



D2D 集成雾无线接入网中的双层分布式缓存

夏骋宇, 蒋雁翔

(东南大学, 江苏 南京 211189)

摘要: 随着移动通信的发展, 人们对低时延的传输网络的要求也在急剧增长。雾无线接入网 (fog-RAN) 作为一种适用于 5G/B5G 通信网络的新型架构, 拥有广阔的应用前景。考虑了集成 D2D 通信的雾无线接入网, 提出了一种基于 D2D 通信和雾无线接入网的双层分布式缓存布置, 希望以此降低用户下载文件的时延。首先得到了这种双层缓存的下载时延的闭式解以及这种布置方案相对于单层缓存布置方案的时延增益的闭式解。在此基础上, 设计了分布式算法以求得最佳缓存布置策略, 使得每个雾无线接入点 (F-AP) 或者用户设备 (UE) 只需要通过邻近设备的信息就能完成缓存。仿真结果显示, 双层分布式缓存能够有效降低下载时延。

关键词: 雾无线接入网; D2D; 双层分布式缓存

中图分类号: TN915

文献标识码: A

doi: 10.11959/j.issn.1000-0801.2018143

Two-layer distributed cache in D2D integrated fog-RAN

XIA Chengyu, JIANG Yanxiang

Southeast University, Nanjing 211189, China

Abstract: With the development of mobile communications, people's demands for low-latency transmission networks are rapidly increasing. Fog-RAN, as a new architecture suitable for 5G/B5G communication networks, has broad application prospects. Considering the fog network of device-to-device communication, a two-layer distributed cache based on device-to-device communication and fog-RAN was proposed to reduce the delay for users to download files. Firstly, a closed-form solution of the download delay of the two-layer cache and a closed-form solution of the delay gain were obtained. On this basis, a distributed algorithm was designed to obtain the best cache strategy, so that each fog-based wireless access point (F-AP) or user equipment (UE) only needed to use the information of the neighboring devices to complete the caching. The simulation results show that the two-layer distributed cache can effectively reduce the download delay.

Key words: fog-radio access network, device to device, two-layer distributed cache

1 引言

近年来, 高质量的无线视频、社交网络等业务和智能应用呈现爆炸式增长。可以预见, 未来

的 5G/B5G 网络中这些业务还将继续增长, 如何在指数增长的业务中为用户提供低时延的体验, 是 5G/B5G 移动通信的重要问题。雾无线接入网 (fog-radio access network, F-RAN) 作为新型的无

收稿日期: 2017-11-27; 修回日期: 2018-04-05



线传输架构,已经引起了学术界和产业界的关注。虽然F-RAN能够将缓存和计算功能下放到边缘节点(F-AP)^[1],但由大量用户下载少数流行文件时产生的通信负担以及随之而来的时延依然是其亟待解决的问题^[2]。此外,F-AP本身缓存容量较小也带来缓存命中率不高的问题^[1]。因此,雾无线接入网需要一种有效的缓存布置算法来缓解用户时延和提高命中率。同时,D2D通信近来已经受到学术界和产业界的关注,但是已有的研究大都关注D2D通信如何有效实现及避免干扰^[3-6]。将D2D通信和雾无线接入网结合,提出一种在F-AP层和用户设备层中的分布式缓存布置。由此带来两个主要的问题:和传统蜂窝基站(cellular base station)相比,F-AP和用户设备的缓存容量都要小得多,因此决定某一个文件应不应该被缓存变得越发重要;如何在用户设备层实现D2D通信也至关重要。

在本文中,希望用分布式算法来解决上面的问题,对于用户设备层的缓存分布,将其转化为背包问题,采用贪心算法求解。文件权重就是它

的流行度,而一个用户设备可以通过搜集周围的信息计算出流行度,从而让所有用户设备分布式地做出缓存决定。将F-AP的缓存分布转化为因子图问题,使用置信度传播算法(belief propagation algorithm)求解。对于如何实现UE间D2D传输,也通过分布式的方式求解。为了降低算法复杂度,在本文的研究中,UE和F-AP都不会预先缓存(proactively cache)任何文件来提高命中率,并且整个网络中UE间的文件交换全部都是单跳通信,不依赖任何多跳通信。这样分布式的解决方案将更简单,更有利于在现实雾无线接入网中的应用。

2 系统模型

如图1所示,考虑一个集成了D2D通信的雾无线接入网,并且假设一个F-AP覆盖范围内的UE不会被其他F-AP覆盖的UE影响^[7]。在这种情况下,一个UE既可以从F-AP下载文件,也可以向其周围的UE获取文件。同时,假设只有同一个F-AP覆盖区域内的UE之间可以进行D2D

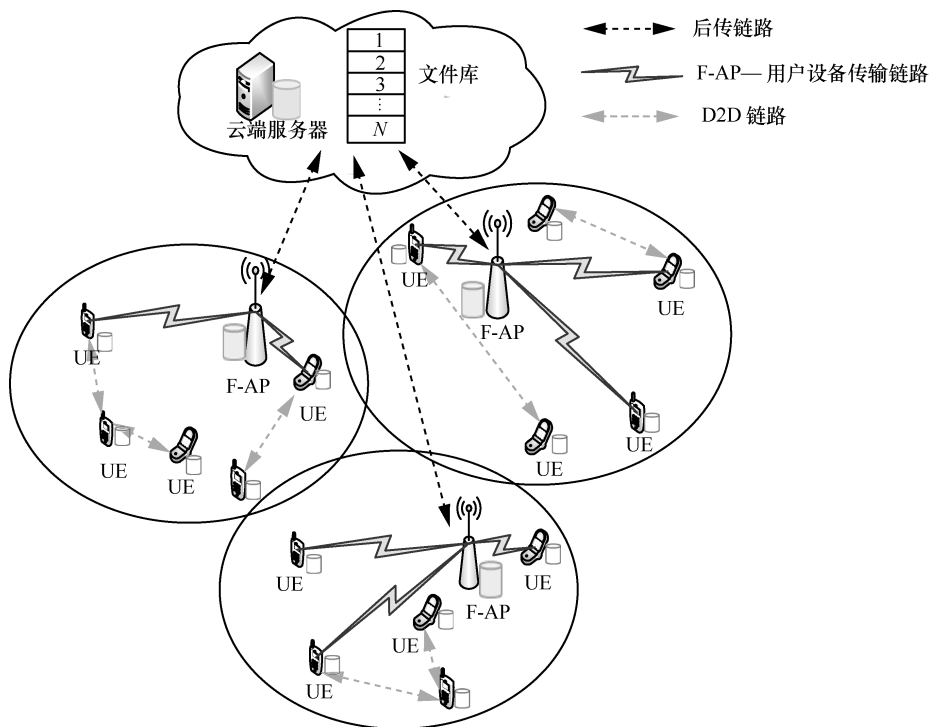


图1 雾无线接入网中的二维缓存

传输^[7]。参考文献[8]已经研究了无线资源分配机制,因此可以假设 F-AP 和 D2D 通信不会互相干涉。正交频分多址 (OFDMA) 技术以及参考文献[8,9]研究的其他技术可以避免 D2D 通信的互相干扰。参考文献[10]对于全双工通信的研究使得可以假设每个 UE 可以在向一个 UE 发送文件的同时从另一个 UE 接收文件。此外,为了降低算法复杂度,假设 F-AP 和 UE 都不会预先缓存某些文件,也就是每个 UE 只会缓存它自己申请过的文件,并将这些文件传输给相邻 UE。每个 F-AP 的缓存一开始也为空,并通过自己收集的本地信息决定缓存。

假设无线网络由 M 个 F-AP 和 K 个 UE 组成,集合 $A = \{a_1, a_2, \dots, a_m\}$ 为 F-AP 的集合, $U = \{u_1, u_2, \dots, u_k\}$ 为 UE 的集合,每个 F-AP 覆盖了 N_m 个 UE。定义 N_k 为可以与 u_k 进行 D2D 通信的 UE 的集合,即 $N_k = N_m$ 。F-AP a_m 所服务的 UE 形成集合 U_m 。

假设总共有 N 个文件,形成集合 $F = \{f_1, f_2, \dots, f_n\}$,存储在云端的服务器中。假设所有的文件大小为 $f_n = |f| (\forall f_n \in F)$ 。某一个 UE u_k 申请文件 f_n 的概率为 q_{nk} 。同时,假设每个 F-AP 的缓存容量为 Q_m 。用变量 x_{nm} 来指明文件 f_n 是否被 F-AP a_m 缓存。假如 f_n 存储在 a_m 中, $x_{nm} = 1$; 否则, $x_{nm} = 0$ 。类似地,用 y_{nk} 来表示文件 f_n 是否被 UE u_k 缓存。

3 方程建立

基于图 1 所示的模型,先考虑缓存在用户层内的分布。假设每一个用户也有一个缓存容量 Q_u 。再定义 z_{nk} 表示文件 f_n 是否缓存在 u_k 的相邻 UE 中。这样, F-AP 中的 UE 向 F-AP 申请文件 f_n 的概率为式 (1):

$$q_{nk}' = (1 - z_{nk})q_{nk} \quad (1)$$

也就是说,如果某个 UE 的周围已经缓存有

文件 f_n , 则它不需要向 F-AP 请求文件;若周围没有此文件,则它仍然要向 F-AP 请求文件。这样就得到了用户向 F-AP 申请文件的概率。同时,用户向相邻用户申请文件 f_n 的概率为 $q_{nk} - q_{nk}'$ 。

这样,通过 D2D 下载的总的时延就可以写为:

$$Z_1 = \sum_{n=1}^N \sum_{k=1}^K (q_{nk} - q_{nk}') d_{D2D} \quad (2)$$

其中, d_{D2D} 为 D2D 通信的平均时延。

从而 F-AP 层下载的总时延为:

$$Z_2 = \sum_{n=1}^N \sum_{k=1}^K (1 - y_{nk}) q_{nk} d_{nk} \quad (3)$$

其中, d_{nk} 为设备 u_k 从 F-AP a_m 下载文件的时延,

且 $d_{nk} = \frac{|f|}{B \ln(1 + h^2 \gamma_m x_{nm})}$ 。其中 h 为 UE 与 F-AP 间的信道增益, γ_m 为 F-AP a_m 的等效传输信噪比 (equivalent transmit SNR)。

这样,就得到总时延的闭式解,为:

$$Z = Z_1 + Z_2 = \sum_{n=1}^N \sum_{k=1}^K (q_{nk} - q_{nk}') d_{D2D} + \sum_{n=1}^N \sum_{k=1}^K (1 - y_{nk}) q_{nk} d_{nk} \quad (4)$$

类似地,可以得到双层缓存对单纯的 F-AP 层缓存的时延增益的闭式解为:

$$\Delta Z = \sum_{n=1}^N \sum_{k=1}^K y_{nk} (d_{nk} - d_{D2D}) \quad (5)$$

缓存布置的优化问题就是求 $\min Z = Z_1 + Z_2$ 。

这是一个 NP 难问题,它同时与 UE 层缓存分布 y_{nk} 和 F-AP 层缓存分布 x_{nm} 有关。下面将用分布式的方法求最优缓存分布。

3.1 UE 层缓存分布

3.1.1 文件流行度

首先研究 UE 层。假设不同 F-AP 覆盖下的 UE 不会互相干扰^[7],因此只需要研究一个 F-AP 覆盖下的 UE。UE 层的缓存问题可以看作一个典型的背包问题:每个文件的流行度可以看作背包问题中每个物品的权重;背包容量就是 UE 的缓



存容量 Q_u 。为了让 UE 可以分布式地估计每个文件的流行度, 设备 u_k 要记住相邻设备向自己请求文件 f_n 的次数, 记为 t_{nk} , 这样, 就能定义某个文件 f_n 的本地流行度, 为:

$$P_{nk} = \frac{t_{nk}}{\sum_n t_{nk}} \quad (6)$$

考虑到突发新闻等因素, 用户的兴趣可能会在短时间内突变, 为了提高 D2D 传输利用率, 让 F-AP 也能通过上述类似的方法计算某个文件的全局流行度, 并使 UE 从 F-AP 下载文件的同时也获得该文件的全局流行度。而全局流行度会在相邻 UE 间快速传播, 这样每个 UE 都能了解到最新的流行文件, 并缓存它。

总之, 对本地流行度可以这样估计: 每次一个相邻 UE 向 u_i 请求文件 f_n , u_i 就会将自己的 t_{ni} 加 1。 u_i 从相邻的 u_j 那里获取一个文件 f_n 的同时, t_{nj} 也会一起传送给 u_i 。这样, u_i 就能更新自己的 t_{ni} :

$$t_{ni} = \delta t_{ni} + (1 - \delta) t_{nj}, \delta \in [0, 1] \quad (7)$$

当 u_i 从 F-AP 下载 1 个文件时, 这个文件的全局流行度 P_n 也会一并传送给 u_i 。 u_i 将 P_n 折算成申请次数, 并更新自己的 t_{ni} :

$$t_{ni} = \delta t_{ni} + (1 - \delta) P_n |N_i|, \delta \in [0, 1] \quad (8)$$

其中, $|N_i|$ 是与 u_i 相邻的 UE 的个数。更新 t_{ni} 之后, u_i 就能计算新的本地流行度。系数 δ 是表征将外来申请次数 (如 t_{nj} , $P_n |N_i|$) 折算成本地申请次数时的影响权重。

3.1.2 优化缓存利用率

然而随之而来的问题是: 经过一段时间后, 每个 UE 都会保存最流行的文件, 导致每个 UE 缓存内容相同, 浪费了缓存资源。为了解决这个问题, UE 不能简单地缓存流行度大的文件。规定 UE 将流行度和缓存率的差作为判断是否缓存的标准, 优先缓存差值大的文件。为了做到这一点, u_i 每次缓存 f_n 之前都要向相邻 UE 发送广播信

号, 询问周围有多少 UE 已经缓存了 f_n , 然后计算 f_n 的缓存率 P'_{ni} 和 $G_{ni} = P_{ni} - P'_{ni}$, 最后缓存 G_{ni} 最大的文件。这样, u_i 每次就能缓存周围 UE 较少缓存的但流行度又较高的文件。

其中, 缓存率 P'_{ni} 定义为 u_i 周围已缓存 f_n 的 UE 个数与 UE 总数的比值, 即:

$$P'_{ni} = \frac{\text{card}\{u_k | u_k \in a_m \& y_{nk} = 1\}}{\text{card}\{u_k | u_k \in a_m\}} \quad (9)$$

3.1.3 UE 层分布式缓存算法

拥有了 G_{ni} , 就能求解缓存分布的背包问题。前面已经提到, 每个 UE 的缓存容量为 Q_u 。当缓存未滿时, UE 总是保存接收到的文件; 如果缓存已滿, 则根据 G_{ni} 的大小, 替换缓存中 G_{ni} 最小的文件。假设 u_i 已经缓存了 Q_u 个文件, 这时它从别处 (F-AP 或别的 UE) 获得了第 $(Q_u + 1)$ 个文件, 记为 f_{Q_u+1} 。在收集所有必要信息之后, u_i 可以计算出所有 $(Q_u + 1)$ 个文件的 G , 进而可以替换缓存中的文件。

用贪心算法来解决这个背包问题, 算法的复杂度由将每个文件的 G 降序排列的过程决定, 为 $O(\text{nlb}(n))$ 。

算法 1 UE 层分布式算法

```

 $H_k \leftarrow G_{ni}$ 
将  $H_k$  降序排列
 $W \leftarrow Q_u, w_n = |f_n|, C_k \leftarrow \Phi$ 
for  $k = 1$  to  $c + 1$  do
    if  $w_n < W$  then
         $W = W - |f_n|$ 
         $C_k = C_k \cup f_n$ 

```

3.2 D2D 通信的实现

前文已经提到, 在 D2D 网络中, 每个 UE 每次最多只能向一个 UE 发送或接收文件, 因此定义状态变量 S_k 表示 u_k 是否正在发送文件, $S_k = 1$ 表示 u_k 正在发送文件, 反之表示没有发送。定义 R_k 表示 u_k 是否正在接收文件, 显然只有当 $R_k = 0$ 时, u_k 才能发送文件请求。 u_i 向 u_j 发出文件申请

的同时会查询 R_j 的值, 只有 $R_j = 0$ 时, u_i 和 u_j 才能进行 D2D 通信。

3.3 F-AP 层缓存分布

上文已经得到了 u_k 向 F-AP 申请文件 f_n 的概率 q_{nk} 以及 F-AP 层的下载总时延 $Z_2 = \sum_{n=1}^N \sum_{k=1}^K (1 - y_{nk}) q_{nk} d_{nk}$ 。那么 F-AP 层内缓存优化问题可以表示为:

$$\min_{\{x_{nm}\}} \frac{1}{K} \sum_{k=1}^K \sum_{n=1}^N q_{nk} d_{nk} \begin{cases} \sum_{n=1}^N x_{nm} \leq Q_m, \forall a_m \in A \\ x_{nm} \in \{0, 1\}, \forall f_n \in F, a_m \in A \end{cases} \quad (10)$$

上文已经提到, 这也是一个 NP 难问题。为了分布式地解决这个问题, 根据 BP (belief propagation) 理论提出一种分布式算法。参考文献[11]阐述了 BP 算法中的信息传递理论和因子图理论。

3.3.1 缓存分布的因子图

为了应用 BP 算法, 必须先将上面带边界条件的优化问题转化为下面的无边界优化问题。可以证明, 原问题等价于求解:

$$\hat{X} = \arg \max_{X \in \{0, 1\}^{NM}} \prod_{(f_n, u_k)} \eta_{nk}(X) \prod_{m=1}^M g_m(X) \quad (11)$$

其中, η_{nk} 和 g_m 是缓存策略 X 的两个函数, 分别等价于 u_k 下载 f_n 的时延以及 F-AP 缓存容量的限制。

$$\eta_{nk}(X) = \exp(-q_{nk} d_{nk}) \quad (12)$$

$$g_m(X) = \begin{cases} 1, \sum_{n=1}^N x_{nm} \leq Q_m \\ 0, \text{其他} \end{cases} \quad (13)$$

这样就能把图 2 所示的因子图应用于求解式 (10)。因为要求解 X , 所以将变量节点 μ_i 设为 x_{nm} , 而 $\eta_{nk}(X)$ 和 $g_m(X)$ 为 X 的函数, 因此将它们设为函数节点 F_j :

$$\mu_i = x_{nm}, i = (m-1)N + n \quad (14)$$

$$F_j = \begin{cases} \eta_{nk}, j = \sum_{k=1}^{K-1} |F_k| + \zeta(n, k) \\ g_m, j = \sum_{k=1}^K |F_k| + m \end{cases} \quad (15)$$

其中, F_k 是用户 u_k 所请求的文件的集合, $|F_k|$ 是集合 F_k 中的元素数量, $\zeta(n, k)$ 指的是集合 F_k 中文件 f_n 的下标。因子图中每个变量节点 μ_i 都和函数节点 F_j 相连, 因此共有 NM 个变量节点、

$M + \sum_{k=1}^K |F_k|$ 个函数节点。

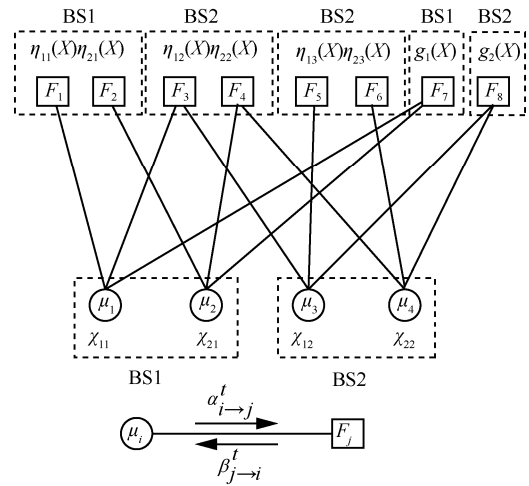


图2 F-AP 层因子图模型

3.3.2 因子图中的置信度传播

在因子图中, 每个变量节点将一个更新后的信息送到一个与它相邻的函数节点中, 并收到一个更新后的、发送自该函数节点的信息, 在多次迭代之后, 就可以计算出寻找优化问题的最优解。而要做的设计一个信息传递算法。用 $m_{\mu_i \rightarrow F_j}^t$ 表示从变量节点发送到函数节点的信息, 用 $m_{F_j \rightarrow \mu_i}^t(x)$ 表示从函数节点发送到变量节点的信息。根据 BP 算法, 信息 $m_{\mu_i \rightarrow F_j}^t(x)$ 可以根据式 (16) 更新为:

$$m_{\mu_i \rightarrow F_j}^{t+1}(X) = \prod_{F_i' \in \Gamma_i^u \setminus \{j\}} m_{\mu_i \rightarrow F_j'}^t(X) \quad (16)$$

其中, $\Gamma_i^u \setminus \{j\}$ 表示除 F_j 以外所有与 μ_i 相连的函数节点。



$m_{F_j \rightarrow \mu_i}^t(x)$ 可以根据式 (17) 更新:

$$m_{F_j \rightarrow \mu_i}^{t+1}(x) = \max_{\Gamma_j^F \setminus \{i\}} \{F_j(X) \prod_{U_m} m_{\mu_i \rightarrow F_j}^t(x_i)\} \quad (17)$$

其中, $\Gamma_j^F \setminus \{i\}$ 表示除 μ_i 外所有与 F_j 相连的变量节点。

因为式 (17) 均为连乘, 因此为了便于计算, 将其化为对数域:

$$\alpha_{i \rightarrow j}^t = \text{lb}\left(\frac{m_{\mu_i \rightarrow F_j}^t(1)}{m_{\mu_i \rightarrow F_j}^t(0)}\right), \beta_{j \rightarrow i}^t = \text{lb}\left(\frac{m_{F_j \rightarrow \mu_i}^t(1)}{m_{F_j \rightarrow \mu_i}^t(0)}\right) \quad (18)$$

$\alpha_{i \rightarrow j}^t$ 可以等价地表示 $m_{\mu_i \rightarrow F_j}^t(x)$, $\beta_{j \rightarrow i}^t$ 可以等价地表示 $m_{F_j \rightarrow \mu_i}^t(x)$ 。这样, 就可以将信息的传递过程表述如下。

$\alpha_{i \rightarrow j}^t$ 按式 (19) 更新:

$$\alpha_{i \rightarrow j}^{t+1} = \sum_{l \in \Gamma_i^t \setminus \{j\}} \beta_{j \rightarrow l}^t \quad (19)$$

当 $F_j = \eta_{nk}$ 时, 信息 $\beta_{j \rightarrow i}^{t+1}$ 由式 (20) 获得:

$$\beta_{j \rightarrow i}^{t+1} = p_{nk}(\bar{D}_{nk}(X_{i,0}^t) - \bar{D}_{nk}(X_{i,1}^t)) \quad (20)$$

当 $F_j = g_m$ 时, 信息 $\beta_{j \rightarrow i}^{t+1}$ 可按式 (21) 更新:

$$\beta_{j \rightarrow i}^{t+1} = \min\{0, -\alpha_{l \rightarrow j}^{(Q_m)}(t)\} \quad (21)$$

其中, $\alpha_{l \rightarrow j}^{(Q_m)}(t)$ 是信息 $\{\alpha_{i \rightarrow j}^{t+1}\}$ 中按降序排列的第 Q_m 个信息。

然后, 就能得出指数域中的置信度比率:

$$\tilde{b}_i^t = \log\left(\frac{b_i^t(1)}{b_i^t(0)}\right) = \sum_{j \in \Gamma_i^t} \beta_{j \rightarrow i}^t \quad (22)$$

至此, 已经得到了置信度, 只需根据式 (23), 就能得出变量节点的值, 为:

$$\mu_i^t = \begin{cases} 1, & \tilde{b}_i^t > 0 \\ 0, & \tilde{b}_i^t < 0 \end{cases} \quad (23)$$

在每个循环迭代中, μ_i^t 根据式 (23) 中的置信度做出估计, 在 0、1 之间交替取值, 直到收敛。

算法 2 F-AP 层分布式算法

初始化 $t=0, \alpha_{i \rightarrow j}^t = \beta_{j \rightarrow i}^t = 0, \forall i, j$

设 $t_{\max}$ 为足够大

while Not convergent and $t \leq t_{\max}$ do

for $m=1:M$ do

for $n=1:N$ do

计算 $\alpha_{i \rightarrow j}^t$

for $k \in U_m$ do

计算 $\beta_{j \rightarrow i}^{t+1}$ ($F_j = \eta_{nk}$)

end for

end for

计算 $\beta_{j \rightarrow i}^{t+1}$ ($F_j = g_m$)

计算 $\tilde{b}_i^t$

得到 μ_i

end for

$t = t + 1$, 检查收敛

end while

得到 X

4 仿真结果及分析

假设场景中有 M 个 F-AP, $M=10$, 有 $K=50$ 个 UE, 它们均匀分布在 10 个 F-AP 范围内。总共有 N 个文件, $N=100$ 。如前文所述, 假设每个 F-AP 能够缓存 Q_m 个文件, 每个用户能够缓存 Q_u 个文件。

取 F-AP 到 UE 的通信链路带宽为 30 MHz^[12], 每个时隙长度为 20 ms, 文件平均大小为 1 GB。F-AP 到 UE 信道增益 $|h_{km}|^2$ 在每个时隙上符合标准指数分布。F-AP 和用户层间传输平均信噪比为 SNR=10 dB, D2D 通信平均信噪比 SNR=20 dB, 从云端下载时延 $D^*=40$ s^[12]。使用 MATLAB 软件进行仿真, 结果如下。

4.1 下载平均时延和 F-AP 缓存容限 Q_m

取 $Q_m=10, 20, \dots, 90$, 并保持 $Q_u=10$ 不变, 得到图 3。图 3 中选择了基于贪心算法的单层集中式缓存^[13]、基于 BP 算法的单层分布式缓存 (即本文的双层缓存去掉 UE 层, 将 UE 层缓存容量加入 F-AP 层) 与本文的双层缓存对比, 设定 3 种缓存

分布的总容量都相同,做出了时延性能曲线。双层缓存对于 F-AP 的容量要求并不是很高,在 Q_m 较小时也能获得较大的时延增益。效果明显优于单层缓存布置。同时可以看出,基于 BP 算法的分布式缓存时延特性基本上与集中式算法相当,这得益于 BP 算法优异的性能。

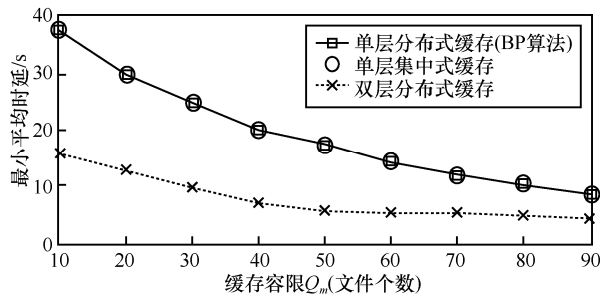


图3 3种缓存方式的时延性能曲线

4.2 算法复杂度

为了研究算法的收敛特性,设 $Q_u=5$ 不变,得到不同 Q_m 的平均时延曲线如图4所示。从图4中可以看出,平均时延一开始有一个初始值,接着在迭代过程中不断改变,最终收敛。算法迭代次数与 Q_m 取值没有明显关系,在 Q_u 不变时,对于所有的 Q_m ,算法都能在10次左右完成收敛。考虑到 UE 网络的贪心算法复杂度为 $O(n \log(n))$, n 为 UE 缓存的文件个数,通常 UE 缓存不会太大,因而总的复杂度依然不高。

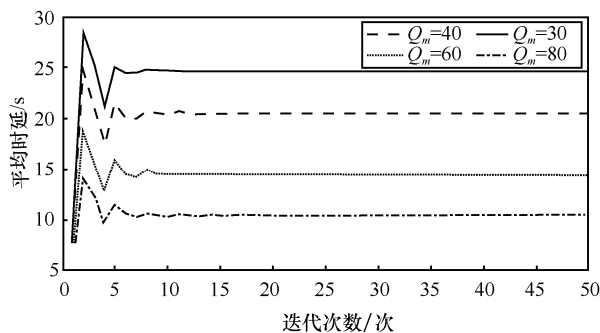


图4 不同 Q_m 的平均时延曲线

4.3 UE 缓存容量 Q_u 和时延性能

本文取 $Q_m=40$, $Q_u=1,5,10,15,20$, 作出了双层缓存对于基于 BP 算法的 F-AP 单层分布式缓存

的时延增益,如图5所示。

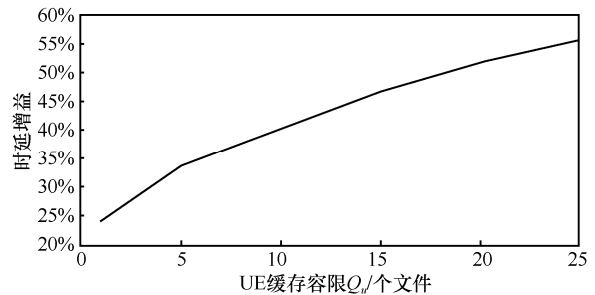


图5 双层缓存对于基于 BP 算法的 F-AP 单层分布式缓存的时延增益

从图5中可以看出,在 Q_u 很小的情况下,时延增益也能提高 25%~50%,这主要得益于 D2D 传输的低时延以及 D2D 层中的分布式背包算法以及对于权重的优化提高了命中率。

5 结束语

本文主要研究了集成了 D2D 通信的雾无线接入网中的缓存分布问题,提出双层缓存布置的分布式算法,仿真结果显示该双层分布式缓存布置可以显著降低用户的下载时延。在将来的研究中,可以考虑实际信道条件或针对链路负载、信令开销等进行优化,最终提出可以工程应用的缓存分布算法。

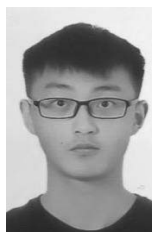
参考文献:

- [1] PENG M, YAN S, ZHANG K, et al. Fog Computing based radio access networks: issues and challenges[J]. IEEE Network, 2016, 30(4): 46-53.
- [2] LIU J, BAI B, ZHANG J, et al. Cache placement in fog-RANs: from centralized to distributed algorithms[J]. IEEE Transactions on Wireless Communications, 2017, PP(99): 1.
- [3] DOPPLER K, RINNEM, WIJTING C, et al. Device-to-device communication as an underlay to LTE-advanced networks[J]. Modern Science & Technology of Telecommunications, 2010, 47(12): 42-49.
- [4] DOPPLER K, YU C H, RIBEIRO C B, et al. Mode selection for device-to-device communication underlaying an LTE-advanced network[C]//2010 IEEE Wireless Communications and Networking Conference (WCNC), April 18-21, 2010, Sydney, NSW, Australia. Piscataway: IEEE Press, 2010: 1-6.
- [5] YU C H, TIRKKONEN O, DOPPLER K, et al. Power optimization of device-to-device communication underlaying cellular



- communication[C]//2009 IEEE International Conference on Communications, June 14-18, 2009, Dresden, Germany. Piscataway: IEEE Press, 2009: 1-5.
- [6] JANIS P. Interference-aware resource allocation for device-to-device radio underlaying cellular networks[C]//2009 IEEE Vehicular Technology Conference (VTC), April 26-28, 2009, Barcelona, Spain. Piscataway: IEEE Press, 2009: 1-5.
- [7] GOLREZAEI N, MOLISCH AF, DIMAKISAG. Base-station assisted device-to-device communications for high-throughput wireless video networks[C]//IEEE International Conference on Communications, June 10-15, 2012, Ottawa, ON, Canada. Piscataway: IEEE Press, 2012.
- [8] ZHU H, WANG J. Chunk-based resource allocation in OFDMA systems-part I: chunk allocation[J]. IEEE Transactions on Wireless Communications, 2009, 57(9): 2734-2744.
- [9] ZHU H, WANG J. Chunk-based resource allocation in OFDMA systems-part II: joint chunk, power and bit allocation[J]. IEEE Transactions on Wireless Communications, 2012, 60(2): 499-509.
- [10] CHOI J, JAIN M, SRINIVASAN K, et al. Achieving single channel, full duplex wireless communication[C]//2010 IEEE International Conference on Mobile Computing and Networking (MobiCom), September 20-24, 2010, Chicago, Illinois, USA. Piscataway: IEEE Press, 2010: 1-12.
- [11] YIN C, YANG D, ZHAO X, et al. State estimation of active distribution system based on the factor graph analysis and belief propagation algorithm[C]//2017 IEEE International Conference on Environment and Electrical Engineering and IEEE Industrial and Commercial Power Systems Europe (IEEE/I&CPS Europe), June 6-9, 2017, Milan, Italy. Piscataway: IEEE Press, 2017: 1-6.
- [12] BAS E, BENNIS M, DEBBAH M, et al. Living on the edge: the role of proactive caching in 5G wireless networks[J]. IEEE Communications Magazine, 2015, 52(8): 82-89.
- [13] BASTUG E, BENNIS M, KOUNTOURIS M, et al. Cache-enabled small cell networks: modeling and tradeoffs[J]. EURASIP Journal on Wireless Communications and Networking, 2015(1): 41.

[作者简介]



夏骋宇 (1997-), 男, 东南大学信息科学与工程学院在读, 主要研究方向为 5G/B5G 新型架构中的缓存应用等。



蒋雁翔 (1977-), 男, 博士, 东南大学移动通信国家重点实验室副教授, 主要研究方向为 5G/B5G 移动通信理论与关键技术、无线大数据智能通信理论与关键技术等。