URSI AP-RASC 2019, New Delhi, India; 09 - 15 March 2019

Transfer Reinforcement Learning based Framework for Energy Savings in Cellular Base Station Network

Shreyata Sharma*, S. J. Darak* and Anand Srivastava*
*Department of Electronics and Communication Engineering, IIIT, Delhi

Abstract

Last few years have witnessed an exponential upsurge in data intensive applications over the communication networks. Energy saving is one of the major aspects in such networks wherein the increased traffic load entails deployment of a large number of base stations (BSs). In this paper, a BS switching scheme is proposed which exploits reinforcement learning (RL) for dynamic sectorization of BSs to increase the energy efficiency of cellular networks. Furthermore, previously estimated traffic statistics is exploited through the process of transfer learning for further improvement in energy savings and speeding up the learning process. The superiority of the proposed framework is depicted through simulations and relevant mathematical analysis. Compared to conventional ON/OFF scheme, proposed framework offers around 40% lower average energy consumption for cellular networks with low to moderate loads.

1 Introduction

Over the past few years, there has been explosive growth in data traffic over the communication networks. According to Ericsson mobility reports, there has been 52% growth in mobile traffic between 2017-2018 [1]. Also, smartphone traffic is expected to grow twelve times between 2015-2021 [2]. To meet such huge demands, network operators are expanding their networks by deployment of large number of base stations (BSs). Studies have reported that the energy consumed by the BSs constitutes more than 55% of the total energy consumption of the communication systems [3]. This means that the massive growth in mobile data traffic is being served at the expense of huge energy consumption and increased carbon footprint. As a result of this, energy saving in communication networks has become a central research area and has to be handled from both ecological and economical perspectives. In the current deployment, the BSs are more or less active all the time with the capacity to serve peak load. The aspect of variation in the traffic load which is a practical scenario is generally not taken into account. Therefore, there is a need to develop an optimal switching scheme such that the BSs are switched ON/OFF according to the traffic load. Furthermore, the existing networks do not fully exploit the past usage statistics for optimal operation of BSs. It is seen that there is a reasonable correlation between the current data traffic and the data traffic in the past. In case of cellular networks, this can be attributed to the typical day-night behavior of daily movements of the users [3]. Therefore, the past data statistics is indeed significant for devising present energy saving scheme [4]. Traffic load based dynamic BS switching has been identified as a promising technique for energy efficient operation of wireless access networks. This technique is implemented in [6, 7, 8] assuming prior information about the traffic load. In [9], RL is applied for optimal BS switching considering traffic load to follow a Poisson distribution. In our previous works [4, 5], the scheme discussed in [9] is used for developing energy saving technique for Wi-Fi AP switching and two state BS switching scheme in HetNets respectively. This work is an extension of our work presented in [4, 5]. In most of the present research works, two state BS switching is considered in which BSs are switched between the active mode and sleep mode according to the traffic load. However, majority of the time, the traffic is at moderate level with few time slots of peak traffic which means existing schemes are not energy efficient. In this case, learning based BS switching schemes could be quite effective considering practical aspects of wireless networks. In the present deployment, when a BS is active it operates in tri-sectorized mode deploying separate power amplified for each sector resulting in high energy consumption.

This paper presents an efficient BS switching scheme for energy saving in cellular networks while taking into account the discussed bottlenecks. Herein, apart from an active state at high traffic and sleep state at low traffic, the BSs are switched to an omnidirectional state at moderate traffic leading to a more efficient energy saving scheme. The BSs are switched based on actor-critic (AC) RL algorithm. It is quite intuitive that with such a scheme, the reduction in energy consumption would be maximum in case of moderate traffic load. Furthermore, previously learned data statistics are well exploited through transfer learning (TL). A brief analysis on trade off between the energy consumption and system delay is also presented.

The research problem formulation and findings in the paper are organized as follows: The proposed work and relevant mathematical analysis is discussed in section 2. The results are discussed in section 3 and section 4 concludes the paper.

2 Proposed Work

To analyze three state BS switching, the AC-RL algorithm is applied on a system containing uniformly distributed

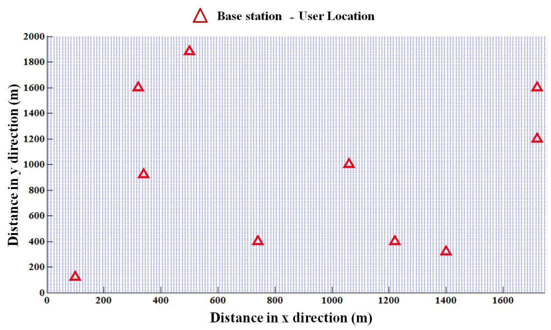

Figure 1. Graphical representation of the base station network.

users in a region served by a set of overlapping BSs distributed randomly as shown in Figure 1.

The traffic load at a given location is considered to follow a Poisson distribution with arrival rate λ [10, 11, 12]. A Markov decision process (MDP) is formulated using the traffic variations. An MDP is characterized by the tuple $M =< S,A,P,C >$, where S is the state space, A is the action space, P is the state transition probability, C is the cost function. At stage k, traffic load is at state $\mathbf{s}^{(k)} = \{s_1^{(k)}, s_2^{(k)}, \ldots\}$, where $s_i^{(k)}$ represent state of i^{th} BS at stage k [4]. When an action $\mathbf{a}^{(k)} = \{a_1^{(k)}, a_2^{(k)}, \ldots\}$ is taken if $a_i^k = 0$ the i^{th} BS is switched OFF else it is switched ON if $a_i^k = 1$. When a BS is ON, if the traffic load is greater than a threshold, it switches to tri-sectorized mode. Else, it switches to omni-directional mode.

The system cost, C in this case is power consumption of the system. Power consumption of an BS consist of two parts: static power consumption which is independent of traffic load and dynamic power consumption which varies proportional to the traffic load. Mathematically,

$$C = \sum_{i \in B'} [(1-q_i)\rho_i P_i + q_i P_i] \tag{1}$$

where q_i is fraction of static power consumption of i^{th} BS, ρ_i is traffic load density, P_i is total power consumption and B' is the set of active BSs [9].

In this work, the solution to the formulated MDP is obtained through AC learning algorithm. AC learning algorithm is a subclass of RL algorithms. In general, RL framework consists of an agent and an environment. There is a continuous interaction between the agent and the environment. At each time step, the agent implements a mapping from states to action, which is called agent's policy. There is a reward (or cost) associated with each action and the expected value of the discounted cost or reward is called the state value function, which is given by,

$$V^{\pi}(s) = E_{\pi}\left[\sum_{k=0}^{\infty} \gamma^k C(s^k, \pi(s^k)) | s^{(0)} = s\right] \tag{2}$$

where E is the expectation operator, $C(s^k, \pi(s^k))$ represents system cost at stage k which depends on state s^k and action $\pi(s^k)$ and γ is the discount factor having value between 0 and 1. It can be seen that the value of γ^k decreases as k increases. This term is included to incorporate the fact that the worth of immediate reward is greater than the later rewards. The goal of RL is to maximize the expected reward (or minimize the cost)[13]. In AC algorithm, the policy structure is called the 'Actor' as it selects the action and the value function acts as a 'Critic' as its value determines how good is the action taken and consequently decides the future course of action. In the present context, the objective of AC algorithm is to find optimal strategy π which maps every state 's' to an action $\pi(s^{(k)})$ such that system cost, C, is minimized. As the learning proceeds, the policy structure tends towards optimal value and at each state optimal action is taken such that the energy consumption is minimized [4].

The BSs are switched following the AC algorithm steps described in [4]. In this case, there is an additional step after action selection i.e. user association and rest of the steps remain the same. After action selection, when certain BSs are switched OFF, users associate themselves with the ON BSs according to the following metric:

$$i^*(x) = argmax_j \frac{c(x,j)}{P_j} \tag{3}$$

where, $c(x, j)$ is the upper-bound on the capacity of the link between user located at x and j^{th} BS calculated according to Shannon's theorem assuming the link between BS and the user to follow COST-231 HATA model. P_j is the power consumed by the j^{th} BS. According to (1), a user located at position x chooses to be served by a BS j if the link between them provide maximum capacity and the BS consumes minimum power.

For transferring the knowledge gained from the past, the overall policy is divided into two parts 'native policy' and 'exotic policy' [9].

$$p_{overall} = (1-\zeta(k))p_{native} + \zeta(k)p_{exotic} \tag{4}$$

where, $\zeta(k) = \theta^k$ is transfer rate that determines fraction of exotic policy that contributes to overall policy, $\theta \in (0,1)$ is the transfer rate factor, p_{native} is the policy which is continuously updated as the learning proceeds and p_{exotic} is the previously learned policy which is transferred to the current task. In the current work, to get the exotic policy the algorithm is executed multiple times. The learned statistics from the previous execution is taken as exotic policy for current execution. This work can be extended in the future to have real data for cellular networks as well as it was done for Wi-Fi network in [4].

2.1 Mathematical Analysis

The comparison between system energy consumption for the conventional two state scheme and the proposed scheme is depicted through underlying mathematical formulation developed in this study.

Let S_i be a random variable which represent the state of i^{th} BS in the set of available BSs. The random variable S_i is a function of BS traffic load which is in-turn a Poisson random variable. S_i is either 0 or 1 depending on whether the BS is ON or OFF and can be described as:

$$S_i = \begin{cases} 0, & \text{when } \Gamma_i < \Gamma_{th} \\ 1, & \text{when } \Gamma_i \geq \Gamma_{th} \end{cases} \tag{5}$$

where, Γ_i is the traffic load corresponding to i^{th} BS and Γ_{th} is the threshold value of the traffic above which the BS has to be switched ON. Therefore, in a two state system expected value of power consumption of a BS can be given as,

$$E = f(S_i = 0)P(S_i = 0) + f(S_i = 1)P(S_i = 1) \quad (6)$$

where, $f(S_i = s_i)$ represents the probability that i^{th} BS is at state s_i and $P(S_i = s_i)$ represents power consumed by the i^{th} BS at state s_i. Now, $f(S_i = 0) = f(\Gamma_i < \Gamma_{th})$ and $f(S_i = 1) = f(\Gamma_i \geq \Gamma_{th})$ therefore,

$$E = f(\Gamma_i < \Gamma_{th})P(S_i = 0) + f(\Gamma_i \geq \Gamma_{th})P(S_i = 1) \quad (7)$$

Putting $P(S_i = 0) = 0$ and $P(S_i = 1) = P_{sectorized}$,

$$E = f(\Gamma_i < \Gamma_{th}) \times 0 + f(\Gamma_i \geq \Gamma_{th}) \times P_{sectorized} \quad (8)$$

Traffic load at a given BS is the sum of traffic due to all users associated with it. Traffic from each user follow Poisson distribution i.e. $F(X = x) = e^{-\lambda}\lambda^x/x!$ $(Poisson(\lambda))$. From the property of Poisson random variable, sum of independent $Poisson(\lambda_i)$ distribution is a $Poisson(\sum \lambda_i)$ distribution. Let $\sum \lambda_i = \lambda$ and $\Gamma_i = x$ therefore,

$$f(\Gamma_i \geq \Gamma_{th}) = \sum_{x=\Gamma_{th}}^{\infty} e^{-\lambda}\lambda^x/x! \quad (9)$$

Therefore, $$E = P_{sectorized} \times [\frac{e^{-\lambda}\lambda^{\Gamma_{th}}}{\Gamma_{th}!} + \frac{e^{-\lambda}\lambda^{\Gamma_{th}+1}}{(\Gamma_{th}+1)!} \cdots] \quad (10)$$

In the three state model presented in this work, at a particular instance of time a given BS can be either in active mode, sleep mode or omnidirectional mode. Hence, in this case the random variable S_i can be defined as,

$$S_i = \begin{cases} 0, & \text{when } \Gamma_i < \Gamma_{th} \\ o, & \text{when } \Gamma_{th} \leq \Gamma_i \leq n \times \Gamma_{th} \\ 1, & \text{when } \Gamma_i \geq n \times \Gamma_{th} \end{cases} \quad (11)$$

where, n is a scalar such that $n > 1$ and o represent the omnidirectional state. In this case, the expected value of power consumption of a BS would be,

$$E_{omni} = f(S_i = 0)P(S_i = 0) + f(S_i = 1)P(S_i = 1) + f(S_i = o)P(S_i = o) \quad (12)$$

$$E_{omni} = f(\Gamma_i < \Gamma_{th})P(S_i = 0) + f(\Gamma_{th} \leq \Gamma_i \leq n \times \Gamma_{th}) P(S_i = o) + f(\Gamma_i > n \times \Gamma_{th})P(S_i = 1) \quad (13)$$

$$E_{omni} = P_{omni} \times \left[\frac{e^{-\lambda}\lambda^{\Gamma_{th}}}{\Gamma_{th}!} + \frac{e^{-\lambda}\lambda^{\Gamma_{th}+1}}{(\Gamma_{th}+1)!} \cdots + \frac{e^{-\lambda}\lambda^{n\times\Gamma_{th}}}{n \times \Gamma_{th}!}\right] + P_{sectorized} \times \left[\frac{e^{-\lambda}\lambda^{(n+1)\times\Gamma_{th}}}{(n+1) \times \Gamma_{th}!} + \cdots\right] \quad (14)$$

In (10) entire range of values are multiplied by $P_{sectorized}$ while in (14) values for which $\Gamma_i > n \times \Gamma_{th}$ are multiplied by $P_{sectorized}$ and rest by P_{omni}. As $P_{omni} < P_{sectorized}$, therefore $E_{omni} < E$. Hence, the expected value of power consumption for omnidirectional mode is always less than or equal to the sectorized mode.

3 RESULTS AND DISCUSSIONS

To evaluate the performance improvement through the proposed scheme the AC algorithm is applied on the system under two scenarios : a) the BSs are switched between active mode and sleep mode according to traffic load b) BSs are switched between active, omnidirectional and sleep modes based on traffic load. As the traffic load is considered to follow a $Poisson(\lambda)$ distribution. Higher the value of λ, higher would be the traffic load. Simulations are performed for varied traffic load. Figure 2 depicts the variation of mean differential improvement with λ i.e. traffic load. Mean differential improvement is the average of difference between energy consumption in a two state model and the proposed three state model. The improvement is highest when the load is moderate as the fraction of BSs going into low power omnidirectional mode would be highest in this case.

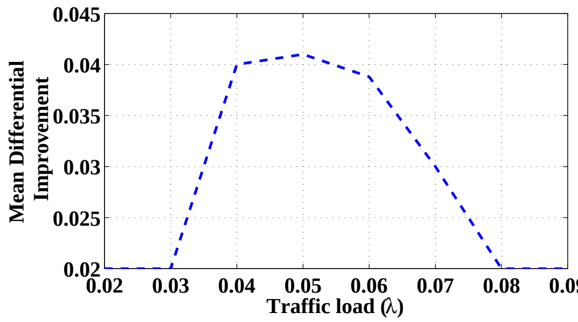

Figure 2. Variation of Mean differential improvement with respect to traffic load.

Further improvement in performance can be achieved by exploiting the past data statistics and using the concept of transfer learning discussed in previous section. This improvement in performance is depicted in Figure 3 for moderate load. The performance is measured in terms of a metric termed as 'Energy Consumption Ratio (ECR)' which is the ratio of energy consumption of the system on a particular instant of learning process to the energy consumption of the system when there is no learning and all BSs are ON at that instant. The learning process proceeds in k stages and each stage is termed as an Episode. It is observed that the proposed three state BS switching in an AC framework leads to 15% additional reduction than the two state switching scheme. Furthermore, the application of transfer learning to this scheme leads to an overall reduction 40% which is a quite significant amount. Power consumption of a BS consist of two parts: static power consumption which is independent of traffic load and dynamic power consumption which varies proportional to the traffic load [4]. Furthermore, a finite delay is incurred in transmitting the overall traffic associated with all the users in the system. To ana-

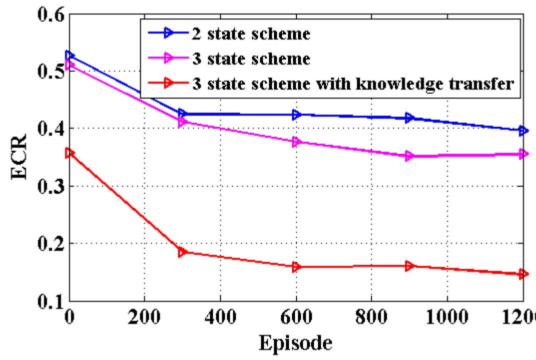

Figure 3. Reduction in system energy consumption at moderate traffic depicted through ECR curve.

lyze the trade off between system delay and energy saving, the system cost described in [4] is modified as,

$$C = \sum_{i \in A'} [(1-q_i)\rho_i P_i + q_i P_i] + \varsigma C_d \tag{15}$$

where, q_i is fraction of static power consumption of i^{th} BS, ρ_i is traffic load density, P_i is total power consumption and A' is the set of active BSs. C_d is the delay equivalent cost given by :

$$C_d = \sum_{i \in A'} \frac{\rho_i}{(1-\rho_i)} \tag{16}$$

As discussed in [11], minimizing C_d is equivalent to minimizing the average delay. A' is the set of active BSs and ς is a scalar that determines the weightage which is to be given to the delay equivalent cost reduction [9]. The variation of mean ECR with delay importance parameter (ς) for the proposed scheme under low, moderate and high traffic is depicted in Table 1 and Figure 4. The simulations are performed by varying the value of ς between 500 and 3000. For each value of ς , the mean ECR and delay are tabulated. Table 1 depicts that a higher value of ς corresponds to a lower delay and a higher energy consumption. This is due to the fact that higher value of ς amounts to a greater importance to delay equivalent cost reduction. If there is a lower tolerance to delay, lesser number of BSs would be turned OFF and hence the energy consumption would be greater which is apparent from Table 1. Therefore, to ensure the required quality of service it is necessary to take care of the trade-off between the system energy consumption and delay.

Table 1. Mean ECR variation with Delay importance parameter

	Delay Importance Parameter	**500**	**1000**	**1500**	**2000**	**2500**	**3000**
High Traffic	**Mean ECR**	0.65	0.70	0.71	0.78	0.85	0.91
	Average Delay	1.31	0.70	0.47	0.39	0.28	0.26
Moderate Traffic	**Mean ECR**	0.57	0.72	0.74	0.86	0.93	1
	Average Delay	1.11	0.74	0.49	0.41	0.36	0.33
Low Traffic	**Mean ECR**	0.58	0.62	0.73	0.75	0.85	1.1
	Average Delay	1.16	0.62	0.48	0.37	0.34	0.33

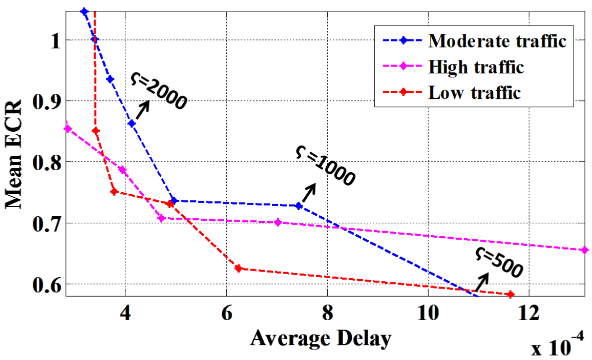

Figure 4. Mean ECR variation with Delay importance parameter

4 Conclusion

This paper presents an important analytical study on energy saving in cellular networks. RL is used to devise an optimal three state BS switching scheme in cellular networks. The scheme results in significant reduction in the energy consumption of the system. A further improvement in the energy saving is achieved by using the concept of knowledge transfer. The simulated results are well supported by the developed mathematical formulation pertaining to the research problem.

Acknowledgment

This work is supported by the DST INSPIRE faculty fellowship awarded by Government of India.

References

[1] Ericsson. (Aug 2018). *Ericsson mobility report Q2 2018 Update* [Online]. Available: https://www.ericsson.com/assets/local/mobility-report/documents/2018/emr-q2-update-2018.pdf

[2] Ericsson. (Feb 2015). *Ericsson mobility report* [Online]. Available: https://www.ericsson.com/assets/local/mobility-report/documents/2015/ericsson-mobility-report-feb-2015-interim.pdf

[3] J. Lorincz, A. Capone, and M. Bogarelli, "Energy savings in wireless access networks through optimized network management," *5th IEEE International Symposium on Wireless Pervasive Computing (ISWPC)*, pp. 449-454, Modena, Italy, May 2010.

[4] S. Sharma, S.J. Darak and Anand Srivastava "A Transfer Learning Framework for Energy Efficient Wi-Fi Networks and Performance Analysis using Real Data," *IEEE Advanced Networks and Telecommunications Systems (ANTS)*, pp. 1-6, Jun. 2016.

[5] S. Sharma, S.J. Darak, A. Srivastava, "Energy saving in heterogeneous cellular network via transfer reinforcement learning based policy," *9th International Conference on Communication Systems and Networks (COMSNETS)*, pp. 397-398, Jan 2017.

[6] 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, Sept. 2011.

[7] E. Oh and B. Krishnamachari, "Energy Savings through Dynamic Base Station Switching in Cellular Wireless Access Networks," *IEEE Global Telecommunications Conference (GLOBECOM 2010)*, pp. 1-5, Miami, FL, Dec. 2010.

[8] S. Zhou, J. Gong, Z. Yang, Z. Niu, and P. Yang, "Green mobile access network with dynamic base station energy saving," *ACM MobiCom*, vol. 9, no. 262, pp. 10-12., Sept. 2009.

[9] R. Li, Z. Zhao, X. Chen, J. Palicot, and H. Zhang, "Tact: a transfer actor-critic learning framework for energy saving in cellular radio access networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 4, pp. 2000-2011, April 2014.

[10] S. Das, H. Viswanathan, and G. Rittenhouse, "Dynamic load balancing through coordinated scheduling in packet data systems," *IEEE INFOCOM,* vol. 1, pp. 786-796, March 2003.

[11] H. Kim, G. de Veciana, X. Yang, and M. Venkatachalam, "Alpha-optimal user association and cell load balancing in wireless networks,"*IEEE INFOCOM,* pp. 1-5, March 2010.

[12] A. Sang, X. Wang, M. Madihian, and R. D. Gitlin, "Coordinated load balancing, handoff/cell-site selection, and scheduling in multi-cell packet data systems,"Wireless Networks, vol. 14, no. 1, pp. 103-120, Feb. 2008.

[13] R. S. and A. G. Barto, Reinforcement learning: An introduction, vol. 1, no. 1.