问题可以表示为:

$$\max_{\boldsymbol{A}} p_{\mathrm{hit}} = 1+\sum_{j=1}^{K}\tau_j\sum_{i=1}^{N}q_i(c_{ji}-1)\exp\left(-\sum_{k=1}^{K}a_k c_{ki}\right) \tag{10}$$

$$\text{s.t.}\ \sum_{i=1}^{N}c_{ki} \leqslant M_k,\ \forall k;\ 0\leqslant c_{ki}\leqslant 1,\ \forall k,i \tag{11}$$

其中,第一个约束表示每一类用户缓存的平均文件个数不能超过其缓存容量大小。

网络拓扑的随机性以及用户终端缓存能力互异的特点使得缓存命中率的求解需要对多个随机变量进行全概率求和,而不同缓存容量用户的缓存布设方案之间的耦合关系使得目标函数关于 $\boldsymbol{A}$ 是非凹的。因此,为了解决不同缓存容量用户策略相互耦合的问题,本文将退而求其次,分析当给定其他类用户策略时,针对某类用户的缓存策略是否具有凹性。将除了第 k 类缓存用户策略 $\boldsymbol{b}_k$ 以外的其他类用户的缓存分布策略 $\{\boldsymbol{b}_1,\cdots,\boldsymbol{b}_{k-1},\boldsymbol{b}_{k+1},\cdots,\boldsymbol{b}_K\}$ 表示为 $\boldsymbol{A}_{-k} = [c_{ji}]_{(K-1)\times N}$。那么,对于任意 k,当给定 $\boldsymbol{A}_{-k}$ 时,p_{hit} 的表达式可以重新表示为:

$$\begin{aligned} p_{\mathrm{hit}} = 1 + &\underbrace{\tau_k\sum_{i=1}^{N}q_i(c_{ki}-1)\mathrm{e}^{-\sum_{k=1}^{K}a_k c_{ki}}}_{j=k} + \\ &\underbrace{\sum_{j=1,j\neq k}^{K}\tau_j\sum_{i=1}^{N}q_i(c_{ji}-1)\mathrm{e}^{-\sum_{k=1}^{K}a_k c_{ki}}}_{j\neq k} \end{aligned} \tag{12}$$

经过两次求导计算,可得 p_{hit} 关于 $\boldsymbol{b}_k$ 的二阶偏导数为 $\dfrac{\partial^2 p_{\mathrm{hit}}}{\partial c_{ki}\partial c_{ki'}} = 0$,以及:

$$\begin{aligned} \frac{\partial^2 p_{\mathrm{hit}}}{\partial c_{ki}^2} &= \frac{\partial^2 p_{\mathrm{hit},j\neq k}}{\partial c_{ki}^2} + \frac{\partial^2 p_{\mathrm{hit},j=k}}{\partial c_{ki}^2} = \\ &\mathrm{e}^{-\sum_{k=1}^{K}a_k c_{ki}}\left(\sum_{j=1,j\neq k}^{K}\tau_j q_i(c_{ji}-1)a_k^2 - \right. \\ &\left. \tau_k a_k q_i(2+a_k-a_k c_{ki})\right) < 0 \end{aligned} \tag{13}$$

由此可知 p_{hit} 的海森矩阵是负定矩阵,这说明 p_{hit} 是关于第 k 类用户缓存布设方案 $\boldsymbol{b}_k$ 的严格凹函数。

根据这一特性,把原问题转化为对每种缓存容量大小的用户缓存策略的序列迭代求解。对于任意给定的缓存分布策略 $\boldsymbol{A}_{-k}$,原问题转化为对第 k 类用户缓存策略优化的子问题,表示为:

$$\max_{\boldsymbol{b}_k} p_{\mathrm{hit}} = 1+\sum_{j=1}^{K}\tau_j\sum_{i=1}^{N}q_i(c_{ji}-1)\mathrm{e}^{-\sum_{k=1}^{K}a_k c_{ki}} \tag{14}$$

$$\text{s.t.}\ \sum_{i=1}^{N}c_{ki}\leqslant M_k,\ 0\leqslant c_{ki}\leqslant 1 \tag{15}$$

首先得到该问题的部分拉格朗日函数为:

$$\begin{aligned} L(\boldsymbol{b}_k,\mu_k) = 1 + \\ \sum_{j=1}^{K}\tau_j\sum_{i=1}^{N}q_i(c_{ji}-1)\mathrm{e}^{-\sum_{k=1}^{K}a_k c_{ki}} + \mu_k\left(M_k-\sum_{i=1}^{N}c_{ki}\right) \end{aligned} \tag{16}$$

其中,μ_k 为保证缓存容量约束的拉格朗日乘子。根据 p_{hit} 的局部凹性可知,利用 Karush-Kuhn-Tucker 最优化条件可以得到子问题的最优解。令 $\dfrac{\partial L(\boldsymbol{b}_k,\mu_k)}{\partial c_{ki}} = 0$ 并经过整理后,可以得到:

$$\mu_k = q_i\mathrm{e}^{-\sum_{j=1}^{K}a_j c_{ji}}\left[\tau_k + a_k\sum_{j=1}^{K}\tau_j(1-c_{ji})\right] \tag{17}$$

令 $x = -\sum_{j=1}^{K}a_j c_{ji}$,式(17)可以进一步化简为 $\mu_k = q_i\tau_k \mathrm{e}^x(1+\bar{a}+x)$;再令 $y = 1+\bar{a}+x$,则有 $y\mathrm{e}^y = \dfrac{\mu_k \mathrm{e}^{1+\bar{a}}}{q_i\tau_k}$,因此,$\sum_{j=1}^{K}a_j c_{ji} = 1+\bar{a}-W\left(\dfrac{\mu_k\mathrm{e}^{1+\bar{a}}}{q_i\tau_k}\right)$。进而,可以得到当给定其他缓存容量用户的缓存分布 $\boldsymbol{A}_{-k}$ 时,第 k 类缓存用户的最优缓存布设方案为:

$$c_{ki}^{*} = \left[\frac{1+\bar{a}}{a_k} - \frac{1}{a_k}W\left(\frac{\mu_k^{*}\mathrm{e}^{1+\bar{a}}}{q_i\tau_k}\right) - \sum_{j=1,j\neq k}^{K}\frac{\tau_j}{\tau_k}c_{ji}\right]_0^1 \tag{18}$$

其中，$\bar{a}=\sum_{j=1}^{K}a_j=\pi R^2(1-\beta)\lambda$，$W(\cdot)$ 为 Lambert-W 函数[13]，$[x]_0^1=\max(0,\min(x,1))$。拉格朗日乘子 μ_k^* 是方程 $\sum_{i=1}^{N}c_{ki}^*(\mu_k^*)=M_k$ 的解，具体可以利用二分法的思想通过数值求解获得。

从式(18)可以看出，不同缓存容量的缓存策略与对应的用户密度有关，对于固定的某种缓存容量来说，不同文件的缓存概率与其流行度 q_i 有关。由于 Lambert-W 函数在非负区间是关于变量的单调增函数，因此 μ_k 越小以及 q_i 越大，则 c_{ki} 越大。为得到 D2D 网络中所有缓存用户缓存布设方案，提出基于坐标梯度的联合缓存布设算法，具体步骤如下：

1) 初始化参数：$\boldsymbol{A}^{(0)}$；ξ_1；ξ_2，$T_{\max}$；
2) For $t=1:T_{\max}$
3) For $k=1:K$
4) 更新矩阵 $\boldsymbol{A}_{-k}^{(t)}=[\boldsymbol{b}_1^{(t)};\cdots;\boldsymbol{b}_{k-1}^{(t)};\boldsymbol{b}_{k+1}^{(t-1)};\cdots;\boldsymbol{b}_K^{(t-1)}]$；
5) While $\left|\sum_{i=1}^{N}c_{ki}-M_k\right|>\xi_2$ do
6) $\mu_k=(\mu_k^{lb}+\mu_k^{ub})/2$；
7) 根据式(18)求得缓存概率 c_{ki}^*，$\forall i$；
8) If $\sum_{i=1}^{N}c_{ki}^*<M_k$ then
9) $\mu_k^{ub}=\mu_k$；
10) else
11) $\mu_k^{lb}=\mu_k$；
12) End If
13) End While
14) 更新矩阵 $\boldsymbol{A}^{(t)}=[\boldsymbol{b}_1^{(t)};\cdots;\boldsymbol{b}_{k-1}^{(t)};\boldsymbol{b}_k^{(t)};\boldsymbol{b}_{k+1}^{(t-1)};\cdots;\boldsymbol{b}_K^{(t-1)}]$；
15) End For
16) 根据式(7)求得缓存命中率 p_{hit}^t；
17) If $|p_{\text{hit}}^t-p_{\text{hit}}^{t-1}|\le\xi_1$ then
18) break；
19) End If
20) End For

算法中，ξ_1 是外部循环用于判断算法是否达到最大缓存命中率的误差精度，ξ_2 是内部循环用于判断缓存的内容总数是否达到缓存容量的误差精度。一般情况下，误差精度越小，对应的该循环收敛所需要的迭代次数越多，反之亦然。$T_{\max}$ 是为防止算法陷入无限循环而设置的迭代次数门限。由于原问题的可行集合为闭集并且目标函数有下界，可知通过该算法至少可以得到一个原问题的局部最优解。通过第5)步到第13)步的判断，如果算法收敛，将立刻跳出外部循环。因此，算法外部循环的实际迭代次数主要通过第5)步到第13)步来决定。假设外部循环实际的最大迭代次数为 $T(T\le T_{\max})$，算法第5)步到第13)步通过二分法来寻找 μ_k^* 以及对应的最优缓存方案 $\boldsymbol{b}_k^*$ 的最大迭代次数是 I，那么上述算法的计算复杂度为 $O(KNIT)$。

4 仿真结果及分析

本章主要通过数值仿真对所提 JCP 算法进行有效性验证。

根据前文可知，JCP 算法可以保证至少收敛到一个局部最优解，但是由于原问题的非凹性，不同初始值可能会导致 JCP 算法收敛到不同的局部最优解[8]。因此，为了评估初始值设置对算法收敛性的影响，图 2 描绘了在不同文件数以及不同缓存容量的用户占比等多种场景下，算法采用不同缓存布设初始值 $\boldsymbol{A}^0$ 时的收敛曲线。以 $K=2$ 为例，Zero Init. 表示 $\boldsymbol{A}^0=\boldsymbol{0}_{K\times N}$ 为零矩阵；Unif. Init. 表示对于每个 k，$\boldsymbol{b}_k^0=\frac{M_k}{N}\boldsymbol{1}_{1\times N}$。通过图中多条曲线对比可知，即使采用不同的 $\boldsymbol{A}^0$，JCP 算法都可以用几次迭代收敛到相同的局部最优解，这说明 JCP 算法对 $\boldsymbol{A}^0$ 的不同取值并不敏感。

为了评估 JCP 算法的性能，将 JCP 与另外三种缓存布设方案进行对比并分析系统参数对算法有效性以及网络缓存命中率的影响。三种对比算法分别为：1）针对不同的缓存容量用户，各自独立地计算最优缓存布设方案（Separate Cache Placement，SCP）[6]；2）根据内容流行度进行缓存（Adaptive cache placement，ADA）[15]；3）根据缓存容量大小，等概率地缓存不同文件（Uniform cache placement，UNI）[16]。为了使仿真结果更具说服力，以下仿真均考虑 $K=2$ 和 $K=3$ 两种情况。考虑到现实场景中用户终端容量大小均以 2 的幂次方呈现，因此，当 $K=2$ 时，$M_1=1$，$M_2=2$，$\tau_1=\tau_2=0.5$；当 $K=3$ 时，$M_1=1$，$M_2=2$，$M_3=4$，$\tau_1=\tau_2=\tau_3=1/3$。此外，除非特殊说明，以下仿真参数设置不变：$N=5$，$\lambda=0.01$ (users/m^2)，$\xi_1=10^{-4}$，$\xi_2=10^{-4}$[16]，$\beta=0.5$，$\nu=0.8$，$R_d=10$ m。

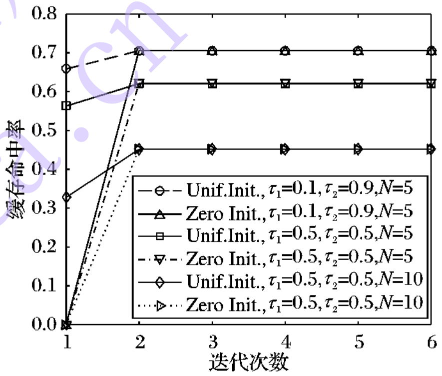

图 2 缓存布设方案初始值对算法收敛性的影响

Fig. 2 Effect of cache placement initialization on convergence

图 3 描绘了不同内容流行度参数 ν 情况下，JCP 与其他三种基准算法的缓存命中率曲线。由图可知，均匀缓存方案由于不受请求概率影响，因此其对应的缓存命中率不随 ν 变化，而其他方案的缓存命中率均随 ν 的增加而提高。纵向对比四种缓存方案，不难发现，JCP 始终要优于其他三种方案。值得一提的是，根据理论分析可知，JCP 的优越性主要来源于多种缓存容量共存的网络中，不同缓存策略的耦合关系对 D2D 卸载方式产生影响。通过图 3 中 JCP 与 SCP 的对比可见：当 $\nu=0$ 时，用户对所有文件的兴趣相同，因此在每种缓存容量的用户缓存内肯定都会均匀缓存各个文件，此时 JCP 与 SCP 效果相同；随着 ν 的增大，JCP 优越性先变大再缩小，在 $\nu=1.8$ 时，只有少量的优势，这是因为此时用户集中请求的文件很少，这种情况下每个请求用户都主要以自卸载方式来满足文件需求。

图 4 描绘了不同请求用户比例 β 情况下，JCP 算法与其他三种算法的缓存命中率曲线。由图可知，请求用户比例越大，系统缓存命中率越小，这是因为 β 越大，潜在的 D2D 发送者越少，缓存用户之间无法充分利用邻近用户的缓存资源。当 $\beta=1$ 时，所有的用户都发送文件请求，此时他们只能采用

自卸载模式,而无法参与 D2D 内容分发,此时,缓存命中率等于自卸载概率。由于文件的自给自足使得此时所有用户的缓存和请求之间都是相互独立的,因此 JCP 算法与 SCP 算法性能相同。除此以外,在整个 β 的区间,JCP 算法都可以获得相比其他三种算法更高的缓存命中率。

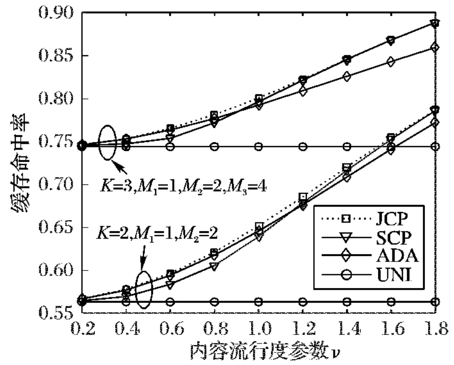

图 3　缓存命中率随内容流行度参数 ν 的变化曲线

Fig. 3　Cache hit ratio versus content popularity factor ν

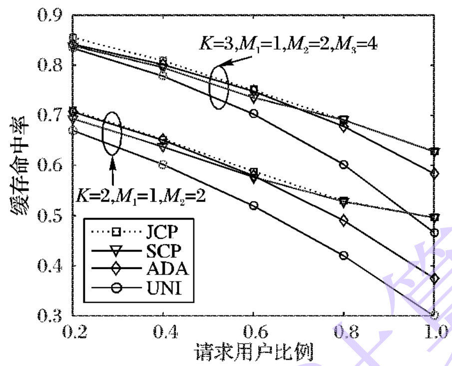

图 4　缓存命中率随请求用户比例 β 的变化曲线

Fig. 4　Cache hit ratio versus requesters' proportion β

图 5 进一步分析了不同的 D2D 通信距离约束 R 情况下,四种算法的性能比较。由图可知,随着 R 变大,每个请求用户周围能够为其提供 D2D 服务的缓存用户数据增多,缓存命中率也随之增大。通过不同算法的对比可知,不同 R 条件下,JCP 仍然始终保持优于其他三种方案。

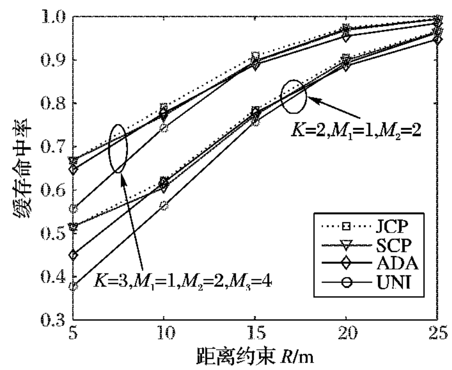

图 5　缓存命中率随 D2D 通信距离约束 R 的变化曲线

Fig. 5　Cache hit ratio versus D2D communication distance constraint R

综上所述,图 2 从侧面说明算法的低复杂度,图 3 到图 5 则从不同的角度揭示了 JCP 算法相比已有方案的优势,即在多用户多缓存容量场景下,本文所提 JCP 算法可以获得更高的缓存命中率。

5　结语

本文针对 D2D 缓存网络中用户存在多种不同缓存容量的实际情况,以最大化网络缓存命中率为目标,优化不同缓存容量用户的缓存布设方案。为了达到这一目的,提出了基于坐标梯度的联合缓存布设算法并通过仿真证明了该算法的优越性以及对现实缓存网络中缓存布设的指导意义。但是本文假设 D2D 通信在一定距离范围内可以无差错传输,但在实际传输过程中,信号无疑会受到信道衰落和干扰的影响,因此在考虑信道条件的情况下优化多缓存容量情况的缓存布设方案是下一步要解决的问题。

参考文献 (References)

[1]　钱志鸿,王雪. 面向 5G 通信网的 D2D 技术综述[J]. 通信学报, 2016, 37(7): 1 – 14. (QIAN Z H, WANG X. Reviews of device-to-device technology for 5G communication networks[J]. Journal on Communications, 2016, 37(7): 1 – 14.)

[2]　黄小柠. 面向无线缓存的 D2D 技术研究[D]. 成都: 电子科技大学, 2016. (HUANG X N. Research on device-to-device technology for wireless caching [D]. Chengdu: University of Electronic Science and Technology of China, 2016.)

[3]　蓝瑞宁. 终端直通蜂窝系统中的边缘缓存技术[D]. 杭州: 浙江大学, 2016. (LAN R N. Edge caching for cellular systems with device-to-device communications [D]. Hangzhou: Zhejiang University, 2016.)

[4]　KANG H J, PARK K Y, KUMIN C, et al. Mobile caching policies for Device-to-Device (D2D) content delivery networking [C]// INCOFOM 2014: Proceedings of the 2014 IEEE Conference on Computer Communications Workshops. Piscataway, NJ: IEEE, 2014: 299 – 304.

[5]　MALAK D, AL-SHALASH M. Device-to-Device content distribution: optimal caching strategies and performance bounds [C]// ICCW 2015: Proceedings of the 2015 IEEE International Conference on Communications Workshop. Piscataway, NJ: IEEE, 2015: 664 – 669.

[6]　CHEN Z, PAPPAS N, KOUNTOURIS M. Probabilistic caching in wireless D2D networks: cache hit optimal versus throughput optimal [J]. IEEE Communications Letters, 2017, 21(3): 584 – 587.

[7]　GUO Y, DUAN L, ZHANG R. Cooperative local caching under heterogeneous file preferences[J]. IEEE Transactions on Communications, 2017, 65(1): 444 – 457.

[8]　PAN Y, PAN C, ZHU H, et al. On consideration of content preference and sharing willingness in D2D assisted offloading[J]. IEEE Journal on Selected Areas in Communications, 2017, 35(4): 978 – 993.

[9]　LIAO J, WONG K K, KHANDAKER M R A, et al. Optimizing cache placement for heterogeneous small cell networks[J]. IEEE Communications Letters, 2017, 21(1): 120 – 123.

[10]　BRESLAU L, CAO P, FAN L, et al. Web caching and Zipf-like distributions: evidence and implications[C]// INFOCOM 1999: Proceedings of the 1999 IEEE Conference on Computer Communications. Piscataway, NJ: IEEE, 1999: 126 – 134.

[11]　CHA M, KWAK H, PODRIGUEZ, et al. Analyzing the video popularity characteristics of large-scale user generated content systems[J]. IEEE/ACM Transactions on Networking, 2009, 17(5): 1357 – 1370.

(下转第 1462 页)

wireless body area networks: technologies and design challenges [J]. IEEE Communications Surveys Tutorials, 2014, 16(3): 1635 –1657.

[2] KOBO H, ABU-MAHFOUZ A, HANCKE G. A survey on software-defined wireless sensor networks: challenges and design requirements[J]. IEEE Access, 2017, 5(3): 1872 –1899.

[3] 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.

[4] BASAR E. Index modulation techniques for 5G wireless networks [J]. IEEE Communications Magazine, 2016, 54(7): 168 –175.

[5] DATTA T, ESHWARAIAH H S, CHOCKALINGAM A. Generalized space and frequency index modulation[J]. IEEE Transactions on Vehicular Technology, 2016, 65(7): 4911 –4924.

[6] BASAR E. On multiple-input multiple-output OFDM with index modulation for next generation wireless networks[J]. IEEE Transactions on Signal Processing, 2016, 64(15): 3868 –3878.

[7] MAHAPATRA R, NIJSURE Y, KADDOUM G, et al. Energy efficiency tradeoff mechanism towards wireless green communication: a survey[J]. IEEE Communications Surveys Tutorials, 2016, 18(1): 686 –705.

[8] MESLEH R Y, HAAS H, SINANOVIC S. Spatial modulation[J]. IEEE Transactions on Vehicular Technology, 2008, 57(4): 2228 – 2241.

[9] JEGANATHAN J, GHRAYEB A, SZCZECINSKI L. Space shift keying modulation for MIMO channels[J]. IEEE Transactions on Wireless Communications, 2009, 8(7): 3692 –3703.

[10] JEGANATHAN J, GHRAYEB A, SZCZECINSKI L. Generalized space shift keying modulation for MIMO channels[C]// Proceedings of IEEE the 19th International Symposium on Personal, Indoor and Mobile Radio Communications. Piscataway, NJ: IEEE, 2011: 1 –5.

[11] YOUNIS A, SERAFIMOVSKI N, MESLEH R, et al. Generalized spatial modulation[EB/OL]. [2017-06-20]. http://www.research.ed.ac.uk/portal/en/publications/generalized-spatial-modulation(fb24cad4-168c-476b-a40b-eba31828f7fe)/export.html.

[12] 刘健伶, 陈志刚, 王磊. 一种自适应广义空间调制及其低复杂度算法[J]. 西安交通大学学报, 2016, 50(4): 48 –53. (LIU J L, CHEN Z G, WANG L. An adaptive generalized space modulation and its low complexity algorithm[J]. Journal of Xi'an Jiaotong University, 2016, 50(4): 48 –53.)

[13] MESLEH R, IKKI S, AGGOUNE H M. Quadrature spatial modulation[J]. IEEE Transactions on Vehicular Technology, 2015, 64(6): 2738 –2742.

[14] LI J, WEN M, CHENG X, et al. Generalized precoding-aided quadrature spatial modulation[J]. IEEE Transactions on Vehicular Technology, 2017, 66(2): 1881 –1886.

[15] KADDOUM G, AHMED M, NIJSURE Y. Code index modulation: a high data rate and energy efficient communication system[J]. IEEE Communications Letters, 2015, 19(2): 175 –178.

[16] KADDOUM G, SOUJERI E. On the comparison between code-index modulation and spatial modulation techniques[C]// Proceedings of the 2015 International Conference on Information and Communication Technology Research. Piscataway, NJ: IEEE, 2015: 24 –27.

[17] KADDOUM G, NIJSURE Y, TRAN H. Generalized code index modulation technique for high data rate communication systems [J]. IEEE Transactions on Vehicular Technology, 2016, 65(9): 7000 –7009.

This work is partially supported by the National Natural Science Foundation of China (61271251), the Science and Technology Innovation Special Project of Chongqing Social Undertaking and People's Livelihood Security (cstc2017shmsA40002).

JIANG Zhilin, born in 1993, M. S. candidate. His research interests include index modulation, high speed wireless transmission.

GE Lijia, born in 1957, professor. His research interests include high speed wireless transmission, real-time signal processing.

XING Fengying, born in 1991, M. S. candidate. His research interests include index modulation, high speed wireless transmission.

YANG Qin, born in 1982, lecturer. His research interests include high speed wireless transmission, real-time signal processing.

(上接第 1457 页)

[12] 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.

[13] STOYAN D, KENDALL W, MECKE J. Stochastic Geometry and Its Applications[M]. 3rd ed. New York: Wiley, 1999: 35 –63.

[14] CORLESS R, GONNET G, HARE D, et al. On the Lambert W function [J]. Advances in Computational Mathematics, 1996, 5(1): 329 –359.

[15] SONG X. GENG Y, MENG X, et al. Cache-enabled device-to-device networks with contention-based multimedia delivery [J]. IEEE Access, 2017, 5: 3228 –3239.

[16] LONG Y, WU D, CAI Y, et al. Joint cache policy and transmit power for cache-enabled D2D networks[J]. IET Communications, 2017, 11(16): 2498 –2506.

This work is partially supported by the National Natural Science Foundation of China (61671474), the Jiangsu Provincial Science Fund for Excellent Young Scholars (BK20170089).

LONG Yanshan, born in 1989, Ph. D. candidate. Her research interests include D2D caching network, stochastic geometry, resource management.

WU Dan, born in 1983, Ph. D, associate professor. Her research interests include D2D communications, game theory, resource management.

CAI Yueming, born in 1961, Ph. D, professor. His research interests include cooperative wireless network, wireless physical-layer security, signal processing.

WANG Meng, born in 1983, M. S., lecturer. His research interests include D2D communications.

GUO Jibin, born in 1973, M. S., lecturer. His research interests include cooperative communication, resource allocation.