

基于随机几何理论的流行度匹配边缘缓存策略

刘浩洋 王钢* 杨文超 王金龙 许尧 赵东来

(哈尔滨工业大学电子与信息工程学院 哈尔滨 150001)

摘要: 异构网络边缘缓存机制是解决传统回程传输链路负载过大的可靠技术之一, 但已有的缓存策略往往不能与被请求数据的流行度相匹配。为了解决这一问题, 该文提出一种流行度匹配边缘缓存策略(PMCP), 该策略能够根据流行度参数匹配对应的文件缓存概率以最大限度提升通信可靠性并降低回程带宽压力。基站的平面位置通过随机几何建模, 文件的被请求概率则通过齐夫分布建模。蒙特卡罗仿真结果表明缓存机制能够有效降低回程带宽压力, 且所提出缓存策略的可靠性优于对比策略。

关键词: 异构网络; 边缘缓存; 随机几何; 流行度匹配

中图分类号: TN919

文献标识码: A

文章编号: 1009-5896(2021)12-3427-07

DOI: 10.11999/JEIT210493

Popularity Matching Edge Caching Policy Based on Stochastic Geometry Theory

LIU Haoyang WANG Gang YANG Wenchao WANG Jinlong
XU Yao ZHAO Donglai

(School of Electronics and Information Engineering, Harbin Institute of Technology, Harbin 150001, China)

Abstract: Edge caching mechanism for heterogeneous network is one of the reliable technologies to solve the excessive link load of the traditional backhaul mechanism, but the existing caching policies often can not match the popularity of the required data. To solve this problem, a Popularity Matching Caching Policy (PMCP) is proposed in this paper, which can match the corresponding file cache probability according to the popularity parameters to maximize communication reliability and reduce backhaul bandwidth pressure. The plane position of the base station is modeled by stochastic geometry theory. The results of the Monte Carlo simulation show that the proposed caching policy can effectively reduce the backhaul bandwidth pressure, and the reliability of the proposed caching policy is better than the comparison policies.

Key words: Heterogeneous network; Edge caching; Stochastic geometry; Popularity matching

1 引言

随着数据量的爆炸式增长, 用户对于高速数据流需求正迅速增加。然而传统上从核心网获取数据的过程往往会对回程链路造成很大带宽压力, 因此近年来的研究倾向于将文件缓存在异构网络的微基站中^[1-3]。

为了降低资源消耗, Song等人^[4]将缓存空间和回程链路通过平衡参数组合为统一资源进行优化, 获得了最优的微基站密度和最优的缓存空间, 并在保证服务质量的前提下最大限度地减轻了回程链路负载。

针对不同的性能指标, Krishnendu等人^[5]为了最大化缓存命中概率, 将网络假设为包含有限数量的基站、缓存文件和用户数量的数学模型, 并将目标问题近似为凸问题, 最终通过随机舍入算法获得最优缓存策略; Wang等人^[6]则基于马尔科夫链提出一种用于提升缓存命中率的算法; Tamoor-Ul-Hassan等人^[7]则着重于通信中断概率, 比较了多种缓存策略并推导了包含频谱分配策略的中断概率; Liu等人^[8]则将缓存和多层异构网络相结合, 优化了通信成功概率和空间频谱效率, 获得了最优缓存概率。

针对特殊的网络结构, Zhang等人^[9]将 (Device to Device, D2D) 通信技术考虑在内, 提出了两种用于解决用户自私性的算法, 并通过所提出的拍卖模型的自然社会效率和个人理性属性验证了算法的有效性; 杨静等人^[10]则通过D2D共享用户之间的内容提出了一种用于降低时延的缓存策略, Hua等

收稿日期: 2021-06-01; 改回日期: 2021-09-10; 网络出版: 2021-09-25

*通信作者: 王钢 gwang51@hit.edu.cn

基金项目: 国家自然科学基金(62071146, 62071147)

Foundation Items: The National Natural Science Foundation of China (62071146, 62071147)

人^[11]则通过将边缘用户分类的D2D边缘缓存策略提升了缓存命中率并降低了下载延迟。

本文通过将基站的2维平面分布建模为多个独立的泊松点过程(Poisson Point Process, PPP), 推导了不同缓存策略对应的平均中断概率。针对文献[7]所采用的传统缓存策略中未考虑数据流行度的缺点, 本文提出一种流行度匹配缓存策略, 使用该策略能够在降低回程带宽压力的情况下获得比传统策略更低的平均中断概率。此外, 本文还分析了微基站部署密度和缓存容量对平均中断概率的影响。

2 系统模型

如图1所示, 考虑异构网络用户的下行通信链路, 其中宏基站部署了用于从核心网获取数据的回程链路, 微基站部署了用于缓存流行内容的存储设备。根据文献[12], 当慢衰落的标准差足够大时, 尽管基站的空间位置是固定的, 但从用户的角度分析, 由于异构网络的基站可以以高密度部署在任何位置, 每个时刻基站的相对位置都可以看作未知且随机的。而采用PPP随机模拟基站的位置分布相当于将多种不同的应用场景进行叠加后取平均, 因此基站的齐次PPP模型可以用于模拟现实世界中确定部署基站下的用户信干噪比(Signal to Interference plus Noise Ratio, SINR)分布。

假设宏基站和微基站的平面分布分别服从两个独立的2维PPP, 记作 ϕ_M 和 ϕ_S , 对应的基站密度分别为 λ_M 和 λ_S 。由于本文仅考虑网络中单个用户的通信可靠性, 因此假设宏基站和微基站使用相同频段。宏基站和微基站的功率分别为 P_M 和 P_S 。微基站的服务半径为 R_{max} 。为了便于理解, 本文将同宏基站连接的用户称作M类用户, 将同微基站连接的用户称作S类用户。

图2所示为使用PPP建模仿真的基站位置分布

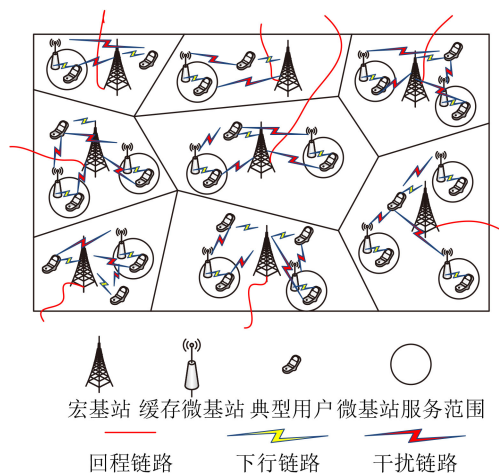


图1 支持微基站缓存的异构网络模型

示意图。可以看到, 在一块平面区域内同时分布着相对较高密度的微基站和相对较低密度的宏基站。

假设每个用户从一个容量为 N 的数据库中获得文件, 每个微基站的存储容量为 R ($0 < R \leq N$)。假设文件大小均为单位值, 并根据它们的流行度进行降序排列。文件的被请求概率服从齐夫分布, 其中文件 i (第 i 流行的文件)的被请求概率为

$$p_i = \frac{i^{-\delta}}{\sum_{j=1}^N j^{-\delta}} \quad (1)$$

其中, δ 是齐夫分布的偏斜参数, 由式(1)可知, 高流行度文件和低流行度文件的请求概率差距会随着 δ 增大而增大。

假设每个微基站独立选择被缓存的文件, 文件 i 的被缓存概率记作 q_i , 因为所有文件大小均为单位值, 因此文件 i 所占据的平均存储空间为 q_i , 那么 q_i 和 R 之间满足

$$R = \sum_{i=1}^N q_i \quad (2)$$

本文的对比缓存策略为平均缓存策略(Uniform Caching Policy, UCP)和流行度缓存策略(Popularity-based Caching Policy, PCP), 缓存率分别为

$$q_{i,UCP} = \frac{R}{N}, \quad i = 1, 2, \dots, N \quad (3)$$

$$q_{i,PCP} = \begin{cases} 1, & i = 1, 2, \dots, R \\ 0, & i = R + 1, R + 2, \dots, N \end{cases} \quad (4)$$

由式(3)可知, 在UCP策略下, 所有文件的被缓存概率相同, 均为 R/N ; 由式(4)可知, PCP策略只缓存前 R 个文件, 即前 R 个文件的被缓存概率均为1, 其余文件的被缓存概率则为0。

由于每个微基站独立选择被缓存的文件, 因此缓存文件 i 的基站集合记作 ϕ_{S_i} , 其部署密度为 $\lambda_{S_i} = \lambda_S q_i$ 。同理, $\phi_{S_i'}$ 和 $\lambda_{S_i'} = \lambda_S(1 - q_i)$ 则分别表示不缓存文件 i 的基站集合和其部署密度。

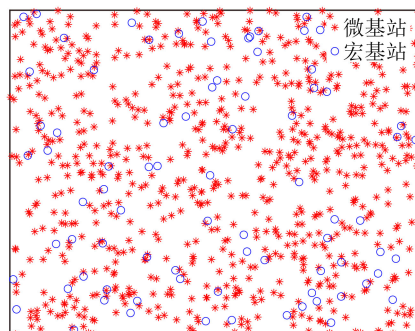


图2 PPP建模的基站位置分布示意图

M类用户和S类用户请求文件*i*时的下行链路 SINR分别为

$$\text{SINR}_M = \frac{P_M h_M R_M^{-\alpha}}{I_{M_i} + \sigma^2} \quad (5)$$

$$\text{SINR}_{S_i} = \frac{P_S h_{S_i} R_{S_i}^{-\alpha}}{I_{S_i} + \sigma^2} \quad (6)$$

其中, h_M 和 h_{S_i} 为用户与连接的宏基站(M0)和微基站(S_{i0})之间的信道系数, 假设微基站-用户和宏基站-用户之间为参数为1的独立同分布瑞利衰落信道, 即 $h_M \sim \exp(1)$, $h_{S_i} \sim \exp(1)$ ($i = 1, 2, \dots, N$); R_M 为用户和与之最近的宏基站的距离, R_{S_i} 为用户和与之最近且缓存文件*i*的微基站的距离; α 为路径损耗系数, 为了便于下文的推导, 假设 α 的值为4。 $I_{M_i} = I_{M_iM} + I_{M_iS_i} + I_{M_iS_i'}$ 表示M类用户所受到的同频干扰, 其中 $I_{M_iM} = \sum_{j \in \phi_M \setminus M0} P_M h_{M_j} R_{M_j}^{-\alpha}$, 代表宏基站对M类用户的干扰; $I_{M_iS_i} = \sum_{j \in \phi_{S_i}} P_S h_{S_{ij}} R_{S_{ij}}^{-\alpha}$, 代表缓存了文件*i*的微基站对M类用户的干扰; $I_{M_iS_i'} = \sum_{j \in \phi_{S_i'}} P_S h_{S_{ij}} R_{S_{ij}}^{-\alpha}$ 代表未缓存文件*i*的微基站对M类用户的干扰。 $I_{S_i} = I_{S_iM} + I_{S_iS_i} + I_{S_iS_i'}$ 表示S类用户所受到的同频干扰, 其中 $I_{S_iM} = \sum_{j \in \phi_M} P_M h_{M_j} R_{M_j}^{-\alpha}$, 代表宏基站对S类用户的干扰; $I_{S_iS_i} = \sum_{j \in \phi_{S_i} \setminus S_{i0}} P_S h_{S_{ij}} R_{S_{ij}}^{-\alpha}$, 代表缓存了文件*i*的微基站对S类用户的干扰; $I_{S_iS_i'} = \sum_{j \in \phi_{S_i'}} P_S h_{S_{ij}} R_{S_{ij}}^{-\alpha}$, 代表未缓存文件*i*的微基站对S类用户的干扰。由于异构网络通常是干扰受限的^[13], 因此本文忽略加性噪声的影响, 即假设 $\sigma^2 = 0$ 。

3 流行度匹配缓存策略

3.1 接入策略

为了减轻回程链路的带宽压力, 采用微基站优先的接入策略。当一个用户请求文件*i*时, 用户将首先在与之相距 R_{\max} 的范围内搜索距之最近且缓存了文件*i*的微基站, 如果有则与之连接, 反之则搜索与之最近的宏基站, 通过回程链路从核心网获取所需文件。在该策略下, 回程链路只有在微基站无法提供服务时才会被占用, 从而最大限度减轻了回程链路的带宽压力。

当用户请求文件*i*时, 与微基站连接的概率记为 p_{S_i} , 此概率等价于与用户相距 R_{\max} 的范围内至少存在一个缓存了文件*i*的微基站的概率, 表示为

$$p_{S_i} = p_{R_{S_i} < R_{\max}} = 1 - e^{-q_i \lambda_S \pi R_{\max}^2} \quad (7)$$

用户对微基站的使用率即为用户的平均缓存命中率, 同时该指标也代表回程链路的带宽释放率

$$p_{\text{release}} = \sum_{i=1}^N p_i p_{S_i} \quad (8)$$

S类用户同微基站之间距离的概率密度函数为

$$\begin{aligned} f_{R_{S_i} | R_{S_i} < R_{\max}}(r) &= \frac{d(F_{R_{S_i} | R_{S_i} < R_{\max}}(r))}{dr} \\ &= \frac{d\left(\frac{p_{R_{S_i} < r}}{p_{S_i}}\right)}{dr} \\ &\stackrel{(a)}{=} \frac{2\pi q_i \lambda_S r e^{-q_i \lambda_S \pi r^2}}{1 - e^{-q_i \lambda_S \pi R_{\max}^2}} \end{aligned} \quad (9)$$

其中, 步骤(a)中使用了式(7)中对 p_{S_i} 的推导。

与式(9)同理, M类用户同宏基站距离的概率密度函数为

$$f_{R_M} = 2\pi \lambda_M r e^{-\lambda_M \pi r^2} \quad (10)$$

3.2 缓存策略

使用UCP时, 即使不同文件的被请求概率差异较大(δ 较大), 高低排名文件的被缓存概率也会保持一致。与之相反, 当使用PCP时, 如果 δ 较小, 即使高低排名文件的被请求概率差异不大, 低排名文件也不会被缓存。可见PCP和UCP仍有很大缺陷, 会造成缓存命中率的降低和中断概率的上升。

造成该问题的主要原因是文件的被缓存概率 q_i 和被请求概率 p_i 的不匹配, 为了解决这个问题, 本文提出了流行度匹配缓存策略(Popularity Matching Caching Policy, PMCP)。在该策略中, 排名第1的被缓存概率最高, 即 $\arg \max_i q_i = 1$, 为了差异化高流行度文件和低流行度文件的被缓存概率, 将 q_i 和 q_{i-1} 设定为等比例关系, 比例系数为 d ($0 < d < 1$), 即 $q_i = q_{i-1} d$ 。根据式(2)可得

$$\begin{aligned} \sum_{i=1}^N q_i &= \sum_{i=1}^N q_1 d^{i-1} \\ &\stackrel{(a)}{=} \frac{q_1 (1 - d^N)}{1 - d} \\ &= R \end{aligned} \quad (11)$$

其中, 步骤(a)为等比数列求和公式, 由式(11)可知 q_1 的取值范围在 $(R/N, 1)$ 之间, 且 q_1 与 d 正相关, 因此 q_1 越大, 缓存概率的差异性就越大。为了将文件缓存概率的差异性和被请求概率的差异性关联起来, 将 q_1 和 δ 相关联, 关系为

$$q_1 = \frac{R}{N} + \left(1 - \frac{R}{N}\right) \arctan(10\delta) \quad (12)$$

其中, $\arctan(10\delta)$ 是 q_1 在 $(R/N, 1)$ 之间的正相关滑动系数, 随着 δ 在 $(0, \infty)$ 范围内增大, q_1 也会在 $(R/N, 1)$ 范围内增大, 即随着文件之间的被请求概率差异性增大, 文件之间的被缓存概率差异性也将增大。

将式(12)代入式(11)中可得

$$R = \left[\frac{R}{N} + \left(1 - \frac{R}{N} \right) \arctan(10\delta) \right] \frac{(1-d^N)}{1-d} \quad (13)$$

使用二分法求解式(13), 可以在 $(0, 1)$ 之间求得 d 的唯一解, 至此可以获得每个文件的被缓存概率 $q_1, q_2 = q_1d, q_3 = q_1d^2, \dots, q_N = q_1d^{N-1}$, 以上即为PMCP的理论基础。

3.3 平均中断概率

中断概率的含义为用户可达速率的下限达成失败率, 即用户的可达速率高于某一下限的概率, 根据香农定理, 可达速率和用户的SINR成正相关, 因此中断概率可以等价为用户的SINR高于给定门限 γ 的概率。

平均中断概率的概念为不同用户通过不同种类的基站请求不同文件时的中断概率的加权平均值

$$\mathbb{P}_{\text{out}} = \sum_{i=1}^N p_i [p_{Si} \mathbb{P}_{\text{out}, Si} + (1 - p_{Si}) \mathbb{P}_{\text{out}, Mi}] \quad (14)$$

其中, $\mathbb{P}_{\text{out}, Si}$ 和 $\mathbb{P}_{\text{out}, Mi}$ 分别为S类用户请求文件 i 时的中断概率和M类用户通过宏基站请求文件 i 时的中断概率。

3.3.1 微基站中断概率

式(14)中 $\mathbb{P}_{\text{out}, Si}$ 的解析表达式如式(15)所示

$$\begin{aligned} \mathbb{P}_{\text{out}, Si} &= \mathbb{E}_{R_{Si}} [p(\text{SINR}_{Si} < \gamma | R_{Si} = r)] \\ &= 1 - \int_0^{R_{\max}} p(\text{SINR}_{Si} > \gamma | R_{Si} = r) f_{R_{Si} | R_{Si} < R_{\max}}(r) dr \\ &= 1 - \int_0^{R_{\max}} p(\text{SINR}_{Si} > \gamma | R_{Si} = r) \\ &\quad \frac{2\pi q_i \lambda_S r e^{-q_i \lambda_S \pi r^2}}{1 - e^{-q_i \lambda_S \pi R_{\max}^2}} dr \\ &\stackrel{(a)}{=} 1 - \int_0^{R_{\max}} L_{I_{SiM}} \left(\frac{\gamma}{P_S r^{-\alpha}} \right) L_{I_{SiSi}} \left(\frac{\gamma}{P_S r^{-\alpha}} \right) \\ &\quad \cdot L_{I_{SiSi'}} \left(\frac{\gamma}{P_S r^{-\alpha}} \right) \frac{2\pi q_i \lambda_S r e^{-q_i \lambda_S \pi r^2}}{1 - e^{-q_i \lambda_S \pi R_{\max}^2}} dr \quad (15) \end{aligned}$$

其中, $\mathcal{L}_{I_{SiM}} \left(\frac{\gamma}{P_S r^{-\alpha}} \right)$, $\mathcal{L}_{I_{SiSi}} \left(\frac{\gamma}{P_S r^{-\alpha}} \right)$, $\mathcal{L}_{I_{SiSi'}} \left(\frac{\gamma}{P_S r^{-\alpha}} \right)$ 分别代表 $I_{SiM}, I_{SiSi}, I_{SiSi'}$ 的概率密度函数在参数 $\frac{\gamma}{P_S r^{-\alpha}}$ 下的拉普拉斯变换。

式(15)步骤(a)的证明: 由于 $I_{SiM}, I_{SiSi}, I_{SiSi'}$ 相互独立, 因此以上3个随机变量之和所组成的随机变量的概率密度函数为此3个随机变量的概率密度函数的卷积, 即 $f_{I_{Si}}(u) = f_{I_{SiM}}(u) * f_{I_{SiSi}}(u) * f_{I_{SiSi'}}(u)$, 因此有

$$\begin{aligned} p(\text{SINR}_{Si} > \gamma | R_{Si} = r) &= p \left(\frac{P_S h_{Si} r^{-\alpha}}{I_{Si}} > \gamma \right) \\ &= \mathbb{E}_{I_{Si}} \left[p \left(h_{Si} > \frac{\gamma I_{Si}}{P_S r^{-\alpha}} \right) | I_{Si} \right] \\ &\stackrel{(a)}{=} \mathbb{E}_{I_{Si}} \left[e^{-\frac{\gamma I_{Si}}{P_S r^{-\alpha}}} | I_{Si} \right] \\ &= \int_0^{\infty} e^{-\frac{\gamma u}{P_S r^{-\alpha}}} f_{I_{Si}}(u) du \\ &= \mathcal{L}_{I_{Si}} \left(\frac{\gamma}{P_S r^{-\alpha}} \right) \\ &\stackrel{(b)}{=} \mathcal{L}_{I_{SiM}} \left(\frac{\gamma}{P_S r^{-\alpha}} \right) \mathcal{L}_{I_{SiSi}} \left(\frac{\gamma}{P_S r^{-\alpha}} \right) \\ &\quad \cdot \mathcal{L}_{I_{SiSi'}} \left(\frac{\gamma}{P_S r^{-\alpha}} \right) \quad (16) \end{aligned}$$

其中, 步骤(a)是根据 $h_M \sim \exp(1)$, 步骤(b)则是根据拉普拉斯变换的卷积定理。

式(15)中的 $\mathcal{L}_{I_{SiM}} \left(\frac{\gamma}{P_S r^{-\alpha}} \right)$ 的解析表达式为

$$\begin{aligned} \mathcal{L}_{I_{SiM}} \left(\frac{\gamma}{P_S r^{-\alpha}} \right) &= \mathbb{E}_{I_{SiM}} \left[\exp \left(-\frac{\gamma I_{SiM}}{P_S r^{-\alpha}} \right) | I_{SiM} \right] \\ &= \mathbb{E}_{\phi_M, h_M} \left[\exp \left(-\frac{\gamma}{P_S r^{-\alpha}} \left(\sum_{j \in \phi_M} P_M h_{Mj} R_{Mj}^{-\alpha} \right) \right) \right] \\ &\stackrel{(a)}{=} \mathbb{E}_{\phi_M} \left[\prod_{j \in \phi_M} \frac{P_S r^{-\alpha}}{P_S r^{-\alpha} + \gamma P_M R_{Mj}^{-\alpha}} \right] \\ &\stackrel{(b)}{=} \exp \left(2\pi \lambda_M \int_0^{\infty} \left(\frac{P_S r^{-\alpha}}{P_S r^{-\alpha} + \gamma P_M x^{-\alpha}} - 1 \right) x dx \right) \\ &= \exp \left(\pi \lambda_M \int_0^{\infty} \left(\frac{P_S r^{-\alpha}}{P_S r^{-\alpha} + \gamma P_M u^{-\frac{\alpha}{2}}} - 1 \right) du \right) \\ &\stackrel{(c)}{=} \exp \left(-\frac{\pi^2}{2} \lambda_M \sqrt{\frac{\gamma P_M}{P_S}} r^2 \right) \quad (17) \end{aligned}$$

其中, 步骤(a)是根据 $h_M \sim \exp(1)$, 步骤(b)是根据PPP的概率生成函数^[14], 步骤(c)则是根据前文 $\alpha = 4$ 的假设。

与式(17)同理, $\mathcal{L}_{I_{SiSi}} \left(\frac{\gamma}{P_S r^{-\alpha}} \right), \mathcal{L}_{I_{SiSi'}} \left(\frac{\gamma}{P_S r^{-\alpha}} \right)$ 可分别表示为

$$\begin{aligned}
 & \mathcal{L}_{I_{SiSi}} \left(\frac{\gamma}{P_S r^{-\alpha}} \right) \\
 &= \mathbb{E}_{\phi_{Si}} \left[\prod_{j \in \phi_{Si}} \frac{r^{-\alpha}}{r^{-\alpha} + \gamma R_{Sij}^{-\alpha}} \right] \\
 &= \exp \left(2\pi q_i \lambda_S \int_r^\infty \left(\frac{r^{-\alpha}}{r^{-\alpha} + \gamma x^{-\alpha}} - 1 \right) x dx \right) \\
 &= \exp \left(-\pi q_i \lambda_S \sqrt{\gamma} r^2 \left(\frac{\pi}{2} - \arctan \left(\frac{1}{\sqrt{\gamma} r^2} \right) \right) \right) \quad (18)
 \end{aligned}$$

$$\begin{aligned}
 & \mathcal{L}_{I_{SiSi}'} \left(\frac{\gamma}{P_S r^{-\alpha}} \right) \\
 &= \mathbb{E}_{\phi_{Si}'} \left[\prod_{j \in \phi_{Si}'} \frac{P_S r^{-\alpha}}{P_S r^{-\alpha} + \gamma P_S R_{Sij}^{-\alpha}} \right] \\
 &= \exp \left(2\pi (1 - q_i) \lambda_S \int_0^\infty \left(\frac{r^{-\alpha}}{r^{-\alpha} + \gamma x^{-\alpha}} - 1 \right) x dx \right) \\
 &= \exp \left(-\frac{\pi^2}{2} (1 - q_i) \lambda_S \sqrt{\gamma} r^2 \right) \quad (19)
 \end{aligned}$$

3.3.2 宏基站中断概率

与式(15)同理，用户通过宏基站获取文件*i*时的中断概率 $\mathbb{P}_{out, Mi}$ 为

$$\begin{aligned}
 \mathbb{P}_{out, Mi} &= \mathbb{E}_{R_{Mi}} [p(\text{SINR}_{Mi} < \gamma | R_{Mi} = r)] \\
 &= \int_0^\infty \left(1 - L_{I_{MiM}} \left(\frac{\gamma}{P_M r^{-\alpha}} \right) L_{I_{MiSi}} \left(\frac{\gamma}{P_M r^{-\alpha}} \right) \right. \\
 &\quad \left. \cdot L_{I_{MiSi}'} \left(\frac{\gamma}{P_M r^{-\alpha}} \right) \right) 2\pi \lambda_M r e^{-\lambda_M \pi r^2} dr \quad (20)
 \end{aligned}$$

其中， $\mathcal{L}_{I_{MiM}} \left(\frac{\gamma}{P_M r^{-\alpha}} \right)$ ， $\mathcal{L}_{I_{MiSi}} \left(\frac{\gamma}{P_M r^{-\alpha}} \right)$ ， $\mathcal{L}_{I_{MiSi}'} \left(\frac{\gamma}{P_M r^{-\alpha}} \right)$ 分别代表 I_{MiM} ， I_{MiSi} ， I_{MiSi}' 的概率密度函数在参数 $\frac{\gamma}{P_M r^{-\alpha}}$ 下的拉普拉斯变换，解析表

达式为

$$\begin{aligned}
 & \mathcal{L}_{I_{MiM}} \left(\frac{\gamma}{P_M r^{-\alpha}} \right) \\
 &= \mathbb{E}_{\phi_M \setminus M0} \left[\prod_{j \in \phi_M \setminus M0} \frac{r^{-\alpha}}{r^{-\alpha} + \gamma R_{Mj}^{-\alpha}} \right] \\
 &= \exp \left(\frac{1}{\sqrt{\gamma} r^2} \left(\frac{\pi}{2} - \arctan \left(\frac{1}{\sqrt{\gamma} r^2} \right) \right) \right) \quad (21)
 \end{aligned}$$

$$\begin{aligned}
 & \mathcal{L}_{I_{MiSi}} \left(\frac{\gamma}{P_M r^{-\alpha}} \right) \\
 &= \mathbb{E}_{\phi_{Si}} \left[\prod_{j \in \phi_{Si}} \frac{P_M r^{-\alpha}}{P_M r^{-\alpha} + \gamma P_S R_{Sij}^{-\alpha}} \right] \\
 &= \exp \left(-\pi q_i \lambda_S \sqrt{\frac{\gamma P_S}{P_M}} r^2 \left(\frac{\pi}{2} - \arctan \left(\frac{R_{\max}}{\sqrt{\frac{\gamma P_S r^4}{P_M}}} \right) \right) \right) \quad (22)
 \end{aligned}$$

$$\begin{aligned}
 & \mathcal{L}_{I_{MiSi}'} \left(\frac{\gamma}{P_M r^{-\alpha}} \right) \\
 &= \mathbb{E}_{\phi_{Si}'} \left[\prod_{j \in \phi_{Si}'} \frac{P_M r^{-\alpha}}{P_M r^{-\alpha} + \gamma P_S R_{Sij}^{-\alpha}} \right] \\
 &= \exp \left(-\pi \lambda_M \sqrt{\gamma} r^2 \left(\frac{\pi}{2} - \arctan \left(\frac{1}{\sqrt{\gamma} r} \right) \right) \right) \quad (23)
 \end{aligned}$$

至此，可以通过式(14)获得平均中断概率。

4 仿真分析

本节的仿真基于一个5 km×5 km大小的异构蜂窝网络，在该区域内进行两次独立的PPP撒点，并通过蒙特卡罗仿真模拟10000次用户的数据请求，仿真参数为 $\gamma=20$ dBm， $P_M=50$ dBm， $P_S=30$ dBm， $N=100$ ， $R_{\max}=30$ m， $\lambda_M=10$ (km²)⁻¹。

图3所示为偏斜参数与回程带宽释放率关系图，其中 $\lambda_S=1000$ (km²)⁻¹， $R=70$ ，通过与无缓存的方案对比可知，使用微基站缓存的方案可以大大降低回程链路的带宽压力，其中PMCP的带宽释放率最高，根据文献[15]， δ 现实中通常在0.5~1.0之间取值，本文取 $\delta=0.6$ 处作为参考点，此处PMCP的带宽释放率为88.30%，相较于UCP和PCP分别提高了2.12%和8.41%，可知PMCP在现实情况下相较于对比策略而言更具有优势。

图4所示为偏斜参数与平均中断概率关系图，其中 $\lambda_S=1000$ (km²)⁻¹， $R=70$ ，由图4可见仿真结果和此前的理论分析结果相吻合。注意到无缓存

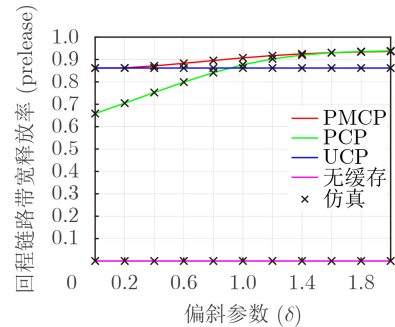


图3 偏斜参数与回程带宽释放率关系图

策略时, 中断概率不随偏斜参数变化, 这是因为根据式(15), 当微基站无缓存部署时, 即 $q_i = 0$ ($i = 1, 2, \dots, N$) 时, $\mathbb{P}_{out, S_i} = 1$ ($i = 1, 2, \dots, N$), 且用户与宏基站的连接和用户请求的文件无关, 因此 SINR 不会随 δ 变化, \mathbb{P}_{out, M_i} 保持不变, 因此平均中断概率保持不变。

当使用UCP时, 无论请求哪个文件, 发生中断的概率都是相同的, 因此代表UCP平均中断概率的蓝色曲线同样不随 δ 变化。

相较于UCP, $\delta = 0$ 时, $q_{i, PMCP} = R/N$ ($i = 1, 2, \dots, N$), 此时PMCP等价于UCP, 因此在 $\delta = 0$ 处二者的可靠性相同, 随着 δ 上升, 文件的被请求概率差异性增加, 此时PMCP的性能逐渐优于UCP; 相较于PCP, 因为高低排名文件在 δ 处于 (0,1) 区间时的被请求概率差距不大, 且PMCP在请求低排名文件时可以获得比PCP更低的中断概率, 因此PMCP的平均中断概率在这一区间内低于PCP。取 $\delta = 0.6$ 处作为参考点, 此处PMCP的平均中断概率为38.97%, 相较于UCP和PCP分别降低了2.63%和1.51%, 由此可知PMCP在现实情况下相较于对比策略而言更具有优势。

图5所示为微基站密度 λ_s 和平均中断概率关系图, 其中 $\delta = 0.6$, $R = 70$ 可以看到当使用不同的缓存策略时, 平均中断概率首先随 λ_s 上升, 这是因为当 $\lambda_s = 0$ 时, 不存在由微基站产生的干扰, 因此在 λ_s 较小时, 由于微基站产生的干扰不断增加, SINR

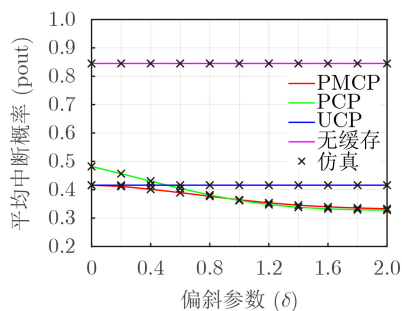


图4 偏斜参数与平均中断概率关系图

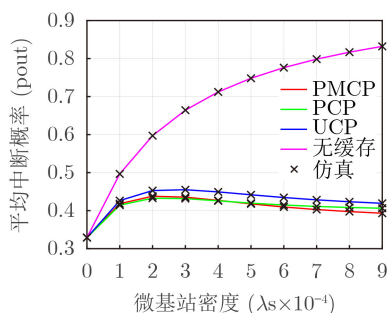


图5 微基站密度与平均中断概率关系图

降低, 中断概率不断上升。随着 λ_s 的增长, 平均中断概率由上升转为下降, 这是因为用户同与之连接的微基站的平均距离降低, 使得接收信号的功率提升, SINR 不断上升, 而当密度继续增大时, 中断概率仍比 $\lambda_s = 0$ 时大, 说明采用缓存异构网络会在一定程度上降低传输的可靠性以换取回程链路的带宽压力释放。与以上不同的是, 当没有部署缓存时, 中断概率随 λ_s 持续上升, 这是因为此时仅有由微基站产生的干扰随着基站数量的增加而上升, 造成 SINR 的持续下降。

图6所示为存储容量数据库容量比与平均中断概率关系图, 其中 $\lambda_s = 1000$ (km^2)⁻¹, $\delta = 0.6$, 可以发现随着存储容量不断上升, 缓存命中率不断上升, 中断概率不断下降。当存储容量和数据库容量相等时, 所有文件都被缓存, 缓存策略失去意义, 所有曲线交于一点。

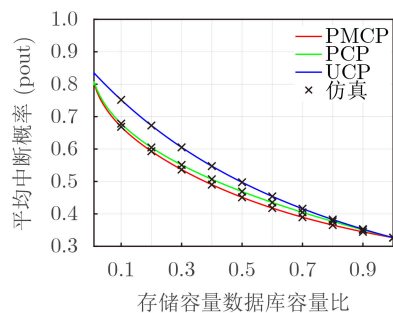


图6 存储容量数据库容量比与平均中断概率关系图

5 结束语

本文提出一种基于缓存异构网络的流行度匹配缓存策略。通过将基站的平面位置建模为多个独立的泊松点过程, 推导获得了用户请求的平均中断概率。仿真结果表明所提出策略在 $\delta = 0.6$ 时能够获得低于UCP 2.63%和低于PCP 1.51%的平均中断概率。此外, 本文通过分析微基站密度和平均中断概率的关系, 得出了部署密集微基站会在损失小部分传输可靠性的情况下大幅降低回程链路的带宽压力的结论。最终, 本文通过分析存储容量和平均中断概率的关系, 得出了存储容量与传输可靠性正相关的结论。本文强调了在缓存异构网络下, 新的缓存策略比传统策略的性能更好, 并强调了一个事实, 即使用高密度网络和差异性更大的缓存策略可以获得更好的性能。

参考文献

- [1] MAO Yuji, YOU Changsheng, ZHANG Jun, et al. A survey on mobile edge computing: The communication perspective[J]. *IEEE Communications Surveys & Tutorials*,

- 2017, 19(4): 2322–2358. doi: [10.1109/COMST.2017.2745201](https://doi.org/10.1109/COMST.2017.2745201).
- [2] BHARATH B N, NAGANANDA K G, and POOR H V. A learning-based approach to caching in heterogenous small cell networks[J]. *IEEE Transactions on Communications*, 2016, 64(4): 1674–1686. doi: [10.1109/TCOMM.2016.2536728](https://doi.org/10.1109/TCOMM.2016.2536728).
- [3] SONG J, SONG H, and CHOI W. Optimal content placement for wireless femto-caching network[J]. *IEEE Transactions on Wireless Communications*, 2017, 16(7): 4433–4444. doi: [10.1109/TWC.2017.2698447](https://doi.org/10.1109/TWC.2017.2698447).
- [4] SONG J and CHOI W. Minimum cache size and backhaul capacity for cache-enabled small cell networks[J]. *IEEE Wireless Communications Letters*, 2018, 7(4): 490–493. doi: [10.1109/LWC.2017.2787765](https://doi.org/10.1109/LWC.2017.2787765).
- [5] KRISHNENDU S, BHARATH B N, and BHATIA V. Cache enabled cellular network: Algorithm for cache placement and guarantees[J]. *IEEE Wireless Communications Letters*, 2019, 8(6): 1550–1554. doi: [10.1109/LWC.2019.2926726](https://doi.org/10.1109/LWC.2019.2926726).
- [6] WANG Hongman, Li Yingxue, ZHAO Xiaoqi, *et al.* An algorithm based on markov chain to improve edge cache hit ratio for blockchain-enabled IoT[J]. *China Communications*, 2020, 17(9): 66–76. doi: [10.23919/JCC.2020.09.006](https://doi.org/10.23919/JCC.2020.09.006).
- [7] TAMOOR-UL-HASSAN S, BENNIS M, NARDELLI P H J, *et al.* Caching in wireless small cell networks: A storage-bandwidth tradeoff[J]. *IEEE Communications Letters*, 2016, 20(6): 1175–1178. doi: [10.1109/LCOMM.2016.2543698](https://doi.org/10.1109/LCOMM.2016.2543698).
- [8] LIU Dong and YANG Chenyang. Caching policy toward maximal success probability and area spectral efficiency of cache-enabled HetNets[J]. *IEEE Transactions on Communications*, 2017, 65(6): 2699–2714. doi: [10.1109/TCOMM.2017.2680447](https://doi.org/10.1109/TCOMM.2017.2680447).
- [9] ZHANG Tiankui, FANG Xinyuan, LIU Yuanwei, *et al.* D2D-enabled mobile user edge caching: A multi-winner auction approach[J]. *IEEE Transactions on Vehicular Technology*, 2019, 68(12): 12314–12328. doi: [10.1109/TVT.2019.2947334](https://doi.org/10.1109/TVT.2019.2947334).
- [10] 杨静, 李金科. 带有特征感知的D2D内容缓存策略[J]. *电子与信息学报*, 2020, 42(9): 2201–2207. doi: [10.11999/JEIT190691](https://doi.org/10.11999/JEIT190691).
- YANG Jing and LI Jinke. Feature-aware D2D content caching strategy[J]. *Journal of Electronics & Information Technology*, 2020, 42(9): 2201–2207. doi: [10.11999/JEIT190691](https://doi.org/10.11999/JEIT190691).
- [11] HUA Haojiang and CHU Xiaoli. Content caching policy with edge caching user classification in fog radio access networks[C]. 2021 IEEE Wireless Communications and Networking Conference, Nanjing, China: IEEE, 2021: 1–7. doi: [10.1109/WCNC49053.2021.9417284](https://doi.org/10.1109/WCNC49053.2021.9417284).
- [12] BŁASZCZYSZYN B, KARRAY M K, and KEELER H P. Using Poisson processes to model lattice cellular networks[C]. IEEE INFOCOM, Turin, Italy: IEEE, 2013: 773–781. doi: [10.1109/INFCOM.2013.6566864](https://doi.org/10.1109/INFCOM.2013.6566864).
- [13] JO H S, SANG Y J, XIA Ping, *et al.* Heterogeneous cellular networks with flexible cell association: A comprehensive downlink SINR analysis[J]. *IEEE Transactions on Wireless Communications*, 2012, 11(10): 3484–3495. doi: [10.1109/TWC.2012.081612.111361](https://doi.org/10.1109/TWC.2012.081612.111361).
- [14] ANDREWS J G, BACCELLI F, and GANTI R K. A tractable approach to coverage and rate in cellular networks[J]. *IEEE Transactions on Communications*, 2011, 59(11): 3122–3134. doi: [10.1109/TCOMM.2011.100411.100541](https://doi.org/10.1109/TCOMM.2011.100411.100541).
- [15] BRESLAU L, CAO Pei, FAN Li, *et al.* Web caching and Zipf-like distributions: Evidence and implications[C]. Eighteenth Annual Joint Conference of the IEEE Computer and Communications Societies, New York, USA: IEEE, 1999: 126–134.
- 刘浩洋：男，1993年生，博士生，研究方向为异构边缘计算与缓存。
王 钢：男，1962年生，教授，博士生导师，研究方向为数据通信、物理层网络编码、通信网理论与技术。
杨文超：男，1976年生，副教授，研究方向为数据通信、无线网络架构、软件无线电。
王金龙：男，1988年生，博士，研究方向为无线携能通信技术。
许 尧：男，1993年生，博士生，研究方向为非正交多址接入技术。
赵东来：男，1991年生，博士生，研究方向为超密集组网技术。

责任编辑：余 蓉