

Research Article

Energy Efficient Caching in Backhaul-Aware Cellular Networks with Dynamic Content Popularity

Jiequ Ji , Kun Zhu , Ran Wang, Bing Chen , and Chen Dai

College of Computer Science and Technology, Nanjing University of Aeronautics and Astronautics, Nanjing, China

Correspondence should be addressed to Kun Zhu; zhukun@nuaa.edu.cn

Received 29 December 2017; Accepted 5 March 2018; Published 8 April 2018

Academic Editor: Zheng Chang

Copyright © 2018 Jiequ Ji et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Caching popular contents at base stations (BSs) has been regarded as an effective approach to alleviate the backhaul load and to improve the quality of service. To meet the explosive data traffic demand and to save energy consumption, energy efficiency (EE) has become an extremely important performance index for the 5th generation (5G) cellular networks. In general, there are two ways for improving the EE for caching, that is, improving the cache-hit rate and optimizing the cache size. In this work, we investigate the energy efficient caching problem in backhaul-aware cellular networks jointly considering these two approaches. Note that most existing works are based on the assumption that the content catalog and popularity are static. However, in practice, content popularity is dynamic. To timely estimate the dynamic content popularity, we propose a method based on shot noise model (SNM). Then we propose a distributed caching policy to improve the cache-hit rate in such a dynamic environment. Furthermore, we analyze the tradeoff between energy efficiency and cache capacity for which an optimization is formulated. We prove its convexity and derive a closed-form optimal cache capacity for maximizing the EE. Simulation results validate the proposed scheme and show that EE can be improved with appropriate choice of cache capacity.

1. Introduction

With the proliferation of new wireless devices and smart phones, as well as bandwidth-intensive applications, user demand for wireless data traffic has been growing tremendously, and the corresponding network load increases exponentially [1, 2]. For more effectively handling the explosive growth of mobile data services, caching at the edge of wireless network has been proposed and has become a hot research issue [3, 4]. Recent studies have shown that caching popular contents at the edge of wireless networks can not only reduce the backhaul cost, access latency, and energy consumption but also boost the throughput [5, 6].

Meanwhile, energy efficiency (EE) has become a major performance metric for future 5G cellular networks [7]. However, only few works in the literature consider the energy efficiency analysis for cache-enabled cellular networks. In general, the EE of a cellular network with caching can be improved from two aspects: improving the cache-hit rate and optimizing the cache size. The work in [8] takes content popularity and request arrival density into account for

deriving the EE. By analyzing the EE for the caching-enabled network, [9] elaborates whether and how caching at BSs can improve the EE. The authors in [10] show that introducing cache can improve the EE if the content catalog size is not infinitely large and placing cached contents at small cell base stations (SBSs) is more energy efficient. Furthermore, a user-centric cooperative model is proposed in [11], and the relationship between EE and the cache-hit rate is analyzed. In [12], the authors analyzed the impact of various factors on EE such as interference level, backhaul and cache capacity, user and BS density, power consumption parameters, BS work state, and content popularity. Also, the condition when the EE can benefit from caching is derived. This work also points out that the backhaul power consumption and caching power consumption are not negligible for analyzing EE. It is worth noting that there are tradeoffs between backhaul power consumption and caching power consumption. Specifically, caching more contents could reduce the backhaul load and correspondingly reduce the backhaul power consumption, at the cost of increasing the caching power consumption. However, it is still not clear what is the optimal amount of

cached popular contents at the BS which could maximize the EE of the downlink system.

Furthermore, all the above works are based on an ideal assumption that the content popularity is known and fixed. Popularity refers to the cumulative number of requests for each content, or the requested probability, or the requested arrival rate over a specific area and period of time [13]. In reality, content popularity is dynamic which needs to track an ever changing popularity profile of contents. Substantial number of experimental results have shown that requests for contents frequently appear during a certain period of time at a specific location, while the number of requests outside this time period is small [14]. Apparently, considering dynamic content popularity will pose some challenges, such as how to predict content popularity and how to analyze cache performance in such a dynamic environment [15].

In this work, we investigate the energy efficient caching problem in backhaul-aware cellular networks by jointly improving cache-hit rate and deriving optimal cache size. And dynamic content popularity is considered. To timely estimate the dynamic content popularity, we propose a method based on shot noise model (SNM), which can predict the request hit rate of contents. Then we propose a distributed caching policy to improve the cache-hit rate in such a dynamic environment. Also, the content redundancy can be reduced by the proposed policy. Finally, we analyze the tradeoff between energy efficiency and cache capacity for which an optimization is formulated. Analytically closed-form solutions are obtained. And the simulation results show effectiveness of the proposed scheme under limited backhaul and cache capacity.

The remainder of this paper is organized as follows: in Section 2, we introduce the system model. The dynamic caching policy and distributed cache policy are proposed in Section 3. The EE optimization problem of a cache-enabled cellular network is formulated and analyzed in Section 4. Numerical results and analysis are provided in Section 5, and the conclusions are summarized in Section 6.

2. System Model

In this section, we describe our network, transmission, and BS sleep pattern models. For ease of reading, we summarize the notation used throughout in Notations.

2.1. Network Model. We consider a backhaul-aware cellular network consisting of N_b base stations. These BSs are randomly distributed on a two-dimensional Euclidean plane \mathbb{R}^2 by a homogeneous spatial Poisson point process (HPPP) $\Phi_b = \{d_i\}$ with density λ_b [16]. Each BS is equipped with N_t transmit antennas and serves multiple users with single antenna. Each BS has a cache device which is able to cache N_c contents and is connected to the core network through a backhaul link with limited backhaul capacity C_{bh} . An illustration of the network model is depicted in Figure 1.

In the network model, the users are modeled by an independent homogeneous Poisson Point Process with density λ_u , where $\lambda_u > \lambda_b$. In the paper, we assume that user associating

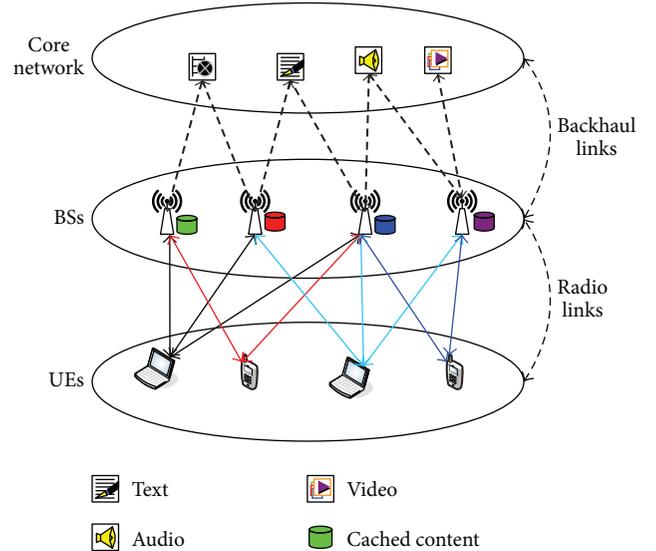


FIGURE 1: An illustration of the network model. A BS is deployed in a circle area with radius D to serve multiple users uniformly distributed in the area.

with BS is based on both content and signal-to-interference-plus-noise ratio (SINR). Specifically, if there are multiple BSs which have cached the requested content, a user will select the nearest one. Otherwise, a user will be associated with an BS with maximal receiving SINR. For mathematical tractability, we use a circular area with a radius of D to approximate the hexagonal shaped area, where a number of single antenna users are evenly distributed. The average user density of each cell is λ_u ; then the probability density of u users in each cell can be expressed as

$$P_{x(u)} = \frac{(\lambda_u/N_b)^u}{\Gamma(u)} \exp\left(-\frac{\lambda_u}{N_b}\right). \quad (1)$$

Denoting $C_i = \{1, 2, \dots, N_c\}$ as the set of contents cached at the i th BS, $i = 1, 2, \dots, N_b$. Each user requests different content, such as video, music, and news. We build the contents of the user request into a content catalog, that is, $F_f = \{1, \dots, F_f\}$. We assume that each content has equal size F bits in the content catalog, and the cache capacity of each BS is $F \cdot N_c$ bits. If the content requested by the user is cached in the local BS, the base station will directly deliver the content to the user from the caching device. And this user is called the cache-hit user, where the transmission rate is determined by the SINR of the downlink channel [17]. On the contrary, when the user requested content is not cached in the local BS, the BS firstly obtains the content from the core network through the backhaul link and then transmits it to the user through the radio link. Similarly, this user is called the cache-miss user.

2.2. Transmission Model. Spatial multiplexing is taken into consideration. At the same time, the number of simultaneously served users can not exceed the number of antennas N_t . And a round-robin scheduling policy is applied [18]. To capture the essence of the problem and simplify the analysis,

we assume that each BS serves users with zero-forcing beam-forming (ZFBB) [19], which is a spatial signal processing technique to eliminate multiuser interference. And the downlink transmission power for all users is the same. We introduce a parameter β to reflect the percentage of intercell interference (ICI) that can be eliminated by interference management methods (i.e., coordinated beam-forming and serial interference suppression) [20].

The signal transmitted from a given BS is affected by two propagation losses before reaching the user:

- (1) Path-loss: a distance dependent path-loss $p_{(r_{ub})} = r_{ub}^{-\alpha}$, where r_{ub} is the distance from the BS to user and α is the path-loss exponent.
- (2) Rayleigh fading: $h_{ub} \in C^{N_r \times 1}$ is the small-scale Rayleigh fading channel vector from the BS to user.

Particularly, we describe the channel gain between the u th user and the b th BS as G_{ub} , which can be expressed as

$$G_{ub} = r_{ub}^{-\alpha} |h_{ub}|^2, \quad (2)$$

where $|h_{ub}|^2$ denotes the integrated channel power gain between the u th user and the b th BS due to Rayleigh fading. With the above expressions, we can get the SINR of a typical user u as

$$\text{SINR}(r_{ub}) = \frac{Pr_{ub}^{-\alpha} |h_{ub}|^2}{\sum_{j=1, j \neq b}^{N_b} \beta Pr_{ub,j}^{-\alpha} \|h_{ub,j}\|^2 + \sigma^2}, \quad (3)$$

where P is the transmit power of each base station and the random variable $h_{ub,i}$ denotes the small-scale Rayleigh fading of the wireless channel and follows the exponential distribution. $I_{BS \rightarrow U} = \sum_{j=1, j \neq b}^{N_b} \beta Pr_{ub,j}^{-\alpha} \|h_{ub,j}\|^2$ denotes the cumulative interference experienced from all BSs except the i th BS, and $r_{ub,i}^{-\alpha}$ is the distance between the user and all base stations except the i th BS. σ^2 is the variance of the white Gaussian noise.

2.3. BS Sleep Pattern of the Network. The energy consumption of a BS is determined by its working state [21]. When a BS is in active state, the components (e.g., power amplifier, signal processing unit, and antennas) will generate energy consumption. On the other hand, if the BS switches to sleep state, some units will be closed, and parts of the energy consumption can be saved. Once a BS has no active user, it will be switched into the sleeping state. According to the system model, the probability of a BS being in sleep state can be expressed as

$$P_s = e^{(\lambda_b / (\lambda_u + \lambda_b))}. \quad (4)$$

3. Distributed Caching with Dynamic Content Popularity

Appropriate modeling has important impacts on the performance analysis of caching policies. However, most existing works do not take the dynamic changes of popular content

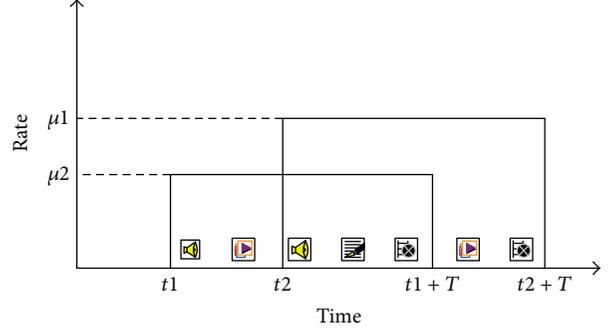


FIGURE 2: An illustration of the Poisson Shot Noise Model. This figure shows the two different request arrival rates and times during the lifetime and sets the active lifetime and shape of all content fixed, where μ_1 and μ_2 are the request arrival rate of a content 1 at time t_1 .

distribution into consideration. In this work, we consider dynamic content popularity and propose a method to handle the time-varying popularity based on Poisson Shot Noise Model (SNM) [22].

Shot Noise Model is a “content-level” business model, which can reveal the dynamic changing of content catalogs and the arrival rates of different contents in the same area, and can be used to evaluate the performance of a cached network. SNM describes the process of each content generation (i.e., content from active to inactive) and each content request process [23]. In [24], the author assumes that the process of generating contents is a Poisson process, and each content is taken out after being active for certain period. The request process of each content is an independent nonhomogeneous Poisson process. In this paper, we assume that the process of generating contents and the request process follows the process of [24]. Figure 2 shows the Poisson Shot Noise Model, which includes two different request arrival rates and time during the lifetime. We consider the active lifetime and shape of all contents to be fixed in this model.

The lifetime of each content f is associated with a time-varying request, which can be described by following characteristics:

- (1) *The start time of request content t_f* : the request start time of contents follows a Poisson process with average arrival rate λ_f (i.e., for a unit of time there is an average of λ_f contents start to be active). For rectangular pulses, the average total number of contents generated in time T_0 is $\bar{N}_f = \lambda_f T_0$ and the average number of active content is $\lambda_f T$, where T is the file life-cycle.
- (2) *Shape $\wedge_f(t)$* : the shape describes the regularity of the request arrival rate of file m changing with time (i.e., rectangular pulse). It is reported that the this feature has a smaller impact on the hit probability under the Least Recently Used (LRU) cache management policy.
- (3) *The total number of requests N_f* : this feature represents the total number of times that content f is requested by a user during the active time.

- (4) *The life-cycle of content* T_f : this feature describes the length of the request for the content f 's arrival time, which can be defined by rectangular pulses, so the file life-cycle T_f can be expressed as

$$T_f = \frac{1}{\int_0^{\infty} \Lambda_f^2(t) dt}. \quad (5)$$

We assume that all contents have the same life-cycle. Denote with t_m the arrival time of each request content f . At time t the lifetime content catalog is expressed as

$$F_f(t) = \{f : t_f < t < t_m + T\}. \quad (6)$$

We can describe the alive content $f \in F_f$ by the active ‘‘age’’ of each content and the total number of requests N_f . The ‘‘age’’ of active content f at time t can be expressed by $\tau_f(t) = t - t_f$, where the ‘‘age’’ refers to the active time of content f at time t since the content appears in the system. According to the above characteristics, for the shot noise model with file life-cycle T and rectangular pulse Λ_f , the request arrival rate of a content f at time t can be expressed as $\mu_f(t) = \Lambda_f N_f(t - t_f)$.

3.1. Request Hit Rate with Estimated Popularity. In this section, we assume that we do not know the popularity of each content; then we estimate the dynamic content popularity by observations of the past user's requests to the file in the caching controller and try to maximize the request hit rate in lifetime T . However, considering that there is no traffic power for transmitting a content from the core network to the BSs, we can achieve the time horizon request hit rate optimization via solving the individual problem of maximizing the request hit rate in each time instance.

In this model, we choose the pointwise traversal method to study the request hit rate in each time instance. For ease of analysis, we choose to study the case where $t = 0$ and observe the random events in the time interval $[-T, 0]$. We use binary caching vector $X_f(t)$ to determine whether the content is cached at time t , where $x_f(t) = 1$ represents the fact that content f is cached in the BS at time t , and $x_f(t) = 0$ otherwise. Due to the limited cache capacity, the constraint $\sum_{f=1}^{f=F_f} x_f(t) \leq C_{ca}$ must be satisfied in each time interval.

In the following, we consider using binary caching vector $X_f(t)$ to maximize the instantaneous request hit probability at the origin. Taking the fact into account that contents popularity μ_f is unknown, the instantaneous request hit probability at the origin can be expressed as $R(x) = \sum_{f=1}^{f=F_f} x_f \mu_f$. Next we would like to optimize the expected request hit rate which can be expressed as follows:

$$\mathbb{E} [R(x) | (N_f), (\tau_f)] = \sum_{f=1}^{f=F_f} x_f \mathbb{E} [\mu_f | N_f, \tau_f]. \quad (7)$$

With SNM, predictive popularity can be expressed as

$$\begin{aligned} \mathbb{E} [\mu_f | (N_f), (\tau_f)] \\ = \frac{\int_0^{\mu_f} \mu_f (\mu_f \tau_f)^{N_f} (e^{-\mu_f \tau_f / N_f}) f(\mu_f) d\mu_f}{\int_0^{\mu_f} (\mu_f \tau_f)^{N_f} (e^{-\mu_f \tau_f / N_f}) f(\mu_f) d\mu_f}, \end{aligned} \quad (8)$$

where $f(\mu_f)$ is the power-law density and τ_f is the ‘‘age’’ of active content f at time t . Accordingly, the request hit rate maximization problem based on predicted popularity can be formulated as

$$\begin{aligned} R^*(N_f, \tau_f) = \arg \max_{\forall f, x_f \in \{0,1\}} & \sum_{f=0}^{f=F_f} x_f \mathbb{E} [\mu_f | N_f, \tau_f] \\ \text{subject to} & \sum_0^{\mathcal{F}} x_f \leq C_{ca}. \end{aligned} \quad (9)$$

It can be seen from the above expression that we can calculate the instantaneous request hit rate by the value of N_f and τ_f .

3.2. Age-Based Threshold Caching Policy. To solve the optimal request hit probability at every time instant, we design a caching policy with age-based threshold, and this policy does not need to calculate the estimation $\widehat{\mu}_f = \mathbb{E}[\mu_f | (N_f), (\tau_f)]$. Intuitively, the ‘‘age’’ τ_f affects the estimation $\widehat{\mu}_f$. Due to the uncertainty in the estimation $\widehat{\mu}_f$ is low; then if we want to maximize the request hit rate we must cache more contents with larger ages. Accordingly, we propose an age-based threshold \bar{A}_τ which depends on the content age τ . Under this policy, if the content satisfies $N_f \geq \bar{A}(\tau_f)$, it will be cached. When the content is fixed and the age is known, the cache decision depends only on the total number of requests N_f , which makes caching decision more simpler.

Note that an important progress of the age-based threshold estimation is to compute \bar{A}_τ . We can compute the age threshold function \bar{A}_τ by iteratively filling cache capacity C_{ca} in the algorithms, which make the marginal request hit rate improvements $\mathbb{E}[\mu_f | (N_f) = \bar{A}_\tau, (\tau_f = \tau)]$ for each request age τ approximately equal. The details can be found in Algorithm 1.

3.3. The Correlation of Content Request Popularity. Due to differences in social and cultural backgrounds among users, restrictions on specific office areas, or the impact of certain group activities, content popularity among BSs may vary. However, there are still possible correlations. We assume that the contents exhibit geographical correlations and hence there exist groups of contents which are very popular in subset of local caches [25–27]. For local caches with the same popularity, we can more effectively constrain the aggregation to a subset of caches that witness similar traffic patterns by aggregating all cached request contents.

Request Model for Correlated Popularity. We design a model for the correlation of content request popularity

Step 1: Selecting Parameter. (i) C_{ca} is the cache capacity, (ii) λ is the average request rate, (iii) $T\lambda$ is the average total number of request content at any time instant, (iv) $\gamma_{ca} = C_{ca}/T\lambda$ is the fraction of caching content catalog, (v) W_{μ_f} is the function of distribution. (vi) Define $\theta(\gamma_{ca})$ is the γ_{ca} th tipper-percentile of W_{μ_f} , where $\theta(\gamma_{ca}) = W_{\mu_f}^{-1}(1 - \gamma_{ca})$.

Step 2: Setting the Age-Based Threshold. Denote $\tilde{N}_f(\tau)$ is the age-based threshold.

$$\tilde{N}_f(\tau) = \min\{f \in \mathbb{N} : \mathbb{E}[\mu_f | N_f = f, \tau_f = \tau] \geq \theta(\gamma_{ca})\}$$

Step 3: Caching Content Vector. Using binary caching vector $X_f(t)$ to determine whether the content is cached at time t .

$$X_f(t) = \begin{cases} 1, & \text{if } N_f \geq \tilde{N}_f(\tau_f), \\ 0, & \text{if otherwise,} \end{cases}$$

Step 4: Caching Capacity. We define a constraint condition for the capacity of the cache, if $\sum_0^F x_f > C_{ca}$, we will select $\sum_0^F x_f = C_{ca}$, $\forall f \in F$ with $x_f = 1$, otherwise set $x_f = 0$ for content.

ALGORITHM 1: Age-based threshold policy.

μ_f^i . The aggregate total popularity is expressed as $\mu_f^{N_b} = N_b \mu_f^i(\tau)$, $\forall f, i \in (1 \cdots N_b)$, and the global average request hit rate is $\mathbb{E}[\mu_f^Z] = \bar{\mu}$. We obtain the correlation between the intercell request arrival rate μ_m^i through community detection [28] and inhomogeneous random graphs [29] as follows:

- (i) Each content f is associated with a property vector M_f , where M_f is the property of the content f (e.g., the probability of the content as comedy). In order to simplify the model we consider that M_f is an independent uniformly distributed random variable taking values in $[0, 1]$.
- (ii) Each BS i is also associated with a property vector L_i , where L_i is the property of the BSs i (e.g., the probability of the user request comedy in intercell i). The vector L_i is selected uniformly and randomly in $[0, 1]$.
- (iii) We assume that M_f and L_i are uniformly distributed in $[0, 1]$. We define $K(x, y) = g(|x - y|)$, where $g(\cdot)$ is a symmetric and continuous function with period of 1, strictly decreasing on $[0, 1/2]$, and $\int_0^1 g(|x - y|) dy = 1$ for all $x \in [0, 1]$.

When the content popularity of each intercell is different, according to [30], the request arrival rate of content f in BS can be expressed as

$$\mu_f^i = \mu_f^Z \frac{K(M_f, L_i)}{\sum_0^Z \in N_b K(M_f, L_i)}, \quad \forall f, i, \quad (10)$$

where Z are the contents in the core network. As the number of caching contents Z increases in the core network and the kernel function $g(\cdot)$ satisfies some base conditions, the normalization constant almost surely becomes deterministic:

$$\frac{1}{Z} \sum_{i=0}^{i=Z} K(M_f, L_i) \rightarrow \lim_{Z \rightarrow \infty} \int_0^1 g(|M_f - y|) dy = 1 \quad (11)$$

$$\mu_f^i \rightarrow \lim_{Z \rightarrow \infty} \frac{\mu_f^Z K(M_f, L_i)}{Z}$$

In the section, the request hit rate of the future contents is predicted according to the request information of the historical contents of each BS, and then each BS caches the most popular N_c different contents.

3.4. BS Cache Policy of the Network. Research shows that multiple users could access the same content, like some videos, news, music, and so on. In this case, the network will be flooded with requests of same contents, which can largely increase the latency, eventually leading to network congestion. Therefore, under the limited cache capacity, assuming that the user's choice distribution matches with the dynamic content popularity, then each BS can cache most popular content to reduce service latency and backhaul congestion in such a dynamic environment.

A content catalog is defined as $F_f = \{1, 2, \dots, F_f\}$, and the size of each content f in the catalog has a length of $L(f)$ Mbytes. According to the description of Section 3, we can predict the content popularity, and the instantaneous popularity follows the Zipf distribution [31]. Each user requests a content from the content catalog F_f , and the probability of requesting the f th content is

$$P_{F(f)} = \frac{\left(\sum_{i=1}^{F_f} (1/i^{\lambda_u/\lambda_b})\right)^{-1}}{f^{\lambda_u/\lambda_b}}, \quad (12)$$

where λ_u/λ_b characterizes the steepness of the distribution. The value of λ_u/λ_b depends on the average number of users per BS, and higher values of λ_u/λ_b correspond to steeper distributions which means that the higher the number of users attached to a BS is, the more accurately the trend is sampled.

As mentioned above, each BS caches most popular N_c content according to the dynamic popularity distribution. Hence, the cache-hit rate of each BS can be obtained by the Zipf distribution:

$$H = \sum_{f=1}^{N_c} \frac{f^{\lambda_u/\lambda_b}}{\sum_{i=1}^{F_f} i^{\lambda_u/\lambda_b}}. \quad (13)$$

It can be proved from (13) that H converges to 1 when the content catalog F_f in the BS goes to infinity. Consider the relation between maximal EE of the network and the cache size. To this end, we provide the cache-hit rate H for large values of N_c and F_f . In order to show the impact of the cache capacity and content library F_f on cache-hit rate, we analyze a normalized cache size $\eta \triangleq N_c/F_f$, $0 < N_c \leq F_f$, $\eta \in [0, 1]$. Thus (13) is normalized as

$$H = \frac{\ln N_c + \Theta(N_c^{-1})}{\ln F_f + \Theta(F_f^{-1})} \approx \frac{\ln N_c}{\ln F_f} \approx 1 + \frac{\ln \eta}{\ln F_f}, \quad (14)$$

where the approximation in (14) is accurate when $N_c \gg 1$ and $F_f \gg 1$.

Cache-hit rate affects the performance of wireless networks, like throughput, energy efficiency, user quality of service (QoS), and so on. In order to increase cache-hit rate, we propose a distributed content caching strategy in backhaul-aware cellular networks, where four adjacent BSs cache different popular contents. According to the system model, we can find that each user is associated with a BS based on both content and SINR. Meanwhile, if there are several BSs caching the requested content, the user selects one whose distance to him is the shortest and delivers the content via radio link directly. Otherwise, user associates with a BS whose SINR is maximal from its neighbouring BSs and then gets the requested content through the backhaul link from the core network. As illustrated in Figure 3, each hexagon represents a BS with cached popular content. This distributed content caching strategy can reduce content redundancy by caching different contents in different BS. When each BS caches N_c contents with the distributed cache strategy, each user can requests $4N_c$ popular contents.

4. Energy Efficiency Optimization

In recent years, the research on green communications has made great progress. It shows that the energy efficiency of the cellular network has become an essential factor in the study of its performance, which is the key to reduce operating costs and is aiming to provide wireless access in the way of sustainable development and rational utilization of resources [32]. We define the EE as the ratio of the average throughput of the network to the average total power consumption at the BS, which can be expressed as

$$EE = \frac{\bar{R}}{\bar{P}_t}, \quad (15)$$

where \bar{R} is the average throughput and \bar{P}_t is the average power consumption.

4.1. Average Throughput of the Network. If user (u_1, u_2, \dots, u_c) request contents are cached at local BS, the user can immediately obtain the requested content from the local cache via the radio links, where the data rate is limited only by the capacity of the wireless channel. On the contrary, if the contents requested by users (u_{c+1}, \dots, u_b) are not cached at local BSs,

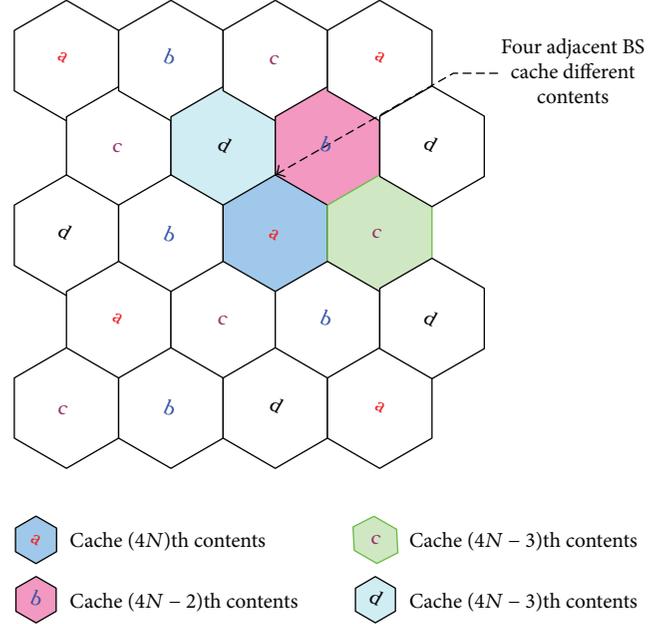


FIGURE 3: An illustration of distributed caching policy. The BS marked with “a,” “b,” “c,” and “d” caches different popular contents and each user associates with the nearest BS that caches the user’s requested contents.

the BSs need to obtain the contents from the core network via the backhaul links and then transmit the contents to users through radio links, so the data rate is subject to backhaul link capacity and radio channel capacity. Considering the above two cases, the downlink average throughput of the b th cell can be expressed as

$$\begin{aligned} \bar{R} &= (1 - P_s) \left(\bar{R}_2 + (\bar{R}_1 - \bar{R}_2) \left(1 + \frac{\ln \eta}{\ln F_f} \right) \right), \\ \bar{R}_1 &= \mathbb{E} \left\{ B \sum_{u=1}^{u_c} \log_2 (1 + \text{SINR}_{u_b}) \right\}, \\ \bar{R}_2 &= \mathbb{E} \left\{ \min \left(B \sum_{U=u_c+1}^{U_b} \log_2 (1 + \text{SINR}_{u_b}), C_{\text{bh}} \right) \right\}, \end{aligned} \quad (16)$$

where \bar{R}_1 is the average sum rate of the cache-hit users and \bar{R}_2 is the average sum rate of the cache-miss users.

4.1.1. The Average Sum Rate of Cache-Hit Users. Considering that the SINR for every user shown in (3) are identically distributed, the average sum rate of the cache-hit users \bar{R}_1 can be approximated as

$$\begin{aligned} \bar{R}_1 &= u_c B \mathbb{E} \left\{ \log_2 \left(1 \right. \right. \\ &\quad \left. \left. + \frac{r_{u_b}^{-\alpha} |h_{u_b}|^2}{\sum_{j=1, j \neq b}^{N_b} \beta u_b r_{u_b, i}^{-\alpha} \|h_{u_b, i}\|^2 + \sigma^2 / P} \right) \right\} \end{aligned}$$

$$\begin{aligned}
&\approx u_c B \left(\log_2 \frac{N_t - u_b + 1}{u_b} + \int_0^D \log_2 r_{ub}^{-\alpha} \frac{2r_{ub}}{D^2} d_{r_{ub}} \right) \\
&- u_c B \left(\log_2 \left(e^{(\lambda_b u / (\lambda_u + \lambda_b u))} \beta 2^\phi D^{-\alpha} + \frac{\sigma^2}{P} \right) \right) \\
&\approx u_c B \left(\frac{\alpha}{2 \ln 2} + \log_2 \right. \\
&\quad \left. \frac{(N_t - u_b + 1)}{u_b (e^{u \lambda_b / (\lambda_u + u \lambda_b)} \beta 2^\phi + (\sigma^2 / P) D^\alpha)} \right), \tag{17}
\end{aligned}$$

where ϕ is a constant only depending on the path-loss exponent α when $N_b \rightarrow \infty$, r_{ub} is the distance from b th BS to the u th user.

4.1.2. The Average Sum Rate of Cache-Miss Users. Considering the transmit power of base stations to be the same, the cumulative interference from all BS approaches to its expectation $\mathbb{E}\{I_{BS \rightarrow U}\}$ when $N_b \rightarrow \infty$ according to the law of the ultradense BSs deployment. This suggests that the comparison of C_{bh} and $B \sum_{U=u_c+1}^{u_b} \log_2(1 + \text{SINR}_{u_b})$ depends on the distance from each user to the local base station when the number of BSs is very large. Therefore, the average sum of the cache-miss users can be approximated as

$$\begin{aligned}
\bar{R}_2 &= \mathbb{E} \left\{ \min \left(B \sum_{U=u_c+1}^{u_b} \log_2(1 + \text{SINR}_{u_b}), C_{bh} \right) \right\} \\
&\approx \mathbb{E} \left\{ \min \left((u_b - u_c) R_3 \right. \right. \\
&\quad \left. \left. + \frac{\alpha B}{2 \ln 2} \sum_{u_c+1}^{u_b} 2 \ln \frac{D}{r_{ub}}, C_{bh} \right) \right\}, \\
R_3 &\approx \log_2 \frac{N_t - u_b + 1}{e^{u \lambda_b / (\lambda_u + u \lambda_b)} \beta 2^\phi + (\sigma^2 / P) D^\alpha}.
\end{aligned} \tag{18}$$

Given that the user's distributions are independent and random, we can apply the probability density function of $2r_{ub}/D^2$ to proving that $(2 \ln(D/r_{ub}), u \in \{1 \dots u_b\}, b \in \{1 \dots N_b\})$ are random variables which are independently and exponentially distributed. Thus, we can get that $x = \sum_{u_c+1}^{u_b} 2 \ln(D/r_{ub})$ is a random variable following Gamma distribution $\mathbb{G}(u_b - u_c, 1)$. In this case, when the probability density function $\sum_{u_c+1}^{u_b} 2 \ln(D/r_{ub}) > 0$, we can derive $f(x) = x^{u_b - u_c - 1} e^{-x} / (u_b - u_c - 1)!$. Therefore, when $C_{bh} \leq (u_b - u_c) R_3$, (18) becomes $\bar{R}_2 = C_{bh}$. When $C_{bh} > (u_b - u_c) R_3$, (18) can be approximated as

$$\begin{aligned}
\bar{R}_2 &\approx \int_0^\infty \min \left(AR_3 + \frac{\alpha B}{2 \ln 2} x, C_{bh} \right) f(x) dx \\
&\approx \int_0^Z \left(AR_3 + \frac{\alpha B}{2 \ln 2} x \right) f(x) dx + \int_Z^\infty C_{bh} f(x) dx \tag{19} \\
&\approx A \left(\frac{\alpha B}{2 \ln 2} \gamma B + \bar{R}_3 \gamma C + C_{bh} \Gamma C \right),
\end{aligned}$$

where $A = (u_b - u_c)$, $B = (u_b - u_c + 1, Z)$, $C = (u_b - u_c, Z)$, $Z = (2 \ln 2 / \alpha B)(C_{bh} - AR_3)$, $\Gamma(a, b) = e^{-b} \sum_{i=0}^{b-1} (b^i / i!)$, and $\gamma(a, b) = 1 - \Gamma(a, b)$.

In summary, the average sum rate of the cache-miss users can be approximated as

$$\bar{R}_2 \approx \begin{cases} A \left(\frac{\alpha B}{2 \ln 2} \gamma B + \bar{R}_3 \gamma C + C_{bh} \Gamma C \right), & \text{if } C_{bh} > AR_3, \\ C_{bh}, & \text{if } C_{bh} \leq AR_3. \end{cases} \tag{20}$$

- (i) When λ_b is large and $N_t = u_b$, by simplifying (17), the average number of bits totally transmitted can be approximated as follows:

$$\bar{R}_1 \approx B \left(\frac{\alpha}{2 \ln 2} + \log_2 \frac{N_t}{e^{u \lambda_b / (\lambda_u + u \lambda_b)} \beta 2^\phi + (\sigma^2 / P) D^\alpha} \right). \tag{21}$$

- (ii) When $u_b - u_c = 1$ and $N_t = u_b$, we can get $z = (2 \ln 2 / \alpha B)(C_{bh} - \bar{R}_e)$, $\gamma(B) = 1 - e^{-z}(1 + z)$, $\gamma(C) = 1 - e^{-z}$ and $\Gamma(C) = e^{-z}$, $R_3 \approx R_4 \approx \log_2(N_t / (e^{u \lambda_b / (\lambda_u + u \lambda_b)} \beta 2^\phi + (\sigma^2 / P) D^\alpha))$, $R_5 = (\alpha / 2 \ln 2) 2^{-((R_4 - C_{bh}) / \alpha)}$. So the formula \bar{R}_{bh} can be approximated as

$$\bar{R}_2 \approx \begin{cases} B \left(\frac{\alpha}{2 \ln 2} + R_3 - R_4 \right), & \text{if } C_{bh} > R_4, \\ C_{bh}, & \text{if } C_{bh} \leq AR_4. \end{cases} \tag{22}$$

4.2. Average Total Power Consumption of the Network. In this work, we investigate the performance of EE in cellular networks with caching at BSs. To this end, we need to calculate the overall energy consumption caused by the transmission and storage of contents.

We have known that when the BS is in sleeping state, some units are turned off, such that the BS reduces part of the power consumption. As a result, we can divide the circuit power consumption and transmit power consumption into two states according to the working state of BS, respectively.

Finally, the overall power consumption of the network can be divided into four parts:

- (i) The average total transmit power of the base station in the entire downlink can be expressed as

$$P_t = \begin{cases} N_b \rho P (1 - P_s), & \text{if } \zeta_{BS} = 1, \\ 0, & \text{if } \zeta_{BS} = 0. \end{cases} \tag{23}$$

- (ii) The average total power consumption of the power supply circuit can be expressed as

$$P_c = \begin{cases} N_b P_1 (1 - P_s), & \text{if } \zeta_{\text{BS}} = 1, \\ N_b P_s P_2, & \text{if } \zeta_{\text{BS}} = 0, \end{cases} \quad (24)$$

where ζ_{BS} is the active state of the base station, ρ reflects the impact of power amplifier, P is the transmit power of the BS in active mode, P_1 is the supply power of the BS in active mode, and P_2 is the circuit power of the BS in sleeping mode.

- (iii) The average caching power consumption: in this paper, we consider DRAM as cache hardware. According to the power model in [33], if the power coefficient of the cache hardware is fixed, the average caching power consumption is related to the size and the number of content cached in the local caching device. The average caching power consumption of BS can be expressed as

$$P_3 = N_c F \omega_1 = \eta F_f F \omega_1, \quad (25)$$

where ω_1 is the power coefficient of caching hardware and η is the normalized cache capacity.

- (iv) The average backhaul power consumption: we use fiber as the backhaul link (with capacity 1 Gbps). When we download contents, the backhaul link will generate power consumption, and the average backhaul power consumption of BS can be expressed as

$$P_4 = (1 - P_s) (1 - H) \omega_2 \bar{R}_2 = - (1 - P_s) \frac{\ln \eta}{\ln F_f} \omega_2 \bar{R}_2, \quad (26)$$

where $\omega_2 = P^1 / C_{\text{bh}}^1$ is the power coefficient of backhaul equipment and P^1 denotes the power consumed by the backhaul for supporting the maximum data rate C_{bh}^1 . From formula (26), we can find that the energy consumption for backhauling will reduce sharply with a little decrease of the cache size, but decreasing more and more slowly with the growth of cache size. Therefore, we need to find the best cache capacity to minimize power consumption.

Then, the average total power consumption at all BSs can be expressed as

$$\bar{P}_t = P_t + P_c + P_3 + P_4. \quad (27)$$

4.3. Energy Efficiency of the Network. Substituting (16) and (27) into (15), we can approximate the network EE as

$$\begin{aligned} \text{EE} & \\ & \approx \frac{(1 - P_s) (\bar{R}_1 \ln F_f + (\bar{R}_1 - \bar{R}_2) \ln \eta)}{(P_t + P_c) \ln F_f + \omega_1 F \eta F_f \ln F_f - (1 - P_s) \omega_2 \bar{R}_2 \ln \eta}. \end{aligned} \quad (28)$$

The goal of this paper is to maximize the energy efficiency of the downlink wireless network by finding out the optimal cache content in the local base station. Mathematically, we have the following optimization problem:

$$\begin{aligned} & \underset{\eta \in (0,1)}{\text{maximize}} && \text{EE}(\eta) \\ & \text{subject to} && \frac{\sum_{f=1}^{\eta F_f} f^{(-\lambda_u/\lambda_b)}}{\sum_{j=1}^{F_f} j^{(-\lambda_u/\lambda_b)}} < 1, \quad \forall N_c \in F_f, \\ & && C_{\text{bh}} \leq C'_{\text{bh}}, \quad \forall N_b \in \text{BS}, \\ & && \frac{P_s \omega_2 \bar{R}_2}{\omega_1 F} > \sum_{i=1}^{F_f} i^{-1}, \quad \forall N_c \in F_f, \end{aligned} \quad (29)$$

where $\sum_{f=1}^{\eta F_f} f^{(-\lambda_u/\lambda_b)} / \sum_{j=1}^{F_f} j^{(-\lambda_u/\lambda_b)} < 1$ represents the probability of the cache-hit rate, $C_{\text{bh}} \leq C'_{\text{bh}}$ is the limit of backhaul capacity, and $P_s \omega_2 \bar{R}_2 / \omega_1 F > \sum_{i=1}^{F_f} i^{-1}$ is the condition that the cache can increase energy efficiency. The following result can be obtained for the maximization of EE and the analysis is carried out later.

Proposition 1. Consider that the BS has limited cache capacity and let η_0 denote the optimal cache size that maximizes EE. Then

$$\eta_0 = \frac{Q}{F_f W(Qe^M)}, \quad (30)$$

where $W(Qe^M)$ is a Lambert-W function satisfying $f(W_{(x)}) = W_{(x)} \exp W_{(x)}$.

Proof.

- (i) We first derive the derivative of η in (28); we obtain

$$\begin{aligned} & \left. \frac{d\text{EE}}{d\eta} \right|_{\eta=\eta_0} = 0 \implies \\ & \frac{Q}{\eta_0 F_f} + \ln \frac{1}{\eta_0 F_f} - M = 0, \\ & Q = \frac{(\bar{R}_1 \bar{R}_2 / (\bar{R}_1 - \bar{R}_2)) \omega_2 (1 - P_s) + P_t + P_c}{\omega_1 F}, \end{aligned} \quad (31)$$

$$M = \frac{\bar{R}_2}{\bar{R}_1 - \bar{R}_2} \ln F_f - 1,$$

where M and Q are considered as constants.

- (ii) Let $f(\eta) = Q/\eta F_f + \ln(1/\eta F_f) - M$, and the derivative of η in the equation is

$$f'(\eta) = \frac{-(M + \eta F_f)}{F_f \eta^2}. \quad (32)$$

TABLE 1: Numerical calculation parameter settings.

Parameters	Value
Number of BSs: N_b	36.
Round radius of cells: D	40.
Number of antennas: N_t	4.
The transmission bandwidth: B	21 MHz.
The noise power: σ^2	-90.
The path-loss model: α	1.
The size of each content: F	30 MB
The size of catalog: F_f	10^5
ρ, P, P_1, P_2	15.2, 22, 3.95, 10.3
The power coefficient of caching: ω_1	6.25×10^{-7}
The power coefficient of backhaul: ω_2	5.35×10^{-5}

- (iii) So $f'(\eta) < 0$ when $\eta > 0$; then $f(\eta)$ decreases with the increase of the value η . When $\eta = \eta_0$, the equation $f(\eta) = 0$; then we can get

$$\begin{aligned} f(\eta) &< 0, & \text{if } \eta > \eta_0, \\ f(\eta) &> 0, & \text{if } \eta < \eta_0. \end{aligned} \quad (33)$$

- (iv) Since $f(\eta)$ decreases with η , we can obtain that EE monotonically increases when $\eta < \eta_0$ and EE monotonically decreases when $\eta > \eta_0$. Through the above descriptions, we find that the EE can be maximized when $\eta = \eta_0$. \square

5. Numerical and Simulation Results

In this section, we evaluate the energy efficiency of downlink networks with caching at BS. Meanwhile, we investigate the impacts of some system parameters on throughput and energy efficiency, for example, interference level β , the ratio of user density to BS density λ_u/N_b , and backhaul capacity C_{bh} .

In the work, the distribution of BS and user depends on HPPP in the whole downlink network, and the user density is set to be $\lambda_u = 35$. In order to facilitate calculation, we record the parameters in Table 1 that will be used throughout the numerical analysis.

5.1. Caching and Backhaul Power Consumption. In this section, we validate whether both the energy consumption for content update and the power consumption for transmit content influence the total power consumption, where the transmission power involves the power consumed by the core network transmitting the content to the BS through backhaul link.

- (i) Caching power consumption: in interval T , we assume that x percent of the cached dynamic contents are updated in BSs. According to (25), the percentage

TABLE 2: Cache and backhaul power consumption.

x	T	η	E_{ca}	E_{bh}
10%	12	0.01	3.5%	10.5%
15%	12	0.05	4.2%	10.9%
20%	12	0.05	4.4%	12%
20%	12	0.1	4.9%	11.6%

of caching energy consumption with respect to the average total energy consumption at all BSs during T can be expressed as

$$E_{ca} = \frac{x F \eta F_f \omega_1}{T \bar{P}_t}, \quad (34)$$

where $x F \eta F_f \omega_1$ is the caching energy consumed for updating contents. In Section 3, we consider that many contents are dynamic changes at lifetime T , and the popularity of many contents changes slowly. We set up several sets of data in Table 2 to validate the assumption that the cache energy consumption for content updates can not be ignored when the content popularity changes slowly. According to the result of E_{ca} in the table, we find that $E_{ca} > 3\%$. Then, the caching energy consumption for content update can not be ignored when content popularity is dynamic changes in lifetime T .

- (ii) Backhaul power consumption: we assume that x percent of the requested contents are transmitted over the backhaul link; the percentage of backhaul power consumption to the total power consumption in interval T is

$$E_{bh} = \frac{-x(1 - P_s)(\ln \eta / \ln F_f \omega_2 \bar{R}_2)}{T \bar{P}_t}. \quad (35)$$

According to the above parameters, we can obtain $E_{bh} > 10\%$; then we must consider the backhaul power consumption for the total power consumption in all BS.

Based on the above analysis, we have a knowledge that the power consumption of cache and backhaul can not be ignored in the total power consumption when analyzing EE.

5.2. Impact of Parameters on Throughput. We compare the simulation results of the average throughput per cell with the numerical results obtained from (16) versus C_{bh} in Figure 4. We can see that the throughput increases with increasing backhaul capacity. Moreover, the throughput increases with C_{bh} more sharply when β is small.

To show the impact of different cache size on the average throughput per cell, we provide the numerical results derived from (16) versus η in Figure 5. It can be seen when the interference level is the same and the throughput of the downlink increases with the cache capacity. Moreover, we can also find when β is small, the average throughput increases

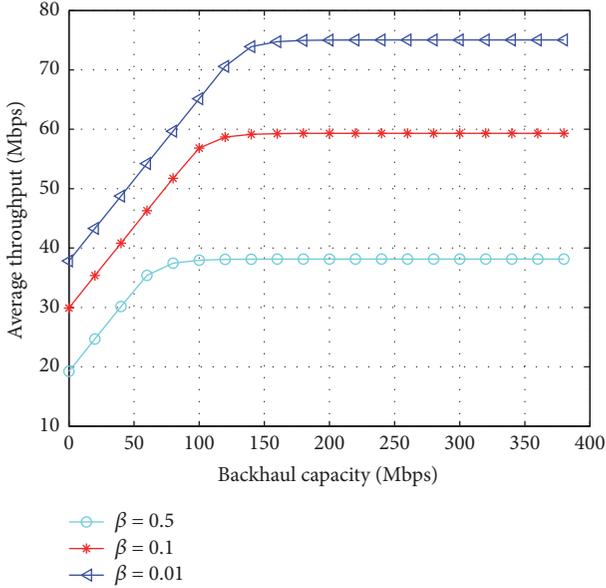


FIGURE 4: Average throughput versus backhaul capacity.

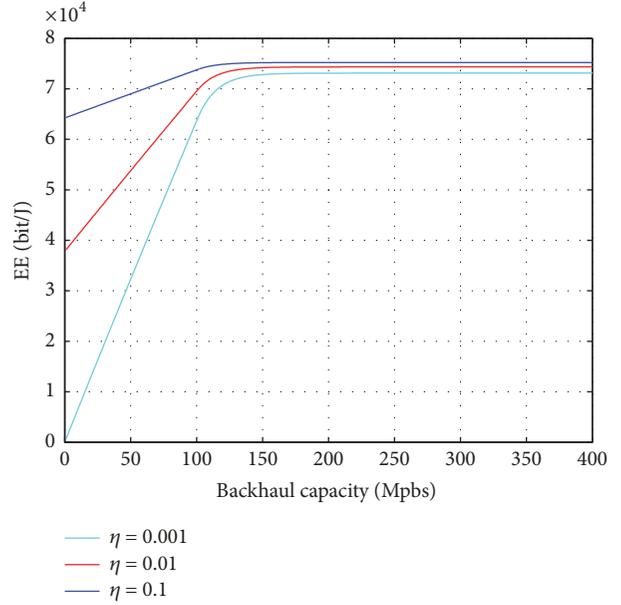


FIGURE 6: EE versus backhaul capacity.

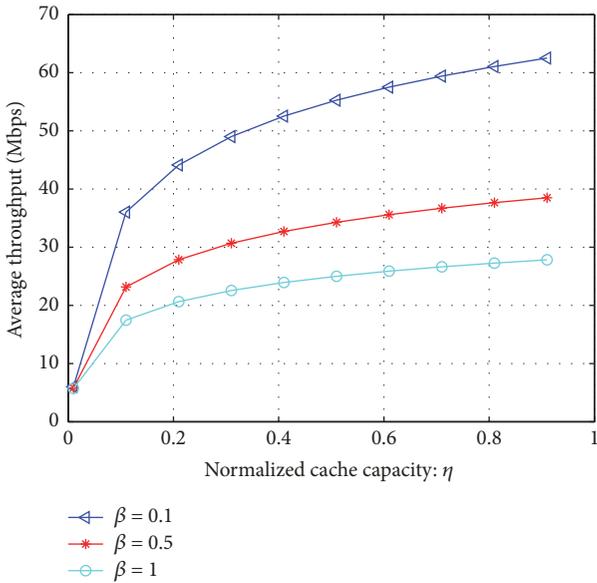


FIGURE 5: Average throughput versus cache capacity.

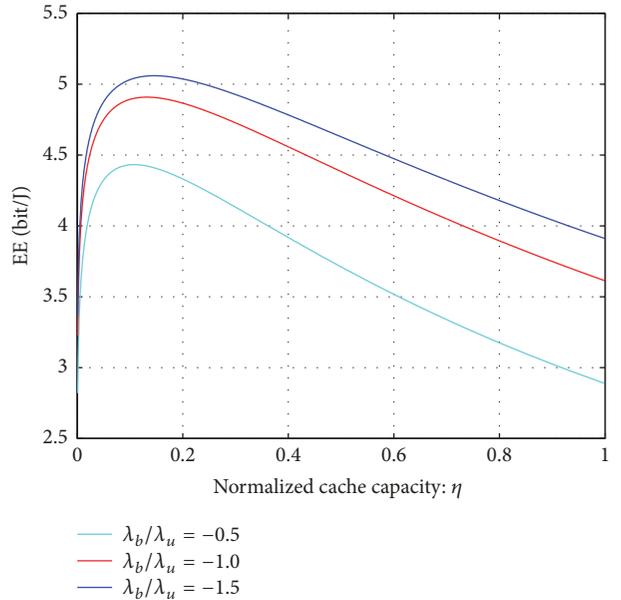


FIGURE 7: EE versus cache capacity η .

with η more sharply. This suggests that the throughput can be boosted more efficiently by caching at the BS if the ICI level can be reduced.

5.3. Impact of Parameters on EE. According to the derivation of the optimal cache capacity for EE maximization in Section 3, we derive numerical results of EE with both backhaul capacity and cache capacity. As shown in Figure 6, when only a small portion of contents are cached at each BS, the EE firstly increases rapidly with the increasing backhaul capacity (i.e., $\eta = 0.001$); then the performance gain approaches a constant. This is due to the fact that though the throughput is increased

and the backhaul power consumption is reduced by caching, ICI limits the final throughput.

In Figure 7, we show the numerical results of EE obtained from (28) versus the normalized cache capacity with different ratios of BS density to user density. It can be seen that the EE firstly increase with η quickly and then decreases gradually. This suggests that there is an optimal normalized cache capacity that maximizes EE. Moreover, with the same cache capacity, EE increases with λ_b/λ_u .

In Section 3, we propose a distributed caching strategy to cache different contents at four adjacent BSs. Figure 8 shows the simulation results of EE for nondistributed caching

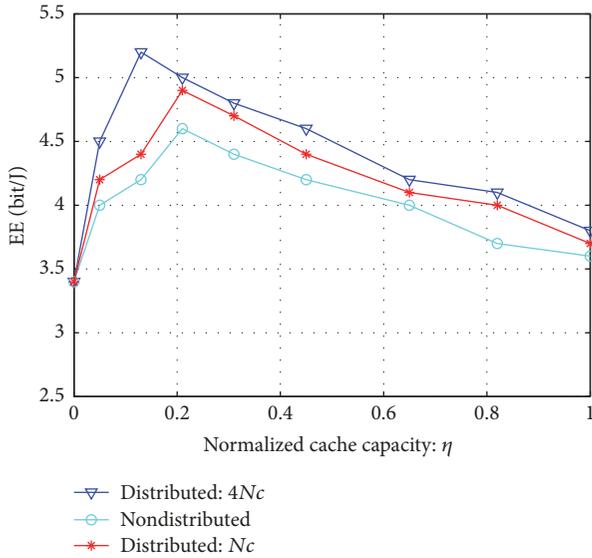


FIGURE 8: Impact of user with distributed caching.

strategy, distributed caching strategy which caches different contents at two adjacent base stations, and distributed caching strategy which caches different contents at four neighbouring BSs. We can see that distributed caching can achieve higher EE as shown in the figure. This is because when each BS uses a distributed cache strategy to cache N_c contents, each user can access $4N_c$ cache contents; that is, the cache-hit rate is increased. No matter what kind of cache policy is adopted, energy efficiency firstly increases and then decreases, which indicates the existence of an optimal cache capacity to maximize energy efficiency.

6. Conclusion

In this paper, we have investigated the energy efficient caching problem in backhaul-aware cellular networks jointly considering the cache-hit rate and the optimal cache. Considering that the content popularity is dynamic, we have proposed a policy with “age” threshold based on the shot noise model to timely estimate content popularity. Then, we have further proposed a distributed caching policy to improve the hit rate. Finally, based on the analysis of the tradeoff between energy efficiency and cache capacity, we have formulated an optimization problem and have derived a closed-form solution of optimal cache capacity for maximizing the EE. The numerical results show that the EE can be improved with appropriate choice of cache capacity. Meanwhile, compared with the nondistributed cache scheme, the distributed cache scheme can further improve the energy efficiency by increasing the cache-hit rate.

Notations

λ_b, λ_u : Distribution density of BSs and users
 N_b : Number of BSs
 N_t : Number of antennas
 F_f : Content catalog
 N_C : Number of content cached at the i th BS

F : Each content size (in bits)
 C_{ca}, C_{bh} : Limited cache capacity and backhaul capacity
 D : Round radius of cells
 r_{ub} : Distance from BS to user
 h_{ub} : Rayleigh fading channel vector
 P : Transmit power of each BS
 P_1 : The supply power of BS in active mode
 P_2 : Circuit power of the BS in sleeping mode
 σ^2 : White Gaussian noise
 β : The percentage of intercell interference
 α : Path-loss exponent
 η : Normalized cache size
 ω_1, ω_2 : Power consumption per bit (in Watt/bit)
 λ_f : Average arrival rate
 μ_f : The request arrival rate of a content f at time t
 $x_f(t) = 1$: Represent that content f is cached in BS at time t
 τ_f : The “age” of active content f at time t .

Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

Acknowledgments

This work is supported by Natural Science Foundation of Jiangsu Province (BK20170805, BK20160874, and BK20160812), Natural Science Foundation of China (61701230), and by research funding (2016-PYS/K-KY-J061).

References

- [1] K. Wang, Z. Chen, and H. Liu, “Modeling and analysis of push-based wireless converged networks,” in *Proceedings of the 2013 IEEE Globecom Workshops, GC Wkshps 2013*, pp. 304–309, USA, December 2013.
- [2] C. Xu, C. Gao, Z. Zhou, Z. Chang, and Y. Jia, “Social network-based content delivery in device-to-device underlay cellular networks using matching theory,” *IEEE Access*, vol. 5, pp. 924–937, 2017.
- [3] Z. Chang, Y. Gu, Z. Han, X. Chen, and T. Ristaniemi, “Context-aware data caching for 5G heterogeneous small cells networks,” in *Proceedings of the 2016 IEEE International Conference on Communications, ICC 2016*, Malaysia, May 2016.
- [4] S. Yunas, M. Valkama, and J. Niemelä, “Spectral and energy efficiency of ultra-dense networks under different deployment strategies,” *IEEE Communications Magazine*, vol. 53, no. 1, pp. 90–100, 2015.
- [5] N. Golrezaei, A. Molisch, A. G. Dimakis, and G. Caire, “Femto-caching and device-to-device collaboration: A new architecture for wireless video distribution,” *IEEE Communications Magazine*, vol. 51, no. 4, pp. 142–149, 2013.
- [6] S. Woo, E. Jeong, S. Park, J. Lee, S. Ihm, and K. Park, “Comparison of caching strategies in modern cellular backhaul networks,”

- in *Proceedings of the 11th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '13)*, pp. 319–332, Taipei, Taiwan, June 2013.
- [7] E. Bastug, M. Bennis, and M. Debbah, “Living on the edge: the role of proactive caching in 5G wireless networks,” *IEEE Communications Magazine*, vol. 52, no. 8, pp. 82–89, 2014.
- [8] B. A. Ramanan, L. M. Drabeck, M. Haner, N. Nithi, T. E. Klein, and C. Sawkar, “Cacheability analysis of HTTP traffic in an operational LTE network,” in *Proceedings of the 12th Annual Wireless Telecommunications Symposium: Global Wireless Communications - Future Directions, WTS 2013*, USA, April 2013.
- [9] X. Wang, M. Chen, T. Taleb, A. Ksentini, and V. C. M. Leung, “Cache in the air: exploiting content caching and delivery techniques for 5G systems,” *IEEE Communications Magazine*, vol. 52, no. 2, pp. 131–139, 2014.
- [10] M. Ji, G. Caire, and A. F. Molisch, “Wireless device-to-device caching networks: basic principles and system performance,” *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 1, pp. 176–189, 2016.
- [11] K. Poularakis, G. Iosifidis, V. Sourlas, and L. Tassiulas, “Multicast-aware caching for small cell networks,” in *Proceedings of the 2014 IEEE Wireless Communications and Networking Conference, WCNC 2014*, pp. 2300–2305, Turkey, April 2014.
- [12] D. Liu and C. Yang, “Energy Efficiency of Downlink Networks with Caching at Base Stations,” *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 907–922, 2016.
- [13] Y. Xu, Y. Li, Z. Wang, T. Lin, G. Zhang, and S. Ci, “Coordinated caching model for minimizing energy consumption in radio access network,” in *Proceedings of the 2014 1st IEEE International Conference on Communications, ICC 2014*, pp. 2406–2411, Australia, June 2014.
- [14] A. Liu and V. K. Lau, “Mixed-timescale precoding and cache control in cached MIMO interference network,” *IEEE Transactions on Signal Processing*, vol. 61, no. 24, pp. 6320–6332, 2013.
- [15] M. Dehghan, A. Seetharam, B. Jiang et al., “On the complexity of optimal routing and content caching in heterogeneous networks,” in *Proceedings of the 34th IEEE Annual Conference on Computer Communications and Networks, IEEE INFOCOM 2015*, pp. 936–944, Hong Kong, May 2015.
- [16] J. Hachem, N. Karamchandani, and S. Diggavi, “Content caching and delivery over heterogeneous wireless networks,” in *Proceedings of the 34th IEEE Annual Conference on Computer Communications and Networks, IEEE INFOCOM 2015*, pp. 756–764, Hong Kong, May 2015.
- [17] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, “Web caching and Zipf-like distributions: evidence and implications,” in *Proceedings of the 18th Annual Joint Conference of the IEEE Computer and Communications Societies (INFOCOM '99)*, pp. 126–134, March 1999.
- [18] C. Li, J. Zhang, and K. B. Letaief, “Throughput and energy efficiency analysis of small cell networks with multi-antenna base stations,” *IEEE Transactions on Wireless Communications*, vol. 13, no. 5, pp. 2505–2517, 2014.
- [19] G. Y. Li, Z. Xu, C. Xiong et al., “Energy-efficient wireless communications: tutorial, survey, and open issues,” *IEEE Wireless Communications Magazine*, vol. 18, no. 6, pp. 28–34, 2011.
- [20] S. Traverso, M. Ahmed, M. Garetto, P. Giaccone, E. Leonardi, and S. Niccolini, “Temporal locality in today’s content caching: Why it matters & how to model it,” *Sigcom Computer. Communication. Review*, vol. 43, no. 5, pp. 5–12, 2013.
- [21] H. Kim, J. Park, M. Bennis, S.-L. Kim, and M. Debbah, “Ultra-dense edge caching under spatio-temporal demand and network dynamics,” in *Proceedings of the 2017 IEEE International Conference on Communications, ICC 2017*, France, May 2017.
- [22] B. Dai and W. Yu, “Energy Efficiency of Downlink Transmission Strategies for Cloud Radio Access Networks,” *IEEE Journal on Selected Areas in Communications*, vol. 34, no. 4, pp. 1037–1050, 2016.
- [23] F. Olmos, B. Kauffmann, A. Simonian, and Y. Carlinet, “Catalog dynamics: Impact of content publishing and perishing on the performance of a LRU cache,” in *Proceedings of the 2014 26th International Teletraffic Congress, ITC 2014*, swe, September 2014.
- [24] S. Traverso, M. Ahmed, M. Garetto, and P. Giaccone, “Temporal locality in today’s content caching: Why it matters & how to model it,” *ACM SIGCOMM Computer Communication Review*, vol. 43, no. 5, 2013.
- [25] M. Zink, K. Suh, Y. Gu, and J. Kurose, “Watch global, cache local: YouTube network traffic at a campus network - Measurements and implications,” in *Proceedings of the Multimedia Computing and Networking Conference, MMCN 2008*, USA, January 2008.
- [26] S. Scellato, C. Mascolo, M. Musolesi, and J. Crowcroft, “Track globally, deliver locally: Improving content delivery networks by tracking geographic social cascades,” in *Proceedings of the 20th International Conference on World Wide Web, WWW 2011*, pp. 457–466, India, April 2011.
- [27] K. Huguenin, A. Kermarrec, K. Kloudas, and F. Taïani, “Content and geographical locality in user-generated content sharing systems,” in *Proceedings of the 22nd International Workshop on Network and Operating System Support for Digital Audio and Video (NOSSDAV '12)*, pp. 77–82, Toronto, Canada, June 2012.
- [28] M. Lelarge, L. Massoulie, and J. Xu, “Reconstruction in the labeled stochastic block model,” in *Proceedings of the 2013 IEEE Information Theory Workshop, ITW 2013*, esp, September 2013.
- [29] J. Liu, V. Yadav, H. Sehgal, J. M. Olson, H. Liu, and N. Elia, “Phase transitions on fixed connected graphs and random graphs in the presence of noise,” *Institute of Electrical and Electronics Engineers Transactions on Automatic Control*, vol. 53, no. 8, pp. 1817–1825, 2008.
- [30] D. Liu and C. Yang, “Will caching at base station improve energy efficiency of downlink transmission?” in *Proceedings of the 2014 IEEE Global Conference on Signal and Information Processing, GlobalSIP 2014*, pp. 173–177, Atlanta, Ga, USA, December 2014.
- [31] S. Pant, N. Chauhan, and P. Kumar, “Effective Cache based Policies in Wireless Sensor Network: A Survey,” *IEEE International Journal of Computer Application*, vol. 11, no. 10, 2010.
- [32] C. Yang, Z. Chen, Y. Yao, B. Xia, and H. Liu, “Energy efficiency in wireless cooperative caching networks,” in *Proceedings of the 2014 1st IEEE International Conference on Communications, ICC 2014*, pp. 4975–4980, Australia, June 2014.
- [33] Z. Yan, S. Chen, Y. Ou, and H. Liu, “Energy Efficiency Analysis of Cache-Enabled Two-Tier HetNets under Different Spectrum Deployment Strategies,” *IEEE Access*, vol. 5, pp. 6791–6800, 2017.



Hindawi

Submit your manuscripts at
www.hindawi.com

