Energy Saving through a Learning Framework in Greener Cellular Radio Access Networks

Rongpeng Li*†, Zhifeng Zhao*†, Xianfu Chen*†‡, and Honggang Zhang*†
*York-Zhejiang Lab for Cognitive Radio and Green Communications
†Dept. of Information Science and Electronic Engineering
Zhejiang University, Zheda Road 38, Hangzhou 310027, China
Email: {lirongpeng, zhaozf, chenxianfu, honggangzhang}@zju.edu.cn
‡VTT Technical Research Centre of Finland, P.O. Box 1100, FI-90571 Oulu, Finland

***Abstract*—Recent works have validated the possibility of energy efficiency improvement in radio access networks (RAN), depending on dynamically turn on/off some base stations (BSs). In this paper, we extend the research over BS switching operation, matching up with traffic load variations. However, instead of depending on the predicted traffic loads, which is still quite challenging to precisely forecast, we formulate the traffic variation as a Markov decision process (MDP). Afterwards, in order to foresightedly minimize the energy consumption of RAN, we adopt the actor-critic method and design a reinforcement learning framework based BS switching operation scheme. In the end, we evaluate our proposed scheme by extensive simulations under various practical configurations and prove the feasibility of significant energy efficiency improvement.**

I. Introduction

The explosive popularity of smartphones and tablets has ignited a surging traffic load demand for radio access and has been incurring massive energy consumption and huge greenhouse gas (GHG) emission [1]. Specifically speaking, the information and communication technologies (ICT) industry accounts for 2% to 10% of the world's overall power consumption [2] and has emerged as one of the major contributors to the world-wide CO_2 emission. Besides that, there are also economical benefits for cellular network operators to reduce the power consumption of their networks. It's envisioned that the power bill will doubly enlarge in fives years for China Mobile [3]. Meanwhile, the energy expenditure accounts for a significant proportion of the overall cost. Therefore, it's quite essential to improve the energy efficiency of ICT industry.

Currently, over 80% of the power consumption takes place in the radio access networks (RAN), especially the base stations (BSs) [4]. The reason behind this is largely due to that the present BS deployment is on the basis of peak traffic loads and generally stays active irrespective of the traffic load [5] while the traffic loads virtually vary heavily [6]. Recently, there has been a substantial body of work towards traffic load-aware BSs adaptation and the authors have validated the possibility of energy efficiency improvement from different perspectives. Luca Chiaraviglio et al. [7] showed the possibility of energy saving by simulations. [8] and [9] proposed how to dynamically adjust the working status of BS, depending on the predicted traffic loads. However, to reliably predict the traffic loads is still quite challenging, which makes these works suffering. On the other hand, [10] and [11] presented dynamic BS switching algorithms with the traffic loads a prior and preliminarily proved the effectiveness of energy saving.

Besides, it is also found that turning on/off some of the BSs will immediately affect the BS, with which a mobile terminal (MT) should be associated. Moreover, subsequent user's association choice in turn leads to the traffic load differences of BSs. Hence, any two consecutive BS switching operations are correlated with each other and current BS switching operation will also further influence the overall energy consumption in the long run. In other words, the expected energy saving scheme must be *foresighted* while minimizing the energy consumption. It should concern its effect on both the current and future system performance to deliver a visionary BS switching operation solution.

[5] presented a partially foresighted energy saving scheme which combines BS switching operation and user association by giving a heuristic solution on the basis of a stationary traffic load profile. In this paper, we try to solve these problem from a different perspective. In a nutshell, we apply Markov decision process (MDP) to model the traffic load variation. Afterwards, we can attain a solution to the formulated MDP model, i.e., BS switching operation (and corresponding user association as well) policy, by taking advantage of actor-critic method, a reinforcement learning approach [12] without a prior knowledge about the traffic loads within the BSs. Within the reinforcement learning framework, a BS switching operation controller[1] firstly estimates the traffic loads variation based on the on-line experience. Consequently, the controller can select one of the possible BS switching operations under the estimated circumstance and then decreases or increases the probability of the same action to be selected lately based on the needed cost. Here, the cost refers to the energy consumption due to such a BS switching operation. After repeating the actions and getting the corresponding cost, the controller would know how to choose the active BSs under one specific traffic load circumstance. Moreover, with the MDP model the

[1]In practice, such a centralized BS switching operation can be conducted by the base station controller (BSC) in second generation (2G) cellular networks or the radio network controller (RNC) in third generation (3G) or long term evolution (LTE) cellular networks. In this paper, we generalize it as a BS switching operation controller.

978-1-4673-0921-9/12/$31.00 ©2012 IEEE

resulting BS switching strategy is foresighted, which would improve energy efficiency in the long run. To the best of our knowledge, our work is the first attempt to apply reinforcement learning framework to energy saving scheme in RANs.

The remainder of the paper is organized as follows. In Section II, we introduce the system model and formulate the traffic variation as an MDP. In Section III, we talk about energy saving scheme by the proposed learning framework. Section IV evaluates the proposed schemes and presents the simulation results. Section V concludes this paper with a summary.

II. System model and problem formulation

A. System model

An RAN usually consists of multiple BSs while the traffic loads of BSs are usually fluctuating, thus often making BSs under-utilization. In this paper, we assume there exists a BS switching operation controller which can timely know the traffic loads in these BSs at current stage and determine the working status of any BS (i.e., active/sleeping mode) at next stage. Besides, our focus is on downlink communication, i.e., from BSs to MTs. We assume there exists a region $\mathcal{L} \in \mathbb{R}^2$ served by a set of BSs $\mathcal{B} = \{1, \ldots, N\}$ as Fig. 1 depicts, where the coverage of these BSs is overlapped. Moreover, we can turn on/off some BSs in a centralized way by the BS switching operation controller. Beyond that, we assume that file transfer requests at a location $x \in \mathcal{L}$ arrive following a Poisson point process with arrival rate per unit area $\lambda(x)$ and file size $\frac{1}{\mu(x)}$. After that, we define *traffic load density* as $\gamma(x) = \lambda(x)/\mu(x) < \infty$ [5]. By the way, the traffic load density also captures spatial traffic variations. For example, a hot spot can be characterized by a high arrival rate and/or possibly large file sizes. Hence, when the set of BSs $\mathcal{B}_{on}$ is turned on, the traffic loads severed by BS $i \in \mathcal{B}_{on}$ can be represented as $\Gamma_i = \int_{\mathcal{L}} \gamma(x) I_i(x, \mathcal{B}_{on})\, \mathrm{d}x$, whereas $I_i(x, \mathcal{B}_{on}) = 1$ is a user association indicator and denotes location x is served by BS $i \in \mathcal{B}_{on}$ and vice versa. We define the traffic load for a sleeping BS i as zero, namely $\Gamma_i = 0,$ if $i \in \mathcal{B} \setminus \mathcal{B}_{on}$. In this paper, we use finite state Markov process (FSMC) to demonstrate the traffic load variation condition, i.e., $p(\Gamma_i^{k+1} | \Gamma_i^k)$. Moreover, the traffic load Γ_i for BS i is partitioned into two parts by a boundary point Γ_b. Here, Γ_b can be the average traffic loads in one BS over a certain period, thus feasible to be known in advance based on the historical records. Therefore, the traffic loads for a specific BS have merely two states, i.e., $s_i = 0$ if $\Gamma_i < \Gamma_b$ and $s_i = 1$ if $\Gamma_i > \Gamma_b$. Subsequently, we construct a state vector $\mathbf{s} = \{s_1, \cdots, s_N\} \in \mathbb{S} = S_1 \times \cdots \times S_N$ to model the traffic load variation for the region of interest. Furthermore, we denote $\mathbf{s}^k$ as the state of stage k.

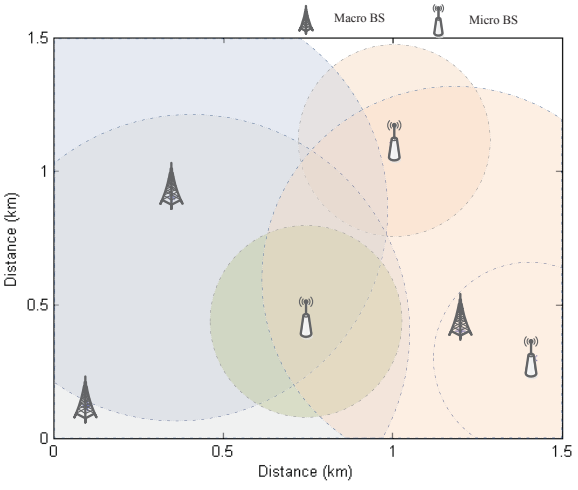

Fig. 1. Illustration of BS deployment in our simulation scenario.

Let's denote the transmission rate of a user located at x and served by BS $i \in \mathcal{B}_{on}$ as $c_i(x, \mathcal{B}_{on})$. For analytical convenience, we assume that $c_i(x, \mathcal{B}_{on})$ does not change over time, i.e., we do not consider fast fading or dynamic inter-cell interferences. Instead, $c_i(x, \mathcal{B}_{on})$ is assumed as a time-averaged transmission rate in this paper, based on the fact that the time scale of user association is commonly much larger than the time scale of fast fading or dynamic inter-cell interferences. Hence, the inter-cell interference is considered as static Gaussian-like noise, which is feasible under interference randomization or fractional frequency reuse, also consistent with the model in [5][13]. Beyond that, though $c_i(x, \mathcal{B}_{on})$ is location-dependent, it is not necessarily determined by the distance from the BS i due to the shadowing effect.

Furthermore, we can naturally define *system load density* as the fraction of time required to deliver traffic load $\gamma(x)$ from BS $i \in \mathcal{B}_{on}$ to location x, namely $\varrho_i(x) = \gamma(x)/c_i(x, \mathcal{B}_{on})$. Similarly, the system load for BS $i \in \mathcal{B}_{on}$ can be represented as $\rho_i = \int_{\mathcal{L}} \varrho_i(x) I_i(x, \mathcal{B}_{on})\, \mathrm{d}x$. Besides, we define the system load for a sleeping BS i as zero, namely $\rho_i = 0,$ if $i \in \mathcal{B} \setminus \mathcal{B}_{on}$. Hence, the indicator set $\mathbb{I} = \{I_i(x, \mathcal{B}_{on}) | i \in \mathcal{B}, x \in \mathcal{L}\}$ is feasible [14] if one BS can serve $\rho_i < 1, \forall i \in \mathcal{B}$. Eventually, our goal is to choose certain active BSs and find a feasible user association indicator set to minimize the overall energy consumption. By exploiting the proposed learning framework, the controller can know the BS switching operation policy at last without the prior knowledge of traffic loads. We will give the details in Section III.

B. Problem formulation

In this paper, we primarily aim to minimize the whole-scale energy consumption of BSs in RANs. Our previous work [9] has shown the energy consumption of BS is not linearly proportional to the traffic load within its coverage area. Moreover, the energy consumption of BSs consists of two categories: constant one and variant one that is proportional to BS's traffic load. Hence, we adopt the generalized energy consumption model [5], which can be summarized as

$$\psi(\rho, \mathcal{B}_{on}) = \sum_{i \in \mathcal{B}_{on}} \left[(1 - q_i) \rho_i P_i + q_i P_i \right], \qquad (1)$$

where $\rho = \{\rho_1, \cdots, \rho_N\}$. Besides, $q_i \in (0, 1)$ is the portion of constant power consumption for BS i, and P_i is the maximum

power consumption of BS i when it is fully utilized.

Above all, our problem is to find an optimal set of active BSs and corresponding user association that minimizes the function of the energy consumption, namely

$$\begin{aligned} &\min_{\mathcal{B}_{on},\rho} \{\psi(\rho, \mathcal{B}_{on})\}, \\ &s.t.\ \rho_i \in [0,1)\ \forall i \in \mathcal{B}. \end{aligned} \tag{2}$$

III. Stochastic BS switching operation with Actor-critic approach

A. Markov decision process

An MDP is defined as a tuple $M =< \mathbb{S}, \mathbb{A}, p, C >$, where $\mathbb{S}$ is the state space, $\mathbb{A}$ is the action space, p is a state transition probability function, and C is a cost function[2]. Specifically, at stage k, the traffic load state is $\mathbf{s}^k$. The controller choose to turn some BSs into sleeping mode (Action $\mathbf{a}^k$) and the users correspondingly associate themselves with the left BSs according to an indicator set $\mathbb{I}^k$[3]. Thereafter, the traffic load state will transform into $\mathbf{s}^{k+1}$ with probability $p(\mathbf{s}^{k+1}|\mathbf{s}^k, \mathbf{a}^k)$. Meanwhile, the immediate cost generated by the environment (computed by (1)) is fed back to the agent, i.e., the BS switching operation controller.

The goal is to find a strategy π, which maps a state $\mathbf{s}$ to an action $\pi(\mathbf{s})$, i.e., $\mathbf{a}^k$, to minimize the discounted accumulative cost starting from the state $\mathbf{s}$. Formally, this accumulative cost is called as a state value function, which can be calculated by [12]

$$\begin{aligned} V^{\pi}(\mathbf{s}) &= \sum_{k=0}^{\infty} \gamma^k C^k(\mathbf{s}^k, \pi(\mathbf{s}^k)|\mathbf{s}^0 = \mathbf{s}) \\ &= C(\mathbf{s}, \pi(\mathbf{s})) + \gamma \sum_{\mathbf{s}' \in \mathbb{S}} p(\mathbf{s}'|\mathbf{s}, \pi(\mathbf{s})) V^{\pi}(\mathbf{s}'), \end{aligned} \tag{4}$$

where γ is the discount factor that maps the future cost to the current state. Given the diminishing importance of future cost than the current one, γ is smaller than 1. The optimal strategy π^* satisfies the Bellman equation [12]:

$$\begin{aligned} V^*(\mathbf{s}) &= V^{\pi^*}(\mathbf{s}) \\ &= \min_{\mathbf{a} \in \mathbb{A}} \left\{ C(\mathbf{s}, \mathbf{a}) + \gamma \sum_{\mathbf{s}' \in \mathbb{S}} p(\mathbf{s}'|\mathbf{s}, \mathbf{a}) V^{\pi^*}(\mathbf{s}') \right\}. \end{aligned} \tag{5}$$

Since the optimal strategy not only minimizes the current cost, but the cumulative cost from the beginning, it contributes to design a foresighted energy saving scheme.

[2]It may be a reward function R on the basis of specific research scenarios. Moreover, it's worthwhile to note here that we use the lowercased $c_i(x, \mathcal{B}_{on})$ to denote transmission rate from BS i to location x while the uppercased C denotes the cost function.

[3]In this paper, we adopt and modify the approach for user association in [5]. At stage k, the user association set $\mathbb{I}^k$ that achieves the minimization of total cost would be that users at location x choose to join BS i^*, while i^* satisfies

$$i^*(x) = \arg\max_{j \in \mathcal{B}_{on}} \frac{c_j(x, \mathcal{B}_{on})}{(1 - q_j)P_j},\ \forall x \in \mathcal{L}. \tag{3}$$

It's worthwhile to note here that this user association scheme may degrade the quality of experience (QoE), such as increasing the delay, etc. We leave how to strike the balance between the user QoE and energy consumption as future work.

B. The actor-critic learning framework for energy saving scheme

There have been some well-known methods to solve the MDP issues such as policy iteration and value iteration of dynamic programming [12]. Unfortunately, these methods heavily depends on prior knowledge of the environmental dynamics. However, it's challenging to know the future traffic load in advance. Therefore, in this paper, we employ an actor-critic method, one kind of reinforcement learning to solve the MDP problem. The reasons to adopt actor-critic method are twofold [15]: (i) since it generates the action directly from the stored policy, it requires little computation to select an action to perform; (ii) it can learn an explicitly stochastic policy which may be useful in non-Markov traffic variation environment of RAN.

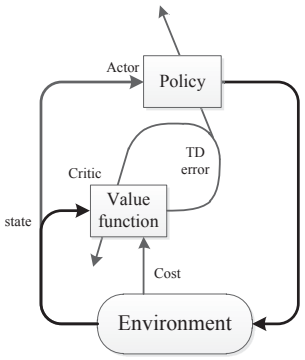

Fig. 2. Classical architecture of actor-critic method.

As the name suggests, the actor-critic method has three components: actor, critic, and environment as illustrated in Fig. 2. At a given state, the actor selects an action in a stochastic way and then executes it. This execution transforms the state of environment to a new one with a certain probability, and feeds back the cost to the actor. Then, the critic criticizes the action executed by the actor through a time difference (TD) error. After the criticism, the actor will prefer to select the action yielding a smaller cost with a higher tendency, and vice versa. The method repeats the above procedure until convergence.

We design an actor-critic learning framework for energy saving scheme as illustrated in Fig. 3.

1) Action selection: Beforehand, we assume the system is at the beginning of stage k, while the traffic load state is $\mathbf{s}^k$. Thereafter, the controller selects an action according to a stochastic policy. The purpose of employing a stochastic policy is to improve performance while explicitly balancing two competing objectives: a) searching for better BS switching operation (exploration) and b) taking as little cost as possible (exploitation), such that the controller not only performs the good BS switching operation based on its past experience but also is able to explore new one. The most common method is to use a Boltzmann distribution. The controller chooses action

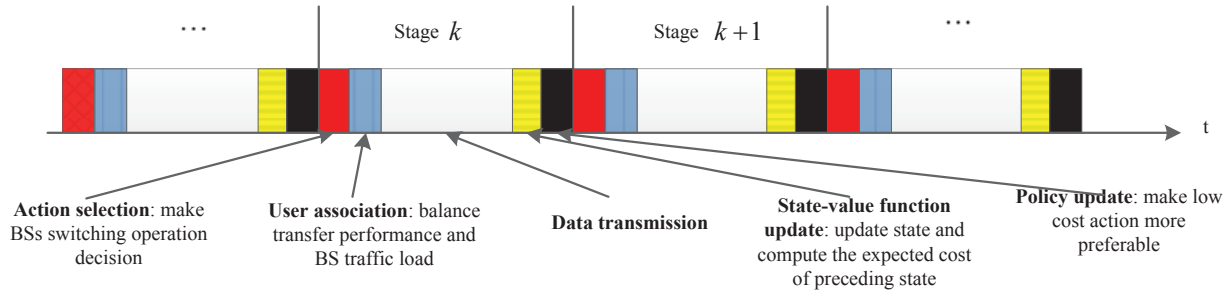

Fig. 3. Illustration of actor-critic learning framework for energy saving scheme.

$\mathbf{a}^k$ in state $\mathbf{s}^k$ of stage k with probability [12]

$$\pi^k(\mathbf{s}^k, \mathbf{a}^k) = \frac{\exp\{p(\mathbf{s}^k, \mathbf{a}^k)\}}{\sum_{\mathbf{a}^k \in \mathbb{A}} \exp\{p(\mathbf{s}^k, \mathbf{a}^k)\}}, \tag{6}$$

where $p(\mathbf{s}^k, \mathbf{a}^k)$ indicates the tendency to select action $\mathbf{a}^k$ at the state $\mathbf{s}^k$, and it will update itself after every action. It's worthwhile to note that though there exists the possibility that the remaining active BSs are not enough to serve the traffic loads in the next stage $k+1$, we can start an emergent response paradigm to quickly turn on some BSs in this case as the conventional energy saving scheme commonly does, which is out of the scope of this paper. Hence, in this paper, we assume the action $\mathbf{a}^k$ the controller finally chooses can satisfy the traffic load requirement.

(2) User association and data transmission: After the controller chooses to turn some of BSs into sleeping mode, the users at location x choose to connect one BS according to (3) and start the data communication.

(3) State-value function update: After the transmission part of stage k, the traffic loads in each BS will change, thus transforming the system to state $\mathbf{s}^{k+1}$. Meanwhile, the total cost for the transmission would be $C^k(\mathbf{s}, \mathbf{a})$. Consequently, a TD error $\delta(\mathbf{s})$ would be computed by the difference between the state-value function $V(\mathbf{s}^k)$ estimated at the preceding state and the one $C^k(\mathbf{s}, \mathbf{a}) + \gamma \cdot V(\mathbf{s}^{k+1})$ at the critic, namely

$$\delta(\mathbf{s}^k) = C^k(\mathbf{s}, \mathbf{a}) + \gamma \cdot V(\mathbf{s}^{k+1}) - V(\mathbf{s}^k). \tag{7}$$

Afterwards, the TD error would feed back to the actor. By the way, the state-value function would be updated as

$$V(\mathbf{s}^k) \leftarrow V(\mathbf{s}^k) + \alpha \cdot \delta(\mathbf{s}^k), \tag{8}$$

where α is a positive step-size parameter which affects the convergence rate.

(4) Policy update: At the end of stage k, we would employ the TD error to "criticize" the selected action, which is implemented as

$$p(\mathbf{s}^k, \mathbf{a}^k) \leftarrow p(\mathbf{s}^k, \mathbf{a}^k) - \beta \cdot \delta(\mathbf{s}^k), \tag{9}$$

where β is a positive step-size parameter. (6) and (9) ensure one action under a specific state can be selected with higher probability if the "foresighted" cost it takes is comparatively smaller or $\delta(\mathbf{s}^k) < 0$.

Now, the procedures which concern our proposed learning framework for energy saving scheme are summarized as Algorithm 1.

Algorithm 1 Algorithm of Energy Saving Scheme through a Learning Framework

Initialization:
for each $\mathbf{s} \in \mathbb{S}$, each $\mathbf{a} \in \mathbb{A}$ **do**
Initialize state-value function $V(\mathbf{s})$, policy function $p(\mathbf{s}, \mathbf{a})$
end for
Repeat until convergent
1) Choose an action according to (6);
2) Users connect some BSs by (3) and then start data transmission;
3) Calculate the cost function $C(\mathbf{s}, \mathbf{a})$ by (1);
4) Identify the traffic loads and accordingly update state $\mathbf{s} \rightarrow \mathbf{s}^{k+1}$ and compute the TD error by (7);
5) Update the state-value function $V(\mathbf{s})$ by (8);
6) Update the policy function $p(\mathbf{s}, \mathbf{a})$ by (9).

IV. Numerical Analysis

We validate the energy efficiency improvement of our proposed scheme by extensive simulations under practical configurations. Here, we simulate under a region consists of three macro BSs and three micro BSs in an area of $1.5km \times 1.5km$ as Fig. 1 shows. Moreover, we assume that file transfer requests at location $x \in \mathcal{L}$ follow a Poisson point process with arrival rate $\lambda(x)$ and file size $1/\mu(x) = 100$ kbyte. Beyond that, we assume the maximum transmission powers for BSs, i.e., 20W and 1W for macro and micro BSs, respectively. Based on the linear relationship in [5], the maximum operational powers for macro BS and micro BS are 865W and 38W, respectively. We set other main parameters in the propagation model according to the COST-231 modified Hata model [16] as summarized in Table I.

By the way, we define *cumulative energy consumption ratio* as the metric to test how much energy saving can be achieved due to the application of our proposed scheme. Specifically, we define the cumulative energy consumption ratio as: the ratio between accumulative energy consumptions when certain BSs

TABLE I
USED SIMULATION PARAMETERS

Parameter description		Value
Simulation area		1.5km × 1.5km
Maximum transmission power	Macro BS	20W
	Micro BS	1W
Maximum operational power	Macro BS	865W
	Micro BS	38W
Height	Macro BS	32m
	Micro BS	12.5m
Channel bandwidth		1.25MHz
Intra-cell interference factor		0.01
File requests	Arrival rate	$5 \times 10^{-6} \sim 10^{-4}$
	File size	100kbyte
Constant power percentage		$0.1 \sim 0.9$

[a] For simplicity, we don't consider fast fading effect and noise influence in our simulation.

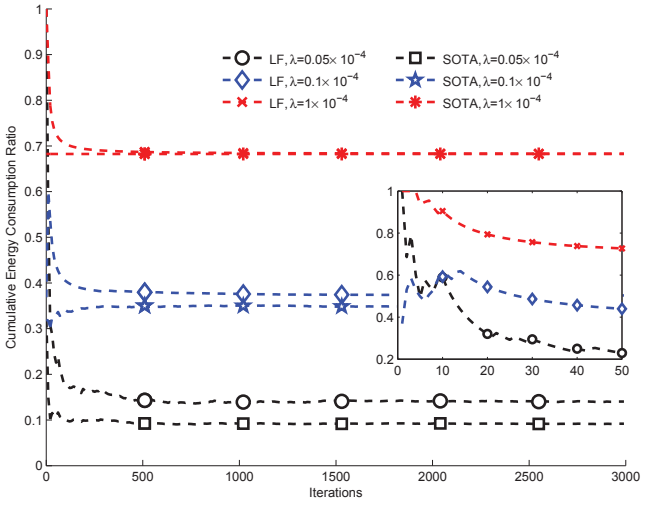

Fig. 4. Performance comparison between learning framework (LF) based energy saving scheme and the state of the art (SOTA) scheme under various homogeneous traffic arrival rates.

are turned off (as our scheme runs) and when all the BSs stay active since our simulation starts. Our definition is reasonable since the definition can show the foresighted energy efficiency improvement, which is exactly the goal of an energy saving scheme. Besides, we compare the performance of our proposed scheme (learning framework based energy saving scheme, LF) with that of the state of the art (SOTA) scheme, which assumes the controller can obtain a full knowledge of traffic loads in prior and find the optimal BS switching solution by exhausting all the possible ones.

A. Effect of traffic loads with static arrival rates on energy saving scheme

We firstly examine how much energy saving can be achieved versus different static traffic load arrival rates. [5] shows a homogeneous traffic distribution of $\lambda(x) = 10^{-4}$ for all $x \in \mathcal{L}$, which offers load corresponding to about 10% of BSs utilizations when all BSs are turned on. Therefore, we vary the homogeneous traffic arrival rate $\lambda(x)$ from 5×10^{-6} to 10^{-4}. Meanwhile, to compute the traffic load boundary points Γ_b, we record the average of traffic loads, i.e., Γ_a, in the whole region and then compute Γ_b for macro BSs and micro BSs by $\Gamma_{b,macro} = \frac{\Gamma_a}{3}$, $\Gamma_{b,micro} = \frac{1}{10}\Gamma_{b,macro}$, respectively.

Fig. 4 shows the effect of traffic load on energy savings when the portion of fixed power consumption q_i equals 0.5. With the decrease of traffic load arrival rate λ from 10^{-4} to 5×10^{-6}, we can expect more significant energy conservation. Moreover, the cumulative energy consumption ratio continues decreasing as the simulation runs since the controller will have a better understanding of the traffic loads, thereby knowing whichever action has better energy efficiency. As a result, when $\lambda = 10^{-4}$, we can expect an 70% of cumulative energy consumption ratio after 500 iterations, which is quite approximate to the SOTA scheme. On the other hand, since the proposed learning scheme is performed without the knowledge of traffic loads a prior, the performance of it is inferior to that of the SOTA scheme, especially at the beginning of the simulations. However, we can see that the gap compensated for the absent knowledge is quite small, which in turn proves the effectiveness of the proposed scheme.

B. Effect of energy consumption models of BSs on energy saving scheme

In this part, we vary the portion of fixed power consumption q_i between 0 and 1, in order to cover various types of BSs with different energy consumption models. Fig. 5 shows the effect of energy consumption models of BSs on energy saving schemes when the traffic file request follows a homogeneous distribution with arrival rate $\lambda(x)$ equaling 10^{-4} and 10^{-5}. The performance of our proposed learning scheme is the result after 5000 iterations. As Fig. 5 depicts, the proposed LF scheme and the SOTA scheme will both perform better when the constant power consumption accounts for a larger proportion of the whole energy consumption. The reason lies in that when the constant power consumption takes a larger percentage, i.e., $q_i = 0.9$, turning off one under-utilized BS will make a clearer difference and save more energy. On the other hand, more than half of the overall energy consumption usually takes place on the constant power, i.e., cooling, idle-mode signaling and processing in the present RAN infrastructure [6]. Therefore, our proposed scheme can render a strong positive effect in saving energy. Moreover, the performance of the proposed LF scheme is quite close to the SOTA scheme however the portion of fixed power consumption q_i varies.

C. Performance of learning framework-based energy saving scheme in periodic traffic load scenario

In this section, we explore the performance of the proposed scheme when traffic loads periodically fluctuates. [10] shows practical traffic load profile is periodical and can be approximated by a sinusoidal function $\overline{\lambda}(t) = \lambda_V \cdot \cos(2\pi(t+\phi)/D) + \lambda_M$, where t is the index of time, D is the period of a traffic load profile, λ_V is the variance of traffic profile and

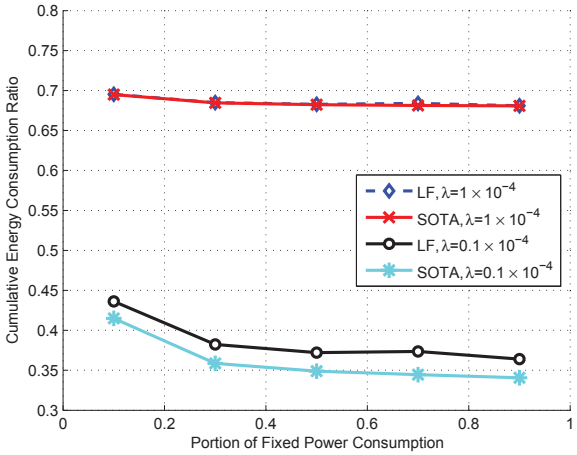

Fig. 5. Performance comparison between learning framework (LF) based energy saving scheme and the state of the art (SOTA) scheme under different energy consumption models. The performance of learning framework (LF) based energy saving scheme is the result after 5000 iterations.

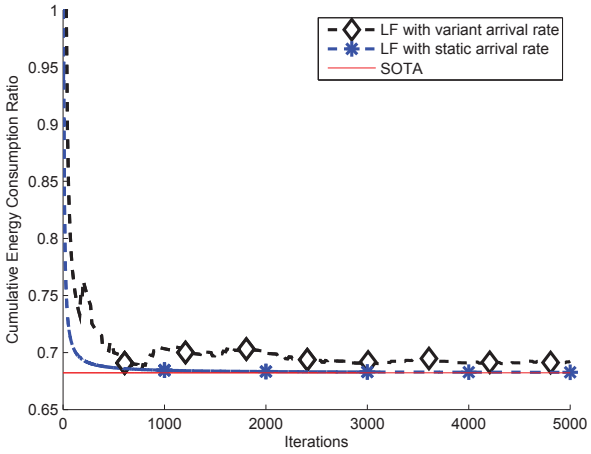

Fig. 6. Performance comparison of learning framework (LF) based energy saving scheme in homogeneous variant traffic arrival rate $\overline{\lambda}(t,x) = (0.99 \cdot \cos(2\pi(t+10)/24)+1) \times 10^{-4}$ and static traffic arrival rate $\lambda(x) = 10^{-4}$.

λ_M is the mean arrival rate. Therefore, we employ $\overline{\lambda}(t,x) = (0.99 \cdot \cos(2\pi(t+10)/24)+1) \times 10^{-4}$ to approximate the traffic load arrival rate in one day (24 hours) at location $x \in \mathcal{L}$. Fig. 6 shows the performance of the proposed LF scheme when the variant traffic loads arrival rate $\overline{\lambda}(t,x)$ equals 10^{-4}. Unfortunately, we can find the cumulative energy consumption ratio is higher than that under the same homogeneous static traffic arrival rate. In other words, it's more challenging to choose an action in an uncertain traffic scenario, thus leading to some performance degradation. Fortunately, we should also notice our proposed scheme still yields approximate performance to the SOTA scheme under static arrival rate after 3000 iterations.

V. Conclusion

In this paper, we developed a learning framework for BS energy saving scheme. We specifically formulated the BS switching operation under a variant traffic load as a Markov decision process. Afterwards, we adopt the actor-critic method, a reinforcement learning approach to give the BS switching solution to decrease the overall energy consumption. Finally, the extensive simulation results manifest the effectiveness and robustness of our energy saving schemes under various practical configurations.

Acknowledgment

This paper is partially supported by the National Basic Research Program of China (973Green, Program No. 2012CB316000) and the National Natural Science Foundation of China (NSFC) under grant number 61071130.

References

[1] H. Zhang, A. Gladisch, M. Pickavet, Z. Tao, and W. Mohr, "Energy efficiency in communications," *IEEE Communications Magazine*, vol. 48, no. 11, pp. 48–49, Nov. 2010.

[2] M. Marsan, L. Chiaraviglio, D. Ciullo, and M. Meo, "Optimal energy savings in cellular networks," in *Proceedings of IEEE ICC 2009*, Dresden, Germany, Jun. 2009.

[3] C. M. R. Institute, "C-RAN: Road towards green radio access network," White Paper, V1.0.0, Tech. Rep., 2010.

[4] G. P. Fettweis and E. Zimmermann, "ICT energy consumption-trends and challenges," in *Proceedings of WPMC 2008*, Lapland, Finland, Sep. 2008.

[5] K. Son, H. Kim, Y. Yi, and B. Krishnamachari, "Base station operation and user association mechanisms for energy-delay tradeoffs in green cellular networks," *IEEE Journal on Selected Areas in Communications*, vol. 29, no. 8, pp. 1525 –1536, Sep. 2011.

[6] C. Peng, S.-B. Lee, S. Lu, H. Luo, and H. Li, "Traffic-driven power savings in operational 3G cellular networks," in *Proceedings of ACM Mobicom 2011*, Las Vegas, Nevada, USA, Sep. 2011.

[7] L. Chiaraviglio, D. Ciullo, M. Meo, M. A. Marsan, and I. Torino, "Energy-aware UMTS access networks," in *Proceedings of WPMC 2008*, Lapland, Finland, Sep. 2008.

[8] Z. Niu, Y. Wu, J. Gong, and Z. Yang, "Cell zooming for cost-efficient green cellular networks," *IEEE Communication Magazine*, vol. 48, no. 11, pp. 74–79, Nov. 2010.

[9] R. Li, Z. Zhao, Y. Wei, X. Zhou, and H. Zhang, "GM-PAB: a grid-based energy saving scheme with predicted traffic load guidance for cellular networks," in *Proceedings of IEEE ICC 2012*, Ottawa, Canada, Jun. 2012.

[10] E. Oh and B.Krishnamachari, "Energy savings through dynamic base station switching in cellular wireless access networks," in *Proceedings of IEEE Globecom 2010*, Miami, Florida, USA, Dec. 2010.

[11] S. Zhou, J. Gong, Z. Yang, Z. Niu, and P. Yang, "Green mobile access network with dynamic base station energy saving," in *Proceedings of ACM Mobicom 2009*, Beijing, China, Sep. 2009.

[12] R. S. Sutton and A. G. Barto, *Reinforcement learning: an introduction*, 2005.

[13] A. Sang, M. MAdihian, X. Wang, and R. D. Gitlin, "Coordinated load balancing, handoff/cell-site selection, and scheduling in multi-cell packet data systems," in *Proceedings of ACM Mobicom 2004*, Philadelphia, PA, Sep. 2004, pp. 302–314.

[14] H. Kim, G. de Veciana, X. Yang, and M. Venkatachalam, "α-optimal user association and cell load balancing in wireless networks," in *Proceedings of IEEE INFOCOM 2010*, San Diego, CA, USA, Mar. 2010.

[15] K. Zhou, "Robust cross-layer design with reinforcement learning for IEEE 802.11n link adaptation," in *Proceedings of IEEE ICC 2011*, Kyoto, Japan, Jun. 2011.

[16] IEEE 802.16m Broadband Wireless Access Working Group, *IEEE 802.16m Evaluation Methodology Document (EMD)*, Jul. 2008. [Online]. Available: http://ieee802.org/16