# Green and Mobility-Aware Caching in 5G Networks

Min Chen, *Senior Member, IEEE*, Yixue Hao, Long Hu, Kaibin Huang, *Senior Member, IEEE*, and Vincent K. N. Lau, *Fellow, IEEE*

***Abstract*— With the drastic increase of mobile devices, there are more and more mobile traffic and repeated requests for content. In 5G networks, small cell base stations (SBSs) caching and caching in wireless device-to-device network can effectively decrease the mobile traffic during peak hours. Currently, most of the related work is focused on how to cache content on SBSs and on mobile devices, and it is assumed that the user can download the entire requested content through the connected SBSs and mobile devices. However, few works have taken user mobility and the randomness of contact duration into consideration. How to improve the caching strategy by exploiting user mobility is still a challenging problem. Thus, in this paper, we first investigate the problem of how to conduct caching placement on SBS and on mobile devices leveraging user mobility, aiming to maximize the cache hit ratio. Specifically, the caching placement on SBSs and on mobile devices is formulated as an integer programming problem, and submodular optimization is adopted to solve the formulated problem. Then, we give the optimal transmission power of SBSs and mobile devices to deliver the caching content in order to reduce the energy cost. Simulation results prove that our caching strategy is more efficient than other existing caching strategies in terms of both cache hit ratio and energy efficiency.**

***Index Terms*— Content caching, device-to-device communications, 5G networks, human mobility, cache hit ratio.**

## I. Introduction

THE he demand for the capacity of wireless networks has been increasing exponentially over the past few years. According to a report by Cisco [1], the global mobile traffic will exceed 24.3 exabytes per month. As such, it is expected that the future fifth generation (5G) wireless networks have to support 1000x higher capacity and 100x lower delay. In particular, densification of wireless networks using small cells have received tremendous interests in the research communities as an effective means to meet the challenging goals.

Manuscript received November 16, 2016; revised April 5, 2017 and July 19, 2017; accepted September 20, 2017. Date of publication October 13, 2017; date of current version December 8, 2017. This work was supported in part by RGC 16204814, in part by the Applied Basic Research Program through Wuhan Science and Technology Bureau under Grant 2017010201010118, and in part by the National Natural Science Foundation of China under Grant 61572220. The work of K. Huang was supported by Hong Kong Research Grants Council under Grant 17209917 and Grant 17259416. The associate editor coordinating the review of this paper and approving it for publication was W. Chen. *(Corresponding author: Long Hu.)*

M. Chen is with the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China, and also with the Wuhan National Laboratory for Optoelectronics, Wuhan 430074, China (e-mail: minchen@ieee.org).

Y. Hao and L. Hu are with the School of Computer Science and Technology, Huazhong University of Science and Technology, Wuhan 430074, China.

K. Huang and V. K. N. Lau are with the Department of Electrical and Electronic Engineering, The University of Hong Kong, Hong Kong (e-mail: huangkb@eee.hku.hk; eeknlau@ece.ust.hk).

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TWC.2017.2760830

5G networks [2] mainly include macrocell base stations (MBSs) and small cell base stations (SBSs). By bringing the base stations closer to the device in small cell networks, the spectral efficiency can be significantly enhanced due to increased spatial reuse. However, such benefit comes at the expense of potentially intense interference and high backhaul cost. Specifically, as the density of the SBS increases, the inter-cell interference becomes more severe, which limits the spectral efficiency of the ulta-dense small cell networks. To cope with the interference issue, cooperative MIMO or, Cooperative Multi-Point (CoMP) techniques have been proposed [3]. By sharing the payload between SBSs, they can cooperatively serve multiple users and therefore eliminating cross-cell interference. However, CoMP will increase the backhaul load due to payload sharing between the SBS. Since the number of backhaul scales with the number of SBSs, the backhaul can easily become the system bottleneck in small cell networks.

On the other hand, it is reported that most of the capacity demand will originate from high quality multimedia streaming applications [4]–[8]. A few popular contents may be repeatedly requested by nearby users within a window of time. Unlike generic data applications, these content-centric applications are cachable. If the requested contents are cached at the radio access network, i.e., at the SBS or at the devices, they may be requested by nearby users in the immediate future. In this case, the new requests can be served by the cache instead of from the core network via the backhaul. Caching at the radio access networks has the potential to significantly reduce the loading of the backhaul and can alleviate the backhaul bottleneck in ultra-dense small cell networks.

There are a number of recent works on caching in wireless networks. In [9], femto-caching has been proposed. In [10], [11], physical layer (PHY) caching has been proposed to exploit the BS cache to exploit CoMP opportunities. Specifically, when the requested contents of the users exist in the caches of multiple BSs simultaneously, the BSs can engage in CoMP without inducing extra loading to the backhaul for payload sharing. In [12], algorithms for distributed caching are proposed for the data dissemination in the downlink of heterogeneous networks. Furthermore, in [13] and [14], the optimal cache scheme was designed in terms of a small cell video caching system, and a data-driven cache strategy was proposed in [15]. In [16]–[18], caching at the devices have been proposed to enhance the capacity of D2D networks. It has been recognized that popular content should be cached on mobile devices and throughput of network may be improved by D2D communications. However, these works are based on the assumption that user remains static, i.e., assumed fixed network topology.

Since user mobility in a caching network is a very critical problem, there are some preliminary works considering user mobility. These works can be classified into two categories, i.e., cache in SBS, and cache in mobile device. *Cache in SBS.* Due to the dense deployment, the handovers between mobile device and SBSs are quite frequent. Such mobility should be considered during the design of a caching policy. Wang *et al.* [19] discuss the femtocaching without file coding under user mobility. Poularakis *et al.* [20] tried to minimize the workload at MBS by making strategy of content caching at SBS when the user mobility is considered. Mobilecacher in [21] describes caching placement which is mobility-aware in SBSs. *Cache in mobile device.* Through transmission opportunity by D2D links, the D2D caching delivers content to contacted user by opportunistic connections. Lan *et al.* [22] proposed mobility-aware caching in D2D networks. Wang *et al.* [23], [24] further discussed the mobility-aware coded caching strategy in D2D network.

However, all of above work assume that the whole cached file [22] or encoded segments [20], [24] can be successfully delivered within the contact duration when mobile device request file from SBS or other mobile device. However, according to the study in [25], the transmission data is related to the contact duration. Thus, takeing into account the contact duration randomness caused by user mobility, the above assumptions are not practical. Considering the caching system designed is consisted of two separated phases [26], i.e., content placement phase and delivery phase. Thus, the challenges of caching content on SBSs and on mobile devices under condition of the randomness of contact duration are mainly two-fold: *mobility-aware content placement* and *green content delivery*. For the former, we need to optimally determine how to cache in the mobile device and the SBSs in order to maximize the cache hit ratio. For the latter, how can we improve the energy efficiency of 5G network during the content delivery.

In order to overcome the above challenges, we present the green and mobility-aware caching model for SBSs and mobile devices, which includes mobility-aware content placement model and green content delivery model. For mobility-aware caching model, we derive a solution to maximize the cache hit ratio on the SBS and mobile devices by using submodular optimization. For green content delivery model, we analyze the optimal transmission power of SBSs and mobile devices.

In summary, the contributions of this paper include as follows.

- *How to Cache:* we considered how to conduct caching on SBSs and mobile device leveraging user mobility, aiming to maximize the cache hit ratio. This problem is formulated as a mixed integer nonlinear programming (MINLP) problem, which is NP-hard. Submodular optimization is introduced and a greedy algorithm is developed to solve this problem.
- *How to Delivery:* we analyzed the energy efficiency when SBS and mobile device deliver the cached content under condition of the randomness of contact duration. Furthermore, we put forward the optimal transmit power of SBSs and mobile device in order to reduce the delivery energy cost.
- *Performance Analysis:* Simulation results prove that our caching strategy is more efficient than other existing caching strategies in terms of both cache hit ratio and energy efficiency. Through the experiments, we find that more popular contents are cached on SBSs and mobile devices when user mobility is low, when the mobility of users is high, it is necessary to consider the diversity of files.

The remainder of this article is organized as follows. The system model is described in Section II. We present the green and mobility-aware caching scheme in Section III and propose the solution for green and mobility-aware caching strategy optimization in Section IV. Our experimental results and discussions are provided in Section V. Finally, Section VI concludes this paper.

## II. System Model

In this paper, we consider a 5G network with macro cell base station (MBS), small cell base stations (SBSs), and mobile devices. A specific scenario is illustrated as Fig. 1, there are five mobile device users *Rachel*, *Eva*, *Tommy*, *Suri* and *Cindy*. The file requested by Rachel includes four encoded segments, i.e., $s_1$, $s_2$, $s_3$, $s_4$. Considering the limitation of storage capacity at user's mobile device, we assume each user only stores one coded segment. Given an example shown in Fig. 1, *Eva*, *Tommy*, *Suri* and *Cindy* store $s_1$, $s_2$, $s_3$ and $s_4$, respectively. Considering user mobility to make a caching strategy, *Rachel* obtains $s_1$, $s_2$ and $s_3$ from *Eva*, *Tommy*, *Suri*, while retrieving $s_4$ from SBS2. The example shows the reduction of the traffic load at MBS. From the above discuss, it is observed that user mobility has significant influence on the content cache of mobile device and SBS. Below we will propose a model for how to cache on mobile device and SBSs according to the user mobility. Denote by $l$ and $n$ the number of SBS and mobile device, respectively. Let $\mathcal{S} = \{S_1, S_2, \cdots, S_l\}$ and $\mathcal{D} = \{D_1, D_2, \cdots, D_n\}$ be the set of SBSs and mobile devices, respectively. The MBS can connect to the SBSs and mobile devices, and each user requests the content independently. The main notations used in this paper are given in Table I.

### A. Mobility Model

The mobility model to be used in this paper is a peer-to-peer connectivity model [27] which is widely investigated in wireless network [23]–[25] and delay tolerant network [28]. Similar to [23], [25], [27], we assume pairwise contact process is independent Poisson process.

*1) The Mobility Model for Users in the Small Cell Network:* The condition for successful communication between mobile device $D_i$ and SBS $S_k$ is that $D_i$ is within the communication radius of $S_k$. We define the contact duration $T_{i,k}$ when mobile device $D_i$ is within communication range of SBS $S_k$ as $T_{i,k} = \{(t - t_0) : ||\mathcal{L}_k^t - \mathcal{L}_i^t|| < R_S, t > t_0\}$, where $t_0$ stands for the most recent time when mobile device $D_i$ just

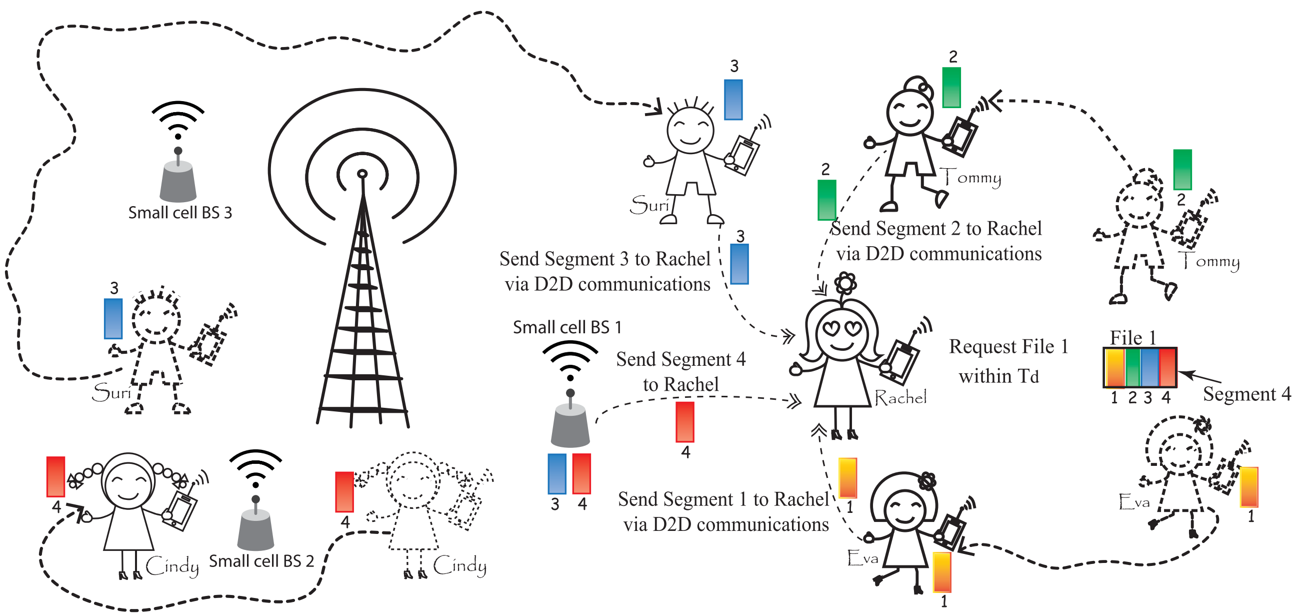



Fig. 1. An illustration of mobility-aware caching in 5G networks.

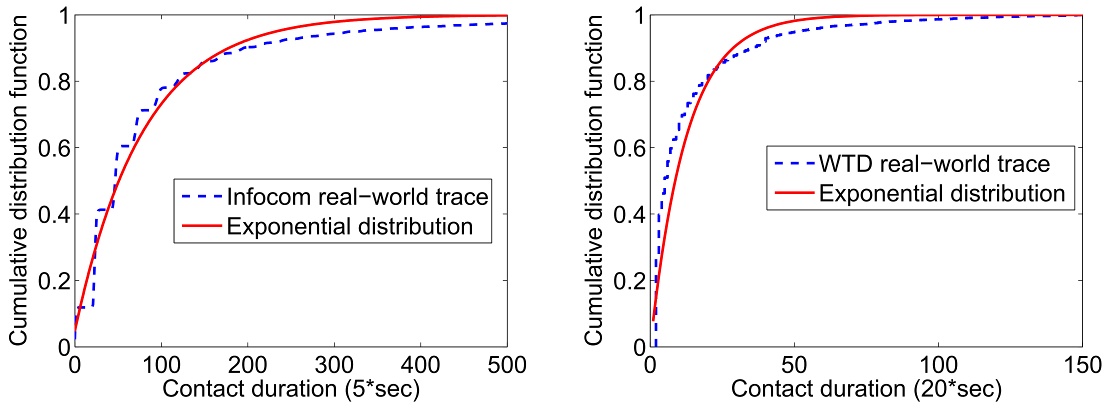
TABLE I
LIST OF COMMONLY USED NOTATIONS

| Notation | Explanation |
|---|---|
| $\mathcal{S}$ | The set of SBSs. |
| $\mathcal{D}$ | The set of mobile devices. |
| $\mathcal{F}$ | The set of files. |
| $C_k^S$ | Cache size of SBS $S_k$. |
| $C_i^D$ | Cache size of mobile device $D_i$. |
| $p_f$ | Probability that a request for file $F_f$. |
| $\lambda_{i,k}^B$ | Exponential distribution parameter of pairwise contact duration between mobile device $D_i$ and SBS $S_k$. |
| $\lambda_{i,j}^D$ | Exponential distribution parameter of pairwise contact duration between mobile device $D_i$ and $D_j$. |
| $\mathcal{A}_{i,k}$ | The amount of data delivered can be sent from SBS $S_k$ to mobile device $D_i$ during a contact duration. |
| $\mathcal{B}_{i,j}$ | The amount of data delivered can be sent from mobile device $D_i$ to $D_j$ during a contact duration. |
| $T_d$ | The deadline for user requests. |
| $P_T^D$ | The transmission power of mobile device. |
| $P_T^B$ | The transmission power of SBS. |
| $P_C^D$ | The circuit power consumption of the mobile device. |
| $P_C^B$ | The constant power of SBS. |
| $W_D$ | The D2D channel bandwidth. |
| $W_B$ | The downlink transmission bandwidth from SBS to mobile device. |

enter the communication range $R_S$ of SBS $S_k$. $\mathcal{L}_k^t$ and $\mathcal{L}_i^t$ stand for the locations of SBS $S_k$ and mobile device $D_i$ at time $t$, respectively. Since the contacts process between SBS and mobile devices follows independent Poisson processes, the pairwise contact duration $T_{i,k}$ of mobile device $D_i$ and SBS $S_k$ follows the exponential distribution with parameter $\lambda_{i,k}^B$, where $\lambda_{i,k}^B$ is named as the contact rate of mobile device $D_i$ and SBS $S_k$. In consideration of limited storage capacity of SBSs, we define cache size of SBS $S_k$ as $C_k^S$.

*2) The Mobility Model for Users in the D2D Network:* Communication can only be successful when the shortest distance between any two mobile devices is within $R_{D2D}$. Define the contact duration $T_{i,j}$ between any two users $D_i$ and $D_j$ as $T_{i,j} = \{(t - t_0) : ||\mathcal{L}_i^t - \mathcal{L}_j^t|| < R_{D2D}, t > t_0\}$. Similarly, $t_0$ stands for the most recent moment when mobile device $D_i$ just enter the communication range $R_{D2D}$ of mobile device $D_j$; $\mathcal{L}_i^t$ and $\mathcal{L}_j^t$ stand for the locations of user $D_i$ and $D_j$ at moment $t$, respectively. Similar to the above, the contact duration between any two mobile device $D_i$ and $D_j$ also follows the exponential distribution with parameter $\lambda_{i,j}^D$, where $\lambda_{i,j}^D$ is termed the contact rate of mobile device $D_i$ and $D_j$. In a similar way, in consideration of limited storage capacity of mobile device, we denote the cache size of mobile device as $C_i^D$.

*3) The Mobility Pattern:* For user's mobility, we utilize the real dataset to verify that the contact duration between user and user as well as between user and SBS follows the exponential distribution. For the contact duration between user and user, the Infocom in Cambrige/Haggle dataset provided in [29] is plotted in Fig. 2(a). The comparison therein with the exponential distribution validates the current model for peer-to-peer contact duration. For the distribution of the contact




(a) Cambridge/Haggle Infocom trace (b) Wireless Topology Discovery project trace

Fig. 2. Distribution of contact duration between user and user as well as between user and SBS.

duration between user and SBS, our model of contact duration between users and SBS has been developed using the dataset from the Wireless Topology Discovery (WTD) project at UCSD [30]. Similar to [20], we treat WiFi APs in the dataset as SBS and chose 2002/10/16 as the day in the experiment. This results in the curve of the contact duration of all users with SBS as shown in Fig. 2(b). It is observed from the figure that the contact duration can be also suitablly modeled by the exponential distribution.

In [23], the experiments show that user contact rate is linearly proportion to the user's moving speed. A higher mobility leads to a larger contact rate. Thus, user's moving speed can be deemed as the user's contact rate. However, the contact duration is decreased accordingly with the increase of user's moving speed. When the contact duration is not enough to finish the data transmission of cached file, it will cause the failure of file transmission. Thus, it is impractical to assume 100% delivery ratio [20], [22], [24] in mobile networks. The contact duration follows exponential distribution. Thus, we assume the throughput of content delivery during the contract duration also follows exponential distribution. Let the random variable $\mathcal{A}_{i,k}^{\omega}$ denote the maximum amount of content delivered from SBS $S_k$ to $D_i$ during the $\omega$th contact duration. Let the random variable $\mathcal{B}_{i,j}^{\omega}$ denote the maximum amount of content delivered from $D_j$ to $D_i$ during the $\omega$th contact duration.

### *B. Content Request Model*

Consider a content library (also called the file library) with $m$ contents. Let $\mathcal{F} = \{F_1, F_2, \cdots, F_m\}$ be the set of all files in the content library.This paper adopts coded caching strategy, i.e. file $F_f$ can be encoded into $s_f$ encoded segments through rateless Fountain coding [31] and each encoded segments have same size. We also assume that $F_f$ can be restored through the $s_f$ code segments [23], [32]. Denote size of $F_f$ as $|F_f|$. Then, the size of each encoded segment can be calculated as $g_f = \frac{|F_f|}{s_f}$. Contents are ordered according to their popularity, i.e., ranking from the most popular file ($F_1$) to the least popular file ($F_m$). Each user requests content $F_f$ randomly and independently from the content library with a probability of $p_f$, and we assume $p_f$ follows the Zipf distribution with parameter $\gamma$ [33]. Then, We have:

$$p_f = \frac{f^{-\gamma}}{\sum_{i=1}^{m} i^{-\gamma}}, \quad f = 1, 2, \cdots, m, \tag{1}$$

where $\gamma$ indicates the uneven distribution of popularity in the contents. Let $T_d$ denote the deadline for user requests. When a user requests content with $T_d$, he or she can obtain content mainly through the following three ways as shown in Fig. 3

- **Local-Caching:** The user would check his or her own mobile device. If the required content is cached on his or her device, then the user would obtain the required content from the local cache.
- **D2D-Caching or SBS-Caching:** If the required content is not cached on the mobile device, then the user may obtain the required content through one of the following two ways within $T_d$. (i) D2D-Caching: Assume there is another mobile device within its D2D range $R_{D2D}$ that stores the required content. Then D2D communication will be established between them, to obtain the required content from the neighboring device. (ii) SBS-Caching: Assume there is a SBS within its communication range $R_S$ that stores the required content. Then SBS communication will be established between them, to obtain the required content from the SBS. In this paper, we assume that the D2D communication is assisted by the SBSs [34] and the SBSs have the global information to make caching decision for all mobile devices.
- **MBS-Caching:** If within the deadline $T_d$ for the requested content, the user cannot obtain the required content with the above-mentioned methods, then his or her request will be processed by the MBS, to obtain the required content from the MBS.

### *C. Energy Consumption Model*

In this section, we present the energy consumption model when mobile device or SBSs send content to content requester. We present the energy cost model for D2D communications and SBS communications, respectively.

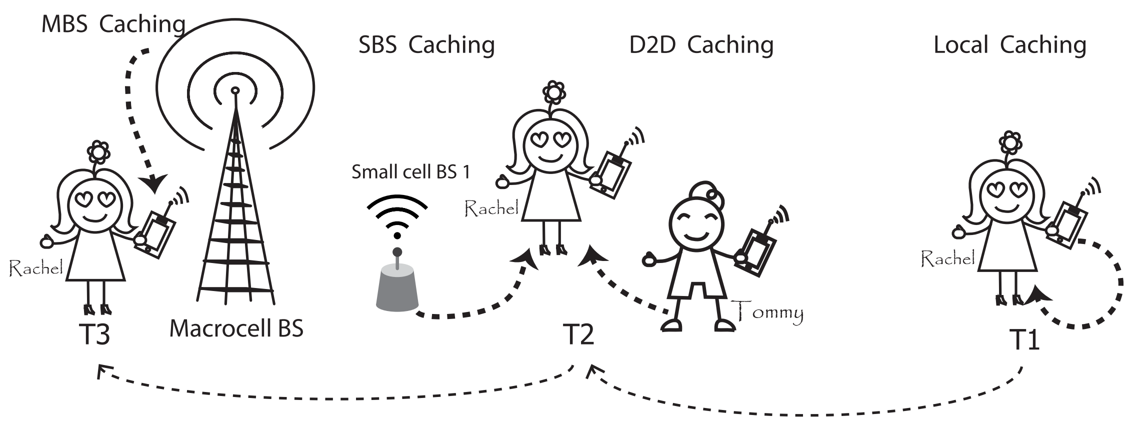



Fig. 3. Illustration of the three ways for content access: local-caching, D2D or SBS caching, and MBS caching.

*1) Energy Cost for D2D Communications:* First, we consider the energy consumption when mobile device $D_j$ transmits cached code segment to $D_i$. For the sake of simplicity, the interference of D2D communications is not considered. According to [35], the average D2D data rate can be obtained:

$$R_D = \mathbb{E}\left\{W_D \log_2\left(1+\frac{P_T^D h_D^2 r_D^{-\alpha}}{\sigma_D^2}\right)\right\} \approx W_D \log_2\left(1+\frac{P_T^D r_D^{-\alpha}}{\sigma_D^2}\right). \tag{2}$$

where $P_T^D$ is the mobile device transmission power, $W_D$ is the channel bandwidth from $D_i$ to $D_j$ $h_D$ represents channel gain, follow zero mean complex Gaussian distribution with unit variance $r_D$ is the distance between $D_i$ and $D_j$. $\sigma_D^2$ denotes white Gaussian noise $\alpha$ is path loss factor.

According to [36], the power consumption of $D_j$ can be modeled as $\beta_D P_T^D + P_C^D + P_H^D$, where $P_C^D$ represents the circuit power consumption of the mobile device $D_j$, and $\beta_D$ is the inverse of power amplifier efficiency factor. $P_H^D$ represents the energy consumption of caching hardware devices. In this paper, we ignore the energy consumed for delivering the cached content. Our focus is to obtain the optimal transmission power within contact duration. Thus, when $D_j$ transmits file $\mathcal{B}_{i,j}^{\omega}$ to $D_i$, the corresponding energy consumption of $D_j$ can be calculated as:

$$E_D = \frac{\mathcal{B}_{i,j}^{\omega}}{R_D}(\beta_D P_T^D + P_C^D), \tag{3}$$

*2) Energy Cost for SBS:* Here, we investigate the energy consumed SBS $S_k$ when $S_k$ delivers coded segment to $D_i$. Similar as before, we do no consider the interference between SBSs. Denoted $P_T^B$ as the SBS transmission power. Then, the downlink speed of mobile device can be calculated as:

$$R_B = \mathbb{E}\left\{W_B \log_2\left(1+\frac{P_T^B h_B^2 r_B^{-\beta}}{\sigma_B^2}\right)\right\} \approx W_B \log_2\left(1+\frac{P_T^B r_B^{-\beta}}{\sigma_B^2}\right), \tag{4}$$

where $P_T^B$ is the SBS transmission power, $W_B$ is the downlink transmission bandwidth from SBS to mobile device, $h_B$ is the channel power gain, follow zero mean complex Gaussian distribution with unit variance $r_B$ represents the distance between $D_i$ and $S_k$, and $\sigma_B^2$ is the variance of additive white Gaussian noise, $\beta$ means the path loss factor. Similarly, we ignore the energy consumption for file storage in SBSs. According to [36], when SBS send file $\mathcal{A}_{i,k}$ to $D_i$, the energy consumption at the SBS can be calculated as:

$$E_B = \frac{\mathcal{A}_{i,k}^{\omega}}{R_B}(\beta_B P_T^B + P_C^B), \tag{5}$$

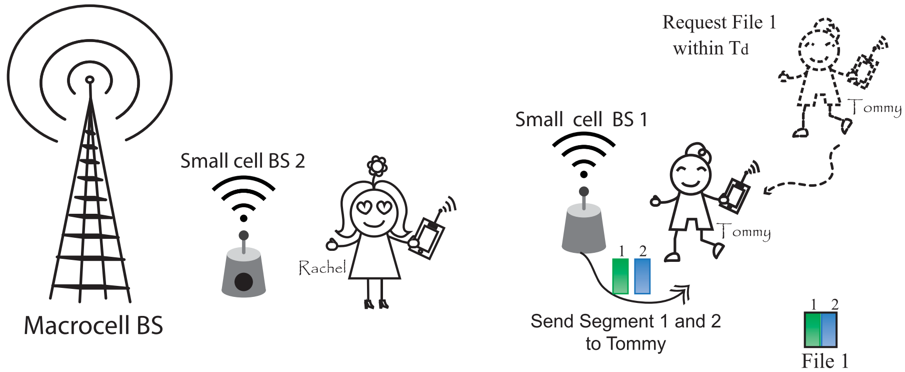
where $P_C^B$ denotes an offset of site power including the baseband processor, the cooling system and etc, and $\beta_B$ is the inverse of power amplifier efficiency factor.

## III. Green and Mobility-Aware Caching Model

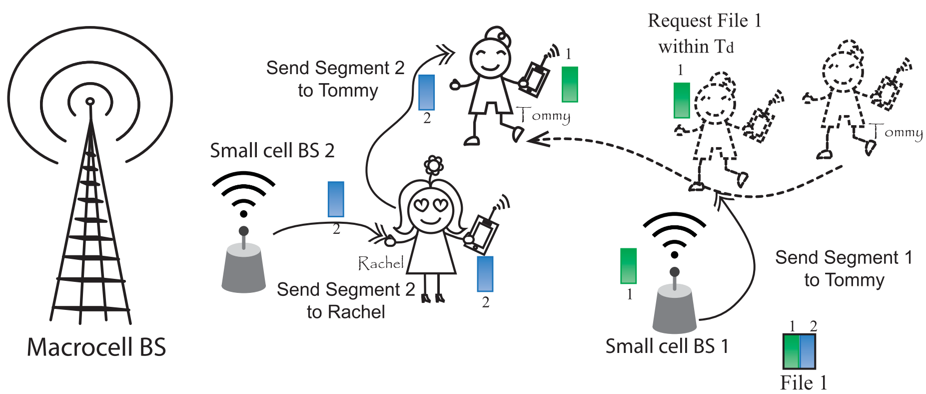
In this section, we present green and mobility-aware caching model. Considering the randomness of contact duration caused by user mobility, we need to decide which contents should be cached at SBSs and mobile devices, in order to maximize the cache hit ratio (the number of requests delivered by the cache server, divided by the total number of requests). Furthermore, in order to ensure the green of 5G network, we describe how to reduce the energy consumption during content delivery.

### A. Motivation

We illustrate how to utilize user mobility with an example, and then design the content placement and delivery for SBSs and mobile devices, as shown in Fig. 4. In Fig. 4(a), user *Tommy* requests content includes two code segments with deadline $T_d$. Due to the low mobility of *Tommy*, the number of contacts with SBS and other users is relatively small, yet the contact duration is long. As shown in Fig. 4, *Tommy* obtained requested file from SBS1. For this case, the code are cached at SBS1. In Fig. 4(b), likewise, user *Tommy* requests content including two code segments with deadline $T_d$. However, the mobility is high for this time. Thus, *Tommy* only finishes the transmission of one code segment during the contact with SBS1. In $T_d$, *Tommy* meets *Rachel* and obtain code segment from *Rachel*'s cache. Thus, the caching strategy is more complicated, since both SBS1 and *Rachel* contribute requested code segments with *Tommy*'s personalized moving speed. In the delivery phase, the SBS1 and user *Rachel* deliver the content to user *Tommy* with the minimum energy consumption,




(a) Case of low user mobility.




(b) Case of high user mobility.

Fig. 4. Illustration for the impact of user mobility on caching.

when they are in the contact duration. So, the cache strategy are closely related to contact duration. Green and mobility-aware caching model which included mobility-aware content placement model and delivery model is shown below.

### B. Mobility-Aware Content Placement Model

Let matrix $\mathbf{X}_{l\times m}$ define the caching strategy of the encoded segments in SBS, where $x_{k,f} \in \mathbf{X}$ represents the number of code segments in SBS $S_k$. Let matrix $\mathbf{Y}_{n\times m}$ denote the caching strategy of the encoded segments in mobile device, where $y_{j,f} \in \mathbf{Y}$ is the number of code segments cached at mobile device $D_j$. Define matrix $\mathbf{Z} = [\mathbf{X}; \mathbf{Y}]$ as the caching strategy matrix, which needs to be solved out. Let $\mathcal{U}_i^f(\mathbf{Z})$ denote the total amount of code segments that $D_i$ can obtain from SBS and mobile device within $T_d$. Thus, the caching strategy with maximal cache hit ratio can be expressed as

$$\underset{\mathbf{Z}}{\text{maximize}}\ \frac{1}{n}\sum_{i=1}^{n}\sum_{f=1}^{m} p_f Pr(\mathcal{U}_i^f(\mathbf{Z}) \geq s_f) \tag{6a}$$

$$\text{subject to}\ \sum_{f=1}^{m} x_{k,f}g_f \leq C_k^S, \quad \forall k \in \{1,\cdots,l\}, \tag{6b}$$

$$\sum_{f=1}^{m} y_{j,f}g_f \leq C_j^D, \quad \forall j \in \{1,\cdots,n\}, \tag{6c}$$

$$x_{k,f} \in \{0,1,\cdots,s_f\}, \quad \forall k, f \tag{6d}$$

$$y_{j,f} \in \{0,1,\cdots,s_f\}, \quad \forall j, f. \tag{6e}$$

where the objective function (6a) represents how to cache in SBS and mobile device, so as to maximize the cache hit ratio. Two constraints (6b) and (6c) mean the total size of cached files cannot exceed the cache capacities of SBS and mobile devices, respectively. Constraints (6d) and (6e) represents the number of cache is integer.

Next, we will give the solution to $Pr(\mathcal{U}_i^f(\mathbf{Z}) \geq s_f)$. According to the mobility model discussed in Section II, the contact number between mobile device with SBS or other mobile device follows Poisson distribution. Within time $T_d$, let $M_{i,k}$ denote the number of contacts between $D_i$ and $S_k$; let $N_{i,j}$ denote the number of contacts between $D_i$ and $D_j$. Both $M_{i,k}$ and $N_{i,j}$ are random variables following Poisson distribution. Then, the total size of file $F_f$, which $D_i$ can obtained within $T_d$ from $D_j$ and $S_k$, can be calculated as:

$$V_{i,j,k}^f = \sum_{\omega=1}^{M_{i,k}} \mathcal{A}_{i,k}^{\omega} + \sum_{\omega=1}^{N_{i,j}} \mathcal{B}_{i,j}^{\omega} \tag{7}$$

where $\mathcal{A}_{i,k}^{\omega}$ follows an exponential distribution with parameter $A_{i,k}^S$, and $\mathcal{B}_{i,j}^{\omega}$ follows an exponential distribution with parameter $B_{i,j}^D$.

Since the number of contacts follows the Poisson distribution, the average contact number for $D_i$ getting contact with $S_k$ can be denoted as $\lambda_{ik}^B T_d$, while the average contact number for $D_i$ getting contact with $D_j$ is denoted by$\lambda_{ij}^D T$. Through such simplification, we can obtain

$$V_{i,j,k}^f = \sum_{\omega=1}^{\lambda_{ik}^B T_d} \mathcal{A}_{i,k}^{\omega} + \sum_{\omega=1}^{\lambda_{ij}^D T_d} \mathcal{B}_{i,j}^{\omega} \tag{8}$$

Denote $V_{i,k}^{f} = \sum_{\omega=1}^{\lambda_{ik}^{B} T_d} \mathcal{A}_{i,k}^{\omega}$, since $\mathcal{A}_{i,k}^{\omega}(\omega = 1, 2, \cdots, \lambda_{ik}^{B} T_d)$ is a collection of independent and identically distributed random variables, according to [37], we obtain $V_{i,k}^{f} \sim \text{Gamma}(\lambda_{ik}^{B} T_d, A_{ik}^{S})$. Denote $V_{i,j}^{f} = \sum_{\omega=1}^{\lambda_{ij}^{D} T_d} \mathcal{B}_{i,j}^{\omega}$, likewise, $V_{i,j}^{f} \sim \text{Gamma}(\lambda_{ij}^{D} T_d, B_{ij}^{D})$. Thus, $D_i$ can obtain the size of the file $F_f$ as follows:

$$V_i^f = \sum_{k=1}^{l} V_{i,k}^f + \sum_{j=1}^{n} V_{i,j}^f \tag{9}$$

Now, we give the probability distribution function (PDF) of $f_{V_i^f}(\upsilon)$. Let $f_{V_{i,k}^f}(\upsilon_1)$ denote the probability distribution function of variable $V_{i,k}$. Let $f_{V_{i,j}^f}(\upsilon_2)$ denote the probability distribution function of $V_{i,j}$. According to above analysis, the probability distribution function of $f_{V_i^f}(\upsilon)$ is the discrete convolution of $f_{V_{i,1}^f}(\upsilon_1), \cdots, f_{V_{i,l}^f}(\upsilon_1), f_{V_{i,1}^f}(\upsilon_2), \cdots, f_{V_{i,1}^f}(\upsilon_2)$. Then,

$$\begin{aligned} f_{V_i^f}(\upsilon) = & \frac{\upsilon^{\lambda_{i1}^B T_d - 1} e^{-\upsilon A_{i1}^S}}{(A_{i1}^S)^{-\lambda_{i1}^B T_d} \Gamma(\lambda_{i1}^B T_d)} \otimes \cdots \otimes \frac{\upsilon^{\lambda_{il}^B T_d - 1} e^{-\upsilon B_{il}^S}}{(B_{il}^S)^{-\lambda_{il}^B T_d} \Gamma(\lambda_{il}^B T_d)} \\ & \otimes \frac{\upsilon^{\lambda_{i1}^D T_d - 1} e^{-\upsilon B_{i1}^D}}{(B_{i1}^D)^{-\lambda_{i1}^D T_d} \Gamma(\lambda_{i1}^D T_d)} \otimes \cdots \otimes \frac{\upsilon^{\lambda_{in}^D T_d - 1} e^{-\upsilon B_{in}^D}}{(B_{i1}^S)^{-\lambda_{i1}^B T_d} \Gamma(\lambda_{i1}^B T_d)} \end{aligned} \tag{10}$$

Due to the intrinsic difficulty feature to solve the convolution, Welch-Satterthwaite estimation [25] is used to obtain

$$f_{V_i^f}(\upsilon) \approx \frac{\upsilon^{\gamma-1} e^{-t\delta}}{\delta^{-\gamma} \Gamma(\gamma)} \tag{11}$$

where

$$\gamma = \frac{(\sum_{k=1}^{l} \lambda_{ik}^B T_d A_{ik}^S + \sum_{j=1}^{n} \lambda_{ij}^D T_d B_{ij}^D)^2}{\sum_{k=1}^{l} \lambda_{ik}^B T_d (A_{ik}^S)^2 + \sum_{j=1}^{n} \lambda_{ij}^D T_d (B_{ij}^D)^2},$$

$$\delta = \frac{\sum_{k=1}^{l} \lambda_{ik}^B T_d (A_{ik}^S)^2 + \sum_{j=1}^{n} \lambda_{ij}^D T_d (B_{ij}^D)^2}{\sum_{k=1}^{l} \lambda_{ik}^B T_d A_{ik}^S + \sum_{l=1}^{n} \lambda_{ij}^D T_d B_{ij}^D}$$

Then, random variable $V_i^f$ can be estimated as a gamma distribution, i.e., $V_i^f \sim \text{Gamma}(\gamma, \delta)$.

Now, we give the computation of $Pr(\mathcal{U}_i^f(\mathbf{Z}) \geq s_f)$. Since $\mathcal{U}_i^f = \min(\lfloor \frac{V_{i,f}}{g_f} \rfloor, g_f)$, and $\mathcal{U}_i^f$ is a discrete random variable, and $V_{i,f}$ is a continuous random variable, we can conclude that:

$$\begin{aligned} Pr(\mathcal{U}_i^f(\mathbf{Z}) \geq s_f) &= 1 - Pr(\mathcal{U}_i^f(\mathbf{Z}) < s_f) \\ &= 1 - \sum_{a=0}^{s_f} Pr(\mathcal{U}_i^f(\mathbf{Z}) = a) \end{aligned} \tag{12}$$

Here, we give the solution to $Pr(\mathcal{U}_i^f(\mathbf{Z}) = a)$ as follows

$$\begin{aligned} & Pr(\mathcal{U}_i^f(\mathbf{Z}) = a) \\ & = \begin{cases} \int_{ga}^{g(a+1)} f_{V_i^f}(\upsilon) d\upsilon, & \text{if } 0 \leq a < (x_{k,f} + y_{j,f}) \\ [1em] \int_{ga}^{\infty} f_{V_i^f}(\upsilon) d\upsilon, & \text{if } a = (x_{k,f} + y_{j,f}) \\ 0, & \text{otherwise} \end{cases} \end{aligned} \tag{13}$$

Finally, the distribution of $Pr(\mathcal{U}_i^f(\mathbf{Z}) \geq s_f)$ can be calculated.

### C. Green Content Delivery Model

When SBS and mobile device deliver the content, energy will be consumed. This section will discuss how the SBS or mobile device deliver the content to the requested mobile device with minimal energy consumption. Given that D2D is used, we formulate the following optimization problem:

$$\begin{aligned} & \underset{P_T^D}{\text{minimize}} \ E_D \\ & \text{subject to: } 0 < P_T^D \leq P_{\max}^D. \end{aligned} \tag{14}$$

The objective function is to find the optimal transmit power of $D_j$, which send the cached content $\mathcal{B}_{ij}$ to requesting $D_i$, to minimize the energy consuming by $D_j$. The constraint is that the transmit power $P_T^D$ of mobile device is within $(0, P_{\max}^D]$, where $P_{\max}^D$ is the maximized transmit power of mobile device.

Similarly, we formulate the following optimization problem for SBSs:

$$\begin{aligned} & \underset{P_T^B}{\text{minimize}} \ E_B \\ & \text{subject to } 0 < P_T^B \leq P_{\max}^B. \end{aligned} \tag{15}$$

The objective function is to find the optimal transmit power of $S_k$, which send the cached content $\mathcal{A}_{ik}$ to requesting $D_i$, to minimize the energy consuming by $S_k$. The constraint is that the transmit power $P_T^B$ of SBSs is within $(0, P_{\max}^B]$, where $P_{\max}^B$ is the maximized transmit power of SBSs.

## IV. Green and Mobility-Aware Caching Policy Solution

In this section, we first give a solution of mobility-aware content placement model using submodular optimization. Then, we derive a solution to the content delivery model.

### A. Solution of Mobility-Aware Content Placement Model

The optimization problem (6) is related with the number of cached code segments at SBS and mobile devices, which is a mixed integer programming (MIP) problem. It is proved as NP-hard problem to cache file in small cell and D2D networks [20], [22]. The problem targeted in this paper not only considers how to cache file at SBS and mobile device, we also consider the content amount per transmission. Therefore this problem is NP-hard.

In this section, we propose to utilize submodular optimization to solve the optimization problem (6). In submodular optimization problem, if object function is monotone submodular function and the constraint is matroid constraint, the greedy algorithm can be used. Furthermore, let $OPT$ represent the optimization solution of original problem; let $\mathbf{Z}^*$ denote the solution by greedy approach, then, $\mathbf{Z}^* \geq (1 - \frac{1}{e})OPT$, i.e., the greedy algorithm can get approximate optimal solution [38]. In the remainder of this section, original problem will be converted into a submodular optimization problem. We then present a greedy algorithm to solve the converted problem.

*1) Submodular Optimization Problem:* Define a set $Z = \{z_{i,f,\nu} | i = 1, \cdots, l+n, f = 1, \cdots, m, \nu = 1, \cdots, s_f\} = Z^1 \cup Z^2$, where

$$Z^1 = \{z_{k,f,\nu} | k = 1, \cdots, l, f = 1, \cdots, m, \nu = 1, \cdots, s_f\}$$
$$Z^2 = \{z_{j,f,\nu} | j = 1, \cdots, n, f = 1, \cdots, m, \nu = 1, \cdots, s_f\}$$

That is, when $i = 1, \cdots, l$ $z_{i,f,\nu} = z_{k,f,\nu}$. When $i = l+1, \cdots, l+n$ $z_{i,f,\nu} = z_{j,f,\nu}$. Let set $A_1$ denote the caching policy at SBS, where $A_1 \subseteq Z^1$. If element $z_{k,f,\nu} \subseteq A_1$, it represents that small cell $S_k$ caches $\nu$ code segments of file $f$. Let set $A_2$ denote the caching strategy at mobile device, where $A_2 \subseteq Z^2$. If element $z_{k,f,\nu} \subseteq$, it means there are $\nu$ code segments of file $f$ stored at mobile device $D_j$. Let set $Z^1_{k,f} = \{z_{k,f,\nu} | \nu = 1, \cdots, s_f\}$ denote the overall configuration of code segments of file $f$ cached at small cell $S_k$. Likewise, set $Z^2_{j,f} = \{z_{j,f,\nu} | \nu = 1, \cdots, s_f\}$ represent the overall configuration of code segments of file $f$ cached at mobile device $D_j$. Then, the original variables $x_{k,f}$ and $y_{j,f}$ can be obtained

$$x_{k,f} = |A_1 \cap Z^1_{k,f}|, \quad y_{j,f} = |A_2 \cap Z^2_{j,f}|$$

where $|\cdot|$ represents the number of elements in the set. Define $A = A_1 \cup A_2$, then we can obtain the object function of original problem as follows:

$$f(A) = \frac{1}{n} \sum_{i=1}^{n} \sum_{f=1}^{m} p_f Pr \left[ \min\left( \lfloor \frac{V_i^f}{g_f} \rfloor, S_f \right) \leq S_f \right].$$

Define $Z^1_k = \{z_{k,f,\nu} | f = 1, \cdots, m, \nu = 1, \cdots, s_f\}$ as all of the cached code segments at SBS $S_k$. Define $Z^2_j = \{z_{j,f,\nu} | f = 1, \cdots, m, \nu = 1, \cdots, s_f\}$. Then, original optimization problem can be rewrite as: $I = I_1 \cup I_2$, where

$$I_1 = \{A_1 | g_f |A_1 \cap Z^1_k| \leq C^s_k, k = 1, \cdots, l\}$$
$$I_2 = \{A_2 | g_f |A_2 \cap Z^2_j| \leq C^s_D, k = 1, \cdots, l\}$$

Therefore, the original problem can be rewritten as:

$$\begin{aligned} &\underset{A}{\text{maximize}} \ f(A) \\ &\text{subject to: } A \subseteq I. \end{aligned} \tag{16}$$

*Theorem 1:* In problem (16), $f(A)$ is a monotone submodular function, and the $(Z, I)$ is matroid constraint.

*Proof: It is proved that $f(A)$ is a monotone submodular function*. If $f(A)$ is a monotonic submodular function, then $\forall A \subset B \subset Z$ and $\forall z_{i,f,\nu} \in Z - B$, $f(A \cup z_{i,f,\nu}) - f(A) \geq f(B \cup z_{i,f,\nu}) - f(B) \geq 0$ can be proven, i.e.,

$$f(A \cup z_{k,f,\nu}) - f(A) \geq f(B \cup z_{k,f,\nu}) - f(B) \geq 0$$
$$f(A \cup z_{j,f,\nu}) - f(A) \geq f(B \cup z_{j,f,\nu}) - f(B) \geq 0$$

can be proven to be correct.

First, we prove $f(A \cup \{z_{k,f,\nu}\}) - f(A) \geq f(B \cup \{z_{k,f,\nu}\}) - f(B) \geq 0$. Since $\forall A \subset B \subset Z$, $E = B - A$, $\forall z_{i,f,\nu} \in Z - B$, define $V_i^f = \sum_{k'=1,k \neq k}^{l} V_{i,k}^f + \sum_{j=1}^{n} V_{i,j}^f$, then where $V_{i,k}^f$ denotes the content amount obtained by $D_i$ from SBS $S_k$ for file $f$. Apply total probability formula, we can obtain, the equation as shown at the top of the next page.

Since $\min\left( \lfloor \frac{V_i^f}{g_f} \rfloor, |B_1 \cap Z^1_{k',f} - E_1| + |B_2 \cap Z^2_{j,f} - E_2| \right) \leq \min\left( \lfloor \frac{V_i^f}{g_f} \rfloor, |B_1 \cap Z^1_{k',f}| + |B_2 \cap Z^2_{j,f}| \right)$ and

$$|B_1 \cap Z^1_{k,f} - E_1| = |B_1 \cap Z^1_{k,f}| - |E_1 \cap Z^1_{k,f}|$$
$$|B_2 \cap Z^2_{j,f} - E_2| = |B_2 \cap Z^2_{j,f}| - |E_2 \cap Z^2_{j,f}|$$

We can further get

Similarly, we have $f(A \cup \{z_{j,f,\upsilon}\}) - f(A) \geq f(B \cup \{z_{k,f,\nu}\}) - f(B)$. Then, we prove that $f(A)$ is a Monotonic submodule function.

*The $\{Z, I\}$ is Matroid:* In other words, we prove that both $(Z^1, I_1)$ and $(Z^2, I_2)$ are matroid. For $(Z^1, I_1)$, we have:

- $\emptyset \in I_1$;
- If $B_1 \subseteq I$ and $A_1 \subseteq B_1$, then $A_1 \subseteq I$;
- If $A_1, B_1 \in I_1$ and $|A_1| < |B_1|$, there exists an element $k \in B - A$ that makes $A \cup \{k\} \in I$.

Therefore, $(Z^1, I_1)$ is a matroid. Similarly, we can show that $(Z^2, I_2)$ is also matroid. Hence, we conclude that $(Z, I)$ is a matroid. ■

Though above proof, we conclude that the object function of the optimization problem (16) is monotone submodular function with constrain of a matroid. Thus, greedy algorithm can be applied for the solution. Next, we will provide the detailed algorithm.

*2) Mobility-Aware Content Placement Algorithm:* Based on Theorem 1, we can know that greedy algorithm could also resolve our optimization problem (16). The specific solution algorithm is provided below: we set an empty cache placement set $A$, in each iteration, we add a file which can maximize the object function until the capacity of SBS or mobile device is full.

**Algorithm 1** Mobility-Aware Content Placement Algorithm

**Input:**
Input All $z_{i,f,\upsilon}$ combinations, $Z$;
Reminding set of $Z$, $Z_r$;
Capacity for SBSs and mobile devices, $C$.
**Output:**
Optimal cache policy, $A$;
1: $A \leftarrow \emptyset$, $Z_r \leftarrow Z$
2: Repeat
3: $z_{i^*,f^*,k^*} = \text{argmax}_{z_{i,f,k} \in Z_r} [f(Z_r + z_{i,f,k}) - f(Z_r)]$
4: $A \leftarrow A + z_{i^*,f^*,\upsilon^*}$
5: $Z_r \leftarrow Z_r - z_{i^*,f^*,\upsilon^*}$
6: If $|A \cap Z_{i^*}| = C$, Then $Z_r \leftarrow Z_r \backslash Z_{i^*}$
7: end if
8: Until $|A| > (\sum_{k=1}^{l} C_k^S + \sum_{i=1}^{n} C_i^D)$

From step 3 we can see that caching $k^*$ encoded segments of file $F_{f^*}$ to SBS or mobile terminal is the optimal cache policy, so $z_{i^*,f^*,k^*}$ is added into optimal cache strategy $A$ in step 4. Step 6 indicates that there could not be more files to be cached in mobile device or SBS once the amount of cached files reaches their storage capacity $C$. So the files $Z_{i^*}$ cached at $i^*$ are deleted from $Z_r$. From step 8, we notice that the iteration proceeds until $|A| > (\sum_{k=1}^{l} C_k^S + \sum_{i=1}^{n} C_i^D)$.

$$
\begin{aligned}
f(A\cup z_{k,f,\upsilon})-f(A) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{f=1}^{m}p_f\\
&\times\left\{Pr\left[\min\left(\lfloor\frac{V_i^f}{g_f}\rfloor,|A_1\cap Z^1_{k',f}|+|A_2\cap Z^2_{j,f}|\right)+\min\left(\lfloor\frac{V_{i,k}^f}{g_f}\rfloor,|A_1\cap Z^1_{k,f}|+1\right)\geq S_f\right]\right.\\
&\quad\left.-Pr\left[\min\left(\lfloor\frac{V_i^f}{g_f}\rfloor,|A_1\cap Z^1_{k',f}|+|A_2\cap Z^2_{j,f}|\right)+\min\left(\lfloor\frac{V_{i,k}^f}{g_f}\rfloor,|A_1\cap Z^1_{k,f}|\right)\geq S_f\right]\right\}\\
f(A\cup z_{k,f,\upsilon})-f(A) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{f=1}^{m}p_f Pr\left[\lfloor\frac{V_{i,k}^f}{g_f}\rfloor\geq|A_1\cap Z^1_{k,f}|+1\right]\\
&\times Pr\left[\min\left(\lfloor\frac{V_i^f}{g_f}\rfloor,|A_1\cap Z^1_{k',f}|+|A_2\cap Z^2_{j,f}|\right)=s_f-|A_1\cap Z^1_{k,f}|-1\right]\\
&= \frac{1}{n}\sum_{i=1}^{n}\sum_{f=1}^{m}p_f Pr\left[\lfloor\frac{V_{i,k}^f}{g_f}\rfloor\geq|B_1\cap Z^1_{k,f}-E_1|+1\right]\\
&\times Pr\left[\min\left(\lfloor\frac{V_i^f}{g_f}\rfloor,|B_1\cap Z^1_{k',f}-E_1|+|B_2\cap Z^2_{j,f}-E_2|\right)=s_f-|B_1\cap Z^1_{k,f}-E_1|-1\right]\\
f(A\cup z_{k,f,\upsilon})-f(A) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{f=1}^{m}p_f Pr\left[\lfloor\frac{V_{i,k}^f}{g_f}\rfloor\geq|B_1\cap Z^1_{k,f}|-|E_1\cap Z^1_{k,f}|+1\right]\\
&\times Pr\Bigg[\min\left(\lfloor\frac{V_i^f}{g_f}\rfloor,|B_1\cap Z^1_{k',f}|-|E_1\cap Z_{k',f}|+|B_2\cap Z^2_{j,f}|-|E_2\cap Z^2_{j,f}|\right)\\
&\qquad=s_f-|B_1\cap Z^1_{k,f}|-|E_1\cap Z^1_{k,f}|-1\Bigg]\geq\frac{1}{n}\sum_{i=1}^{n}\sum_{f=1}^{m}p_f Pr\left[\lfloor\frac{V_{i,k}^f}{g_f}\rfloor\geq|B_1\cap Z^1_{k,f}|+1\right]\\
&\times Pr\left[\min\left(\lfloor\frac{V_i^f}{g_f}\rfloor,|B_1\cap Z^1_{k',f}|+|B_2\cap Z^2_{j,f}|\right)=s_f-|B_1\cap Z^1_{k,f}|-1\right]=f(B\cup\{z_{k,f,\upsilon}\})-f(B)
\end{aligned}
$$

*B. Solution of Green Content Delivery Model*

In this section, we provide the solution of optimization problems (14) and (15), i.e., the optimal transmit powers of SBS and mobile device are given as follows:

*Theorem 2:* According to the optimization problems (14) and (15), we can get the optimal transmit powers of SBS and mobile device, $P_T^D$ and $P_T^B$, are as shown in (20) and (21).

*Proof:* We first give the optimal transmit power of mobile devices. For (15), we define $x = P_T^D$, $\delta = \frac{r_D^2}{\sigma_D^2}$, $\theta = \eta_D P_C^D$. Then we can get the objective function in (15) can be transformed as follows:

$$f(x)=\frac{\mathcal{B}_{ij}(x+\theta)}{\eta_D W_D\log_2^{(1+\delta x)}} \tag{17}$$

Take the derivative of (17), Then we have

$$f'(x)=\frac{\mathcal{B}_{ij}[(1+\delta x)\log_2^{(1+\delta x)}\ln 2-(1+\delta x)-(\delta\theta-1)]}{\eta_D W_D\log_2^{2(1+\delta x)}(1+\delta x)\ln 2} \tag{18}$$

From the (18), we can obtain that the denominator of $f'(x)$ is always positive, when the numerator of $f'(x)$ is greater than 0, $f'(x)>0$, and vise versa. Due to $x\in(0,P^D_{\max}]$, $1+\delta x\in(1,1+\delta P^D_{\max}]$, define $y=1+\delta x$, $\xi=\delta\theta-1$, then the (18) can be rewritten as:

$$f'(y)=\frac{\mathcal{B}_{ij}[y\log_2^y\ln 2-y-\xi]}{\eta_D W_D\log_2^{2y}(y)\ln 2} \tag{19}$$

Define $g(y)=\mathcal{B}_{ij}\,y\log_2^y\ln 2-y-\xi$, and taking the derivative of $g(y)$ on $y$, we can obtain $g'(y)=\mathcal{B}_{ij}\log_2^y\ln 2$, thus we know that when $y\in(1,1+\delta P_{\max}]$, $g(y)$ is monotonic increasing function. Further, when $y\to 1$, $g(y)<0$. And if $g(1+\delta P^D_{\max})<0$, $f'(y)<0$, then we can know $f(x)$ is monotonic decreasing function, so it gets minimum when $x=P^D_{\max}$. Besides, if $g(1+\delta P^D_{\max})>0$, there exists zero point $y_0$ making $g(y_0)=0$, so we conclude that function $f(x)$ is first monotonic increasing, then monotonic decreasing, and achieves the minimum at $y_0$. Finally, we obtain the optimal $P_T^{D*}$ as follows:

$$P_T^{D*}=\begin{cases}P^D_{\max}, & \text{if } g(1+\delta P^D_{\max})<0,\\ P_{y_0}, & \text{if } g(1+\delta P^D_{\max})\geq 0\end{cases} \tag{20}$$

In the same way, we can get $P_T^{B*}$ as follows:

$$P_T^{B*}=\begin{cases}P^B_{\max}, & \text{if } g(1+\delta P^B_{\max})<0,\\ P_{y_0}, & \text{if } g(1+\alpha P^B_{\max})\geq 0\end{cases} \tag{21}$$

■

By solving (14) and (15), we can obtain the optimal transmit power of mobile device and SBS (denoted by $P_T^{D*}$ and $P_T^{B*}$, respectively). Furthermore, the minimum energy consumption of mobile device and SBS can be calculated, denoted as $E_D^*$ and $E_B^*$, respectively.

Thus, the user average energy consumed when requesting file can be obtained as follows:

$$\bar{E}^* = \frac{1}{n}\sum_{i=1}^{n}\sum_{f=1}^{m} p_f \left(\sum_{\omega=1}^{M_{i,k}} E_B^*(\mathcal{A}_{i,k}^{\omega}) + \sum_{\omega=1}^{N_{i,j}} E_D^*(\mathcal{B}_{i,j}^{\omega})\right) \tag{22}$$

Since the number of contacts follows the Poisson distribution, the average time count for $D_i$ getting contact with SBS $S_k$ can be expressed as $\lambda_{i,k}^B T_d$, while the average time count for $D_i$ getting contact with $D_j$ can be expressed as $\lambda_{i,j}^D T_d$. Then, we can obtain the average minimum energy consumption of network when the content is delivered as follows:

$$\bar{E}^* = \frac{1}{n}\sum_{i=1}^{n}\sum_{f=1}^{m} p_f \left(\sum_{\omega=1}^{\lambda_{i,k}^B T_d} E_B^*(\mathcal{A}_{i,k}^{\omega}) + \sum_{\omega=1}^{\lambda_{i,j}^D T_d} E_D^*(\mathcal{B}_{i,j}^{\omega})\right) \tag{23}$$

## V. Performance Evaluation

In this section, we evaluate the performance of the proposed green and mobility-aware caching system. We consider a ultra-dense cellular network including 5 SBSs and 60 mobile devices. According to [20] and [36], we set the size of each file to 10 MB and file library contains $10^4$ files, each file can be encoded into 2 code segments [23]. We assume that file request probability follows a Zipf distribution with $\gamma = 0.5$ and the deadline $T_d = 600s$. For simplicity, assume the cache size for all mobile devices is the same and can cache 1 percent of file library size in maximum. Similarly, assume the cache size for all SBSs is the same and can cache 10 percent of file library size in maximum.

For user mobility parameter settings according to [23], we use Gamma distribution to represent contact rate, i.e., the contact rate $\lambda_{i,j}^D$ between mobile device $D_i$ and $D_j$ follows the $\Gamma(4.43, 1/1088)$, and the contact rate $\lambda_{i,k}^B$ between mobile device $D_i$ and SBS $S_k$ follows the $\Gamma(10, 1/100)$. Wang *et al.* [23] show that the contact rate increases with the increasing use moving speed. Thus, larger $\lambda_{i,j}^D$ and $\lambda_{i,k}^B$ are, faster a user moves. According to [20] and [23], we set the amount of data that SBS and mobile device can deliver to a requested mobile device during a contact duration follows exponential distribution with means 10 MB and 5 MB, respectively.

In our experiments, the parameters regarding energy consumption at SBS and mobile devices are referred to [35]. The simulation parameters and their values are summarized in Table II. To evaluate the system, we measure the *cache hit ratio* and *energy consumption*.

### A. Comparison With Other Methods

Green and mobility-aware caching (GM-caching), proposed in this paper, is compared with three different caching strategies: *Popular caching* [39], *Random caching* [23] and *mobility-aware caching with fixed amount data can be delivered (MCF)* [20], [21], [23].

TABLE II
Simulation Parameters of Energy Cost

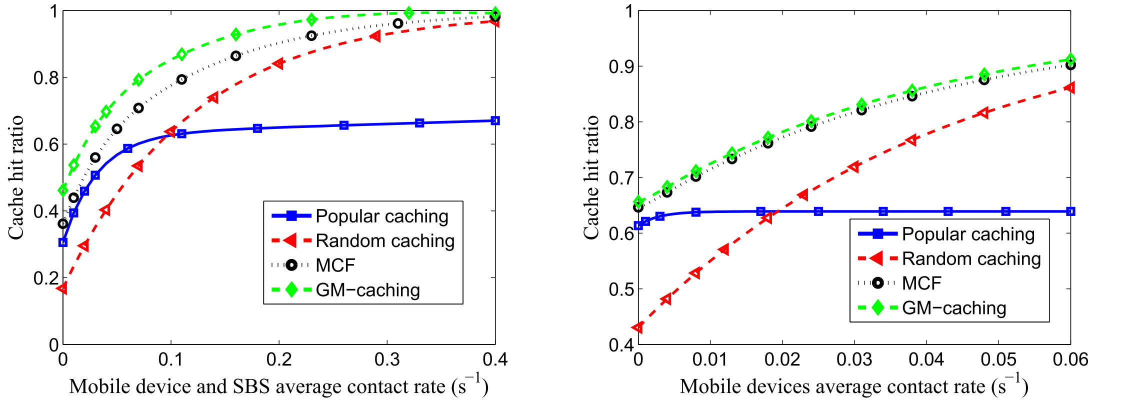
| **Parameters** | **Value** |
|---|---|
| Maximize transmit power of SBS, $P_{\max}^B$ | 6.3 W |
| Constant power of SBS, $P_C^B$ | 56 W |
| Transmission bandwidth of SBS, $W_B$ | 20 MHz |
| Gaussian noise of SBS, $\sigma_B^2$ | -104 dBm |
| Power amplifier efficiency of SBS, $1/\beta_B$ | 0.38 |
| Transmission bandwidth of D2D, $W_D$ | 20 MHz |
| Maximize transmit power of user mobile device, $P_{\max}^D$ | 0.2W |
| Circuit power consume of user mobile device, $P_C^D$ | 115.9 mW |
| Gaussian noise ,$\sigma_D^2$ | -95dBm |
| Power amplifier efficiency of user mobile device, $1/\beta_D$ | 0.2 |
| Path loss factor, $\alpha$, $\beta$ | 4 |

- *Popular Caching:* The popular caching strategies on SBSs and on mobile devices are as follows, (i) caching strategy on SBSs: $C_k^S$ most popular contents should be cached on each SBS $S_k$. (ii) caching strategy on mobile devices: $C_j^D$ most popular contents should be cached on each mobile device $D_j$.
- *Random Caching:* The random caching caching strategies on SBSs and on mobile devices are as follows, (i) caching strategy on SBSs: randomly choose $C_k^S$ contents to be cached on each SBS $S_k$. (ii) caching strategy on mobile device: randomly choose $C_j^D$ contents to be cached on each mobile device $D_j$.
- *MCF:* Under this strategy, we assume the amount of content delivered is the average of $\mathcal{A}_{i,k}^{\omega}$ and $\mathcal{B}_{i,k}^{\omega}$.

### B. Evaluation Results

*Impact of User Mobility:* In Fig. 5, the influence of user mobility on caching strategy with three different caching strategies is demonstrated. In this experiment, for sake of convenient representation, let abscissa be $\overline{\lambda}_{i,k}^B$ and $\overline{\lambda}_{i,j}^D$, corresponding to the mean value of Gamma distribution that user mobility follows. In other words, $\overline{\lambda}_{i,k}^B$ stands for the average contact rate of mobile device $D_i$ and SBSs $S_k$, and $\overline{\lambda}_{i,j}^D$ stands for the average contact rate of mobile device $D_i$ and $D_j$. It can be seen from Fig. 5 that the proposed GM-caching strategy is superior to the popular caching strategy, random caching strategy and MCF caching strategy. This is because the popular caching strategy, the most popular contents are cached by SBS and mobile device, so there are less opportunities for cooperation and sharing. The random caching strategy does not consider user mobility, so it is not as good as the GM-caching. Although the MCF cache strategy takes into account the user mobility, it does not consider whether the encoded segments can be transmitted in a contact duration. Thus, the performance of the green and mobility cache strategy is the best.

From Fig. 5(a) and Fig. 5(b), we can obtain that when the user mobility is low, the cache hit ratio of GM-caching and popular caching is not significant. This is because when




(a) The impact of average contact rate of mobile devices and SBSs ($\overline{\lambda}_{i,k}^{B}$) on cache hit ratio.

(b) The impact of average contact rate of mobile devices ($\overline{\lambda}_{i,j}^{D}$) on cache hit ratio.

Fig. 5. The impact of user mobility on cache hit ratio.

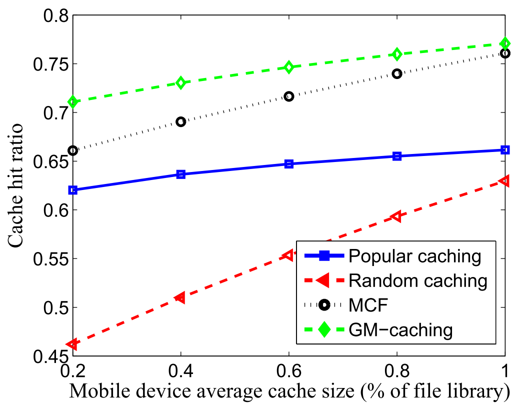



Fig. 6. The impact of cache size of mobile device on the cache hit ratio.

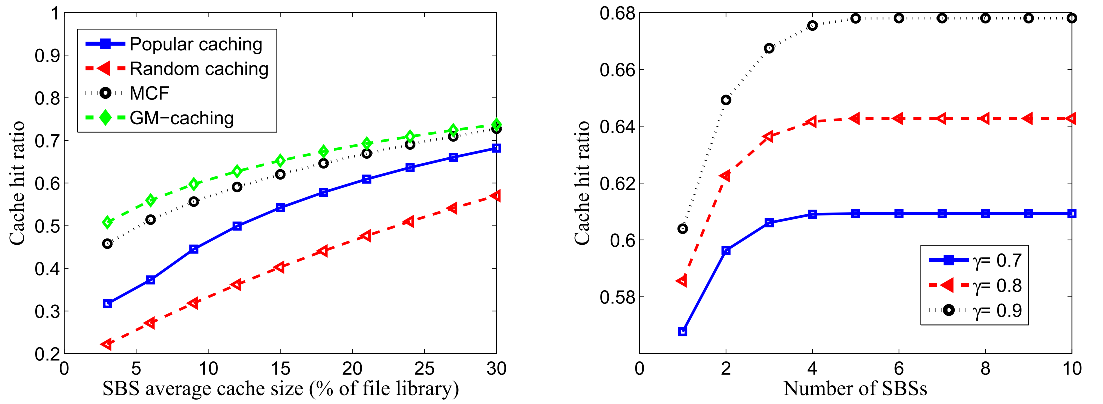
the user mobility is low, mobile device not only less handoff between different SBS, but also less opportunity to meet other mobile devices. However, there is great demand for popular contents. Thus, the popular contents should be cached on SBSs and on mobile devices to improve the cache hit ratio. When the mobility of user is high, we can see that the gap between the GM-caching and the random caching is small. This is because the cache network is relatively active, user have more contact opportunities with SBSs, so the probability of the user access to popular file become large. Thus, in this case, the cache hit ratio can be improved by considering the diversity of files.

*Impact of Mobile Devices:* As shown in Fig. 6, the relationship between cache size of mobile devices and the cache hit ratio is examined. In Fig. 6, the abscissa stands for the ratio of mobile device cache capacity to the entire file library size. For instance, abscissa 0.6 represents mobile device can store up to 0.6 percent of the entire file library size. It can be concluded from the figure that the cache hit ratio increases when the user device average cache size increases. This is because more contents can be cached when cache capacity for mobile device is larger, so as to achieve a higher cache hit ratio. Besides, it can also be seen from the figure that the influence on cache hit ratio with random caching is greater than that with popular caching, when the cache size for mobile devices is larger. This is because the diversity of contents that can be cached with random caching is superior to that with popular caching, when storage capacity for mobile devices is larger, thus to achieve a higher cache hit ratio.

*Impact of SBSs:* As shown in Fig. 7(a), the influence of cache size for SBSs on cache hit ratio is presented. Similarly, when the average cache size for SBSs is larger, the cache hit ratio is higher. This is because more contents can be cached when the cache capacity for SBSs is larger, so as to achieve a higher cache hit ratio.

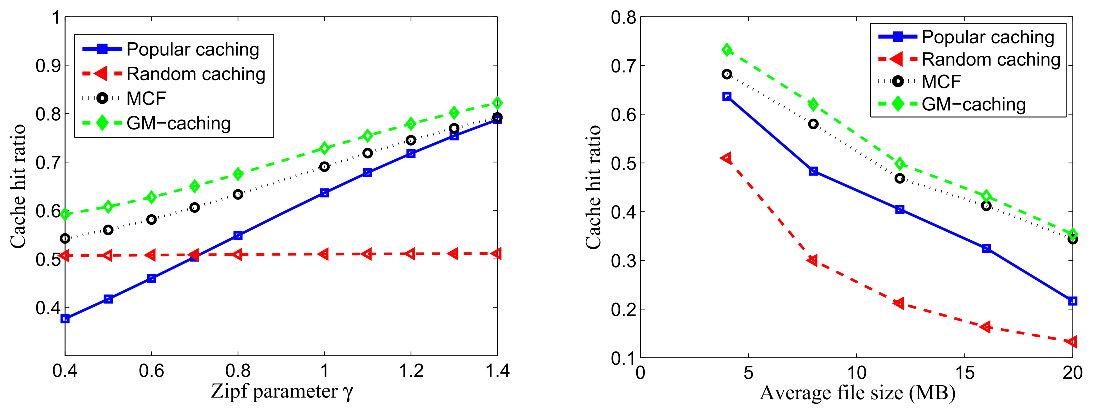
We explore how the number of SBSs impacts the cache hit ratio in Fig. 7(b). From the figure, we observe that when the number of SBS is less than 6, the cache hit ratio increase with the increase of the number of SBS. While when the number of SBS is more than 6, this is not significant. Furthermore, it can also seen from the figure that, when $\gamma = 0.7$, the number of SBS to 4 when the cache hit ratio is not changed, and $\gamma = 0.9$, the number of SBS reached a little change when the number of up to 6. Thus, the deployment of SBSs should be more decentralized as popularity $\gamma$ increase, i.e., the quantity of SBSs is large but the capacity for each SBS is small. This is because mobile device would be more concerned about popular content as $\gamma$ increases. With decentralized deployment, the most popular content will be cached on each SBS, thus to meet the requirements of most users, so as to achieve a higher hit ratio.

*Impact of Files:* In Fig. 8(a), the cache strategies under different Zipf parameter $\gamma$ is evaluated. It can be seen from the figure that as popularity $\gamma$ increases, the cache hit ratio with GM-caching is lager. It can also be seen from the figure that with the increase of $\gamma$, the cache hit ratio of popular caching strategy increase gradually, but the cache hit ratio of random caching strategy does not change much, i.e., the influence of $\gamma$ on popular caching is greater while that on random caching is less. This is because the requirements of user for popular contents are greater as $\gamma$ increases, the requirements of users can be better met with popular caching. Therefore, popular caching increases as $\gamma$ increases.




(a) The impact of SBSs cache size on the cache hit ratio. (b) The impact of SBSs number ($l$) on the cache hit ratio.

Fig. 7. The impact of SBS average cache size and density on the cache hit ratio.




(a) The impact of Zipf parameter ($\gamma$) on the cache hit ratio. (b) The impact of average file size ($|F_f|$) on the cache hit ratio.

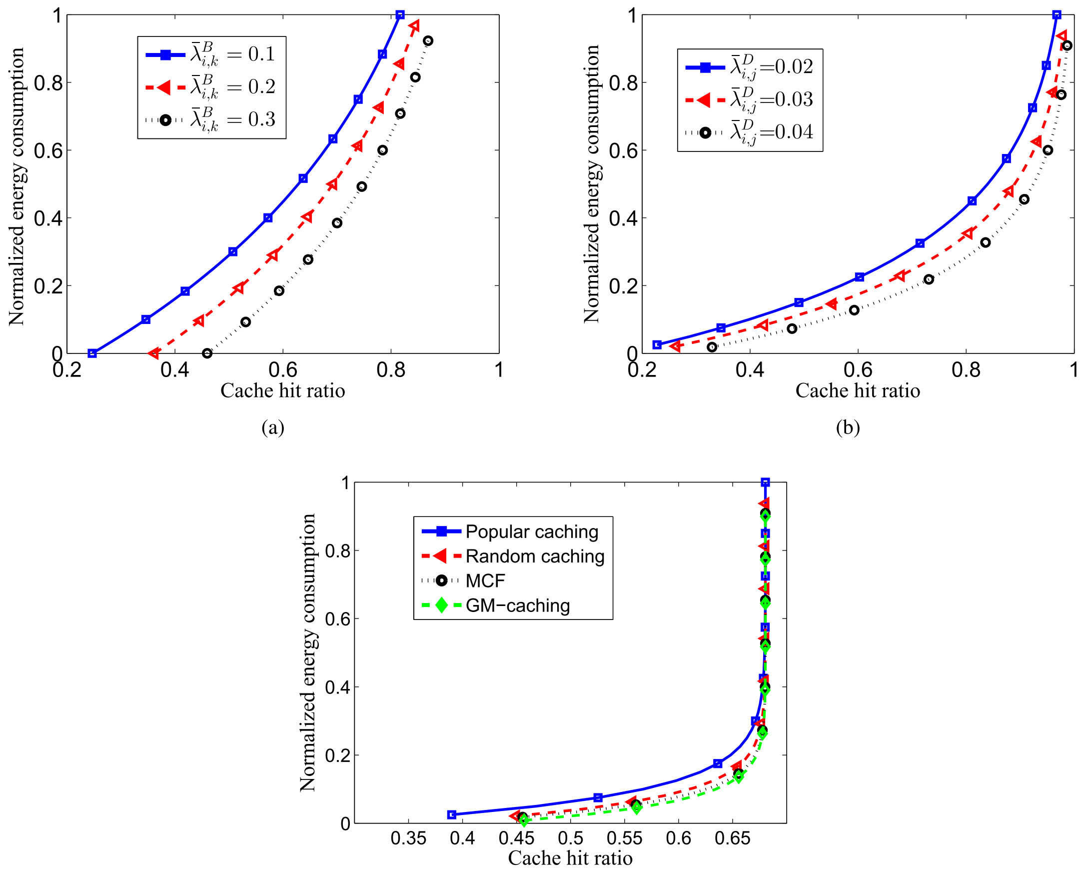
Fig. 8. The impact of Zipf parameter and file average size on cache hit ratio.

In Fig. 8(b), the influence of average file size on the cache hit ratio is presented. In this experiment, consider the entire file library size is fixed, i.e., the large size of file, the less number of file. The abscissa stands for the size of each file. From the figure, we observe that when the the size of each file is large, the cache hit ratio is smaller. This is because as the size of file is large and the cache capacity for SBSs and mobile devices is fixed, the amount of files that can be cached on SBSs and on mobile devices would be smaller, so that the cache hit ratio becomes lower.

*Energy Consumption:* Fig. 9(a) and Fig. 9(b) discuss the relationship between cache hit ratio and energy consumption. In this experiment, the energy consumption is normalized. From the Fig. 9(a) and Fig. 9(b), we can observe that with the increase of the cache hit ratio, energy consumption is also rising. This is because with the increase in cache hit ratio, the probability that the requested file download from SBS and mobile device increase, which increases the energy consumption of the SBS and mobile device. Furthermore, Fig. 9(a) shows that, when the mobile device $D_i$ and SBS $S_k$ average contact rate $\overline{\lambda}_{i,j}^{B}$ increase, more energy will consumes under the same cache hit ratio. Similarly, Fig. 9(b) has the same conclusion. Thus, under the same cache hit ratio, the higher user mobility, the less energy consumed. Fig. 9(c) shows that, under the same cache hit ratio, GM-caching is more energy efficient than popular caching, random caching and MCF. This is because GM-caching not only takes into account the user mobility, but also considers the randomness of the contact duration.

## C. Discussion

In this article, for user's mobility, we describe the user's mobility by utilizing the peer-to-peer connectivity model from the two perspectives: contact duration and contact frequency. The model reflects a primitive level of social interaction between users. However, the effects of more elaborate social relations and behaviours on user's mobility has not been taken into account. These effects of social networks on caching design have been studied in existing work (see e.g., [40]–[43]). It is an interesting direction for further investigation by extending the current work to include a mobility model account for more elaborate effects by social networks. Another direction warranting future research is to consider joint user-and-SBS caching powered by energy harvesting [44]. The technology can reduce transmission energy consumption but the energy randomness makes the optimal design more challenging.




(a) (b) (c)

Fig. 9. Energy consumption under different caching strategies.

Furthermore, it is also a very interesting topic to consider caching and computing offloading together in edge cloud [45] to meet the delay sensitive application requirements [46].

## VI. Conclusion

In this paper, we first studied the influence of mobility on caching on SBSs and on mobile devices, and proposed a mobility-aware content placement model, where specific caching strategies were developed for SBSs and mobile device leverages user mobility, aiming to maximize the cache hit ratio. Then, we discussed the energy consumption for delivery the caching content and given the optimal transmission power of SBS and mobile device. The simulation results showed that the proposed GM-caching strategy outperforms other caching strategies in prior work, in terms of cache hit ratio and energy efficiency of 5G networks.

## References

[1] Cisco. (Feb. 2017). *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update 2016–2021*. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[2] T. Taleb, A. Ksentini, and R. Jantti, "'Anything as a service' for 5G mobile systems," *IEEE Netw.*, vol. 30, no. 6, pp. 84–91, Nov./Dec. 2016.

[3] A. Liu and V. K. N. Lau, "Cache-enabled opportunistic cooperative MIMO for video streaming in wireless systems," *IEEE Trans. Signal Process.*, vol. 62, no. 2, pp. 390–402, Jan. 2014.

[4] K. Yang, S. Ou, and H.-H. Chen, "On effective offloading services for resource-constrained mobile devices running heavier mobile Internet applications," *IEEE Commun. Mag.*, vol. 46, no. 1, pp. 56–63, Jan. 2008.

[5] T. Taleb, B. Mada, M.-I. Corici, A. Nakao, and H. Flinck, "PERMIT: Network slicing for personalized 5G mobile telecommunications," *IEEE Commun. Mag.*, vol. 55, no. 5, pp. 88–93, May 2017.

[6] K. Yang, S. Ou, K. Guild, and H. H. Chen, "Convergence of Ethernet PON and IEEE 802.16 broadband access networks and its QoS-aware dynamic bandwidth allocation scheme," *IEEE J. Sel. Areas Commun.*, vol. 27, no. 2, pp. 101–116, Feb. 2009.

[7] M. Chen, Y. Zhang, M. Qiu, N. Guizani, and Y. Hao, "SPHA: Smart personal health advisor based on deep analytics," *IEEE Commun.*, vol. 55, no. 12, Dec. 2017.

[8] K. Yang, S. Ou, H. H. Chen, and J. He, "A multihop peer-communication protocol with fairness guarantee for IEEE 802.16-based vehicular networks," *IEEE Trans. Veh. Technol.*, vol. 56, no. 6, pp. 3358–3370, Nov. 2007.

[9] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "FemtoCaching: Wireless content delivery through distributed caching helpers," *IEEE Trans. Inf. Theory*, vol. 59, no. 12, pp. 8402–8413, Dec. 2013.

[10] A. Liu, V. Lau, and W. Han, "Randomized user-centric clustering for cloud radio access network with PHY caching," in *Proc. IEEE Global Conf. Signal Inf. Process. (GlobalSIP)*, Dec. 2015, pp. 948–952.

[11] A. Liu and V. K. N. Lau, "Asymptotic scaling laws of wireless ad hoc network with physical layer caching," *IEEE Trans. Wireless Commun.*, vol. 15, no. 3, pp. 1657–1664, Mar. 2016.

[12] J. Li, Y. Chen, Z. Lin, W. Chen, B. Vucetic, and L. Hanzo, "Distributed caching for data dissemination in the downlink of heterogeneous networks," *IEEE Trans. Commun.*, vol. 63, no. 10, pp. 3553–3568, Oct. 2015.

[13] J. Li, W. Chen, M. Xiao, F. Shu, and X. Liu, "Efficient video pricing and caching in heterogeneous networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 10, pp. 8744–8751, Oct. 2016.

[14] J. Li, H. Chen, Y. Chen, Z. Lin, B. Vucetic, and L. Hanzo, "Pricing and resource allocation via game theory for a small-cell video caching system," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 8, pp. 2115–2129, Aug. 2016.

[15] M. Chen, Y. Qian, Y. Hao, Y. Li, and J. Song, "Data-driven computing and caching in 5G networks: Architecture and delay analysis," *IEEE Wireless Commun.*, vol. 25, no. 1, pp. 11–18, Feb. 2018.

[16] M. Ji, G. Caire, and A. F. Molisch, "Wireless device-to-device caching networks: Basic principles and system performance," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 1, pp. 176–189, Jan. 2016.

[17] H. Wang, X. Wang, K. Li, J. Ren, X. Zhang, and T. Jiang, "A measurement study of device-to-device sharing in mobile social networks based on spark," *Concurrency Comput., Pract. Exper.*, vol. 29, no. 16, pp. 1–10, 2017.

[18] X. Wang, Z. Sheng, S. Yang, and V. C. M. Leung, "Tag-assisted social-aware opportunistic device-to-device sharing for traffic offloading in mobile social networks," *IEEE Wireless Commun.*, vol. 23, no. 4, pp. 60–67, Aug. 2016.

[19] T. Wang, L. Song, and Z. Han, "Dynamic femtocaching for mobile users," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2015, pp. 861–865.

[20] K. Poularakis and L. Tassiulas, "Code, cache and deliver on the move: A novel caching paradigm in hyper-dense small-cell networks," *IEEE Trans. Mobile Comput.*, vol. 16, no. 3, pp. 675–687, Mar. 2017.

[21] Y. Guan, Y. Xiao, H. Feng, C.-C. Shen, and L. J. Cimini, "MobiCacher: Mobility-aware content caching in small-cell networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2014, pp. 4537–4542.

[22] R. Lan, W. Wang, A. Huang, and H. Shan, "Device-to-device offloading with proactive caching in mobile cellular networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2015, pp. 1–6.

[23] R. Wang, J. Zhang, S. H. Song, and K. B. Letaief, "Mobility-aware caching in D2D networks," *IEEE Trans. Wireless Commun.*, vol. 16, no. 8, pp. 5001–5015, Aug. 2017.

[24] R. Wang, X. Peng, J. Zhang, and K. B. Letaief, "Mobility-aware caching for content-centric wireless networks: Modeling and methodology," *IEEE Commun. Mag.*, vol. 54, no. 8, pp. 77–83, Aug. 2016.

[25] Z. Lu, X. Sun, and T. La Porta, "Cooperative data offloading in opportunistic mobile networks," in *Proc. IEEE 35th Annu. Int. Conf. Comput. Commun. (INFOCOM)*, Apr. 2016, pp. 1–9.

[26] S. Retal, M. Bagaa, T. Taleb, and H. Flinck, "Content delivery network slicing: QoE and cost awareness," in *Proc. IEEE ICC*, Paris, France, May 2017, pp. 1–6.

[27] Y. Li and W. Wang, "Can mobile cloudlets support mobile applications?" in *Proc. IEEE INFOCOM*, Apr./May 2014, pp. 1060–1068.

[28] S. Arabi, E. Sabir, T. Taleb, and M. Sadik, "The right content for the right relay in self-organizing delay tolerant networks: A matching game perspective," in *Proc. IEEE ICC*, Paris, France, May 2017, pp. 1–6.

[29] J. Scott, R. Gass, J. Crowcroft, P. Hui, C. Diot, and A. Chaintreau. (Sep. 2006). *CRAWDAD Dataset Cambridge/Haggle (Version 2006-09-15)*. [Online]. Available: http://crawdad.org/cambridge/haggle/20090529

[30] M. McNett and G. M. Voelker, "Access and mobility of wireless PDA users," *ACM SIGMOBILE Mobile Comput. Commun. Rev.*, vol. 9, no. 2, pp. 40–55, 2005.

[31] D. J. C. MacKay, "Fountain codes," *IEE Proc.-Commun.*, vol. 152, no. 6, pp. 1062–1068, Dec. 2005.

[32] J. Zhao, X. Zhuo, Q. Li, W. Gao, and G. Cao, "Contact duration aware data replication in DTNs with licensed and unlicensed spectrum," *IEEE Trans. Mobile Comput.*, vol. 15, no. 4, pp. 803–816, Apr. 2016.

[33] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, "Web caching and Zipf-like distributions: Evidence and implications," in *Proc. 18th Annu. Joint Conf. IEEE Comput. Commun. Soc. (INFOCOM)*, vol. 1. Mar. 1999, pp. 126–134.

[34] T. Taleb, K. Samdanis, B. Mada, H. Flinck, S. Dutta, and D. Sabella, "On multi-access edge computing: A survey of the emerging 5G network edge cloud architecture and orchestration," *IEEE Commun. Surveys Tuts.*, vol. 19, no. 3, pp. 1657–1681, 3rd Quart., 2017.

[35] B. Chen and C. Yang, "Energy costs for traffic offloading by cache-enabled d2d communications," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2016, pp. 1–6.

[36] D. Liu and C. Yang, "Energy efficiency of downlink networks with caching at base stations," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 4, pp. 907–922, Apr. 2016.

[37] S. M. Ross, *Introduction to Probability Models*. San Francisco, CA, USA: Academic, 2014.

[38] G. Calinescu, C. Chekuri, M. Pál, and J. Vondrák, "Maximizing a monotone submodular function subject to a matroid constraint," *SIAM J. Comput.*, vol. 40, no. 6, pp. 1740–1766, 2011.

[39] H. Ahlehagh and S. Dey, "Video-aware scheduling and caching in the radio access network," *IEEE/ACM Trans. Netw.*, vol. 22, no. 5, pp. 1444–1462, Oct. 2014.

[40] N. Vastardis and K. Yang, "Mobile social networks: Architectures, social properties, and key research challenges," *IEEE Commun. Surveys Tuts.*, vol. 15, no. 3, pp. 1355–1371, 3rd Quart., 2013.

[41] X. Wang, Y. Zhang, V. C. M. Leung, N. Guizani, and T. Jiang, "D2D big data: Content deliveries over wireless device-to-device sharing in realistic large scale mobile networks," *IEEE Wireless Commun.*, vol. 25, no. 1, pp. 1–10, Feb. 2018.

[42] T. Taleb, A. Ksentini, M. Chen, and R. Jantti, "Coping with emerging mobile social media applications through dynamic service function chaining," *IEEE Trans. Wireless Commun.*, vol. 15, no. 4, pp. 2859–2871, Apr. 2016.

[43] S. Wang, Y. Zhang, H. Wang, Z. Wang, X. Wang, and T. Jiang, "Large scale measurement and analytics on social groups of device-to-device sharing in mobile social networks," *Mobile Netw. Appl.*, vol. 23, no. 4, pp. 1–13, Dec. 2017.

[44] K. Yang, Q. Yu, S. Leng, B. Fan, and F. Wu, "Data and energy integrated communication networks for wireless big data," *IEEE Access*, vol. 4, pp. 713–723, 2016.

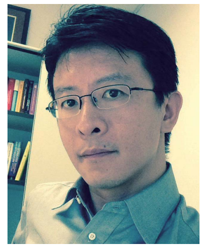
[45] S. Lal, I. Oliver, S. Ravidas, and T. Taleb, "Assuring virtual network function image integrity and host sealing in telco cloud," in *Proc. IEEE ICC*, Paris, France, May 2017, pp. 1–6.

[46] I. Farris, T. Taleb, M. Bagaa, and H. Flinck, "Optimizing service replication for mobile delay-sensitive applications in 5G edge network," in *Proc. IEEE ICC*, Paris, France, May 2017, pp. 1–6.

**Min Chen** (SM'09) was an Assistant Professor with the School of Computer Science and Engineering, Seoul National University (SNU). He has been a Full Professor with the School of Computer Science and Technology, Huazhong University of Science and Technology (HUST), since 2012. He is the Director of the Embedded and Pervasive Computing Laboratory, HUST. He was a Post-Doctoral Fellow with SNU for one and half years. He was a Post-Doctoral Fellow with the Department of Electrical and Computer Engineering, The University of British Columbia (UBC) for three years. He is a Keynote Speaker for CyberC 2012, Mobiquitous 2012, Cloudcomp 2015, IndustrialIoT 2016, and the 7th Brainstorming Workshop on 5GWireless. He has over 300 paper publications, including over 200 SCI papers, over 80 IEEE transactions/journal papers, 16 ISI highly cited papers, and 8 hot papers. He has authored four books: the *OPNET IoT Simulation* in 2015, the *Big Data Inspiration* in 2015, the *5G Software Defined Networks* in 2016, and the *Introduction to Cognitive Computing* in 2017, with HUST Presss, a book on big data: the*Big Data Related Technologies* in 2014 and a book on 5G: the *Cloud Based 5G Wireless Networks* in 2016 with Springer Series in Computer Science. His latest book (co-authored with Prof. K. Hwang) entitled the *Big Data Analytics for Cloud/IoT and Cognitive Computing* (Wiley, U.K.) appears in 2017. His Google Scholars Citations reached over 10,500 with an h-index of 50. His top paper was cited over 1050 times. His research focuses on Cyber physical systems, IoT sensing, 5G networks, mobile cloud computing, SDN, healthcare big Data, medica cloud privacy and security, body area networks, emotion communications and robotics. He received the Best Paper Award from QShine 2008, the IEEE ICC 2012, the ICST IndustrialIoT 2016, and the IEEE IWCMC 2016. He received the IEEE Communications Society Fred W. Ellersick Prize in 2017. He is the Chair of the IEEE Computer Society Special Technical Communities on Big Data. He is the Co-Chair of the IEEE ICC 2012-Communications Theory Symposium, and the Co-Chair of IEEE ICC 2013-Wireless Networks Symposium. He is the General Co-Chair of the IEEE CIT-2012, Tridentcom 2014, Mobimedia 2015, and Tridentcom 2017. He serves as an Editor or an Associate Editor for the *Information Sciences*, *Information Fusion*, and the IEEE ACCESS. He is a Guest Editor of the IEEE Network, the IEEE Wireless Communications, and the IEEE Transactions Service Computing.

**Yixue Hao** received the B.E. degree from Henan University, China, and the Ph.D. degree in computer science from the Huazhong University of Science and Technology (HUST), China, 2017. He is currently a Post-Doctoral Scholar with the School of Computer Science and Technology, HUST. His research includes 5G network, Internet of Things, edge caching, and mobile edge computing.

**Long Hu** was a visiting student with the Department of Electrical and Computer Engineering, The University of British Columbia, from 2015 to 2017. He has been a Lecturer with the School of Computer Science and Technology, Huazhong University of Science and Technology, China, since 2017. His research includes the Internet of Things, software defined networking, caching, 5G, body area networks, body sensor networks, and mobile cloud computing.



**Kaibin Huang** (M'08–SM'13) received the B.Eng. (Hons.) and M.Eng. degrees in electrical engineering from the National University of Singapore, and the Ph.D. degree in electrical engineering from The University of Texas at Austin (UT Austin). Since 2014, he has been an Assistant Professor with the Department of Electrical and Electronic Engineering (EEE), The University of Hong Kong. He is an Adjunct Professor with the School of EEE, Yonsei University, South Korea. He was a Faculty Member with the Department of Applied Mathematics (AMA), The Hong Kong Polytechnic University (PolyU), and the Department of EEE, Yonsei University. His research interests focus on the analysis and design of wireless networks using stochastic geometry and multi-antenna techniques. He frequently serves on the technical program committees of major IEEE conferences in wireless communications. In 2015, he edited a JOURNAL ON SELECTED AREAS IN COMMUNICATIONS Special Issue on Communications Powered by Energy Harvesting. He was an elected member of the SPCOM Technical Committee of the IEEE Signal Processing Society from 2012 to 2015. He received the 2015 IEEE ComSoc Asia Pacific Outstanding Paper Award, the Outstanding Teaching Award from Yonsei, Motorola Partnerships in Research Grant, the University Continuing Fellowship from UT Austin, and the Best Paper Award from the IEEE GLOBECOM 2006 and PolyU AMA in 2013. He held the University Visiting Scholarship at Kansai University, Japan, in 2017. He served as the Lead Chair for the Wireless Communication Symposium of the IEEE Globecom 2017 and the Communication Theory Symposium of the IEEE GLOBECOM 2014 and the TPC Co-Chair for IEEE PIMRC 2017 and the IEEE CTW 2013. He was an Editor of the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS Series on Green Communications and Networking from 2015 to 2016, the IEEE WIRELESS COMMUNICATIONS LETTERS from 2011 to 2016, and the IEEE/KICS JOURNAL OF COMMUNICATION AND NETWORKS from 2009 to 2015. He is currently an Editor of the IEEE TRANSACTIONS ON GREEN COMMUNICATIONS AND NETWORKING and the IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS.

**Vincent K. N. Lau** (SM'04–F'12) received the B.Eng. (Hons.) degree from The University of Hong Kong in 1992 and the Ph.D. degree from Cambridge University in 1997. He was with Bell Labs from 1997 to 2004 and the Department of Electronic and Computer Engineering, The Hong Kong University of Science and Technology (HKUST), in 2004. He is currently a Chair Professor and the Founding Director of Huawei–HKUST Joint Innovation Laboratory, HKUST. His current research focus includes robust and delay-optimal cross layer optimization for MIMO/ OFDM wireless systems, interference mitigation techniques for wireless networks, massive MIMO, M2M, and network control systems.