

RESEARCH

Open Access



A novel energy efficiency algorithm in green mobile networks with cache

Haipeng Yao^{1*}, Chao Fang¹, Chao Qiu², Chenglin Zhao² and Yunjie Liu¹

Abstract

With more devices and emerging data-intensive services, mobile data traffic grows explosively, which makes the energy efficiency issue in current and future wireless networks be a growing concern. Recently, the advantages of bringing caching scheme in wireless networks have been widely investigated. Although additional power consumption is incurred by deploying caching equipments, the traffic loading balance of radio access network (RAN) and total network delay will be improved. In this paper, we study the energy efficiency problem in future wireless network. Particularly, we give an energy-delay tradeoff algorithm by using the sleeping control and power matching scheme. In the proposed algorithm, we consider a N -based sleeping control scheme, where the base station (BS) is in the sleep state whenever there is no arriving user, and works again when N users arrive. Moreover, the proposed algorithm is suitable for multi-BS scenario; to simplify the analysis, we mainly consider the single case. Simulation results shows that the proposed algorithm can obviously reduce the network power and delay in heterogeneous network conditions. Moreover, we find that a larger cache size does not always obtain a less network cost, because more cache power is consumed as cache size increases.

Keywords: Future wireless network; Caching scheme; Sleeping control; Power matching; Energy-delay tradeoff

1 Introduction

The growing popularity of smart phones, machine to machine devices place an increasing demand for differentiated services from wireless networks. The number of wireless devices is predicted to reach 7 trillion by 2020. With more devices and emerging data-intensive services, mobile data traffic is estimated to increase by 1000-fold in 2010-2020 [1]. Moreover, by use of the wireless networks, M2M communications are rapidly developing based on the large diversity of machine type terminals, including sensors, mobile phones, consumer electronics, utility metering, vending machines, and so on. With the dramatic penetration of embedded devices, M2M communications will become a dominant communication paradigm in the communication network, which currently concentrates on machine-to-human or human-to-human information production, exchange, and processing [2]. With the explosively grows of mobile data traffic, it brings a lot of challenges in designing and deployment of future

wireless networks, including the architecture, complexity control scheme, energy assumption, etc. One of the most important challenges is the energy efficiency problem [3]. Moreover, because of the centralized architecture of current wireless networks, the bandwidth as well as the wireless link capacity of the RAN and the backhaul network cannot practically deal with the explosive growth in mobile traffic [4].

Recently, more and more researchers have proved the advantages of caching scheme in wireless networks, which can speed up content distribution and improve network resource utilization. In [5], a comprehensive overview of the recently proposed in-network caching mechanisms for information-centric networking (ICN) is proposed. In [6], an embedding cache in wireless networks is analyzed to achieve significant reduction in response time and thus provide exceptional wireless user experience. In [7], the authors present a comprehensive survey of state-of-art techniques aiming to address the challenges to ICN caching technologies, with particular focus on reducing cache redundancy and improving the availability of cached content, and also points out several interesting yet challenging research directions in the

* Correspondence: yaohaipeng@bupt.edu.cn

¹State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, People's Republic of China

Full list of author information is available at the end of the article

caching study area. In [8], the authors propose an economic model that can be used to analyze the cost saving and benefit when cache function is placed at different places of mobile network, and a real mobile network is studied according to the proposed model. In [9], the potential of forward caching in 3G cellular networks is explored, and the authors develop a caching cost model to realize the tradeoffs between deploying forward caching at different levels in the 3G network hierarchy. In [10], the authors put forward a link quality-based cache replacement technique in mobile ad hoc network. In particular, the source obtains multiple paths to the destination through multipath routing, and the acquired paths are stored in route cache. The cache replacement technique estimates the link quality using received signal strength (RSS) value. Links that possess low RSS value are removed from the route cache. In [11], the authors demonstrate the feasibility and effectiveness of using micro-caches at the base-stations of the RAN, coupled with new caching policies based on video preference of users in the cell and a new scheduling technique that allocates RAN backhaul bandwidth in coordination with requesting video clients.

Although existing excellent works have been successfully done on caching in the cellular networks, some basic questions still remain unanswered. The most important one is energy-delay tradeoff problem. In this paper, we systematically study the energy efficiency problem in future wireless network. Particularly, we give an energy-delay tradeoff algorithm by using the sleeping control and power matching scheme. In the proposed algorithm, we consider a N -based sleeping control scheme, where the BS is in the sleep state whenever there is no arriving user, and works again when N users arrive. Moreover, the proposed algorithm is suitable for multi-BS scenario; to simplify the analysis, we mainly consider the single case. We also evaluate the proposed algorithm in different network conditions and compare it with the scenario without caching scheme. Simulation results reveal that the proposed model obviously reduces the network power and delay, and makes an energy-delay tradeoff. In addition, we find that a large cache size does not always mean a less network cost because of the more cache power consumption.

The rest of the paper is organized as follows. Section 2 gives a brief description of the system model and formulates the problems. Section 3 depicts the details of the N -based sleeping control with power matching with caching scheme. Our simulation results are given in Section 4. Section 5 concludes the paper.

2 System model

In this section, we analyze the energy-delay tradeoff problem in future wireless network. The network we considered here is shown in Fig. 1. We model a situation where the wireless devices including the mobile phones

and M2M devices coexist and operate in a local region with the BSs. The wireless devices satisfy uniform distribution in this region. We also assume each BS has a limited cache capacity C . The cache capability of BSs here is different from the traditional BSs which are not equipped with cache. In the proposed scenario, the BSs are responsible for managing the whole wireless network, and the network related functionalities are implemented in the BSs, including connectivity establishment, access control, and QoS management. In particular, the BSs provide the connectivity to the wireless devices, and the connection between the wireless network and other network, and the devices within the wireless network could directly communicate with the BS based on the established connectivity to upload and download information.

In the proposed scenario, the wireless terminals can send requests to obtain their interested contents, and the BS can satisfy parts of the requests by using the contents buffered in the BS cache. Moreover, we assume that the packet error has been guaranteed by the physical layer technique, and the retransmission is not considered [12].

2.1 Content popularity model

We assume that the number of different contents provided by a content provider is F , ranked from 1 (the most popular) to F (the least popular) based on its popularity [13, 14]. By letting the total number of requests in a given time duration t be R , the number of requests to the content of popularity k , R_k , follows a Zipf-like distribution as

$$R_k = R \frac{k^{-\alpha}}{\sum_{k=1}^F k^{-\alpha}} = \frac{R}{a_0} k^{-\alpha}, k = 1, 2, \dots, F \quad (1)$$

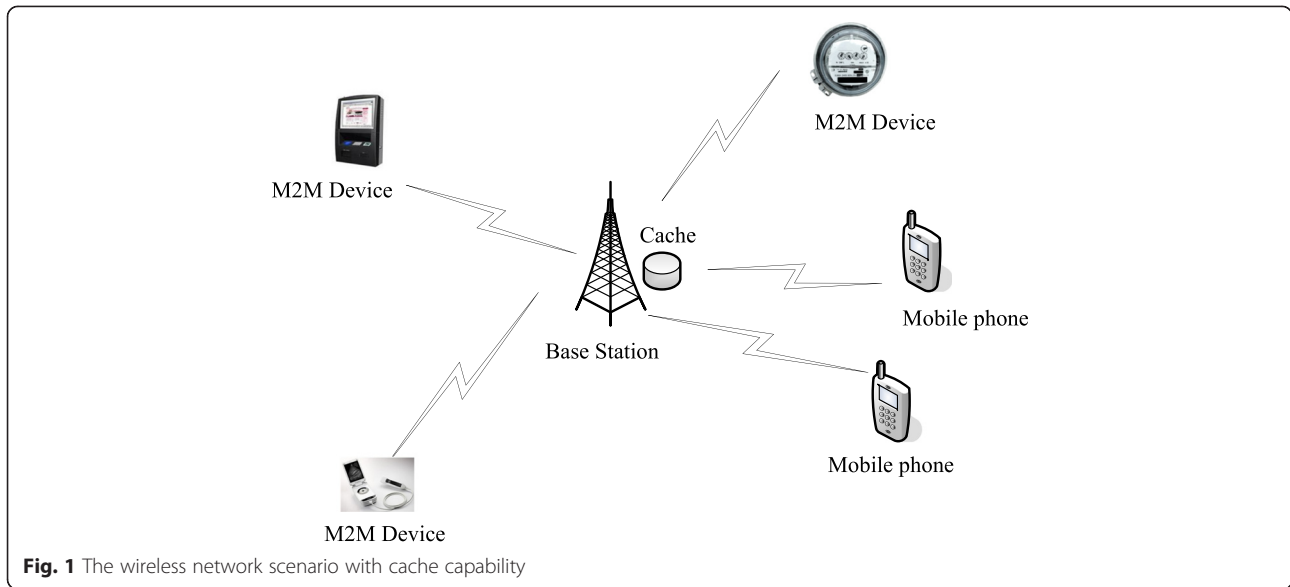
where a_0 is a constant to normalize the request rate [11, 12]. A large value for α indicates that the number of popular contents increases.

2.2 Wireless Devices Distribution Model

In this paper, we assume that the wireless devices satisfy uniform distribution in the proposed region and the number of wireless devices is distributed according to a homogeneous spatial Poisson process with density λ . Thus, the probability that there exist N devices in a region of measure S is given by

$$\Pr(k) = \frac{e^{-\lambda S} (\lambda S)^N}{k!} \quad (2)$$

To evaluate the performance of the proposed future wireless network, we assume that the BSs hold the cognitive capability to detect the existence of wireless devices. Thus, the main metric is minimizing either the



miss probability for a target false alarm probability or the false alarm probability for a target miss probability [2]. Therefore, for a targeted detection probability $p_{d,i}^T$, the probability of the false alarm for the device i is written as

$$P_{f,i}(\tau_i) = Q\left(Q^{-1}(p_{f,i}^T)\sqrt{2\gamma_i|h_i|^2 + 1} + \gamma_i|h_i|^2\sqrt{\tau_i f_s}\right), \tag{3}$$

where τ_i , γ_i , h_i , and f_s are the sensing duration, average received signal to noise (SNR), target channel gain between the primary transmitter, and sampling frequency, respectively. However, for a targeted false alarm probability $p_{f,i}^T$, the probability of the detection is given by

$$P_{d,i}(\tau_i) = Q\left(\frac{Q^{-1}(p_{f,i}^T) - \gamma_i|h_i|^2\sqrt{\tau_i f_s}}{\sqrt{2\gamma_i|h_i|^2 + 1}}\right). \tag{4}$$

2.3 Power model

In the proposed scenario, we assume that the wireless devices arrive according to a Poisson process with arrival rate r . A random amount of downlink service with average length l bits is required by each device, e.g., non-real-time file download with average file size l , and then the user leaves after being served [3, 12]. Assuming the arrival rate can be well estimated [12, 15], with time varying traffic intensity in practice, we just need to operate according to the current arrival rate. In the following, we use an energy-proportional model [16] to model

the power consumption in the wireless network with caching scheme.

2.3.1 Cache power

We assume the cache in the future wireless network has active mode and sleep mode, just like the traditional BS. Thus, cache scheme may be modeled as a binary hypothesis problem:

- H_0 : The cache is active.
- H_1 : The cache is sleep.

Therefore, the total power consumed by the cache is given by [12–14]

$$P_{ca} = \begin{cases} P_0^{ca} + N_c l \omega_{ca}, & H_0 \\ P_{sl}^{ca}, & H_1 \end{cases} \tag{5}$$

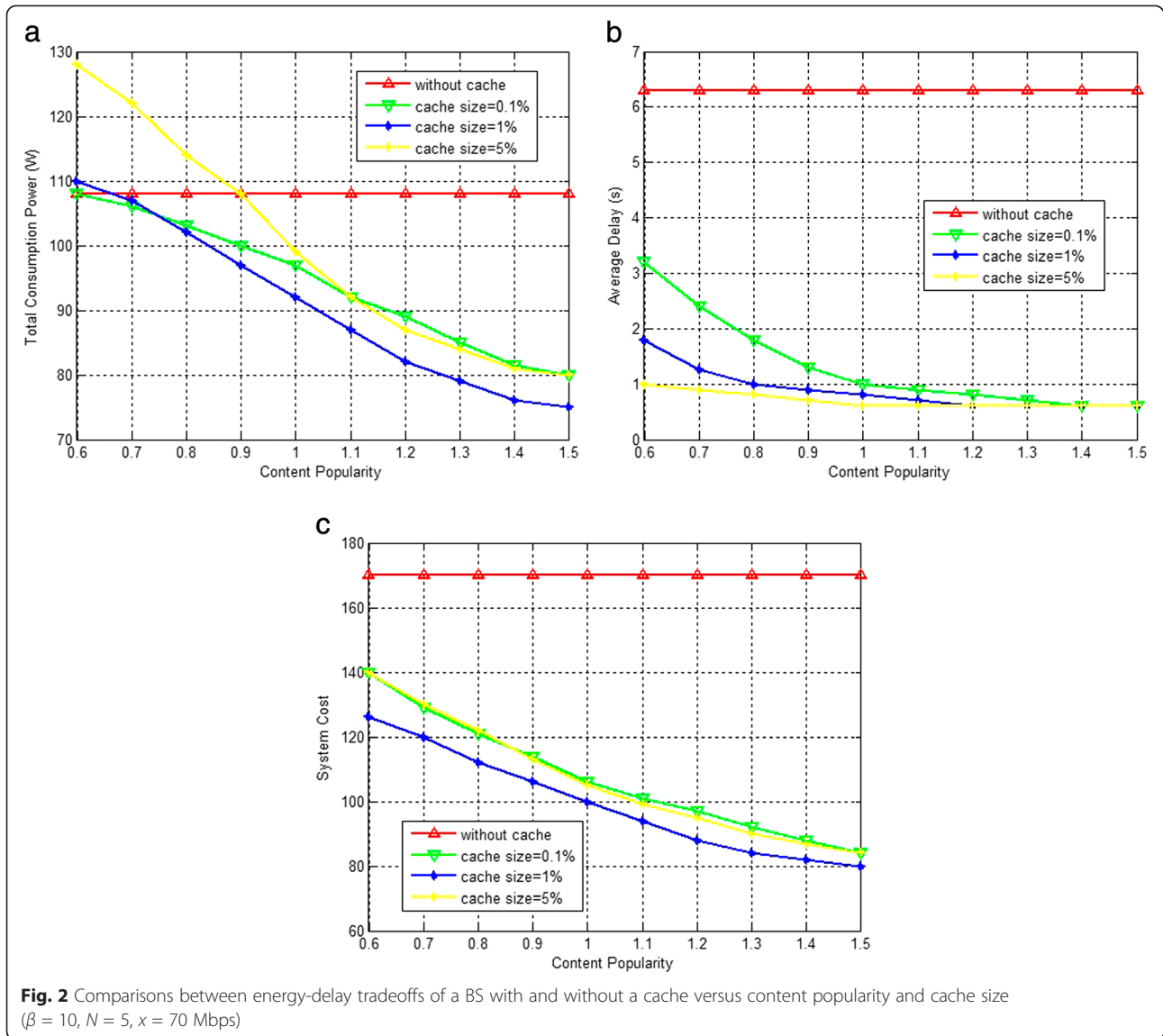
where P_0^{ca} and P_{sl}^{ca} are the static power consumption of cache in H_0 and H_1 , respectively, $N_c \in [0, C]$ represents the number of contents cached in the BSs, and ω_{ca} is the power efficiency of caching. We assumed that there is a fixed cache switching energy cost E_s^{ca} for each mode transition.

2.3.2 Base station power

The $M/G/1$ processor-sharing (PS) model is used here [17, 18]. We also assume the each BS have active mode and sleep mode, with the traditional BS power consumption P_{BS} as follows:

$$P_{BS} = \begin{cases} P_0^{BS} + \Delta p P_t, & H_0 \\ P_{sl}^{BS}, & H_1 \end{cases} \tag{6}$$

where P_0^{BS} and P_{sl}^{BS} are the static power consumption of each BS in H_0 and H_1 , respectively, ΔP is the slope of



the load-dependent power consumption, and the transmit power P_t adapts to the system traffic load [12].

Assume that the BS service capacity or service rate is q bits per second, which adapts to the system traffic load and is equally shared by all users being served. So the user departure rate is $\mu = q/l$, and the traffic load (or busy probability) at the BS is $\rho = \lambda/\mu = \lambda l/q$ [16]. The relationship between the service rate q and the transmit power P_t is

$$P_t = \frac{\rho}{\gamma} \left(2^{\frac{\rho}{B}} - 1 \right), P_t \in [0, P_t^{\max}] \quad (7)$$

where γ is the received instantaneous signal-to-noise ratio (SNR), B is the bandwidth, N_0 is the additive white Gaussian noise (AWGN), H is the channel gain, P_p

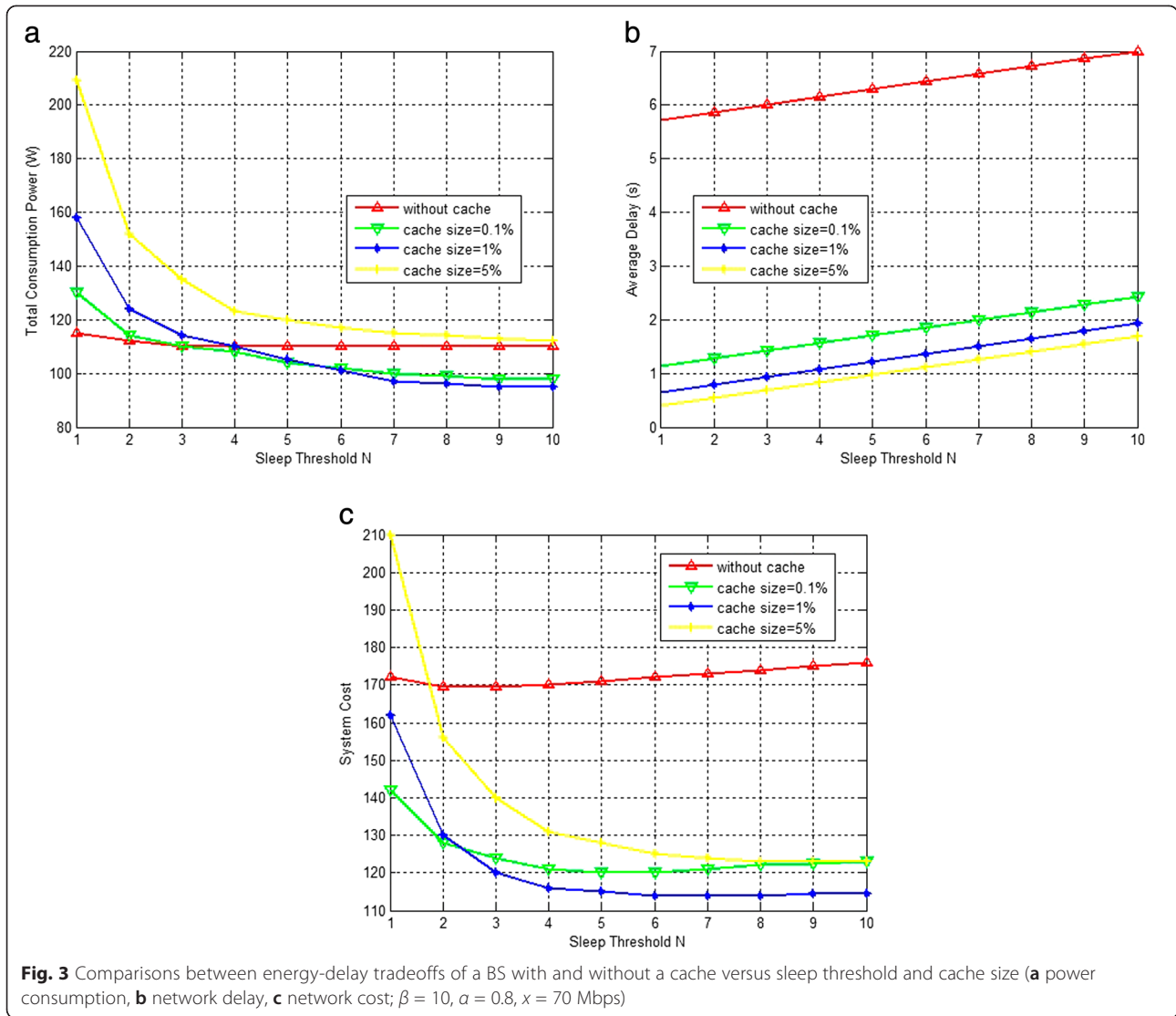
denotes the transmitted power of wireless device, and γ is defined as follows:

$$\gamma = \frac{P_p}{N_0} H \quad (8)$$

Therefore, with service rate q , the traditional BS power expression (4) [12] can be rewritten as

$$P_{BS} = \begin{cases} P_0^{BS} + \frac{\Delta p \lambda l}{\gamma q} \left(\frac{q}{2B} - 1 \right), & H0 \\ P_{sl}^{BS}, & H1 \end{cases} \quad (9)$$

However, with a cache, all the user requests to the BS will be partly satisfied by it. According to [3], the widely used caching decision policies in ICN include



Leave Copy Everywhere (LCE), FIX(P), Random Caching (RND), and Unique Caching (UniCache) [19], and the most common cache replacement policy is the popular Least Recently Used (LRU) [20]. Unlike these above common cache policies, in this paper, we use the assumption adopted in [3], that is, the contents with content popularity of the top N_c are buffered in the cache. Then, the number of requests unsatisfied by the cache is

$$\rho' = \frac{\rho}{a_0} \sum_{k=N_c+1}^F k^{-\alpha} = \frac{\lambda l}{a_0 q} \sum_{k=N_c+1}^F k^{-\alpha} \quad (10)$$

Therefore, based on (5), the total BS power expression with cache can be written as

$$P'_{BS} = \begin{cases} P_0^{BS} + \frac{\Delta p}{\gamma} \frac{\lambda l}{a_0 q} \left(\frac{q}{2B-1} \right) \sum_{k=N_c+1}^F k^{-\alpha}, & H0 \\ P_{sl}^{BS}, & H1 \end{cases} \quad (11)$$

2.4 Delay model

According to the property of M/G/1 PS queue [21] and the Little's Law [22], the average delay of a BS without a cache [12] is

$$D_{BS} = \frac{l}{q - \lambda l} = \frac{\rho}{\lambda(1-\rho)} \quad (12)$$

Similarly, the average delay of a BS with a cache [3] is

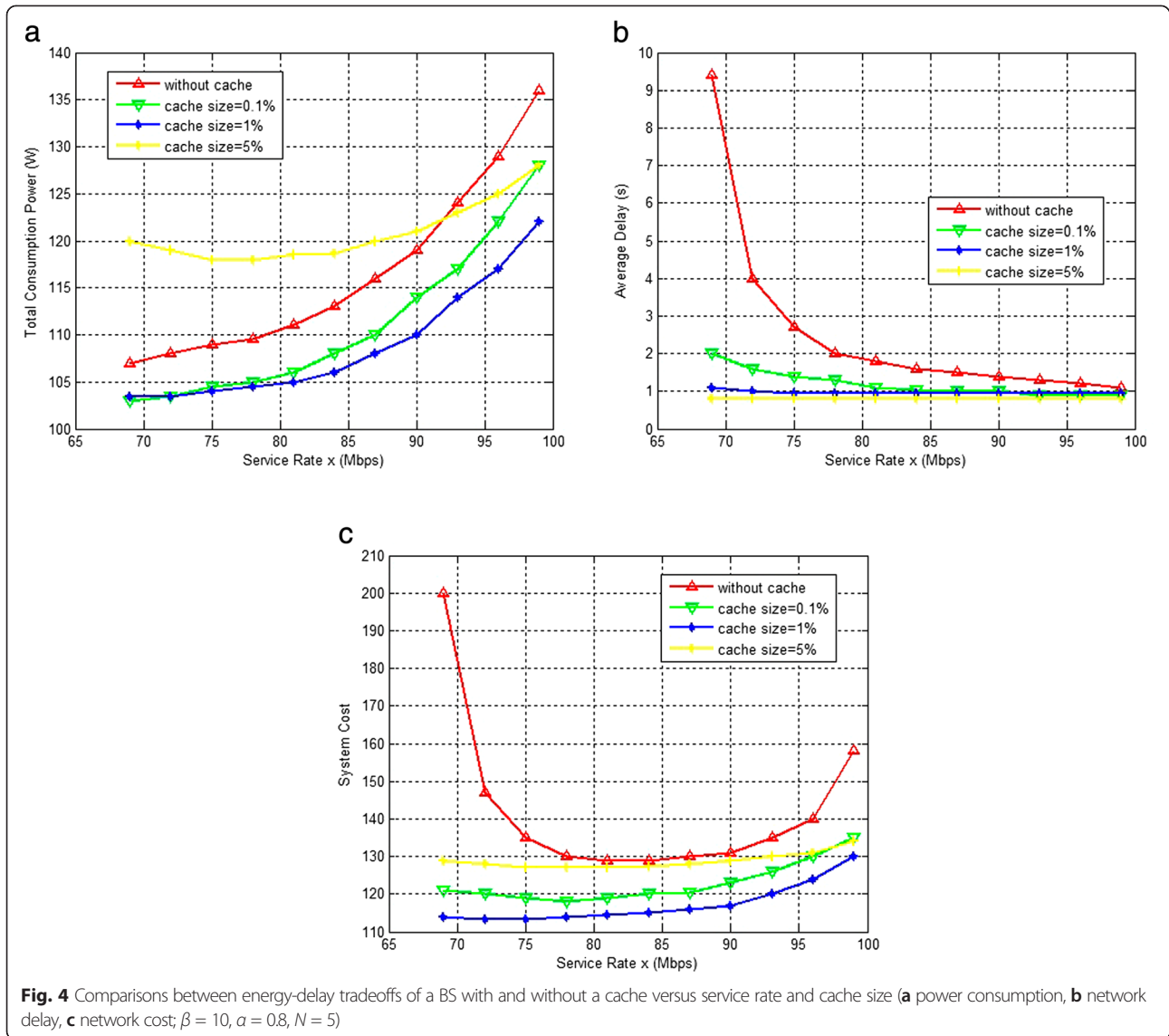


Fig. 4 Comparisons between energy-delay tradeoffs of a BS with and without a cache versus service rate and cache size (a power consumption, b network delay, c network cost; $\beta = 10$, $\alpha = 0.8$, $N = 5$)

$$D'_{BS} = \frac{\rho'}{\lambda(1-\rho')} = \frac{l \sum_{k=N_c+1}^F k^{-\alpha}}{a_0 q - \lambda l \sum_{k=N_c+1}^F k^{-\alpha}} \quad (13)$$

2.5 System cost

The objective is to minimize the system cost, which is a weighted combination of average total power $P^{\text{tot}}_{T_{\text{sleep}}x}$ and average delay $D_{T_{\text{sleep}}x}$ [12]. Here T_{sleep} represents the parameter in the BS sleeping control: For the N -based sleeping control, $T_{\text{sleep}} = N$; $T_{\text{sleep}} = 0$ denotes that there is no sleeping control [12]. Therefore, we seek to find the service rate q and sleeping control parameter T_{sleep} that minimize

$$z(T_{\text{sleep}}, q) = P^{\text{tot}}_{T_{\text{sleep}}, q} + \beta D_{T_{\text{sleep}}, q} \quad (14)$$

The positive weighting factor β indicates the relative importance of the average delay over the average power which can be thought of as a Lagrange multiplier on an average delay constraint [12, 23, 24]. The “delay” we consider in this paper is the average response time from the user’s service request arriving at the BS until this request is finished [12]. From the Little’s Law, we know that the mean delay is directly related to the average queue length.

3 Sleeping control with power matching

We assume the BS is in sleep mode when there is no service request and works again when N user requests

arrive. Based on an extended Markov chain with departure rate $\mu = q/l$, we define an extended state space $\{(i, j); i = 0, j = 0, 1, \dots, N - 1; i = 1, j = 1, 2, \dots\}$. If $i = 0$, it means that the BS is in sleep mode, and j is the number of users in the network; if $i = 1$, the BS is in active mode, and j represents the number of users in the network. Therefore, the state transition probability [12] can be written as

$$P_r(i, j) = \begin{cases} \frac{q-\lambda}{Nx}, & \text{if } i = 0 \\ \frac{\lambda}{Nq} \left(1 - \left(\frac{\lambda l}{q}\right)^j\right), & \text{if } i = 1, 1 \leq j \leq N \\ \frac{\lambda}{Nq} \left(\left(\frac{\lambda l}{q}\right)^{j-N} - \left(\frac{\lambda l}{q}\right)^j\right), & \text{if } i = 1, j > N \end{cases} \quad (15)$$

After some calculation, we obtain the probability in sleep mode [12] as

$$\sum_{j=0}^{N-1} P_r(0, j) = 1 - \frac{\lambda l}{q}, \quad (16)$$

and the average queue length [12] is

$$\sum_{j=0}^{N-1} j P_r(0, j) + \sum_{j=0}^{\infty} j P_r(1, j) i = \frac{\lambda l}{q-\lambda l} + \frac{N-1}{2} \quad (17)$$

By introducing the sleeping control, we have to consider the energy cost for mode transitions. The system goes to sleep when no user request arrives and wakes again until N users arrive with the average assembling

time to be N/λ . At the beginning of each active period, there are N users in the system, and thus the average working time is $N/(\mu - \lambda)$. Therefore, the mode transition frequency F_m which is defined as the mode transition times between active mode and sleep mode per unit time [12], is

$$F_m = \frac{2}{\frac{N}{\lambda} + \frac{N}{\mu-\lambda}} = \frac{2\lambda}{N} \left(1 - \frac{\lambda}{\mu}\right) \quad (18)$$

where the mode switching energy cost per unit time is $E_s^{BS} F_m$.

Then the total power consumption $P_{N,q}^{tot}$ under the N -based sleeping control with service rate q [12] is

$$\begin{aligned} P_{N,q}^{tot} &= \frac{\lambda l}{q} \left[P_0^{BS} + \frac{\Delta P}{\gamma} \left(\frac{q}{2B} - 1\right) \right] + \left(1 - \frac{\lambda l}{q}\right) \left[P_{sl}^{BS} + \frac{2\lambda E_s^{BS}}{N} \right] \\ &= \rho \left[P_0^{BS} + \frac{\Delta P}{\gamma} \left(\frac{q}{2B} - 1\right) \right] + (1-\rho) \left[P_{sl}^{BS} + \frac{2\lambda E_s^{BS}}{N} \right] \end{aligned} \quad (19)$$

Using the Little's Law, the average delay $D_{N,q}$ [12] is given by

$$D_{N,q} = \frac{l}{q-\lambda l} + \frac{N-1}{2\lambda} \quad (20)$$

However, with a cache, all the user requests to the BS will be partly satisfied by it. Therefore, the total BS power $P_{N,q}^{tot}$ with a cache can be written as

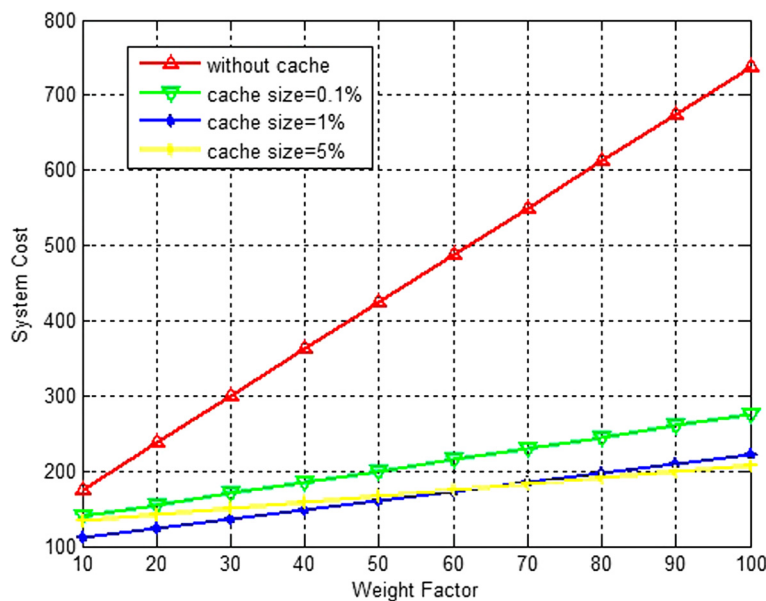


Fig. 5 Comparisons between total cost of a BS with and without a cache versus weight factor ($\alpha = 0.8, N = 5, x = 70$ Mbps)

$$\begin{aligned}
P_{N,q}^{\text{tot}'} &= \rho' \left[P_0^{ca} + N_c l w_{ca} + P_0^{\text{BS}} + \frac{\lambda \Delta P}{a_0 q \gamma} \left(\frac{q}{2B} - 1 \right) \sum_{k=N_c+1}^F k^{-\alpha} \right] \\
&+ (1-\rho') \left[P_{\text{sleep}}^{ca} + P_{\text{sleep}}^{\text{BS}} + \frac{2\lambda}{N} (E_s^{ca} + E_s^{\text{BS}}) \right] \\
&= \left(\frac{\lambda}{a_0 q} \sum_{k=N_c+1}^F k^{-\alpha} \right) [P_0^{ca} + N_c l w_{ca} + P_0^{\text{BS}} \\
&+ \frac{\lambda \Delta P}{a_0 q \gamma} \left(\frac{q}{2B} - 1 \right) \sum_{k=N_c+1}^F k^{-\alpha}] + \left(1 - \frac{\lambda}{a_0 q} \sum_{k=N_c+1}^F k^{-\alpha} \right) \\
&\quad \left[P_{\text{sl}}^{ca} + P_{\text{sl}}^{\text{BS}} + \frac{2\lambda}{N} (E_s^{ca} + E_s^{\text{BS}}) \right]
\end{aligned} \tag{21}$$

Similarly, the average delay $D'_{N,q}$ of a BS with a cache is

$$D'_{N,q} = \frac{l \sum_{k=N_c+1}^F k^{-\alpha}}{a_0 q - \lambda \sum_{k=N_c+1}^F k^{-\alpha}} + \frac{N-1}{2\lambda} \tag{22}$$

Based on (21) and (22), the system cost can be written as

$$\begin{aligned}
z &= P_{N,q}^{\text{tot}'} + \beta D'_{N,q} = \left(\frac{\lambda}{a_0 q} \sum_{k=N_c+1}^F k^{-\alpha} \right) [P_0^{ca} + N_c l w_{ca} \\
&+ P_0^{\text{BS}} + \frac{\lambda \Delta P}{a_0 q \gamma} \left(\frac{q}{2B} - 1 \right) \sum_{k=N_c+1}^F k^{-\alpha}] \\
&+ \left(1 - \frac{\lambda}{a_0 q} \sum_{k=N_c+1}^F k^{-\alpha} \right) \\
&\quad \left[P_{\text{sl}}^{ca} + P_{\text{sl}}^{\text{BS}} + \frac{2\lambda}{N} (E_s^{ca} + E_s^{\text{BS}}) \right] \\
&+ \beta \frac{l \sum_{k=N_c+1}^F k^{-\alpha}}{a_0 q - \lambda \sum_{k=N_c+1}^F k^{-\alpha}}
\end{aligned} \tag{23}$$

4 Simulation results

In this section, we use computer simulations to evaluate the performance of the proposed energy-delay tradeoff model. We first describe the simulation settings and then compare it with the traditional network cost model, which does not have a cache.

4.1 Simulation settings

The simulation is assumed to be carried out in a single urban micro-cell scenario. According to the ITU test environments [25] and the related introduction in [3], the system bandwidth $B = 10$ MHz, the maximum transmit power $P_t^{\text{max}} = 15$ W, and the channel gain $\gamma = 10$ (dB) in the simulation. We take the micro BS energy consumption parameters $P_0 = 100$ W and $\Delta P = 7$ [26]. Moreover,

users arrive according to a Poisson process with arrival rate $\lambda = 4$, and each user requests exactly one file whose size is exponentially distributed with mean $l = 2$ MB from the BS and leaves the system after being served [3, 12]. The system operates in a time-slotted fashion, and the BS schedules users in a round robin way, serving one user in each time slot. Moreover, it is assumed that the average user arrival rate x is 70 Mbps.

In the simulation, it is also assumed that there are 10,000 different contents in the system, where values of skewness factor α range from 0.6 [27] to 1.5 [13]. We abstract the cache size for a BS as a proportion that the cache size is defined as the relative size to the total amount of different contents in the network, which varies from 0.1 to 5 % [28].

4.2 Performance evaluation results

Figure 2 shows the energy-delay tradeoffs of the two models versus content popularity with different cache sizes when weight factor $\beta = 10$, sleep threshold $N = 5$, and service rate $x = 70$ Mbps. From Fig. 2, we can observe that the content popularity and cache size have some effect on the energy-delay tradeoffs of the proposed model. When the Zipf skewness parameter α increases, the content diversity decreases, and the number of popular contents increases, which makes the network power consumption, delay, and cost decrease gradually. But these parameters of the traditional model remain the same, because it does not consider the content popularity or have a cache capacity. As for cache capacity, a larger buffer means that more popular contents can be cached, which reduces the network delay, as shown in Fig. 2b. However, a larger cache size does not always mean a better energy-delay tradeoff, as shown in Fig. 2a, c. The reason is that the larger cache size is, the more cache power is consumed, which can influence the energy-delay tradeoff.

Figure 3 shows the energy-delay tradeoffs of the two models versus sleep threshold with different cache sizes when weight factor $\beta = 10$, content popularity $\alpha = 0.8$, and service rate $x = 70$ Mbps. From Fig. 3a, we can observe that the sleep threshold and cache size have some effect on the energy-delay tradeoffs of the proposed model. When the value of sleep threshold increases, the BS and cache will have a longer sleeping time, which makes the network power consumption decrease more obviously than the traditional model. However, with a larger sleep threshold, network delay of both models will increase because of the longer waiting time in the queue, as shown in Fig. 3b. Due to the smaller network delay, the proposed model has a smaller system cost, as shown in Fig. 3c. Similarly, a larger cache size does not always mean a better energy-delay tradeoff, as shown in Fig. 3a, c, which has been analyzed above.

Figure 4 shows the energy-delay tradeoffs of the two models versus service rate with different cache sizes when weight factor $\beta = 10$, content popularity $\alpha = 0.8$, and sleep threshold $N = 5$. From Fig. 4, we can observe that the service rate and cache size have some effect on the energy-delay tradeoffs of the proposed model. When service rate x increases, the BS service capacity is improved, which makes the network power consumption grow and network delay decrease gradually. However, a larger cache size does not always mean a lower power consumption, as shown in Fig. 4a. The reason is that the larger cache size is, the more cache power is consumed, which can influence the energy-delay tradeoff. As for network cost, due to the smaller network delay, the network cost of the proposed model is lower than that of the traditional model. Especially when the service rate is small, the gap between them is more obvious.

Figure 5 shows the energy-delay tradeoffs of the two models versus weight factor β with different cache sizes when content popularity $\alpha = 0.8$, sleep threshold $N = 5$, and service rate $x = 70$ Mbps. From Fig. 5, we can observe that the weight factor and cache size have some effect on the energy-delay tradeoffs of the two models. The increase of weight factor indicates that the average delay is more important than the network power. That is to say, network cost is more sensitive to average delay. However, a larger cache size does not always mean a better energy-delay tradeoff, which has been analyzed above.

5 Conclusions

In this paper, we have studied the sleeping control and power matching problem for energy-delay tradeoffs in the context of single BS, which has a cache capacity to buffer the contents through it. In the proposed model, an N -based sleeping control scheme is considered. Simulation results reveal that, by introducing the cache, the network power and delay can be obviously reduced in different network conditions compared to the scenario without a cache. In addition, we find that a large cache size does not always mean a less network cost because of the more cache power consumption.

Competing interests

The authors declare that they have no competing interests.

Acknowledgements

This work was supported by NSF (61471056), China Jiangsu Future Internet Research Fund (BY2013095-3-1), and BUPT Youth Research and Innovation Plan (2014RC0103).

Author details

¹State Key Laboratory of Networking and Switching Technology, Beijing University of Posts and Telecommunications, Beijing, People's Republic of China. ²Key Laboratory of Universal Wireless Communications, Ministry of Education, Beijing University of Posts and Telecommunications, Beijing, People's Republic of China.

References

- Nokia Siemens Networks, 2020: Beyond 4G Radio Evolution for the Gigabit Experience. White Paper, Feb. 2011
- H Yao, T Huang, C Zhao, X Kang, Z Liu, Optimal power allocation in cognitive radio based machine-to-machine network. *Eurasip Journal of Wireless Communication*,(1), 1-9 (2014)
- C Fang, H Yao, C Zhao, Y Liu, Modeling energy-delay tradeoffs in single base station with cache. Accepted by Int J DistribSens N(2015, 501: 401465)
- X Zhang, M Chen, T Taleb, A Ksentini, V Leung, Cache in the air: exploiting content caching and delivery techniques for 5G systems. *IEEE Commun Mag* **52**(2), 131-139 (2014)
- Meng Zhang, Hongbin Luo, Hongke Zhang, A Survey of Caching Mechanisms in Information-Centric Networking. *IEEE Communications Surveys & Tutorials*, PP(99),1-28, 2015
- H Sarkissian. The business case for caching in 4G LTE networks, (LSI Tech Rep 2012), http://www.wireless2020.com/docs/LSI_WP_Content_Cach_Cv3.pdf
- G Zhang, Y Li, T Bin, Caching in information centric networking: A survey. *Computer Network* **57**(16), 3128-3141 (2013)
- Xuejun Cai, Shunliang Zhang, Yunfei Zhang, Economic analysis of cache location in mobile network(Proc. 2013 IEEE Wireless Communications and Networking Conference (WCNC), Shanghai, China, 2013)
- J Erman, A Gerber, M Hajiaghayi, D Pei, S Sen, O Spatscheck, To cache or not to cache: the 3G case. *Internet Comput* **15**(2), 27-34 (2011)
- J Dhanapal, SK Srivatsa, Link quality-based cache replacement technique in mobile ad hoc network. *IET Information Security* **7**(4), 277-282 (2013)
- H Ahlehagh, S Dey, *Video caching in radio access network: impact on delay and capacity* (Proc. IEEE Wireless Communications and Networking Conference (WCNC), Shanghai, 2012)
- J Wu, S Zhou, Z Niu, Traffic-aware base station sleeping control and power matching for energy-delay tradeoffs in green cellular networks. *IEEE Trans. Wireless Communication* **12**(8), 4196-4209 (2013)
- N Choi, K Guan, D Kilper, G Atkinson, *In-network caching effect on optimal energy consumption in content-centric networking* (Proc. IEEE ICC'12, Ottawa, Canada, 2012)
- K Guan, G Atkinson, D Kilper, E Gulsen, *On the Energy Efficiency of Content Delivery Architectures* (Proc. IEEE ICC'11 Workshops, Kyoto, Japan, 2011)
- Y Shu, M Yu, J Liu, OWW Yang, *Wireless traffic modeling and prediction using seasonal ARIMA models* (Proc. IEEE ICC'03, Alaska, USA, 2003)
- LA Barroso, U Holzle, The case for energy-proportional computing. *IEEE Computer* **40**(12), 33-37 (2007)
- S Borst, *User-level performance of channel-aware scheduling algorithms in wireless data networks* (Proc. IEEE INFOCOM'03, San Francisco, CA, 2003)
- IE Telatar, RG Gallager, Combining queueing theory with information theory for multi-access. *IEEE J Sel Areas Commun* **13**(6), 963-969 (1995)
- Aloannou and S Weber. A taxonomy of caching approaches in information-centric network architectures. Elsevier. J. 2013, <https://www.cs.tcd.ie/publications/tech-reports/reports.15/TCD-CS-2015-02.pdf>
- PR Jelenković, X Kang, Characterizing the miss sequence of the lru cache. *SIGMETRICS Perform Eval Rev* **36**(2), 119-121 (2008)
- J Walrand. An Introduction to Queueing Networks. (Prentice Hall, 1998), http://www.researchgate.net/publication/215562479_An_introduction_to_queueing_networks__Jean_Walrand
- JDC Little, SC Graves, Little's law. *Springer US* **115**, 81-100 (2008)
- R Berry, R Gallager, Communication over fading channels with delay constraints. *IEEE Trans Inf Theory* **48**(5), 1135-1149 (2002)
- S Boyd, L Vandenberghe, *Convex Optimization* (Cambridge Univ. Press, Cambridge, U. K., 2004)
- Ericsson, Radio characteristics of the itu test environments and deployment scenarios r1-091320 (3GPP TSG-RAN1#56bis, Seoul, Korea, 2009)
- MA Imran, E Katranaras, G Auer, O Blume, V Giannini, I Godor, et al., D2.3: energy efficiency analysis of the reference systems, areas of improvements and target breakdown, (ICT-EARTH Tech Rep INFOS-ICT-247733), Nov. 2010. [Online]. Available: <http://cordis.europa.eu/docs/projects/cnect/3/247733/080/deliverables/001-EARTHWP2D23v2.pdf>
- L Breslau, *Web caching and zipf-like distributions: evidence and implications* (Proc. IEEE INFOCOM'99, New York, NY, USA, 1999)
- D. Rossi and G. Rossini, Caching performance of content centric networks under multi-path routing (and more), (Telecom ParisTech Tech Rep, 2011)