Performance Analysis for Caching in Multi-tier IoT Networks with Joint Transmission

Tianming Feng[1], Shuo Shi[1(✉)], Xuemai Gu[1], and Zhenyu Xu[2]

[1] Harbin Institute of Technology, Harbin 150001, China
{fengtianming,crcss,guxuemai}@hit.edu.cn
[2] Huizhou Engineering Vocational College, Huizhou 516023, China
hitusa@126.com

Abstract. The rapid growth of the number of IoT devices in the network has brought huge traffic pressure to the network. Caching at the edge has been regarded as a promising technique to solve this problem. However, how to further improve the successful transmission probability (STP) in cache-enabled multi-tier IoT networks (CMINs) is still an open issue. To this end, this paper proposes a base station (BS) joint transmission scheme in CMIN where the nearest BS that stores the requested files in each tier is selected to cooperatively serve the typical UE. Based on the proposed scheme, we derive an integral expression for the STP, and optimize the content caching strategy for a two-tier network case. The gradient projection method is used to solve the optimization problem, and a locally optimal caching strategy (LCS) is obtained. Numerical simulations show that the LCS achieves a significant gain in STP over three comparative baseline strategies.

Keywords: IoT networks · Cache-enabled networks · Joint transmission

1 Introduction

The rapidly growing number of mobile IoT devices on the network has led to explosive growth in the demand for mobile data traffic. Deploying multi-tier edge base station (BS) to form a heterogeneous network (HetNet) has become an effective solution to meet these demands [1]. However, densely deployed BSs will inevitably cause high inter-cell interference. As one of the downlink coordinated multipoint transmission (CoMP) transmission techniques, joint transmission (JT) [2], where a user is simultaneously served by several BSs, can effectively mitigate the interference. Based on the recent observation that a large portion of the traffic is caused by repeatedly downloading a few popular contents, caching the popular contents at the edge of the IoT network, such as edge BSs, IoT devices, can greatly reduce the traffic burden of IoT network.

X. Wang et al. (Eds.): 6GN 2020, LNICST 337, pp. 345–356, 2020.
https://doi.org/10.1007/978-3-030-63941-9_26

The concept of caching technique has been extensively studied in the recent years. The authors of [3] design an optimal tier-level caching strategy for HetNets. This work is extended to a K-tier multi-antenna multi-user HetNets scenario by [4]. In [5], the authors design an optimal caching strategy in heterogeneous IoT networks to improve the offloading rate for backhaul links. [6] jointly optimizes content placement and activation densities of BSs of different tiers to reduce the energy consumption for heterogeneous industrial IoT networks. [7] investigates the influence of JT on caching strategy and obtains a locally optimal solution in the general case and a globally optimal solution in some special cases. Based on this work, the authors of [8] study the advantage of introducing local channel state information into JT and design an algorithm for locally optimal caching strategy in IoT networks. [9] studies the tradeoff between the content diversity gain and the cooperative gain and proposes an optimal caching strategy to balance the tradeoff.

However, all the aforementioned works that jointly consider JT and caching technique in multi-tier networks only limit the content caching strategy design to the single-tier BSs. That is, only one tier of BSs in the network have cache capacity, the BSs of other tiers do not have the ability to cache. Moreover, the BS cooperation is just limited to the BSs from the same tier. In this paper, the content caching strategy with BS JT is studied in a cache-enabled multi-tier IoT network (CMIN). Different from the existing literature, every tier of BSs in our model has cache capacity to store a limited number of contents. JT scheme is adopted in this work, under which the user is jointly served by the nearest BSs that cache the requested content from each tier. So that the BS cooperation is "vertical" (i.e., cross-tier) rather than "horizontal" (i.e., in-tier) considered in the existing literature. Based on the JT and caching model, integral expressions for the successful transmission probability (STP) are derived for an M-tier IoT network case and a special case of two-tier network. Then, a locally optimal caching strategy (LCS) for the two-tier IoT network is obtained by maximizing the STP of the network using the gradient projection method (GPM). Finally, simulation results demonstrate the LCS achieves significant gains in STP over several comparative baselines with BS JT.

2 System Model

2.1 Network and Caching Model

We consider a downlink large-scale CMIN consisting of M tiers of BSs, where the set of BS tiers is denoted as $\mathcal{M} = \{1, 2, \cdots, M\}$. The BSs in the m-th tier are spatially distributed as an independent homogeneous Poisson point process (PPP) Φ_m, with density λ_m and transmission power P_m, $m \in \mathcal{M}$. The locations of the BSs of all tiers are denoted by Φ, i.e., $\Phi = \bigcup_{m\in\mathcal{M}} \Phi_m$. We focus on the analysis of a typical user equipment (UE) u_0, which is assumed to locate at the origin without loss of generality. When the typical UE is served by a BS located at $x_m \in \mathbb{R}^2$ from the m-th tier, the signal power it receives can be expressed as $P_m \|x_m\|^{-\alpha_m} |h_{x_m}|^2$, where $\|x_m\|^{-\alpha_m}$ and $|h_{x_m}|^2$ correspond to

large-scale fading and small-scale fading, respectively; $\alpha_m > 2$ denotes the path-loss exponent; h_{x_m} models the Rayleigh fading between the typical UE and the BS, i.e., $h_{x_m} \overset{d}{\sim} \mathcal{CN}(0,1)$ or channel power $|h_{x_m}|^2 \overset{d}{\sim} \text{Exp}(1)$ [1].

The CMIN considered in this paper has a content database denoted by $\mathcal{N} = \{1, 2, \cdots, N\}$. There are $N \geq 1$ files in the database, whose sizes are assumed to be equal and normalized to one for ease of analysis. The popularity of file n is denoted as $a_n \in [0, 1]$. Here, we assume the popularity distribution follows a Zipf distribution [3], i.e.,

$$a_n = \frac{n^{-\gamma}}{\sum_{n \in \mathcal{N}} n^{-\gamma}}, \quad \text{for } \forall n \in \mathcal{N}, \tag{1}$$

where $\gamma \geq 0$ is the Zipf exponent. Each UE randomly requests one file according to the file popularity distribution in one time slot.

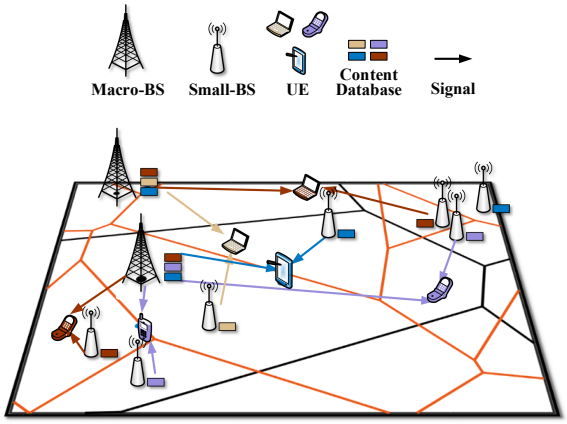

Fig. 1. Illustration of a two-tier cache-enabled IoT network with joint transmission scheme. There are four different files in the database, which are indicated by four different colors. The colors of the UEs represent the files they request. In this scenario, $N = 4$, $C_1 = 3$, $C_2 = 1$.

In tier m, each BS has a limited cache size C_m which can store at most C_m different files out of N. In this paper, we consider a *random caching* scheme [3,10], in which the file n is stored in a BS from tier m randomly with probability $t_{mn} \in [0, 1]$, which is called the *caching probability*. Denote by $\mathbf{T}_m = [t_{m1}, \cdots, t_{mN}]$ the caching probability vector of all N files for tier m, which is identical for all BSs in tier m. Given $\mathbf{T}_m$, one BS from tier m can randomly choose C_m files to store, using the probabilistic content caching policy proposed in [10]. Let $\mathbf{t}_n = [t_{1n}, \cdots, t_{Mn}]^T$ be the caching distribution of file n, and $\mathbf{T} = [\mathbf{t}_1, \cdots, \mathbf{t}_N]$ be the caching probability matrix for all files and all BS tiers. The rows and columns of $\mathbf{T}$ correspond to $\mathbf{T}_m$ and $\mathbf{t}_n$ and further correspond to different tiers and different files, respectively. Due to the limited cache size, we have [3,10]:

$$0 \le t_{mn} \le 1, \quad \forall m \in \mathcal{M}, n \in \mathcal{N} \tag{2}$$

$$\textstyle\sum_{n\in\mathcal{N}} t_{mn} \le C_m, \forall m \in \mathcal{M} \tag{3}$$

Note that the BSs that cache file n from the m-th tier can also be denoted as a homogeneous PPP $\Phi_{m,n}$, with density $\lambda_{m,n} = t_{mn}\lambda_m$, and the remaining BSs that do not cache file n from tier m are denoted as $\Phi_{m,-n}$, with density $\lambda_{m,-n} = (1 - t_{mn})\lambda_m$, according to the thinning theorem for PPP.

2.2 Joint Transmission

Assume that the typical UE u_0 requests file n in current time slot. We adopt a joint transmission scheme in which the typical UE u_0 is served jointly by the *nearest* BSs that cache the requested file from each tier, as shown in Fig. 1. Therefore, the cooperative BS set denoted by $\mathcal{C}_n$ can be defined as

$$\mathcal{C}_n \triangleq \{\{x_{k,n,0}\} | \, x_{k,n,0} = \arg\max_{x\in\Phi_{k,n}} \|x\|^{-\alpha_k}, \forall k \in \mathcal{K}\}, \tag{4}$$

where $x_{k,n,0}$ represents the nearest BS to u_0 from tier k that stores file n; $\mathcal{K} \subseteq \mathcal{M}$ is the set of indexes of tiers to which each cooperative BS belongs, and let $K \triangleq |\mathcal{K}|$ denote the number of elements in $\mathcal{K}$. Note that due to the effect of $\mathbf{T}$, there may be a tier m so that all BSs from this tier do not store file n if $t_{mn} = 0$, i.e., the BSs from this tier do not participate in JT. Therefore, there are up to $K \le M$ BSs to jointly transmit file n to u_0. In order to focus on the performance analysis of the cache-enabled HetNet, we assume that all the BSs from every tiers are not equipped with backhaul links, which means when the file n is not stored in any BS because of the limited BS storage, i.e., $t_{mn} = 0$ for $\forall m \in \mathcal{M}$, the file request can not be satisfied, and a transmission failure occurs, which is referred to as a *cache miss case*. The sum of the desired non-coherent signal yields a received power boost to improve the received signal-to-interference-plus-noise ratio (SINR). Based on the JT scheme described above, the received signal at u_0 when requesting file n can be written as

$$y_n = \underbrace{\sum_{x\in\mathcal{C}_n} P_{\nu(x)}^{1/2} \|x\|^{-\alpha_{\nu(x)}/2} h_x X_n}_{\text{desired signal}} + \underbrace{\sum_{x\in\Phi\setminus\mathcal{C}_n} P_{\nu(x)}^{1/2} \|x\|^{-\alpha_{\nu(x)}/2} h_x X_x}_{\text{interference}} + Z, \tag{5}$$

where $\nu(x)$ returns the index of tier to which a BS located at x belongs, i.e., $\nu(x) = m$ iff $x \in \Phi_m$; X denotes the symbol jointly sent by the cooperative BSs, which is the desired symbol of u_0; X_x denotes the symbol sent by the BSs located outside $\mathcal{C}_n$, which is regarded as interference symbol to u_0; $Z \overset{d}{\sim} \mathcal{CN}(0, N_0)$ models the background thermal noise.

2.3 Performance Metric

Since the strength of interference in HetNets is usually much stronger than that of the thermal noise, it is reasonable to neglect the impact of the thermal noise

and just consider the interference-limited network, i.e., $N_0 = 0$. The SIR of the typical UE u_0 requesting file n is given by

$$\mathrm{SIR}_n = \frac{\left| \sum\limits_{x \in \mathcal{C}_n} P_{\nu(x)}^{1/2} \|x\|^{-\alpha_{\nu(x)}/2} h_x \right|^2}{\sum\limits_{x \in \Phi \setminus \mathcal{C}_n} P_{\nu(x)} \|x\|^{-\alpha_{\nu(x)}} |h_x|^2} = \frac{\left| \sum_{k \in \mathcal{K}} P_k^{1/2} \|x_{k,n,0}\|^{-\alpha_k/2} h_{k,n,0} \right|^2}{\sum\limits_{m=1}^{M} P_m I_m}, \quad (6)$$

where $I_m = \sum_{x \in \Phi_m \setminus \{x_{m,n,0}\}} \|x\|^{-\alpha_m} |h_x|^2$ is the interference caused by all BSs from tier m normalized by the transmission power P_m. In this paper, we employ the STP as the system performance metric. When u_0 requests file n, the transmission will succeed if the received data rate at the UE exceeds a given threshold r [bps/Hz], i.e., $\log_2(1 + \mathrm{SIR}_n) \geq r$. Furthermore, the STP is defined as

$$q(\mathbf{T}) \triangleq \Pr\left[\mathrm{SIR} \geq \tau\right] = \sum_{n \in \mathcal{N}} a_n q_n(\mathbf{t}_n), \quad (7)$$

$q_n(\mathbf{t}_n) \triangleq \Pr\left[\mathrm{SIR}_n \geq \tau\right]$ denotes the STP when u_0 requests file n, and the second equality holds due to the total probability theorem. For notational simplicity, we define the ratios of transmission power and BS density as $P_{ij} \triangleq P_i / P_j$ and $\hat{\lambda}_{ij} \triangleq \lambda_i / \lambda_j$, respectively.

3 Analysis of Performance Metric

In this section, we first derive the expression of STP for a given caching probability matrix $\mathbf{T}$. Then verify the obtained expression using Monte Carlo simulation.

For ease of notation, we first have the following definitions:

$$F(\alpha, x) = {}_2F_1\left(-\frac{2}{\alpha}, 1; 1 - \frac{2}{\alpha}; -x\right) - 1, \quad (8)$$

where ${}_2F_1\left(a, b; c; d\right)$ denotes the Gauss hypergeometric function.

Theorem 1. *The STP for the joint transmission scheme considered in this work is given by*

$$q(\mathbf{T}) = \sum_{n \in \mathcal{N}} a_n q_n(\mathbf{t}_n), \quad (9)$$

where $q_n(\mathbf{t}_n)$ is expressed as

$$q_n(\mathbf{t}_n) = \int_0^\infty \int_0^\infty \cdots \int_0^\infty q_{n,\mathbf{R}_0}(\mathbf{t}_n, \mathbf{r}) \prod_{k \in \mathcal{K}} f_{R_{k,0}}(r_k) \, \mathrm{d}\mathbf{r}. \quad (10)$$

$\mathbf{R}_0 = [R_{1,0}, \cdots, R_{k,0}, \cdots], k \in \mathcal{K}$ is the distance vector between the typical UE and its serving BSs, i.e., $R_{k,0} = \|x_{k,n,0}\|$, and $\mathbf{r}$ is a realization of $\mathbf{R}_0$; $f_{R_{k,0}}(r_k)$ is the probability density function (PDF) of $R_{k,0}$, which can be given by

$$f_{R_{k,0}}(r_k) = 2\pi \lambda_k t_{kn} r_k e^{-\pi \lambda_k t_{kn} r_k^2}. \quad (11)$$

$q_{n,\mathbf{R}_0}(\mathbf{t}_n, \mathbf{r})$ is the STP conditioned on $\mathbf{R}_0 = \mathbf{r}$ and is given by

$$q_{n,\mathrm{R}_0}(\mathbf{t}_n, \mathbf{r}) = \prod_{m\in\mathcal{K}} \exp\left(-F\left(\alpha_m, \frac{\tau}{\sum_{k\in\mathcal{K}} P_{km}\frac{r_m^{\alpha_m}}{r_k^{\alpha_k}}}\right)\pi\lambda_m t_{mn} r_m^2\right) \times \prod_{m=1}^{M} \exp\left(-\frac{\pi\lambda_m(1-t_{mn})}{\mathrm{sinc}(2/\alpha_m)}\left(\frac{\tau}{\sum_{k\in\mathcal{K}} P_{km} r_k^{-\alpha_k}}\right)^{\frac{2}{\alpha_m}}\right), \tag{12}$$

where, $\mathrm{sinc}(x) = \frac{\sin(\pi x)}{\pi x}$.

Proof. According to the total probability theorem, we have (9). Next, we calculate $q_n(\mathbf{t}_n)$. As illustrated before, the interference can be categorized into two types: 1) the interference caused by the BSs in $\Phi_{m,n}$ which are farther away from u_0 than the serving BS from tier m; 2) the interference caused by the BSs in $\Phi_{m,-n}$ which may be closer to u_0 than the serving BS from tier m. For the second case, if there are no BSs that store file n, i.e., $t_{mn} = 0$, all the BSs in the tier m are interfering BSs. Therefore, the interference from tier m can be rewritten as $I_m = \mathbb{1}(t_{mn} > 0)I_{m,n} + I_{m,-n}$, where $\mathbb{1}(\bullet)$ is the indicator function; $I_{m,n} \triangleq \sum_{x\in\Phi_{m,n}\setminus\{x_{m,n,0}\}} \|x\|^{-\alpha_m} |h_x|^2$ and $I_{m,-n} \triangleq \sum_{x\in\Phi_{m,-n}} \|x\|^{-\alpha_m} |h_x|^2$. The received signal power of u_0 is $S = \left|\sum_{k\in\mathcal{K}} P_k^{1/2} \|x_{k,n,0}\|^{-\alpha_k/2} h_{k,n,0}\right|^2$. Conditioning on $\mathbf{R}_0 = \mathbf{r}$, we can obtain the conditional STP as

$$\begin{aligned} q_{n,\mathbf{R}_0}(\mathbf{t}_n, \mathbf{r}) &= \Pr\left[\mathrm{SIR}_n \geq \tau \middle| \mathbf{R}_0 = \mathbf{r}\right] \\ &= \mathbb{E}_{I_m}\left[\Pr\left[S \geq \tau \sum_{m=1}^{M} P_m I_m \middle| \mathbf{R}_0 = \mathbf{r}\right]\right] \\ &\overset{(a)}{=} \mathbb{E}_{I_{m,n}, I_{m,-n}}\left[e^{-\frac{\tau\sum_{m=1}^{M} P_m(\mathbb{1}(t_{mn}>0)I_{m,n} + I_{m,-n})}{\sum_{k\in\mathcal{K}} P_k r_k^{-\alpha_k}}}\right] \\ &\overset{(b)}{=} \prod_{m\in\mathcal{K}} \mathbb{E}_{I_{m,n}}\left[e^{-\frac{\tau P_m I_{m,n}}{\sum_{k\in\mathcal{K}} P_k r_k^{-\alpha_k}}}\right] \prod_{m=1}^{M} \mathbb{E}_{I_{m,-n}}\left[e^{-\frac{\tau P_m I_{m,-n}}{\sum_{k\in\mathcal{K}} P_k r_k^{-\alpha_k}}}\right] \\ &\triangleq \prod_{m\in\mathcal{K}} \mathcal{L}_{I_{m,n}}(sP_m, \mathbf{r}) \prod_{m=1}^{M} \mathcal{L}_{I_{m,-n}}(sP_m, \mathbf{r})\Big|_{s=\frac{\tau}{\sum_{k\in\mathcal{K}} P_k r_k^{-\alpha_k}}}, \end{aligned} \tag{13}$$

where (a) follows from that $S \overset{d}{\sim} \mathrm{Exp}\left(\frac{1}{\sum_{k\in\mathcal{K}} P_k r_k^{-\alpha_k}}\right)$; (b) is due to the independence of the homogeneous PPPs; $\mathcal{L}_{I_{m,n}}$ and $\mathcal{L}_{I_{m,-n}}$ represent the Laplace transforms of the interference $I_{m,n}$ and $I_{m,-n}$, respectively, which can be derived as follows

$$\begin{aligned}\mathcal{L}_{I_{m,n}}(sP_m,\mathbf{r}) &= \mathbb{E}_{\Phi_{m,n},|h_x|}\left[e^{-sP_m\sum_{x\in\Phi_{m,n}\setminus\{x_{m,n,0}\}}\|x\|^{-\alpha_m}|h_x|^2}\right]\\ &\overset{(a)}{=} \mathbb{E}_{\Phi_{m,n}}\left[\prod_{x\in\Phi_{m,n}\setminus\{x_{m,n,0}\}}\mathbb{E}_{|h_x|}\left[e^{-sP_m\|x\|^{-\alpha_m}|h_x|^2}\right]\right]\\ &\overset{(b)}{=} \mathbb{E}_{\Phi_{m,n}}\left[\prod_{x\in\Phi_{m,n}\setminus\{x_{m,n,0}\}}\frac{1}{1+sP_m\|x\|^{-\alpha_m}}\right]\\ &\overset{(c)}{=} e^{-2\pi\lambda_{m,n}\int_{r_m}^{\infty}\left(1-\frac{1}{1+sP_m r^{-\alpha_m}}\right)r\mathrm{d}r}\\ &= e^{-F\left(\alpha_m,\frac{sP_m}{r_m^{\alpha_m}}\right)\pi\lambda_m t_{mn} r_m^2},\end{aligned} \tag{14}$$

where (a) is due to the independence of the channels; (b) follows from $|h_x|^2 \overset{d}{\sim} \mathrm{Exp}(1)$; (c) is from the probability generating functional for a PPP and converting from Cartesian to polar coordinates. Similarly, we have

$$\begin{aligned}\mathcal{L}_{I_{m,-n}}(sP_m,\mathbf{r}) &= \mathbb{E}_{\Phi_{m,-n},|h_x|}\left[e^{-sP_m\sum_{x\in\Phi_{m,-n}}\|x\|^{-\alpha_m}|h_x|^2}\right]\\ &= e^{-2\pi\lambda_{m,-n}\int_0^{\infty}\left(1-\frac{1}{1+sP_m r^{-\alpha_m}}\right)r\mathrm{d}r}\\ &= e^{-\frac{\pi\lambda_m(1-t_{mn})}{\mathrm{sinc}(2/\alpha_m)}(sP_m)^{(2/\alpha_m)}}.\end{aligned} \tag{15}$$

Then, remove the condition $\mathbf{R}_0 = \mathbf{r}$, $q_n(\mathbf{t}_n)$ can be obtained as

$$q_n(\mathbf{t}_n) = \int_0^{\infty}\int_0^{\infty}\cdots\int_0^{\infty} q_{n,\mathbf{R}_0}(\mathbf{t}_n,\mathbf{r})\prod_{k\in\mathcal{K}} f_{R_{k,0}}(r_k)\,\mathrm{d}\mathbf{r}. \tag{16}$$

The proof of Theorem 1 is completed.

From Theorem 1, we can know that even though the expression of the STP is in integral form, the Gaussian hypergeometric function can be effectively calculated with a numerical method when the number of tiers of the network M is small. Figure 2 plots the STP $q(\mathbf{T})$ versus τ with and without JT, respectively. From Fig. 2, we can see that the derived analytical expression of STP with JT matches the corresponding Monte Carlo results perfectly. Besides, the STP of the system with BS JT has been greatly improved compared to the non-cooperation scheme, which confirms the effectiveness of the BS JT.

Next, a special case is considered, where the tier number $M = 2$.

Corollary 1 (STP for the Two-Tier HetNets). *For the two-tier network where $M = 2$, let $\alpha_1 = \alpha_2 = \alpha$, the STP is given by*

$$q^t(\mathbf{T}) = \sum_{n\in\mathcal{N}} a_n\tilde{q}_n(\mathbf{t}_n), \tag{17}$$

where

$$\tilde{q}_n(\mathbf{t}_n) = \int_{\mathbb{R}^{2+}} Q_{1,2}(u_1,u_2,t_{1n})Q_{2,1}(u_2,u_1,t_{2n})\mathrm{d}\mathbf{u}, \tag{18}$$

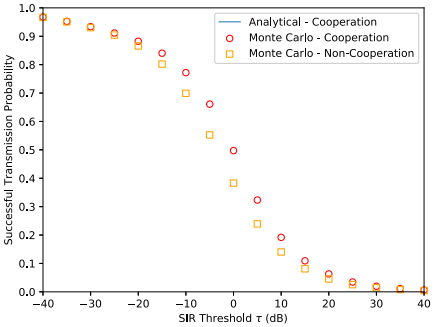

Fig. 2. STP $q(\mathbf{T})$ versus τ. $M = 3$, $N = 9$, $C_1 = 7$, $C_2 = 5$, $C_3 = 3$, $\gamma = 1$, $\alpha_1 = 4.3$, $\alpha_2 = 4$, $\alpha_3 = 3.8$, $\lambda_1 = 1/(200^2\pi)\,\mathrm{m}^{-2}$, $\lambda_2 = 5/(100^2\pi)\,\mathrm{m}^{-2}$, $\lambda_3 = 1/(50^2\pi)\,\mathrm{m}^{-2}$, $P_1 = 46\,\mathrm{dBm}$, $P_2 = 32\,\mathrm{dBm}$, $P_3 = 23\,\mathrm{dBm}$, $\mathbf{T}_1 = [1, 1, 1, 1, 0.8, 0.7, 0.6, 0.5, 0.4]$, $\mathbf{T}_2 = [1, 1, 0.8, 0.7, 0.6, 0.5, 0.4, 0, 0]$, $\mathbf{T}_3 = [0.8, 0.7, 0.6, 0.5, 0.4, 0, 0, 0, 0]$. In the Monte Carlo simulations, the BSs are deployed in a square area of $1,000 \times 1,000\,\mathrm{m}^2$, and the results are obtained by averaging over 10^5 independent realizations.

$$Q_{i,j}\left(x, y, z\right) = z \exp\left(-xz - zA_{i,j}(x, y) - (1 - z)B_{i,j}(x, y)\right), \tag{19}$$

$$A_{i,j}(x, y) = xF\left(\alpha, \frac{\tau}{1 + P_{ji}\left(\hat{\lambda}_{ji}\frac{x}{y}\right)^{\frac{\alpha}{2}}}\right), \tag{20}$$

$$B_{i,j}(x, y) = \frac{x}{\mathrm{sinc}(2/\alpha)}\left(\frac{\tau}{1 + P_{ji}\left(\hat{\lambda}_{ji}\frac{x}{y}\right)^{\frac{\alpha}{2}}}\right)^{\frac{2}{\alpha}}. \tag{21}$$

From Corollary 1, we can observe that the STP is not only influenced by the caching probability t_n, the path-loss exponent α, and the SIR threshold τ but also affected by the ratios of transmission power P_{ij} and BS density $\hat{\lambda}_{ij}$, respectively, when $M = 2$.

4 STP Maximization

In this section, we maximize the STP by optimizing the placement probability matrix $\mathbf{T}$ for the two-tier IoT network. The optimization problem for this case can be formulated as

Problem 1 (Maximization of STP for Two-Tier IoT network).

$$\max_{\mathbf{T}} q^t(\mathbf{T})$$
$$\text{s.t. } (2), (3),$$

where the objective function is given by Corollary 1. Let $\mathbf{T}^\star = [\mathbf{T}_1^{\star T}, \mathbf{T}_2^{\star T}]^T$ be the optimal solution of this problem. From Problem 1, we can observe that the constraint set is convex, whereas the convexity of the objection function is hard to determine because of its complicated expression. Therefore, a locally optimal solution can be obtained using the gradient projection method (GPM) [11], the procedure of which is summarized in Algorithm 1. In Step 3 of Algorithm 1, the stepsize $s(k)$ satisfies

Algorithm 1. Locally Optimal Solution to Problem 1

1: Initialize $\epsilon = 10^{-6}$, $k = 0$, $k_{\max} = 10^6$, and $t_{mn}(0) = \frac{C_m}{N}$, for $\forall n \in \mathcal{N}, m \in \{1, 2\}$.
2: **repeat**
3: For $\forall n \in \mathcal{N}, m \in \{1, 2\}$, compute $\bar{t}_{mn}(k+1) = t_{mn}(k) + s(k)\frac{\partial q^t(\mathbf{T}(k))}{\partial t_{mn}(k)}$, where $s(k)$ satisfies (22).
4: For $\forall n \in \mathcal{N}, m \in \{1, 2\}$, compute the projection $t_{mn}(k+1) = [\bar{t}_{mn}(k+1) - u_m^\star]_0^1$, where the scalar $u_m^\star$ satisfies $\sum_{n\in\mathcal{N}} [\bar{t}_{mn}(k+1) - u_m^\star]_0^1 = C_m$, $[x]_0^1 = \max\{\min\{1, x\}, 0\}$.
5: $k \leftarrow k + 1$.
6: **until** $|t_{mn}(k+1) - t_{mn}(k)| < \epsilon$, for $\forall n \in \mathcal{N}, m \in \{1, 2\}$ or $k > k_{\max}$.

$$\lim_{k\to\infty} s(k) = 0, \quad \sum_{k=0}^{\infty} s(k) = \infty, \tag{22}$$

and the partial derivative $\frac{\partial q^t(\mathbf{T}(k))}{\partial t_{mn}(k)} = a_n \frac{\partial \tilde{q}(\mathbf{t}_n(k))}{\partial t_{mn}(k)}$ is given by

$$\begin{aligned}\frac{\partial \tilde{q}_n(\mathbf{t}_n)}{\partial t_{in}} =& \frac{\tilde{q}_n^0(t_{1n}, t_{2n})}{t_{in}} + \int_{\mathbb{R}^{2+}} (-A_{i,j}(u_i, u_j) + B_{i,j}(u_i, u_j) - u_i) \\ & \times Q_{i,j}(u_i, u_j, t_{in}) Q_{j,i}(u_j, u_i, t_{jn}) \, \mathrm{d}\mathbf{u}, \qquad t_{1n} > 0, t_{2n} > 0,\end{aligned} \tag{23}$$

where $i, j \in \{1, 2\}$, $i \neq j$. By using Algorithm 1, a locally optimal caching strategy (LCS) can be obtained.

5 Numerical Results

In this section, some simulations are conducted to compare the LCS obtained by Algorithm 1 with three baseline strategies, i.e., MPC (most popular caching) [12], IIDC (i.i.d. caching) [13] and UDC (uniform distribution caching) [14]. Note that the three baselines also adopt a joint transmission scheme. We focus on a two-tier IoT network with randomly deployed macro BSs and small BSs. Unless otherwise noted, we set $M = 2$, $N = 100$, $C_1 = 30$, $C_2 = 20$, $\gamma = 1$, $\alpha_1 = \alpha_2 = 4$, $\lambda_1 = 5/(200^2\pi)\,\mathrm{m}^{-2}$, $\lambda_2 = 1/(50^2\pi)\,\mathrm{m}^{-2}$, $P_1 = 43\,\mathrm{dBm}$, $P_2 = 23\,\mathrm{dBm}$, $\tau = 0\,\mathrm{dB}$.

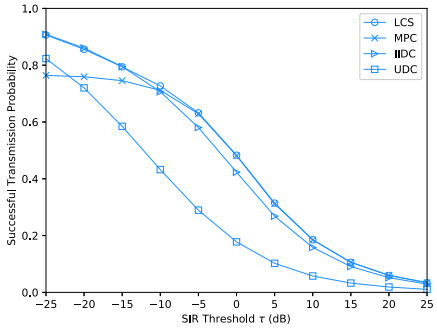

Fig. 3. Comparison of the LCS with three baseline strategies.

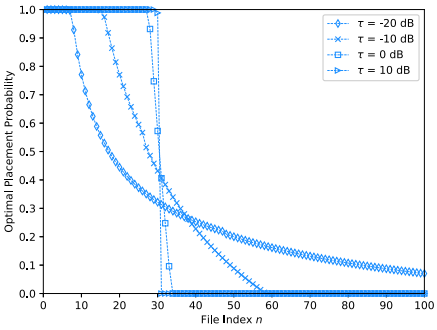

Fig. 4. Locally optimal caching probability vector $\mathbf{T}_1^\star$ versus file index n.

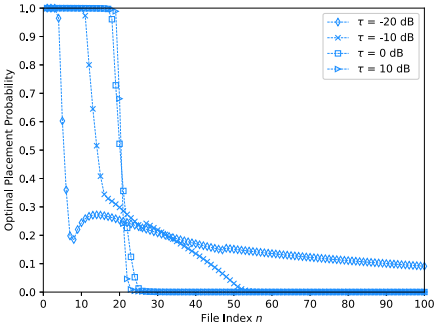

Fig. 5. Locally optimal caching probability vector $\mathbf{T}_2^\star$ versus file index n.

Figure 3 shows the comparison between the LCS and the three baseline strategies. From Fig. 3, we can observe that the LCS outperforms all the three baselines. In particular, when τ is small, the LCS is consistent with the IIDC; however, when τ is large, the LCS is consistent with the MPC. This is because that when τ is small, caching more different files can provide higher file diversity; and when τ is large, caching the most popular files can guarantee the most frequent file requests, which is more meaningful for improving the STP.

Figure 4 and Fig. 5 depict the locally optimal caching vector $\mathbf{T}_1^\star$ and $\mathbf{T}_2^\star$, respectively. From the figures, one can see that files of higher popularity have higher probability to be cached at BSs. Furthermore, there exist a situation where we have $1 \leq N_1 < N_2 \leq N$ that satisfies $t_1^\star = t_2^\star = \cdots = t_{N_1}^\star = 1$, $t_i^\star \in (0,1)$ for all $i \in (N_1, N_2)$ and $t_{N_2}^\star = t_{N_2+1}^\star = \cdots = t_N^\star = 0$ for each tier, as illustrated in [8, Remark 2].

6 Conclusion

In this paper we studied the caching probability optimization in cache-enabled multi-tier IoT networks with base station joint transmission (JT). We developed a JT scheme and derived the STP expression for the K-tier cache-enabled IoT networks when adopting random caching scheme. Then the locally optimal caching strategy (LCS) was obtained by using the gradient projection method. Simulations were conducted to verify the established STP model and compare the LCS with three baseline strategies, i.e., MDP, IIDC and UDC. The results showed that LCS outperforms all the three baselines in terms of STP.

References

1. Nigam, G., Minero, P., Haenggi, M.: Coordinated multipoint joint transmission in heterogeneous networks. IEEE Trans. Commun. (2014). https://doi.org/10.1109/TCOMM.2014.2363660
2. Nigam, G., Minero, P., Haenggi, M.: Spatiotemporal cooperation in heterogeneous cellular networks. IEEE J. Sel. Areas Commun. **33**(6), 1253–1265 (2015). https://doi.org/10.1109/JSAC.2015.2417017
3. Wen, J., Huang, K., Yang, S., Li, V.O.: Cache-enabled heterogeneous cellular networks: optimal tier-level content placement. IEEE Trans. Wireless Commun. (2017). https://doi.org/10.1109/TWC.2017.2717819
4. Kuang, S., Liu, X., Liu, N.: Analysis and optimization of random caching in K -Tier Multi-antenna multi-user HetNets. IEEE Trans. Commun. **67**(8), 5721–5735 (2019). https://doi.org/10.1109/TCOMM.2019.2913378
5. Zhang, S., Liu, J.: Optimal probabilistic caching in heterogeneous IoT networks. IEEE Internet Things J. **7**(4), 3404–3414 (2020). https://doi.org/10.1109/jiot.2020.2969466
6. Yang, J., Ma, C., Jiang, B., Ding, G., Zheng, G., Wang, H.: Joint optimization in cached-enabled heterogeneous network for efficient industrial IoT. IEEE J. Sel. Areas Commun. **8716**(c), 1 (2020). https://doi.org/10.1109/jsac.2020.2980907

7. Wen, W., Cui, Y., Zheng, F.C., Jin, S., Jiang, Y.: random caching based cooperative transmission in heterogeneous wireless networks. IEEE Trans. Commun. **66**(7), 2809–2825 (2018). https://doi.org/10.1109/TCOMM.2018.2808188
8. Feng, T., Shi, S., Gu, S., Xiang, W., Gu, X.: Optimal content placement for cache-enabled IoT networks with local channel state information based joint transmission. IET Commun. (2020, to be published). https://doi.org/10.1049/iet-com.2020.0167
9. Chae, S.H., Quek, T.Q., Choi, W.: Content placement for wireless cooperative caching helpers: a tradeoff between cooperative gain and content diversity gain. IEEE Trans. Wireless Commun. **16**(10), 6795–6807 (2017). https://doi.org/10.1109/TWC.2017.2731760
10. Blaszczyszyn, B., Giovanidis, A.: Optimal geographic caching in cellular networks. In: IEEE International Conference on Communications (2015). https://doi.org/10.1109/ICC.2015.7248843
11. Bertsekas, D.P.: Nonlinear Programming. Athena Scientific, 2nd edn. (1999)
12. Baştuğ, E., Bennis, M., Kountouris, M., Debbah, M.: Cache-enabled small cell networks: modeling and tradeoffs. EURASIP J. Wireless Commun. Netw. **2015**(1), 41–47 (2015)
13. Bharath, B.N., Nagananda, K.G., Poor, H.V.: A learning-based approach to caching in heterogenous small cell networks. IEEE Trans. Commun. **64**(4), 1674–1686 (2016). https://doi.org/10.1109/TCOMM.2016.2536728
14. Tamoor-ul-Hassan, S., Bennis, M., Nardelli, P.H.J., Latva-Aho, M.: Modeling and analysis of content caching in wireless small cell networks. In: Proceedings of IEEE ISWCS, Bussels, Belgium, pp. 765–769, August 2015. https://doi.org/10.1109/ISWCS.2015.7454454