Energy Minimization Transmission for Cache-Enabled Wireless Relay Networks

Bin Ge¹, Ziheng Li¹ and Chao Meng²

¹College of Computer Science and Engineering, AnHui University of Science and Technology, Huainan 232001, China
²School of Networks and Telecommunications Engineering, Jinling Institute of Technology, Nanjing 211169, China
Email: zihenglee@163.com

Abstract. Caching techniques have attracted extensive attention, which can significantly reduce peak traffic load and access latency, and improve energy efficiency and quality of service. In this paper, we strive to proactively caching the content requested by the user in periods with low traffic to minimize the total energy consumption and a low complexity algorithm is proposed. Furthermore, we also propose an optimal offline transmission strategy. Considering the cache size is large enough, we can obtain the optimal cache size, which can achieve our optimization goal. Numerical and simulation results show that the optimal data transmission algorithm can significantly reduce the global energy consumption of the network compared with the traditional not caching algorithm. As cache size increases, the superiority of the proposed algorithm becomes more obvious.

1. Introduction
With the wide application of an increasing number of mobile devices, such as laptops, smart phones and so on, mobile data traffic is experiencing tremendous growth, which presents an unprecedented challenge to service providers. The mobile traffic is forecast to reach 49 Ebytes per month by the end of 2021 [1]. Nevertheless, with the deployment of a large number of BSs, the expensive backhaul links between base stations(BSs) and the core network or among BSs, causes huge energy consumption [2].

Caching is a promising approach to alleviate the explosive traffic, which can reduce the backhaul cost, access latency and energy consumption significantly. If SBSs can predict the requested content, locations and other critical information of users in advance, then prefetching the requested content during off-peak time and storing them in cache memories installed in SBSs or user devices. When requested by users, SBSs are able to directly send the requested content to each users without requiring the use of the backhaul link [3], [4]. In addition, the cost of installing cache memories is much lower compared with that of improving the backhaul capacity, which can lead to significant gains to the future wireless network [2].

In the early 1990s, Caching techniques have been used to alleviate web congestion and obtained important gains [5]. In recent years, the application of cache is considered in many literatures. [6] proposed a user-centric proactive caching policy, which defined a collaboration distance according to the affordable energy cost of users, and investigated the tradeoff between maximizing traffic offloading and minimizing energy cost by considering the collaboration distance and the user demands statistics. Due to the limited cache size, not all files are cached in BSs. An optimal content placement strategy was proposed to minimize the total data distribution energy consumption, which combined file splitting with MDS encoding in [7]. [8] analysed the relationship between the energy efficiency of
the cache-enabled cellular networks with limited backhaul and the content popularity and cache size. It also presented an optimal BS density, which can maximize the energy efficiency with limited cache size. In [9], a “String Visualization” optimal transmission strategy was proposed, which used a calculus approach and considered controlling data transmission rate to realize a continuous-time optimization on the minimization of energy consumption over a finite-time horizon. Furthermore, the optimization problem was illustrated graphically.

In this paper, we assume the requested content are pre-downloaded in the cache. Once users request the content, the content is immediately removed from the cache, regardless of whether the same content will be requested in the future. In order to minimize energy consumption, the additional white Gaussian noise channel model is considered in the case of limited storage capacity, and this energy minimization problem is transformed into a convex optimization problem. In addition, assuming that the storage capacity is large enough, we can get an approximately constant transmission rate over the entire transmission time, which results in minimal energy consumption, and the optimal cache capacity is achieved. Moreover, a new algorithm with lower complexity is proposed, which can minimize energy consumption while meeting the user’s requested rate.

2. System Model

We assume the length of the data transmission time during the entire optimization process is $T$, which consists of $n$ time slots, and define $s_n$ as the starting point of the $n$-th time slot. Let $0 = s_0 < s_1 < \ldots < s_{n-1} < s_n < T$, the length of the $n$-th time slot can be denoted as $t_n = s_{n+1} - s_n$, and the length of each time slot does not have to be equal. The user’s requested rate $d(t)$ remains constant within a time slot and may not be equal in different time slot.

In addition, the data departure curve $D(t, r)$, represents the amount of the cumulative data transmitted by the MBS at time $t (t \geq 0)$ and can be denoted as $D(t, r) = \int_0^t r(\tau) d\tau$, $r(t)$ is the instantaneous transmission rate at time $t$ from the MBS to the SBS. Similarly, the maximum data departure curve $B(t, c) = B + \int_0^t d(\tau) d\tau$, is the maximum amount of data transmitted by the MBSs at time $t (t \geq 0)$, which can ensure the amount of data transmitted by the MBSs does not cause the cache to overflow. The minimum amount of the data must be cumulatively transmitted by the MBSs at time $t (t \geq 0)$ is denoted as $A(t, c) = \int_0^t d(\tau) d\tau$, which aims to meet the user’s demand.

According to the classical Shannon channel capacity formula, the instantaneous transmission rate can be expressed as

$$r(t) = \log_2(1 + \frac{h(t)p(t)}{\sigma^2}),$$ (1)

where $h(t)$ represents the channel power gain from the MBS to the SBS, and $\sigma^2$ denotes the noise power of the channel. For ease of presentation, the additional white Gaussian noise channel model is considered, (1) can be further expressed as

![Figure 1. A system model of heterogeneous wireless cache networks.](image-url)
Therefore, the power consumption \( p(t)=e^{r(t)}-1 \) can be obtained.

We assume the order of requested content is known, and the finite cache capacity is \( B \). We will prefetch the content that has not been requested and store it in cache installed in SBSs. The content is removed from the cache as soon as it is requested by the user, regardless of whether the same content will be requested again in the future.

The goal of this paper is to minimize the total energy consumed by the BS within a given time range \([0, T]\). Therefore, the optimization problem can be expressed as:

\[
\min_{r(t)} E(r(t)) = \int_0^T (e^{r(t)} - 1) du
\]

\text{s.t. } \int_0^T r(u) du \geq \int_0^T d(u) du, \forall t \in [0, T]

\int_0^T r(u) du \leq B + \int_0^T d(u) du, \forall t \in [0, T]

\tag{3b}

r(t) \geq 0
\tag{3d}

where formula (3a) is the objective function, which represents minimizing the energy consumption of the MBSs. The constraint (3b) denotes the instantaneous transmission rate must satisfy the user’s requested rate, which prevents communication interruption. The constraint (3c) restricts the maximum departure curve. The cumulative transmission data of the MBS at any time cannot overflow the cache memory due to the limited cache memory. The constraint (3d) means that the instantaneous transmission rate must be non-negative.

It is obvious that problem 3 is an infinite dimensional optimization problem, which is hard to solve in general. Considering some properties of the user’s requested rate within each time slot and the channel condition, this problem can be transformed into a finite dimensional optimization problem and expressed as:

\[
\min_{r_i} \sum_{i=1}^N t_i \left( e^{r_i} - 1 \right)
\]

\text{s.t. } \sum_{i=1}^n t_i (d_i - r_i) \leq 0, \text{ for } n = 1, \ldots, N
\tag{4b}

\sum_{i=1}^n t_i (r_i - d_i) - B \leq 0 \text{ and } r_i \geq 0, \text{ for } n = 1, \ldots, N.
\tag{4c}

Obviously, the objective function is convex and all constraints are linear. Therefore, this convex optimization problem can be solved easily by using standard optimization tools, which allows performance analysis and optimal design for any system parameters with low complexity. Accordingly, this paper provides some relevant characteristics of the optimal solution by derivation analysis and an optimal transmission strategy is proposed.

### 3. Optimal Transmission Strategy

Problem 4 is a convex optimization problem, the corresponding Lagrangian function can be expressed as:

\[
L = \sum_{i=1}^N t_i (e^{r_i} - 1) + \sum_{j=1}^J \lambda_j \left( \sum_{i=1}^n t_i (d_i - r_i) \right) + \sum_{j=1}^J \mu_j \left( \sum_{i=1}^n t_i (r_i - d_i) - B \right) + \sum_{j=1}^J \eta_j r_j
\]

where \( \lambda_j \geq 0, \mu_j \geq 0, \eta_j \geq 0 \) are the lagrangian multipliers defined to satisfy constraint (4b) - (4c). Furthermore, the additional slackness conditions are given by
\begin{align*}
\lambda_j \left( \sum_{i=1}^{N} t_i (d_i - r_j) \right) &= 0, \quad \forall j, \quad (6) \\
\mu_j \left( \sum_{i=1}^{N} t_i (r_j - d_i) - B \right) &= 0, \quad \forall j, \quad (7) \\
\eta_j r_j &= 0, \quad \forall j. \quad (8)
\end{align*}

When \( \lambda_j \) and \( \mu_j \) are positive, another parameter of the condition in the corresponding slackness condition must be zero to satisfy the slackness condition. Thus, \( \lambda_j \) and \( \mu_j \) cannot be positive at the same time, that is, \( \lambda_j \) and \( \mu_j \) always satisfy the following formula:

\[ \lambda_j \mu_j = 0, \quad \forall j = 1, \ldots, N. \quad (9) \]

According to the KKT conditions, we can obtain

\[ \frac{\partial L}{\partial r_j} = t_j e^0 \sum_{j=1}^{N} \lambda_j t_j + \sum_{j=1}^{N} \mu_j t_j - \eta_i = 0. \quad (10) \]

Considering the case that cache capacity is large enough \( (B \to \infty) \), from the second set of slackness condition in (7), we can find \( \mu_j = 0 \) for \( j = 1, \ldots, N \). The optimal instantaneous transmission rate is given as follows:

\[ r_j^* = \ln \left( \sum_{j=1}^{N} \lambda_j \right). \quad (11) \]

A special case can be obtained within total transmission time \( T \), which the optimal data departure curve tangents to the maximum departure curve and the cache capacity \( B \) at this time is the optimal value.

4. An Optimal Energy-efficient Offline Transmission Strategy

In this section, an optimal offline transmission strategy is proposed according to the analysis of problem (4a) and the optimality for problem (4a) is illustrated.

The maximum departure curve \( B(t, c) \) and the minimum departure curve \( \Lambda(t, c) \) can be obtained based on the constraints (4b) and (4c). We define the slopes of these two curves as \( r_i^* (r_{i,1}^*, r_{i,2}^*, \ldots, r_{i,N}^*) \) and \( r_j^* (r_{1,j}^*, r_{2,j}^*, \ldots, r_{N,j}^*) \), for \( i = 1, 2, \ldots, N \), respectively.

**Table 1.** An optimal offline transmission strategy.

| Step | Description |
|------|-------------|
| 1.   | Let \( t_o = s_i = 0 \), \( D_{op}(0) = 0 \), \( D_{del}(0) = 0 \), the optimal data departure curve \( D_{op}(t) \) is obtained recursively which starts from the starting point (\( t_o = 0 \), \( D_{op}(0) = 0 \)). |
| 2.   | Connect the starting point and the inflection point of the maximum departure curve and the minimum departure curve at time point \( s_2 \). Corresponding positive slope is \( r_{i,1}^*, r_{i,2}^* \) respectively. Let \( r_{i,op} = r_{i,1}^*, r_{i,low} = r_{i,2}^* \). |
| 3.   | Connect the starting point and the inflection point of the maximum departure curve and the minimum departure curve at time point \( s_3 \). Corresponding positive slope is \( r_{s,1}^*, r_{s,2}^* \) respectively. Extend ray \( r_{i,1}^*, r_{i,2}^* \) and intersect \( t = s_3 \) at point \( r_{i,1}, r_{i,2} \). Let \( r_{i,op} = \min \{ r_{i,1}, r_{i,2} \}, r_{i,low} = \max \{ r_{i,1}, r_{i,2} \} \). |
| 4.   | If \( s_3 = T \), stop the recursion. Otherwise, continue execute step2. |
| 5.   | The optimal transmission rate can be obtained with ideal circuit power consumption through Step2–Step4: \( I_{ideal} = \{ I_{ideal}(t_1), I_{ideal}(t_2), \ldots, I_{ideal}(s_N) \} \). |
As a result, the optimal departure curve for the case of the ideal circuit power consumption can be achieved through step 2 - step 4 in Table 1.

5. Numerical Results
This section presents the numerical results of the pre-download caching algorithm for the case of the ideal circuit power consumption. We assume the system bandwidth $W=1\text{MHz}$ and the circuit power consumption $\varepsilon=3$ Watt. The time slot is normalized to $t_i = 1s$, for $i = 1, 2, ..., N$.

The optimal data departure curve corresponding to different cache capacity sizes are presented in Fig. 2 and Fig. 3. The number of time slots is $N = 4$. We assume the order of the data file is $\{f_1, f_2, f_3, f_4\}$, where the sizes of the data file $f_1$, $f_2$, $f_3$ and $f_4$ are 25 Mnats, 5 Mnats, 8 Mnats and 65 Mnats, respectively. In Fig. 2, the cache capacity is 20 Mnats. The cumulative amount of data transferred is 25 Mnats in time slot 1. The distance between the maximum departure curve and the optimal data departure curve is 20 Mnats, which means the remaining cache capacity is 20 Mnats. In time slot 2, the remaining cache capacity is 8.5 Mnats which indicates that not only the file requested by the user is transmitted during this time, but also 8 Mnats data of the file $f_3$ and 3.5Mnats data of the file $f_4$ are transmitted to the cache, which provides prefetching gain for the file $f_3$, $f_4$. In time slot 3, MBS pre-transmits 16.5Mnats data of the file $f_4$ into the cache and the remaining cache capacity is 0. Eventually the transmission of all files is completed in time slot 4.

We assume the cache capacity is large enough and the optimal cache capacity is obtained in Fig. 3. The optimal instantaneous transmission rate for each time period is approximately equal, and the data departure curve $D(t, r)$ approximates a straight line which represents the minimum energy consumption can be obtained. Therefore, the data departure curve is optimal and the minimum energy consumption is obtained when the cache capacity is 39 Mnats.
In Fig. 4, we show the variation curve of the energy consumption and the cache size in the case of not using cache and using cache for pre-downloading contents. We assume the number of the files that may be requested by the users is $M=50$, the number of time slots is $N=5$ and the requested rates $d(t)$ are distributed in the range of $[0.1, 4.0]$ Mbits/s. The request probability of the files $f_k$ which is independent and identically distributed at each time slot according to a Zipf law of parameters:

$$\zeta_k = k^{-\alpha} / \sum_{n=1}^{M} n^{-\alpha}$$

where represents the skewness of the file popularity, which is to measure the frequency that the same file is requested for multiple times [10]. For the no caching algorithm (NCA), the instantaneous transmission rate $r(t)$ are equal to the user’s requested rate $d(t)$. The algorithm does not provide any gains and consumes more energy. On the other hand, for the pre-download caching algorithm (PDCA), the cache is only used to prefetch data, regardless of whether the same content will be requested in the future, that is, the algorithm only provides the prefetching gain. As the cache size increases, the performance of the PDCA algorithm is significantly better than that of the NCA algorithm.

6. Conclusion

In this paper, we investigated minimizing the total data transmission energy consumption by proactively caching the content in periods with low traffic for the cache-enabled wireless networks. We proved the minimization of energy consumption can be transformed into a convex optimization problem, which allows for a low-complexity evaluation of the system performance. Then, an optimal energy-efficient offline transmission strategy was proposed. Numerical and simulation results validated the optimal data departure curve can be obtained with a given cache size, which can minimize the total energy consumption. When the cache size is large enough, we can obtain an optimal cache size, which is a tradeoff between the minimization of energy consumption and cache size. As the cache capacity increases, energy consumption can be significantly reduced. The superiority of the proposed algorithm is obvious by comparison with the no caching algorithm.

7. References

[1] T. Wang, P.C. Li, X.B. Wang, Y.F. Wang, T.H. Guo and Y. Cao, “A comprehensive survey on mobile data offloading in heterogeneous network” 2019 Wireless Networks 25 573-584.

[2] 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” 2018 IEEE Commun. Surveys Tuts. 20 1710-1732.

[3] E. Bastug, M. Bennis and M. Debbah, “Living on the edge: The role of proactive caching in 5G wireless networks” 2014 IEEE Commun. Mag. 52 82-89.
[4] M. Maddah-Ali and U. Niesen, “Fundamental limits of caching” 2014 *IEEE Trans. Inf. Theory*. 60 2856–2867.

[5] J. Wang, “A survey of Web caching schemes for the Internet” 1999 *ACM Comput. Commun. Rev.* 29 36–46.

[6] B. Chen and C. Yang, “Energy costs for traffic offloading by cache-enabled D2D communications” 2016 *IEEE Wireless Communications and Networking Conference*, Doha, 1-6.

[7] F. Gabry, V. Bioglio and I. Land, “On Energy-Efficient Edge Caching in Heterogeneous Networks” 2016 *IEEE J. Select. Areas Commun.* 34 3288-3298.

[8] Congshan Fan, Tiankui Zhang, Zhimin Zeng, and Yue Chen, “Energy Efficiency Analysis of Cache-Enabled Cellular Networks with Limited Backhaul” 2018 *Wireless Communications and Mobile Computing*, vol. 2018, Article ID 6910876, 11 pages, 2018.

[9] M. A. Zafer and E. Modiano, “A Calculus Approach to Energy-Efficient Data Transmission With Quality-of-Service Constraints” 2009 *IEEE/ACM Transactions on Networking* 17 898-911.

[10] L. Breslau, P. Cao, L. Fan, G. Phillips, and S. Shenker, “Web caching and Zipf-like distributions: Evidence and implications” 1999 *INFOCOM* 126–134.