SPECIAL SECTION ON SMART CACHING, COMMUNICATIONS, COMPUTING
AND CYBERSECURITY FOR INFORMATION-CENTRIC INTERNET OF THINGS

IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 19, 2019, accepted June 2, 2019, date of publication June 6, 2019, date of current version June 20, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2921389

An Energy-Efficient Transmission Strategy for Cache-Enabled Wireless Networks With Non-Negligible Circuit Power

CHAO MENG[1], GANG WANG[2], (Member, IEEE), XI DAI[1],
SHIPING CHEN[3], (Senior Member, IEEE), AND WEI NI[3], (Senior Member, IEEE)
[1]School of Networks and Telecommunications Engineering, Jinling Institute of Technology, Nanjing 211169, China
[2]National Mobile Communications Research Laboratory, Southeast University, Nanjing 210096, China
[3]CSIRO Data61, Sydney, NSW 2122, Australia

Corresponding author: Gang Wang (wanggang@seu.edu.cn)

This work was supported in part by the National Natural Science Foundation of China under Grant 61771132 and Grant 61471115, in part by the Natural Science Foundation of the Higher Education Institutions of Jiangsu Province under Grant 16KJB510011, and in part by the Research Foundation of the Jinling Institute of Technology for Advanced Talents under Grant 40620044.

ABSTRACT In this paper, a cache-enabled wireless network with circuit power consumption is considered, and the strategy of joint transmission and local caching is studied. The optimization problem of energy consumption minimization is formulated, where the optimized variables include the transmission rate and the local caching rate. The constraint conditions contain the maximum data departure curve, the minimum data departure curve, and the range of the local caching rate. The original problem is intractable and is transformed into a two-layer optimization one by variable substitution. Under given caching policy, the sufficient and necessary conditions for the optimal offline transmission strategy are deduced by using the transmission characteristics. Then, an energy-efficient offline transmission strategy is proposed, and its optimality is proved. The simulation results are provided to verify the performance of the proposed strategy. The simulation results reveal the impact of the cache capacity and the file popularity on energy consumption. Under the same conditions, the performance of the proposed strategy is better than that of the other known algorithms. The proposed strategy to energy-efficient communication is helpful for future green communication.

INDEX TERMS Cache-enabled wireless networks, caching, energy consumption, offline transmission, energy-efficient, file popularity.

I. INTRODUCTION

With the development of 5G wireless communication technology [1], [2], the Internet of Things (IoT) has been extensively applied, such as smart city, intelligent transportation, driverless, smart home, smart medical, wearable equipment and so on [3]–[5]. IoT enables each device to connect the Internet, so that these devices can easily access the network at any time. When billions of the devices are connected to the Internet, a large amount of data will be generated [6]. It is predicted that by the end of 2021, the mobile traffic generated by IoT-related applications will reach 49 Ebytes per month [7]. To enable the sustainable 5G networks, energy efficiency (EE), which is measured in bits per Joule, has become a new design metric for future green wireless communication systems [8]–[10]. IoT is a heterogeneous resource-constrained network in essence [11]–[13]. The so-called constraints are that the computing power, the battery and the memory of the device is limited [14]. To fulfill the needs of IoT, Information-Centric Networking (ICN) is considered as a promising approach, which has attracted wide attention [15]–[17].

As a novel Internet architecture, ICN is a content-based approach. The user's requirement can be satisfied whatever the location of the information and the characteristic of its source. The main features of ICN include network caching, location-independent content name, enhanced security and better mobility management, which enable ICN to be suitable for the distribution of IoT huge data. Cache memory stores the content, such as video and web pages, in the edge of wireless

The associate editor coordinating the review of this manuscript and approving it for publication was Jun Wu.

networks (the base station (BS), relay node, or user equipment), and can effectively reduce the backhaul traffic load and the energy consumption [18]. When a large number of users request the same popular content at different time, the wireless networks need to transmit the same content to each user repeatedly, which will cause the waste of time and frequency resources [19]. If the BS can store the popular content in the cache memory, the popular content only needs to be sent once through the backhaul link, which can effectively reduce the backhaul traffic load and the energy consumption [15]. The cache-enabled macro-cellular networks are considered in [20], each BS is equipped with a large cache memory, the users can directly achieve the required content from the BS, rather than from the core networks by the backhaul link. For the cache-enabled heterogeneous networks (HetNets), the coverage area of a macro BS (MBS) overlaps that of a small BS (SBS), they share the same spectrum resource. In the cache-enabled HetNets, both the MBS and the SBS are equipped with the cache memory, the users can get the required content from the nearest MBS or the SBS [21].

In cache-enabled wireless networks, the network operation can be divided into two phases: the content placement phase and delivery phase. In the content placement phase, the appropriate content is stored in cache memory during the non-peak periods. When the requested file is available in the local cache memory (at the SBS or the user), the file can be served by the cache instead of the MBS. This improves the quality of service, and reduces the traffic load of the backhaul link, which can achieve the local caching gain [22]. The content delivery phase works in busy traffic periods, the network can transmit the data file directly to the served users by unicast or multicast transmission [23], [24], and the global caching gain can be obtained. The delivery strategy needs to be designed by holistically considering the size of the data, the power allocation and the transmission time slot.

Recently, the joint content placement and delivery has attracted attention [25]–[28]. When the traffic load is low, some data can be prefetched before transmission, and the cost of prefetching the data is also considered. The prefetching gain is helpful to use the better channel conditions, equalize the transmission rate, balance the network load, and reduce the energy consumption. The prefetching transmission strategies which minimize the energy consumption and the bandwidth were considered in [29] and [30], respectively. The cache was only used to store the prefetched content for the users. Once the cached content was transmitted to the users, it would be deleted from the cache, regardless of any possible future requests for the content. Therefore, only the prefetching gain was considered, and there was no local caching gain in [29] and [30]. The joint optimal transmission and caching strategies based on the prefetching gain and local caching gain have been studied in [31] and [32]. The energy consumption minimization of the MSB by jointly designing the transmission strategy of the MBS and the caching strategy of the SBS was studied in [31]. The original infinite-dimensional optimization problem was transformed into a convex optimization problem, the numerical simulation results were given. The HetNets and D2D communication were considered in [32]. In HetNets scenario, the cache was located on the SBS, and the SBS could serve multiple users at the same time. The optimization problem of minimizing the energy consumption could be solved by the dual decomposition method, and a sub-gradient algorithm was proposed.

In the research of the joint optimal transmission and caching strategies in the above literature, only the radio frequency (RF) transmission power consumption was considered, while the circuit power consumption, such as the filter, AC/DC converter, mixer, and other devices, was not taken into account. Without considering the circuit power consumption, the rate-power function is concave and monotonously increasing. Therefore, increasing the transmission time and reducing the transmission rate can reduce the energy consumption [33]–[35]. The circuit power consumption usually occupies a large proportion in the overall power consumption, which cannot be ignored [36], [37]. When considering the circuit power consumption, the intermittent transmission is the optimal transmission scheme [38], [39].

In this paper, the problem of minimizing the energy consumption for cache-enabled wireless networks with circuit power consumption is considered. The aim of the paper is to investigate the energy-efficient transmission strategy by considering the prefetching and local caching gains. Compared to the existing works, the main contributions of the paper are summarized as follows:

(1) Firstly, the offline energy consumption minimization for a cache-enabled wireless network with circuit power consumption is established, and the optimization variables include the transmission rate and the local caching rate. The constraint conditions contain maximum data departure curve, minimum data departure curve, and the range of the local caching rate.

(2) Secondly, the original problem is non-convex, we transform it into a two-layer optimization problem by variable substitution. Under given caching policy, the sufficient and necessary conditions for the optimal offline transmission strategy are derived. Then an energy-efficient offline transmission strategy is proposed, and its optimality is proved. In addition, the computational complexity of the proposed algorithm is analyzed, which is lower than that of the existing algorithm.

(3) Thirdly, the simulation results are provided to verify the performance of the proposed strategy. The simulation results reveal the impact of the cache capacity and the file popularity on the energy consumption. For the same conditions, the performance of the proposed strategy is better than that of the existing other algorithms.

The remainder of this paper is organized as follows. Section II introduces the system model and presents the initial problem formulation. Section III derives the necessary and sufficient conditions for the transformed problem, and proposes an energy-efficient optimal offline transmission strategy. Section IV gives the simulation results and some analysis. Finally, in Section V the conclusion is drawn.

II. SYSTEM MODEL AND PROBLEM FORMULATION

A. DATA FLOW MODEL

A heterogeneous wireless cache network with the downlink is considered as shown in Fig. 1. The MBS accesses the core network, and transmits the users' requested data files to the SBS by a wireless backhaul link with the transmission rate $r(t)$. Then the SBS sends the data to K users in a time division multiple access (TDMA) method. Furthermore, the SBS is equipped with a cache memory with capacity of B bits, which is used to cache the received data from the MBS. It is assumed that the MBS and the SBS all work in the different frequency bands, therefore there is no interference between them.

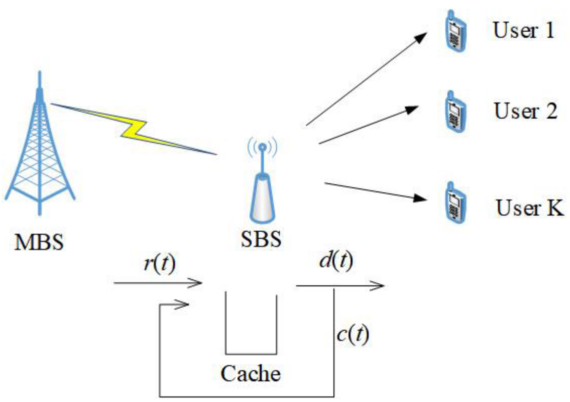


FIGURE 1. A system model of heterogeneous wireless cache network.

The transmission time T consists of N time slots, and the starting point of the n-th time slot is denoted as s_n. Let $0 = s_1 < s_2 < \ldots < s_{N-1} < s_N < T$, and the length of the n-th time slot can be expressed as $t_n = s_{n+1} - s_n$. Define $\mathcal{F} = \{f_0, f_1, f_2, \ldots, f_{|F|}\}$ as the set of all the possible requested data files by the users, where f_i is a specific data file. Specifically, f_0 defines as an empty file, which indicates that there is no requested data file by the user at the corresponding time slot.

It is assumed that the requested data files for the users are known for the MBS, i.e. the offline policy. At the time duration $[0, T]$, the SBS transmits the data file to the corresponding user at each time slot according to the known order of the user's request. In order to formulate the related optimization problem, some definitions are given as follows.

Definition 1: The user's requested rate. The requested rate $d(t)$ is the transmission rate of the data file requested by the user. The SBS must transmit the data file to serve the user with not less than this rate $d(t) \geq 0, t \in [0, T]$.

For simplicity and mathematical expression, we assume that each requested data file has a fixed time slot length and a constant requested rate. The time slot length and the requested rate of a data file are usually determined by the upper protocol stack of the system. For example, the requested rate of a video file is higher than that of a general web browsing. Therefore, the user's requested rate can be expressed as $d(t) = \sum_{n=1}^{N} d_n rect\left((t - (s_n + t_n/2))/t_n\right)$, where $rect\left((t-a)/b\right)$ denotes a rectangular function, the central point is a and the length is b. If the same data file is requested at time slots n and n', there must be $t_n = t_{n'}$ and $d_n = d_{n'}$.

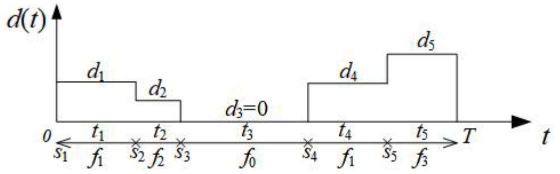


FIGURE 2. Illustration of the requested data files.

Fig. 2 illustrates an example of the users' requested rates. The profile of the users' requested data files are $\{f_1, f_2, f_0, f_1, f_3\}$. Since the data file f_1 is requested at the first and fourth time slots, therefore we have $t_1 = t_4$ and $d_1 = d_4$. It should be noted that the optimal transmission and caching strategy in offline mode is studied in this paper, so the profile of the users' requested data files is available for the MBS.

Definition 2: Local caching rate. The local caching rate $c(t)$ represents the rate at which the SBS caches the data file in the local cache for the future user's request while transmitting the current data file. $0 \leq c(t) \leq d(t), t \in [0, T]$.

As shown in Fig. 1, the cache memory in the SBS can be virtually divided into two parts: the prefetching data from the MBS and the local caching data from the SBS, respectively. The prefetching data is determined by the transmission strategy of the MBS, $r \triangleq r(t)|_{t=0}^{T}$, which provides the prefetching gain. The local caching data is determined by the local caching policy of the SBS, $c \triangleq c(t)|_{t=0}^{T}$, which provides the local caching gain. When the same data files are requested at the future time slots, the caching data of these files does not need to download again from the MBS, which reduces the energy consumption of the backhaul link. In practical system, it is unnecessary to move the data from the cache to be stored again. We can consider that the SBS deletes the user's requested data at the rate $d(t) - c(t)$.

Definition 3: Data departure curve. The data departure curve $D(t, r)$ represents the amount of the cumulative data transmitted by the MBS at time t $(t \geq 0)$, which depends on the transmission strategy of the MBS. Therefore we have $D(t, r) = \int_0^t r(\tau) d\tau$.

Definition 4: Maximum data departure curve. The maximum data departure curve denotes the maximum amount of data transmitted by the MBS at time t, which can ensure that the amount of the transmission data by the MBS does not overflow the cache. It is related to the local caching strategy c of the SBS, and can be expressed as $B(t, c) = B + \int_0^t (d(\tau) - c(\tau)) d\tau$.

The minimum amount of the cumulative data transmitted by the MBS at time t is bounded, which is determined by the data which is not in the cache. If a specific file has been transmitted by the SBS before, the cache stores a part of the data according to the local caching strategy. The MBS does not need to transmit the caching data when the same file is transmitted, and can only transmit the data with the rate $d(t) - c(\rho(t))$, where $\rho(t) : [0, T] \rightarrow [0, T] \cup \{-1\}$ is a mapping function. The mapping function maps a time slot to the corresponding previous time slot for the same data file. If the file has not been requested before, we define $\rho(t) = -1$, and $c(-1) \overset{\Delta}{=} 0$. The offline mode is considered, $\rho(t)$ is known for the MBS. The lower bound of the data departure curve is determined by the minimum data departure curve, which is described as follows.

Definition 5: Minimum data departure curve. The minimum data departure curve $A(t, c)$, is the minimum amount of the data cumulatively transmitted from the MBS, which must satisfy the user's requested data file. The minimum data departure curve is also related to the local caching strategy c, which can be represented as $A(t, c) = \int_0^t (d(\tau) - c(\rho(\tau)))d\tau$.

B. TRANSMISSION MODEL WITH CIRCUIT POWER CONSUMPTION

The RF transmission power of the MBS can be expressed as $p(t)$, and the corresponding real-time transmission rate is $r(t)$. The relationship between the transmission power and the transmission rate can be expressed as

$$p(t) = g(r(t)), \tag{1}$$

which should satisfy the following four assumptions:

(1) If $r(t) \geq 0$, then $g(r(t)) \geq 0$, for $t \in [0, +\infty)$, and $g(0) = 0$;

(2) $g(r(t))$ is a monotonic increasing function of $r(t)$;

(3) $g(r(t))$ is a strictly convex function of $r(t)$;

(4) $g(r(t))$ is continuous and derivative.

It is obvious that the classical Shannon channel capacity formula satisfies the above assumptions, the transmission rate $r(t)$ of the MBS can be expressed as

$$r(t) = \ln(1 + h(t)p(t)) \tag{2}$$

where $h(t)$ represents the channel power gain from the MBS to the SBS. Similarly in [31], [40], [41], the additional white Gaussian noise channel model is considered, formula (2) can be further expressed as $r(t) = \ln(1 + p(t))$. Therefore, we have $p(t) = e^{r(t)} - 1$. In addition to the RF transmission power consumption, there are many other components that consume the power in the actual system, such as AD/DA converters, mixers, filters and analog RF amplifiers, which are commonly referred to as the circuit power consumption. When the RF power of the MBS $p(t) > 0$, the circuit power consumption is denoted as $\varepsilon > 0$; when there is no data transmission, the transmitter does not work, so $p(t) = 0$ and the circuit power consumption $\varepsilon = 0$. Therefore, the actual power consumption model of the MBS can be expressed as [40].

$$P_{total}(t) = \begin{cases} p(t) + \varepsilon, & p(t) > 0; \\ 0, & p(t) = 0. \end{cases} \tag{3}$$

C. PROBLEM FORMATION

Based on the data flow model and the circuit power consumption model, the optimization goal in this paper is to find the joint optimal transmission and caching strategy, so that the energy consumption of the MBS is minimized at the whole time range $[0, T]$. The optimization problem can be formulated as follows

$$\min_{r,c} \; E(r(t)) = \int_0^T P_{total}(t)dt \tag{4a}$$
$$s.t. \; D(t, r) \geq A(t, c), \tag{4b}$$
$$D(t, r) \leq B(t, c), \tag{4c}$$
$$r(t) \geq 0, \tag{4d}$$
$$0 \leq c(t) \leq d(t), \tag{4e}$$

where (4a) is to minimize the energy consumption of the MBS. The constraints (4b) and (4c) denote the feasible region of the cumulative transmission data of the MBS. The cumulative transmission data of the MBS at any time cannot overflow the cache memory, but also must satisfy the user's requested rate at the same time. The constraint (4d) represents that the transmission rate must be non-negative. The constraint (4e) means that the caching rate should not exceed the user's real-time requested rate.

III. AN ENERGY-EFFICIENT OPTIMAL TRANSMISSION STRATEGY

It can be seen that problem (4) is an infinite-dimensional optimization problem, and is intractable in general. In this section, we first give some properties of the optimal solution, so that the original infinite-dimensional optimization problem can be transformed into a two-level finite-dimensional optimization problem [33]. Then, based on the Karush-Kuhn-Tucker (KKT) conditions of the optimization problem, the related properties of the optimal solution of the finite-dimensional optimization problem are revealed. Finally, an energy-efficient optimal transmission strategy is proposed to minimize the energy consumption of the MBS.

A. CONVERSION OF THE INFINITE-DIMENSIONAL OPTIMIZATION PROBLEM

Because function $g(\cdot)$ is strictly convex, therefore the actual power consumption $P_{total}(t)$ also maintains strict convexity. Then, we have the following Lemmas:

Lemma 1: In each transmission slot $t_n, n = 1, 2, \ldots, N$, the constant transmission rate is the optimal transmission rate to minimize the energy consumption of the MBS.

Proof: The specific process can be referred to the proof of Theorem 5 in [33].

Lemma 2: In each transmission slot $t_n, n = 1, 2, \ldots, N$, it is optimal for the cache memory to cache the data at a constant rate.

Proof: The proof follows similarly to the proof of Lemma 1 in [31].

According to Lemma 2, the local caching rate $c(t)$ is a piece-wise function. In addition, it is assumed that the user's real-time requested rate $d(t)$ is also a piece-wise function. Their values may change only at the beginning of each time slot. Therefore, the maximum data departure curve $B(t, c)$ and the minimum data departure curve $A(t, c)$ are the piece-wise linear functions, and their slopes may only change at the beginning of each time slot.

As shown in [42], when the circuit power consumption is not taken into account, the continuous transmission of the MBS at each time slot is optimal. When considering the circuit power consumption, the continuous transmission may not be optimal, and the intermittent transmission become the optimal transmission. It is assumed that the actual transmission duration of the MBS in each slot t_i is τ_i, where $0 \le \tau_i \le t_i$. From Lemma 1 and Lemma 2, the optimization problem (4) can be equivalently transformed into the following finite-dimensional form

$$\min_{\{\tau_i, r_i, c_i\}} \sum\nolimits_{i=1}^{N} \tau_i(e^{r_i} - 1 + \varepsilon) \tag{5a}$$

$$s.t. \quad \sum\nolimits_{i=1}^{n} r_i\tau_i \le B + \sum\nolimits_{i=1}^{n} d_i t_i - \sum\nolimits_{i=1}^{n} c_i t_i, \tag{5b}$$

$$\sum\nolimits_{i=1}^{n} r_i\tau_i \ge \sum\nolimits_{i=1}^{n} d_i t_i - \sum_{i\in\Phi_n} c_i t_i, \tag{5c}$$

$$0 \le \tau_i \le t_i, \tag{5d}$$

$$0 \le c_i \le d_i, \tag{5e}$$

for $n = 1, 2, \ldots, N$, $i = 1, 2, \ldots, N$, and (5a) $-$ (5e) are discrete expressions of (4a) $-$ (4e), respectively. Φ_n is a set of time slot indexes, in which the relevant data files have been requested before. For the optimization problem (5), the optimization variables $\{c_i\}$ only appear in the constraints, but not in the objective function, so problem (5) can be further written as a two-layer optimization problem.

$$\min_{\{c_i\}} \ \psi(\{c_i\})$$
$$s.t. \ 0 \le c_i \le d_i, \tag{6}$$

for $i = 1, 2, \ldots, N$. $\psi(\{c_i\})$ can be expressed as

$$\min_{\{\tau_i, r_i\}} \sum\nolimits_{i=1}^{N} \tau_i(e^{r_i} - 1 + \varepsilon) \tag{7a}$$

$$s.t. \quad \sum\nolimits_{i=1}^{n} r_i\tau_i \le B + \sum\nolimits_{i=1}^{n} d_i t_i - \sum\nolimits_{i=1}^{n} c_i t_i, \tag{7b}$$

$$\sum\nolimits_{i=1}^{n} r_i\tau_i \ge \sum\nolimits_{i=1}^{n} d_i t_i - \sum_{i\in\Phi_n} c_i t_i, \tag{7c}$$

$$0 \le \tau_i \le t_i, \tag{7d}$$

for $n = 1, 2, \ldots, N$, $i = 1, 2, \ldots, N$. The outer-layer optimization variables $\{c_i\}$ of (6) are in the constraints (7b) and (7c) of the inner-layer optimization problem, and affect the feasible region of the data departure curve. The local caching is meaningful only when there are the same requested data files. Therefore, it is possible to search the same type of data files to obtain the optimal local caching rate. In this paper, the exhaustive search (ES) is used to solve the local caching rate $\{c_i\}$, and the computational complexity of ES is related to the number of data files and the files popularity. How to effectively obtain the optimal local caching rate $\{c_i\}$ is still an open issue. Next, we focus on the solution of the inner-layer optimization problem for the given local caching strategy $\{c_i\}$.

B. OPTIMAL TRANSMISSION CONDITIONS FOR GIVEN CACHING STRATEGY

Note that $\tau_i(e^{r_i} - 1 + \varepsilon)$ in (7a) and $r_i\tau_i$ in (7b) and (7c) are non-convex since the Hessian matrices are not positive semi-definite, problem (7) is non-convex. Through variable substitution, it can be converted into a convex problem. We define $\theta_i = r_i\tau_i$ for $i = 1, 2, \ldots, N$. Note that θ_i is the amount of the transmitted data actually by the MBS at time slot t_i. The optimization problem (7) can be rewritten as

$$\min_{\{\tau_i, \theta_i\}} \sum\nolimits_{i=1}^{N} \tau_i(e^{\theta_i/\tau_i} - 1 + \varepsilon) \tag{8a}$$

$$s.t. \quad \sum\nolimits_{i=1}^{n} \theta_i \le B + \sum\nolimits_{i=1}^{n} d_i t_i - \sum\nolimits_{i=1}^{n} c_i t_i, \tag{8b}$$

$$\sum\nolimits_{i=1}^{n} \theta_i \ge \sum\nolimits_{i=1}^{n} d_i t_i - \sum_{i\in B_n} c_i t_i, \tag{8c}$$

$$0 \le \tau_i \le t_i, \tag{8d}$$

$$\theta_i \ge 0, \tag{8e}$$

for $n = 1, 2, \ldots, N$, $i = 1, 2, \ldots, N$. Where the term $\tau_i(e^{\theta_i/\tau_i} - 1 + \varepsilon)$ is the perspective function of the concave function $e^{\theta_i} - 1 + \varepsilon$. When $\tau_i = 0$, we let $\tau_i(e^{\theta_i/\tau_i} - 1 + \varepsilon) = 0$. Since the perspective function has the convexity preservation, the objective function (8a) is also convex [43]. Furthermore, the constraint conditions in (8b) $-$ (8e) are all linear. Therefore, problem (8) is convex, and can be solved by the general convex methods such as primal-dual algorithm and sub-gradient method [43]. The computational complexity of the general algorithms is very high. In this paper, we exploit the specific structure of the optimization problem (8), and present a low complex algorithm, which can provide more insights. The Lagrangian function of problem (8) can be written as

$$\begin{aligned}
&\ell(\tau_i, \theta_i, \lambda_n, \mu_n, \varphi_i, \phi_i, \gamma_i) \\
&= \sum\nolimits_{i=1}^{N} \tau_i(e^{\theta_i/\tau_i} - 1 + \varepsilon) \\
&\quad + \sum\nolimits_{n=1}^{N} \lambda_n \left(\sum\nolimits_{i=1}^{n} \theta_i - B - \sum\nolimits_{i=1}^{n} d_i t_i + \sum\nolimits_{i=1}^{n} c_i t_i\right) \\
&\quad + \sum\nolimits_{n=1}^{N} \mu_n \left(\sum\nolimits_{i=1}^{n} d_i t_i - \sum_{i\in B_n} c_i t_i - \sum\nolimits_{i=1}^{n} \theta_i\right) \\
&\quad + \sum\nolimits_{i=1}^{N} \varphi_i(\tau_i - t_i) - \sum\nolimits_{i=1}^{N} \phi_i\theta_i - \sum\nolimits_{i=1}^{N} \gamma_i\tau_i,
\end{aligned} \tag{9}$$

for $i = 1, 2, \ldots, N$ and $n = 1, 2, \ldots, N$. $\lambda_n \ge 0$, $\mu_n \ge 0$, $\varphi_i \ge 0$, $\phi_i \ge 0$ and $\gamma_i \ge 0$ are the Lagrange multipliers for the corresponding constraints in (8b) $-$ (8e), respectively. We denote θ_i^* and τ_i^* as the optimal amount of the transmitted data by the MBS at time slot t_i and the corresponding optimal transmission duration, respectively. Moreover, let $\{\lambda_n^*, \mu_n^*, \varphi_i^*, \phi_i^*, \gamma_i^*\}$ be the optimal Lagrange multipliers for its dual problem. Therefore, the KKT conditions are

$$\frac{\partial \ell}{\partial \theta_i} = e^{\theta_i^*/\tau_i^*} - \sum\nolimits_{i=1}^{N} (\lambda_n^* - \mu_n^*) - \phi_i^* = 0, \tag{10}$$

$$\frac{\partial \ell}{\partial \tau_i} = e^{\theta_i^*/\tau_i^*} - 1 + \varepsilon - \frac{\theta_i^*}{\tau_i^*} e^{\theta_i^*/\tau_i^*} + \varphi_i^* - \gamma_i^* = 0. \tag{11}$$

Furthermore, the complementary slackness conditions are given by

$$\lambda_n^* \left(\sum\nolimits_{i=1}^{n} \theta_i^* - B - \sum\nolimits_{i=1}^{n} d_i t_i + \sum\nolimits_{i=1}^{n} c_i t_i \right) = 0, \tag{12}$$

$$\mu_n^* \left(\sum\nolimits_{i=1}^{n} d_i t_i - \sum_{i \in B_n} c_i t_i - \sum\nolimits_{i=1}^{n} \theta_i^* \right) = 0, \tag{13}$$

$$\varphi_i^* (\tau_i^* - t_i) = 0, \quad \phi_i^* \theta_i^* = 0, \quad \gamma_i^* \tau_i^* = 0. \tag{14}$$

It should be noticed that since the optimization problem (8) is convex, the KKT conditions are the necessary and sufficient conditions for the optimality of the transmission schedule. Based on the KKT conditions, we will clarify some specific properties of the optimal transmission strategy for problem (8).

Lemma 3: If $\theta_i^* = 0$ *or* $\tau_i^* = 0$*, then the optimal transmission rate* $r_i^* = 0$.

Lemma 4: If $0 < \tau_i^* \leq t_i$ *and* $\theta_i^* > 0$*, then the optimal transmission rate* $r_i^* = \ln \sum_{n=i}^{N} (\mu_n^* - \lambda_n^*)$.

Proof: If $0 < \tau_i^* \leq t_i$ *and* $\theta_i^* > 0$*, according to the complementary slackness conditions in* (14)*, we have* $\phi_i^* = \gamma_i^* = 0$. *The optimal amount of the transmitted data by the MBS at time slot* t_i *can be obtained from* (10) *as* $\theta_i^* = \tau_i^* \ln \sum_{n=i}^{N} (\mu_n^* - \lambda_n^*)$. *By substituting* $\theta_i^* = r_i^* \tau_i^*$ *into the equation, we can obtain the optimal transmission rate as follows*

$$r_i^* = \ln \sum\nolimits_{n=i}^{N} (\mu_n^* - \lambda_n^*). \tag{15}$$

Lemma 5: When the circuit power consumption is negligible, i.e. $\varepsilon = 0$*, we consider this as the case of the ideal circuit power consumption. There are only two possible optimal transmission strategies for the MBS at each time slot* t_i*:*

(1) When $\theta_i^* > 0$, $\tau_i^* = t_i$ *. That is, when the MBS has the data to transmit, the transmission duration lasts the whole time slot.*

(2) When $\theta_i^* = 0$, $\tau_i^* = 0$*. That is, when there is no data to be transmitted by the MBS, the transmission duration is zero.*

Proof: This lemma can be proved by using reduction to absurdity. It is assumed that when $\theta_i^* > 0$, $0 < \tau_i^* < t_i$*. According to the complementary slackness condition in* (14)*, we have* $\phi_i^* = \varphi_i^* = \gamma_i^* = 0$*, it can be obtained from* (11)

$$e^{\theta_i^*/\tau_i^*} - 1 + \varepsilon - \frac{\theta_i^*}{\tau_i^*} e^{\theta_i^*/\tau_i^*} = 0 \tag{16}$$

by substituting $\theta_i^* = r_i^* \tau_i^*$ *into* (16)*, we obtain*

$$e^{r_i^*} - 1 - r_i^* e^{r_i^*} = 0 \tag{17}$$

The solution of (17) *is* $r_i^* = 0$*, and* $\theta_i^* = r_i^* \tau_i^* = 0$*. This is in contradiction with* $\theta_i^* > 0$*, and the hypothesis is not valid. Therefore, when* $\theta_i^* > 0$*, we have* $\tau_i^* = t_i$.

Lemma 6: When the circuit power consumption is non-negligible, we regard this case as the actual circuit power consumption. There is one of the two possible optimal transmission strategies for the MBS at each time slot:

(1) $\tau_i^* < t_i$, $r_i^* = r_{ee}$*;*

(2) $\tau_i^* = t_i$, $r_i^* \geq r_{ee}$*,*

where r_{ee} *is the transmission rate that maximizes the system EE, i.e.* $r_{ee} = \arg\max \frac{r}{g(r)+\varepsilon}$.

Proof: Considering the complementary slackness condition in (14)*, when* $\tau_i^* \neq 0$, $\gamma_i^* = 0$, (11) *can be rewritten as*

$$\frac{\partial \ell}{\partial \tau_i} = e^{\theta_i^*/\tau_i^*} - 1 + \varepsilon - \frac{\theta_i^*}{\tau_i^*} e^{\theta_i^*/\tau_i^*} + \varphi_i^* = 0, \tag{18}$$

by substituting $\theta_i^* = r_i^* \tau_i^*$ *into* (18)*, we can obtain* $r_i^* e^{r_i^*} - e^{r_i^*} - \varepsilon + 1 = \varphi_i^*$*. Let* $f(x) = xe^x - e^x - \varepsilon + 1$*, we have*

$$f(r_i^*) = r_i^* e^{r_i^*} - e^{r_i^*} - \varepsilon + 1 = \varphi_i^*. \tag{19}$$

Meanwhile, the system EE is defined as $\xi_{ee}(r) = \frac{r}{e^r - 1 + \varepsilon}$*, and it is easy to know that* $\xi_{ee}(r)$ *is a quasi-concave function. Therefore, there exists a unique optimal solution* r_{ee}*, which can be obtained by calculating the first derivative of* $\xi_{ee}(r)$.

$$\frac{d\xi_{ee}(r)}{dr} = \frac{-f(r)}{(e^r - 1 + \varepsilon)^2}. \tag{20}$$

It is noted that when $r < r_{ee}$, $\xi_{ee}(r)$ *is a monotonic incremental function about* r*; when* $r > r_{ee}$, $\xi_{ee}(r)$ *is a monotonic decreasing function about* r*; and when* $r = r_{ee}$, $\xi_{ee}(r)$ *reaches the maximum value, that is, the optimal EE. It can be summarized as follows*

$$\begin{aligned} f(r) < 0, \ \frac{d\xi_{ee}(r)}{dr} > 0, \quad & \text{when } r < r_{ee}; \\ f(r) = 0, \ \frac{d\xi_{ee}(r)}{dr} = 0, \quad & \text{when } r = r_{ee}; \\ f(r) > 0, \ \frac{d\xi_{ee}(r)}{dr} < 0, \quad & \text{when } r > r_{ee}. \end{aligned} \tag{21}$$

When $\tau_i^* < t_i$*, by the complementary slackness conditions in* (14)*, we have* $\varphi_i^* = 0$*. We can get* $f(r_i^*) = \varphi_i^* = 0$ *from* (19)*, therefore we obtain* $r_i^* = r_{ee}$ *from* (21)*. Similarly, when* $\tau_i^* = t_i$*, we have* $\varphi_i^* \geq 0$*, and* $f(r_i^*) = \varphi_i^* \geq 0$ *is deduced from* (19)*, and we can have* $r_i^* \geq r_{ee}$ *from* (21).

Considering the actual circuit power consumption, the non-zero optimal transmission rate of the MBS must satisfy $r_i^* \geq r_{ee}$ for any time slot t_i. That is to say, if r_{ee} can satisfy the feasible region of the data transmission, any transmission rate $r_i < r_{ee}$ will be replaced by r_{ee}. Because r_{ee} is the optimal transmission rate for EE, it can also minimize the energy consumption of the system. In this case, the transmission duration of the MBS at each time slot is $\tau_i^* < t_i$, that is, the data transmission is intermittent. If r_{ee} cannot satisfy the feasible region of the data transmission, the MBS will transmit the data at the whole time slot at rate $r_i^* \geq r_{ee}$.

In addition, it should be noted that Lemma (6) only gives the conditions that r_i^* should satisfy, and the relationship between the corresponding transmission duration τ_i^* and the time slot t_i for the optimal transmission strategy, but the optimal transmission strategy that satisfies these conditions is not unique. Any transmission strategy that transmits the data with the rate r_{ee} in the transmission duration $\tau_i^* < t_i$ is a feasible optimal solution.

Theorem 1: Any feasible transmission strategy that satisfies Lemmas (3)−(6) *is the optimal solution of the optimization problem* (8).

Proof: Lemmas (3)−(6) *are derived from the KKT conditions* (10)-(14) *of the optimization problem* (8)*, while problem* (8) *is a convex optimization problem. According to optimization theory, KKT conditions are the sufficient and necessary conditions for the optimal solution. Therefore, any feasible transmission strategy satisfying Lemmas* (3)−(6) *is optimal for the optimization problem* (8).

IV. AN ENERGY-EFFICIENT OPTIMAL OFFLINE TRANSMISSION STRATEGY

In this section, an offline transmission strategy with the optimal EE based on Lemmas 3−6 is proposed, then its optimality for problem (8) is illustrated.

When the local caching strategy $\{c_i\}$ is given, the feasible region of θ_i in problem (8), i.e. the right expressions of constraints (8b) and (8c), can be uniquely determined. So for a given local caching strategy $\{c_i\}$, the maximum data departure curve $B(t, c)$ and the minimum data departure curve $A(t, c)$ of problem (8) can be obtained. Within the feasible region, the optimal data departure curve for the case of the ideal circuit power consumption can be obtained by using the "String Tautening" algorithm [33], [44]. Intuitively it is equivalent to binding one end of a string at the starting point (0, 0) and passing the other end through the terminal point $(T, A(t))$. When the string is pulled tightly between the maximum data departure curve $B(t)$ and the minimum data departure curve $A(t)$, the trajectory of the string is the optimal data departure curve that minimizes the energy consumption of the MBS. When $r_i < r_{ee}$, the optimal data departure curve can be achieved by replaced r_i with r_{ee}, then we can get the optimal data departure curve for the case of the actual circuit power consumption. The specific algorithm is presented in Algorithm 1.

Step 1−step 4 in Algorithm 1 can obtain the minimum energy consumption for the case of the ideal circuit power consumption. According to Lemma (5), the optimal solution for the case of the ideal circuit power consumption is only $\tau_i = 0$ or $\tau_i = t_i$. The time slots for $\tau_i = t_i$ are divided into two sets of $M|_{r_i \geq r_{ee}}$ and $N|_{r_i < r_{ee}}$. Similarly, according to Lemma 6, the optimal solution of problem (8) is one of $\tau_i = 0$, $0 < \tau_i < t_i$, and $\tau_i = t_i$. The case of $\tau_i = 0$ is not considered, and the solution sets can be classified into two sets of $P|_{\tau_i = t_i}$ and $Q|_{0<\tau_i<t_i}$.

For the set $M|_{r_i \geq r_{ee}}$, $\tau_i = t_i$, r_i^* can minimize the objective function of the case of the ideal circuit power consumption. When a constant ε is added to the objective function, r_i^* is still the optimal solution. Therefore we can let $M|_{r_i \geq r_{ee}} = P|_{\tau_i = t_i}$. For the set $N|_{r_i < r_{ee}}$, r_i^* can minimize the objective function of the case of the ideal circuit power consumption, but it does not satisfy the condition of the optimal solution of problem (8). r_i^* in the set $N|_{r_i < r_{ee}}$ is replaced with r_{ee}, which can guarantee the minimal data transmission requirements. Let the transmission duration $\tau_i = \frac{r_i t_i}{r_{ee}}$, the set $Q|_{0<\tau_i<t_i}$ can be obtained, that is the execution process of step 5 in the proposed algorithm. Because $r_i < r_{ee}$, $\tau_i < t_i$. In summary, the proposed algorithm in Algorithm 1 satisfies Lemmas (3)−(6), so it is the optimal solution of problem (8).

Algorithm 1 An Energy-Efficient Offline Transmission Strategy

```
1. Let $t_{st} = s_1 = 0, D_{opt}(0) = 0, D_{opt}(T) = A(T)$, the optimal data departure curve $D_{opt}(t)$ is obtained recursively from the starting point $(t_{st} = 0, D_{opt}(0) = 0)$.
2. Find out $\beta_0 = \inf \mathcal{F}_B = \sup \mathcal{F}_A$ and get the optimal line segment $L_0$. $\mathcal{F}_B(\mathcal{F}_A)$ is a set of rays with positive slope which intersects with $B(t)(A(t))$ from the starting point.
3. Get the first intersection point $t_{end}$ , and $L_0(t_{end}) = B(t_{end})$ or $L_0(t_{end}) = A(t_{end})$. Set $D_{opt}(t)$ as $L_0(t)$, where $t \in (t_{st}, t_{end}]$.
4. If $t_{end} = T$ , go to step 5; else, set the starting point as $(t_{end}, D_{opt}(t_{end}))$ and repeat steps 2 and 3.
5. When $r_i < r_{ee}$ at time slot $t_i$ for $D_{opt}(t)$, the corresponding transmission curve is replaced by the line segment which the starting point is $s_i$, the slope is $r_{ee}$, and the length is $\tau_i = \frac{r_i t_i}{r_{ee}}$.
```

The computational complexity of the whole algorithm is related to the number of the requested files, the file popularity and the cache capacity. It is assumed that there are N requested files, the number of the file types is $|\mathcal{F}| \geq 1$, the number of files f_i is N_i, and we have $\sum_{i=1}^{|F|} N_i = N$. The computational complexity of the local caching rate can be expressed as $\mathcal{O}(N - |\mathcal{F}|)$, which is related to the file popularity. If the files are different, no local caching is required, all local caching rates are 0, and the computational complexity is $\mathcal{O}(1)$. If all files are the same, the computational complexity is $\mathcal{O}(1)$. To sum up, the computational complexity of solving the local caching rate is $\mathcal{O}(1) \sim \mathcal{O}(N - |\mathcal{F}|)$.

The process of each slot in algorithm 1 includes two steps. The first step is to find the optimal transmission rate for the ideal circuit power consumption. The second step is to compare the optimal transmission rate with r_{ee}. The computational complexity in each slot is 2, and the computational complexity of algorithm 1 is $\mathcal{O}(2N)$. Therefore, the computational complexity of the proposed algorithm is $\mathcal{O}(2N) \sim \mathcal{O}(2N(N - |\mathcal{F}|))$.

The computational complexity of the existing inner layer transmission algorithm in [45] is $\mathcal{O}(N^2)$, the computational complexity of the outer layer is $\mathcal{O}(N)$, and the total computational complexity is $\mathcal{O}(N^3)$. Therefore, the computational

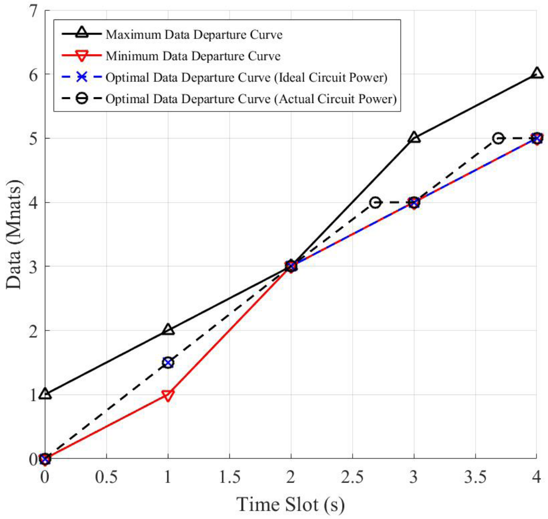


FIGURE 3. **Illustration of the proposed transmission strategy when the cache capacity is 1 Mnats.**

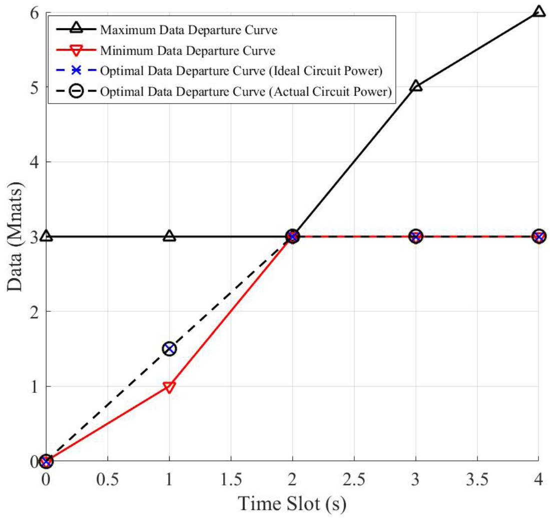


FIGURE 4. **Illustration of the proposed transmission strategy when the cache capacity is 3 Mnats.**

complexity of the proposed algorithm is more efficient than that of the existing one.

V. NUMERICAL RESULTS

In this section, we provide the numerical simulation results of the proposed energy-efficient offline transmission strategy. It is assumed that the system bandwidth $W = 1$ MHz, the circuit power consumption $\varepsilon = 3$ Watt [36], [39]. The time slot is normalized to $t_i = 1$ s, $i = 1, 2, \ldots, N$.

Fig. 3 and Fig. 4 describe the optimal data departure curve of the MBS when the cache capacities are 1 Mnats and 3 Mnats, respectively. The number of the data files is $N = 4$, and the order of the data files is $\{f_1, f_2, f_2, f_1\}$, where the sizes of the data files f_1 and f_2 are 1 Mnats and 2 Mnats, respectively.

In Fig. 3, the cache capacity is 1 Mnats, and the optimal caching rates are $c(t_2) = 1$ Mnats/s, and $c(t_1) = c(t_3) = c(t_4) = 0$ Mnats/s, then the maximum data departure curve and the minimum data departure curve can be obtained. For the ideal circuit power consumption, in time slot t_1, the transmission rate of the MBS is $r_1 = 1.5$ Mnats/s, and the cumulative transmission data is 1.5 Mnats. The MBS not only transmits 1 Mnats data of the file f_1 completely, but also transmits 0.5 Mnats data of the file f_2, which is stored in the cache. It provides prefetching gain for the file f_2. At time instant s_2, the interval between the maximum data departure curve and the optimal data departure curve is 0.5 Mnats, which means that there is still 0.5 Mnats free space available in the cache. At time slot t_2, the transmission rate r_2 is 1.5 Mnats/s, and the 0.5 Mnats data of the file f_2 stored in the cache at time slot t_1 is also transmitted. The cumulative transmission data is 3 Mnats at time instant s_3, and the file f_2 is transmitted completely. The local caching rate is $c(t_2) = 1$ Mnats/s, and there is 1 Mnats data of the file f_2 stored in the cache for the future transmission, which provides the local caching gain. The maximum data departure curve intersects the minimum data departure curve at time instant s_3, and the cache is full. At time slot t_3, the 1 Mnats data of the file f_2 stored in the cache is transmitted to the user. In order to satisfy the user's file request, the MBS must transmit the remaining 1 Mnats data of file f_2 to the SBS with rate $r_3 = 1$ Mnats/s. At time slot t_4, the transmission rate of file f_1 is $r_4 = 1$ Mnats/s, which is the same as the user's requested rate. Considering the actual circuit power consumption, the optimal transmission rate is $r_{ee} = 1.463$ Mnats/s. At time slots t_1 and t_2, $r_1 > r_{ee}$ and $r_2 > r_{ee}$, the data transmission rate of the MBS is still r_1 and r_2. Therefore, the optimal actual data departure curve overlaps the optimal ideal data departure curve in these two time slots. At time slots t_3 and t_4, $r_3 < r_{ee}$ and $r_4 < r_{ee}$, the optimal data transmission rate of the actual circuit power consumption is replaced by r_{ee} at these two time slots. The transmission duration $\tau_3 = \frac{r_3 t_3}{r_{ee}} = 0.684$ s, and $\tau_4 = \frac{r_4 t_4}{r_{ee}} = 0.684$ s. The ideal energy consumption per unit bandwidth at the four time slots is 3.482 J, 3.482 J, 1.718 J and 1.718 J, and the total ideal energy consumption per unit bandwidth is 10.4 J. The corresponding actual energy consumption per unit bandwidth is 6.482 J, 6.482 J, 4.322 J, and 4.322 J, and the total actual energy consumption per unit bandwidth is 21.61 J. It can be seen that the actual energy consumption is greater than the ideal energy consumption, it is mainly because the actual energy consumption includes the circuit power consumption.

In Fig. 4, the cache capacity is 3 Mnats, and the optimal caching rates are $c(t_1) = 1$ Mnats/s, $c(t_2) = 2$ Mnats/s, and $c(t_3) = c(t_4) = 0$ Mnats/s. For the case of the ideal circuit power consumption, the transmission rates are $r_1 = r_2 = 1.5$ Mnats/s, and $r_3 = r_4 = 0$ Mnats/s. Since $r_{ee} = 1.463$ Mnats/s, $r_1(r_2) > r_{ee}$, and $r_3 = r_4 = 0$ Mnats/s, the optimal data transmission curve for the case of the actual circuit power consumption is the same as that for the case of the ideal circuit power consumption. At time slot t_1, the MBS transmits a total of 1.5 Mnats data, which

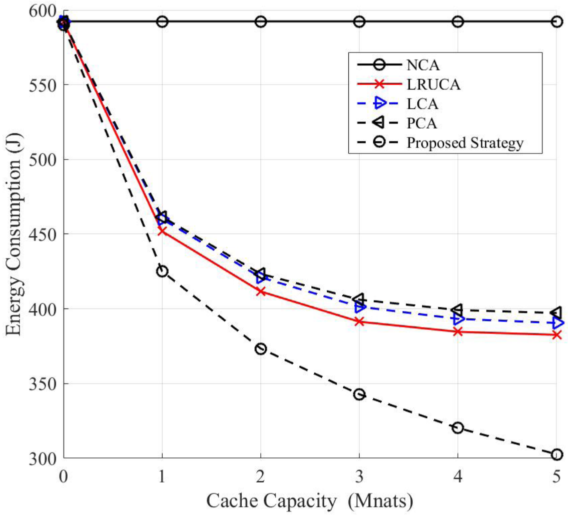


FIGURE 5. The energy consumption with respect to the size of the cache capacity.

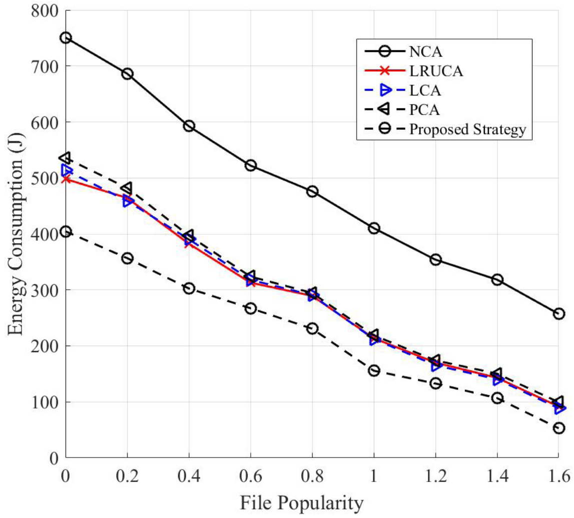


FIGURE 6. The energy consumption with respect to the file popularity.

includes 1 Mnats data of file f_1 and 0.5 Mnats data of file f_2. These two parts of data are stored in the cache, which provide the local caching gain and the prefetching gain. In time slot t_2, the MBS continues to transmit the remaining 1.5 Mnats data of file f_2, and this part of data is also stored in the cache, which provides the local caching gain. At time instant s_3, the data of files f_1 and f_2 are stored in the cache, and the cache is full. At time slots t_3 and t_4, the MBS does not need to transmit any more data, and the SBS can transmit the caching data directly to the corresponding users, thus the energy saving can be obtained. Therefore, the cache capacity has an impact on the energy consumption.

Fig. 5 and Fig. 6 illustrate the energy consumption versus the cache capacity and the file popularity for different algorithms, respectively. Considering a total of $N = 60$ files that may be requested by the users, the requested rates $d(t)$ are distributed in the range of [0.1, 3.5] Mnats/s [31]. The probability of the requested file f_j is θ_j, which is independent and identically distributed at each time slot and obeys the Zipf distribution, i.e., $\theta_j = j^{-\gamma}(\sum_{q=1}^{|F|} q^{-\gamma})$, where γ denotes the skewness of the file popularity [32]. When $\gamma = 0$, the file popularity obeys the uniform distribution. When γ becomes larger, the file popularity begins to skew. For comparison, the performance of the other four transmission algorithms are given. The no caching algorithm (NCA): the MBS transmits the data in real time with the rate $r(t) = d(t)$, and there is no cache at the SBS. The prefetching caching algorithm (PCA): the cache is only used for prefetching data, that is, it only provides the prefetching gain [29]. The local caching algorithm (LCA): for the most popular files, they are cached at the first time, then are stored in the cache for the rest time slots, i.e., only the local caching gain is provided. The least recently used caching algorithm (LRUCA): the cache always stores the most recently requested files [46].

The energy consumption with respect to the cache capacity is shown in Fig. 5, where $\gamma = 0.4$. The NCA does not have any prefetching gain and local caching gain, and consumes the most energy. Its energy consumption does not change with the cache capacity. With the increase of the cache capacity, the energy consumption of PCA decreases slowly, then the curve of the energy consumption tends to be flat. The curves of the energy consumption of LRUCA and LCA are similar to that of PCA, but the performance of LRUCA is better than that of LCA and PCA. The performance of the proposed energy-efficient optimal strategy is obviously superior to these of the other four algorithms. With the increase of the cache capacity, the superiority of the proposed strategy becomes more obvious. This is because the proposed strategy makes full use of the prefetching gain and the local caching gain.

Fig. 6 shows the impact of the file popularity on the energy consumption of the MBS, and the cache capacity is 4 Mnats. With the increase of file popularity, the curves of the energy consumption in the five algorithms are all decreasing. NCA consumes the most energy. The performances of LRUA, LCA and PCA are similar. The performance of the proposed strategy is obviously better than that of the other four algorithms, which validates the superiority of the proposed strategy in energy saving.

VI. CONCLUSION

The energy consumption minimization of the MBS for the cache-enabled wireless network with the circuit power consumption was investigated. The MBS transmitted the wireless data to the SBS, and the SBS equipped with a cache served the users. By utilizing the transmission characteristics of the data files, the infinite-dimension optimization problem was transformed into the finite-dimension optimization problem. Under a given local caching strategy, the maximum data departure curve and the minimum data departure curve were determined. The sufficient and necessary conditions for the optimal transmission strategy were obtained. Then, an optimal energy-efficient offline transmission strategy was proposed. The simulation results were provided to verify the performance of the proposed strategy. The simulation results showed the influence of the cache capacity and the file popularity on the performance of the proposed strategy. By comparing with the other algorithms, the performance of the proposed strategy had the obvious advantage. In the future, we will extend our work to the case of multiple antennas.

REFERENCES

[1] Q. Wu, G. Y. Li, W. Chen, D. W. K. Ng, and R. Schober, "An overview of sustainable green 5G networks," *IEEE Wireless Commun.*, vol. 24, no. 4, pp. 72–80, Aug. 2017.

[2] V. W. S. Wong, R. Schober, D. W. K. Ng, and L.-C. Wang. *Key Technologies for 5G Wireless Systems*. Cambridge, U.K.: Cambridge Univ. Press, 2017.

[3] J. Wu, S. Luo, S. Wang, and H. Wang, "NLES: A novel lifetime extension scheme for safety-critical cyber-physical systems using SDN and NFV," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2463–2475, Apr. 2019.

[4] G. Li, J. Wu, J. Li, K. Wang, and T. Ye, "Service popularity-based smart resources partitioning for fog computing-enabled industrial Internet of Things," *IEEE Trans. Ind. Informat.*, vol. 14, no. 10, pp. 4702–4711, Oct. 2018.

[5] Z. Guan, Y. Zhang, L. Wu, J. Wu, J. Li, Y. Ma, and J. Hu, "APPA: An anonymous and privacy preserving data aggregation scheme for fog-enhanced IoT," *J. Netw. Comput. Appl.*, vol. 125, pp. 82–92, Jan. 2019.

[6] J. Wu, M. Dong, K. Ota, J. Li, and Z. Guan, "Big data analysis-based secure cluster management for optimized control plane in software-defined networks," *IEEE Trans. Netw. Service Manage.*, vol. 15, no. 1, pp. 27–38, Mar. 2018.

[7] "Cisco visual networking index: Global mobile data traffic forecast update, 2016–2021," Cisco, San Jose, CA, USA, White Paper 1454457600805266, 2017. [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/serviceprovider/visual-networking-index-vni/mobile-white-paper-c11-520862.html

[8] C. Li, F. Sun, L. Yang, and J. M. Cioffi, "Energy efficient MIMO relay transmissions via joint power allocations," *IEEE Trans. Circuits Syst. II, Exp. Briefs*, vol. 61, no. 7, pp. 531–535, Jul. 2014.

[9] H. Dai, Y. Huang, Y. Xu, C. Li, B. Wang, and L. Yang, "Energy-efficient resource allocation for energy harvesting-based device-to-device communication," *IEEE Trans. Veh. Technol.*, vol. 68, no. 1, pp. 509–524, Jan. 2019.

[10] C. Li, S. Zhang, P. Liu, F. Sun, J. M. Cioffi, and L. Yang, "Overhearing protocol design exploiting intercell interference in cooperative green networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 1, pp. 441–446, Jan. 2016.

[11] C. Li, H. J. Yang, F. Sun, J. M. Cioffi, and L. Yang, "Multiuser overhearing for cooperative two-way multiantenna relays," *IEEE Trans. Veh. Technol.*, vol. 65, no. 5, pp. 3796–3802, May 2016.

[12] C. Li, P. Liu, C. Zou, F. Sun, J. M. Cioffi, and L. Yang, "Spectral-efficient cellular communications with coexistent one- and two-hop transmissions," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6765–6772, Aug. 2016.

[13] C. Li, H. J. Yang, F. Sun, J. M. Cioffi, and L. Yang, "Adaptive overhearing in two-way multi-antenna relay channels," *IEEE Signal Process. Lett.*, vol. 23, no. 1, pp. 117–120, Jan. 2016.

[14] S. Arshad, M. A. Azam, M. H. Rehmani, and J. Loo, "Recent advances in information-centric networking-based Internet of Things (ICN-IoT)," *IEEE Internet Things J.*, vol. 6, no. 2, pp. 2128–2158, Apr. 2019.

[15] L. Li, G. Zhao, and R. S. Blum, "A survey of caching techniques in cellular networks: Research issues and challenges in content placement and delivery strategies," *IEEE Commun. Surveys Tuts.*, vol. 20, no. 3, pp. 1710–1732, 3rd Quart., 2018.

[16] H. Dai, Y. Huang, J. Wang, and L. Yang, "Resource optimization in heterogeneous cloud radio access networks," *IEEE Commun. Lett.*, vol. 22, no. 3, pp. 494–497, Mar. 2018.

[17] J. Wu, M. Dong, K. Ota, J. Li, and Z. Guan, "FCSS: Fog computing based content-aware filtering for security services in information centric social networks," *IEEE Trans. Emerg. Topics Comput.*, to be published.

[18] E. Zeydan, E. Bastug, M. Bennis, M. A. Kader, I. A. Karatepe, A. S. Er, and M. Debbah, "Big data caching for networking: Moving from cloud to edge," *IEEE Commun. Mag.*, vol. 54, no. 9, pp. 36–42, Sep. 2016.

[19] 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 Commun. Mag.*, vol. 52, no. 2, pp. 131–139, Feb. 2014.

[20] A. Khreishah and J. Chakareski, "Collaborative caching for multicell-coordinated systems," in *Proc. IEEE Conf. Comput. Commun. Workshops*, Apr./May 2015, pp. 257–262.

[21] C. Yang, Y. Yao, Z. Chen, and B. Xia, "Analysis on cache-enabled wireless heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 1, pp. 131–145, Jan. 2016.

[22] M. A. Maddah-Ali and U. Niesen, "Fundamental limits of caching," *IEEE Trans. Inf. Theory*, vol. 60, no. 5, pp. 2856–2867, May 2014.

[23] F. Pantisano, M. Bennis, W. Saad, and M. Debbah, "Match to cache: Joint user association and backhaul allocation in cache-aware small cell networks," in *Proc. IEEE Int. Conf. Commun. (ICC)*, London, U.K., Jun. 2015, pp. 3082–3087.

[24] B. Zhou, Y. Cui, and M. Tao, "Stochastic content-centric multicast scheduling for cache-enabled heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 15, no. 9, pp. 6284–6297, Sep. 2016.

[25] N. Golrezaei, K. Shanmugam, A. G. Dimakis, A. F. Molisch, and G. Caire, "FemtoCaching: Wireless video content delivery through distributed caching helpers," in *Proc. IEEE INFOCOM*, Orlando, FL, USA , Mar. 2012, pp. 1107–1115.

[26] J. Hachem, N. Karamchandani, and S. Diggavi, "Content caching and delivery over heterogeneous wireless networks," in *Proc. IEEE Conf. Comput. Commun. (INFOCOM)*, Kowloon, Hong Kong, Apr./May 2015, pp. 756–764.

[27] S. Tamoor-ul-Hassan, M. Bennis, P. H. J. Nardelli, and M. Latva-Aho, "Modeling and analysis of content caching in wireless small cell networks," in *Proc. IEEE Int. Symp. Wireless Commun. Syst. (ISWCS)*, Brussels, Belgium, Aug. 2015, pp. 765–769.

[28] B. Lv, L. Huang, and R. Wang, "Joint content delivery and caching placement via dynamic programming," Jan. 2018, *arXiv:1801.00924*. [Online]. Available: https://arxiv.org/abs/1801.00924

[29] A. C. Güngör and D. Gündüz, "Proactive wireless caching at mobile user devices for energy efficiency," in *Proc. IEEE Int. Symp. Wireless Commun. Syst. (ISWCS)*, Brussels, Belgium, Aug. 2015, pp. 186–190.

[30] S. Sadr and S. Valentin, "Anticipatory buffer control and resource allocation for wireless video streaming," Apr. 2013, *arXiv:1304.3056*. [Online]. Available: https://arxiv.org/abs/1304.3056

[31] M. Gregori, J. Gómez-Vilardebò, J. Matamoros, and D. Gündüz, "Joint transmission and caching policy design for energy minimization in the wireless backhaul link," in *Proc. IEEE Int. Symp. Inf. Theory (ISIT)*, Hong Kong, China, Jun. 2015, pp. 1004–1008.

[32] M. Gregori, J. Gómez-Vilardebó, J. Matamoros, and D. Gündüz, "Wireless content caching for small cell and D2D networks," *IEEE J. Sel. Areas Commun.*, vol. 34, no. 5, pp. 1222–1234, May 2016.

[33] M. A. Zafer and E. Modiano, "A calculus approach to energy-efficient data transmission with quality-of-service constraints," *IEEE/ACM Trans. Netw.*, vol. 17, no. 3, pp. 898–911, Jun. 2009.

[34] K. Tutuncuoglu and A. Yener, "Optimum transmission policies for battery limited energy harvesting nodes," *IEEE Trans. Wireless Commun.*, vol. 11, no. 3, pp. 1180–1189, Mar. 2012.

[35] O. Ozel, K. Tutuncuoglu, J. Yang, S. Ulukus, and A. Yener, "Transmission with energy harvesting nodes in fading wireless channels: Optimal policies," *IEEE J. Sel. Areas Commun.*, vol. 29, no. 8, pp. 1732–1743, Sep. 2011.

[36] Z. Nan, T. Chen, X. Wang, and W. Ni, "Energy-efficient transmission schedule for delay-limited bursty data arrivals under nonideal circuit power consumption," *IEEE Trans. Veh. Technol.*, vol. 65, no. 8, pp. 6588–6600, Aug. 2016.

[37] X. Chen, W. Ni, X. Wang, and Y. Sun, "Optimal quality-of-service scheduling for energy-harvesting powered wireless communications," *IEEE Trans. Wireless Commun.*, vol. 15, no. 5, pp. 3269–3280, May 2016.

[38] O. Orhan, D. Gündüz, and E. Erkip, "Energy harvesting broadband communication systems with processing energy cost," *IEEE Trans. Wireless Commun.*, vol. 13, no. 11, pp. 6095–6107, Nov. 2014.

[39] Q. Zhou, Z. Yang, N. Liu, Z. Pan, and X. You, "Energy-efficient data transmission with non-FIFO packets with processing cost," *IEEE Access*, vol. 5, pp. 5158–5170, 2017.

[40] J. Xu and R. Zhang, "Throughput optimal policies for energy harvesting wireless transmitters with non-ideal circuit power," *IEEE J. Sel. Areas Commun.*, vol. 32, no. 2, pp. 322–332, Feb. 2014.

[41] J. Yang and S. Ulukus, "Optimal packet scheduling in an energy harvesting communication system," *IEEE Trans. Commun.*, vol. 60, no. 1, pp. 220–230, Jan. 2012.

[42] P. Youssef-Massaad, L. Zheng, and M. Medard, "Bursty transmission and glue pouring: On wireless channels with overhead costs," *IEEE Trans. Wireless Commun.*, vol. 7, no. 12, pp. 5188–5194, Dec. 2008.

[43] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[44] X. Chen, W. Ni, X. Wang, and Y. Sun, "Provisioning quality-of-service to energy harvesting wireless communications," *IEEE Commun. Mag.*, vol. 53, no. 4, pp. 102–109, Apr. 2015.

[45] Y. Jin, J. Xu, and L. Qiu, "Energy-efficient scheduling with individual packet delay constraints and non-ideal circuit power," *J. Commun. Netw.*, vol. 16, no. 1, pp. 36–44, Feb. 2014.

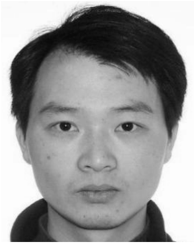
[46] L. Rizzo and L. Vicisano, "Replacement policies for a proxy cache," *IEEE/ACM Trans. Netw.*, vol. 8, no. 2, pp. 158–170, Apr. 2000.

CHAO MENG received the M.S. degree in electronic engineering from Lanzhou Jiaotong University, Lanzhou, China, in 2008, and the Ph.D. degree from the National Mobile Communications Research Laboratory, Southeast University, Nanjing, China, in 2015. He is currently a Lecturer with the Jinling Institute of Technology. His research interests include cooperative communication, energy-efficient communication, physical layer security, and energy harvesting.

GANG WANG received the M.S. degree in circuit and system and the Ph.D. degree in communication and information system from Southeast University, Nanjing, China, in 2003 and 2019, respectively, where he is currently a Lecturer with the National Mobile Communications Research Laboratory. His research interests include green communications, energy harvesting, and wireless communication theory.

XI DAI received the B.E. degree in electronic engineering from the University of Electronic Science and Technology of China (UESTC), Chengdu, China, in 2011, and the M.E. degree in electronic engineering from the Stevens Institute of Technology, NJ, USA, in 2013. She is currently a Lecturer with the Jinling Institute of Technology, Nanjing, China. Her current research interests include signal processing, image processing, and computer vision.

SHIPING CHEN is currently a Principal Research Scientist with CSIRO Data61. He is also an Adjunct Associate Professor with the University of Sydney and the University of New South Wales (UNSW) through teaching and supervising Ph.D. students. He has been working on distributed systems for over 20 years with a focus on performance, security, and trust. He has published over 150 research papers in these areas. He is also actively involved in software engineering and service/cloud computing research communities through publications, journal editorships, and conference TPC services, including WWW, EDOC, ICSOC, and IEEE ICWS/SCC/CLOUD. His current research interests include cybersecurity, blockchain, and service-oriented collaborations. He is a Senior Member of the IEEE.



WEI NI received the B.E. and Ph.D. degrees in electronic engineering from Fudan University, Shanghai, China, in 2000 and 2005, respectively. He is currently a Team Leader with CSIRO, Sydney, NSW, Australia, and an Adjunct Professor with the University of Technology Sydney (UTS). He also holds adjunct positions at the University of New South Wales (UNSW) and Macquarie University (MQ). Prior to this, he was a Postdoctoral Research Fellow with Shanghai Jiaotong University, from 2005 to 2008, a Deputy Project Manager with the Bell Labs R&D Center, Alcatel/Alcatel-Lucent, from 2005 to 2008, and a Senior Researcher with Devices R&D, Nokia, from 2008 to 2009. His research interests include stochastic optimization, game theory, graph theory, as well as their applications to network and security. He has been the Vice Chair of the IEEE NSW VTS Chapter, since 2018, and he was the Secretary of the IEEE NSW VTS Chapter, from 2015 to 2018, the Track Chair for VTC-Spring 2017, the Track Co-Chair for the IEEE VTC-Spring 2016, and the Publication Chair for BodyNet 2015, and a PC Member of the IEEE ICC'14, ICCC'15, EICE'14, and WCNC'10. He is a Senior Member of the IEEE.

• • •