

Energy Optimization of a Base Station using Q-learning Algorithm

Anisha Aggarwal

Department of Mathematics
Indian Institute of Technology Delhi
New Delhi - 110016, India
anishaaggarwal7895@gmail.com

Dharmaraja Selvamuthu

Department of Mathematics
Indian Institute of Technology Delhi
New Delhi - 110016, India
dharmar@maths.iitd.ac.in

Abstract—A sleep strategy with several sleep mode (SM) levels for energy-efficient 5G base stations (BS) is proposed to reduce energy consumption. Energy consumption and Quality of Service (QoS) management are paired as a result of awakening sleeping BSs. Advanced Sleep Modes (ASMs) gradually deactivate BS components to save energy. Based on component transition times (deactivation and activation), different SM levels could be considered. This research proposes a Q-learning based approach for ASMs to present an optimal tradeoff between energy consumption and QoS.

Index Terms—Q-learning, Wireless communication, Power consumption, Delay

I. INTRODUCTION

Wireless networks will need to serve a substantially larger capacity demand due to the fast growth in the number of users and services. Energy consumption (EC) in cellular networks has become an essential pillar in designing future fifth-generation (5G) networks due to academic and industrial interest sparked by economic and environmental concerns over the past decade. However, the Information and communication technology (ICT) sector's energy consumption is rising because of the rising demand for traffic. According to estimates, ICT uses roughly 4.7% of the electrical energy produced worldwide and contributes just about 1.7% of the world's total CO₂ emissions. On the one hand, this reduces the amount of carbon emissions released into the environment. On the contrary, the decrease in EC will have the effect of lowering the operational expenses (OPEX) that are incurred by network operators. These objectives led to the development of technologies that make it possible to create future wireless architectures with improved coverage and increased energy efficiency (EE).

However, compared to current cellular systems, the 5G is anticipated to offer universal internet access with a 1000 times greater data throughput. As a result, 5G systems bring novel network topologies and application scenarios in addition to physical layer methods. In particular, the construction of numerous small-scale base stations (BS), often known as network densification, is a significant trend in 5G networks. It is not sustainable to attempt to accomplish this challenging goal using the paradigm and architecture described above since doing so, will result in an energy crisis with significant

economic and environmental implications.

Numerous research has looked into various strategies to lower the energy usage in mobile networks due to the economic and environmental concerns of mobile operators as well as their desire to support the vision of 5G communication systems. The majority of research focuses on lowering the power to improve the energy efficiency of cellular networks because the radio access network (RAN), and particularly the BS, consumes 75%–80% of the energy used by a telecommunications network.

Because the BSs are the primary energy consumers in the wireless network, accounting for around 80% of the total EC, a significant number of research projects have focused on discovering effective strategies to improve their energy efficiency. In this regard, some studies examined reducing the BS transmit power, others concentrated on the hardware's efficiency, and many others considered the concept of deactivating the BSs, i.e., putting them into sleep mode (SM), whenever it is feasible.

Making wise use of the BSs is becoming more and more crucial because the BS is a cellular network component that consumes the most power. The following techniques can be used in cellular networks to achieve EE: 1). advanced hardware utilization at the BSs [1], 2). choosing the most energy-efficient physical layer technique for a specific data load, and 3). intelligent deployment of BSs by turning off a component of them during times of low data traffic are just a few of the strategies used. These diverse approaches, which address the aforementioned issue at various scales, include network planning and deployment, switching-off techniques [2], radio resource management optimization, component level augmentation, and the use of renewable energy sources. Sleep mode techniques are the most effective energy-saving strategies among the various layers. When considering any energy-saving strategy for the BS, it is crucial to make sure that the chosen approach won't negatively impact the Quality of Service (QoS) that users receive.

In [3], a theoretical paradigm for BS energy conservation is put forth incorporating dynamic BS operation and the related user association problem. A 5G BSs sleep model is recommended to reduce the energy consumption of a cel-

lular network. The proposed model states that the BS can switch between different levels of SMs, each of which has its own activation/deactivation time and power needs. Due to the need for reference signaling and backward compatibility, implementing these several stages of SMs on 4G networks is difficult. However, the cell reference signals (CRS) are eliminated in 5G, allowing a BS to investigate these various SM levels.

Q-learning algorithm is proposed for various problems in wireless networks in literature. A strategy based on Q-learning in combination with logarithmic cooling (LC) is proposed to optimize the connection of user equipment (UE) with a base station or access point (AP) to address energy efficiency issues of device-to-device (D2D) communications in heterogeneous networks [4]. To make local caching effective, which can be seen as a form of recommendation, we introduce a straightforward approach, i.e., informing the users what the BS has cached; Q-learning is used to understand the request likelihood and the statistics of random arrival and departure of mobile users because we don't know the probabilities that the users would request the files [5]. Decentralized Q-learning, a form of real-time multi-agent reinforcement learning, handles the combined interference generated by numerous wireless regional area networks [6]. A distributed Q-learning algorithm for managing energy-harvesting in heterogeneous networks has been proposed [7]. A scheme is proposed using Q-learning in which UEs learn bias values that minimize the number of outage UEs from experience [8].

This paper focuses on the SMs approach and, more specifically, on the advanced sleep modes feature. It refers to a systematic deactivation of the BS's components based on the time required for each component to deactivate and restart (transition time). Based on this time, several types of SMs with distinct features, such as duration and power consumption, are defined.

In this paper, the model is proposed to examine the tradeoff between the average power consumption, service rate, and average delay (i.e., the time for which user requests wait in the buffer before the BS reactivates) using ASMs, in contrast to models that use only one type of SM. This given tradeoff is analyzed using the reinforcement learning algorithm named Q-learning algorithm. An unbalanced traffic cellular network where each BS selects its optimal SM level to strike a balance between QoS and power consumption is considered. For each BS, to determine the appropriate energy-saving strategy, an adaptive sleep scheme algorithm based on reinforcement learning is suggested that manages the states of the BSs.

This paper is organized as follows. Section II illustrates the proposed model description. Section III explains the Q-learning algorithm and how the results can be obtained for the presented model using the Q-learning algorithm. Section IV deals with simulation and numerical results. Finally, concluding remarks are presented in Section V.

II. MODEL DESCRIPTION

The model is proposed to have the optimal tradeoff between QoS provided to the user requests and power consumption in the BS. The QoS is maximized in terms of minimizing the delay to the service requests. For the minimization of energy consumption in the BS, ASMs are considered. The advanced sleep modes consider SM_1 , SM_2 , SM_3 and the states for this model are defined as:

Active state: The state where the service is provided to the user's requests.

Idle state: After providing the service to all the user requests, the system moves to the idle state for some time before moving to sleep mode if there is no user request in the idle state.

SM_1 : SM_1 is the shortest sleep mode of duration $71 \mu s$, which includes both the deactivation and reactivation times. The analog front-end and some of the digital baseband's components and power amplifier are turned off.

SM_2 : The deactivation and reactivation times are both included in the duration of SM_2 , which is 1 ms. Compared to the SM_1 , a few additional analog front-end components are disabled in the SM_2 .

SM_3 : The power amplifier, every component of the digital baseband, and practically every component of the analog front-end (aside from the clock generator) are all turned off during SM_3 's 10 ms of operation.

The ASMs are defined only based on the transition times of the BS's components. Since more and more components are deactivated when BS moves to deeper SM levels, which can result in greater energy savings, i.e., the energy consumption in the BS will be less. But deeper SM levels mean more time the BS will be deactivated. Hence, the delay will be more, which can affect the consumers' QoS. Hence, to maximize the QoS, the delay needs to be minimized as well; therefore, the three ASMs are considered in this model.

Consider the stochastic process $\{X(t), t \geq 0\}$, where $X(t)$ examines the state of a BS at any time t . The state space for this stochastic process is given by $S = \{\text{Idle}, \text{Active}, SM_1, SM_2, SM_3\}$. The system moves to the appropriate state based on the system's traffic. Fig. 1 demonstrates the state transition diagram for the proposed model. Let n be the highest possible user count for the service

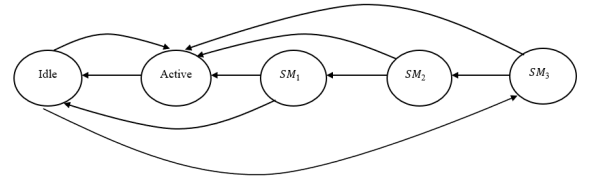


Fig. 1. State transition diagram for the proposed model

in communication with the BS. In a wireless communication system, assume an uplink and downlink communication between a serving BS and multiple mobile receivers [9].

The transmission power constraint $P_{\max}$ is not exceeded

while the power consumption is minimized according to the proposed model. The system resources are shared between all the users available for the service. Let k be the users available for service.

The following equation gives the service rate for i^{th} user over one transmission frame of period τ_{frame} :

$$R_i = \frac{w\tau}{\tau_{frame}} \log_2 \left(1 + \frac{P}{N_0 w} \right), i = 1, 2, \dots, k \quad (1)$$

where P is the power level based on the number of users, with bandwidth, w in MHz, time slot duration, τ in seconds, and noise spectral density, N_0 in W/Hz. The power consumption is rated at a minimum active consumption P_0 , a linear transmission power dependence factor ΔPM (the slope), and power consumption in different sleep modes P_{S_1} , P_{S_2} and P_{S_3} . P_{S_1} , P_{S_2} , and P_{S_3} are the power consumption in sleep mode 1, sleep mode 2, and sleep mode 3, respectively, i.e., if the system is in SM_1 , SM_2 and SM_3 , the power consumption is given by P_{S_1} , P_{S_2} and P_{S_3} . Any time slot t 's transmission power consumption is given by P_t . The figures given in [1] are used to determine the maximum power consumption of a BS operated with a single RF chain while assuming ΔPM and component efficiencies to be unchanged. Formally, the power consumption is $P_{consumption}(P_t) = P_0 + \Delta PM \times P_t, P_t > 0$. When P_t is zero, i.e., there is no transmission power consumption. Hence, the system is in one of the sleep modes.

$$\begin{aligned} & \text{minimize} \quad P_{consumption, frame}(r) = \\ & \frac{1}{T} \left(\sum_{t=1}^{T_{Active}} (P_0 + \Delta PM P_t) + \sum_{t=1}^{T_{S_1}} P_{S_1} + \right. \\ & \left. \sum_{t=1}^{T_{S_2}} P_{S_2} + \sum_{t=1}^{T_{S_3}} P_{S_3} \right) \\ & \text{subject to} \\ & P_t \leq P_{max}, \\ & T_{Active} + T_{S_1} + T_{S_2} + T_{S_3} = T, \\ & R_k \leq \frac{w\tau}{\tau_{frame}} \log_2 \left(1 + \frac{P_a}{N_0 w} \right) \end{aligned} \quad (2)$$

with the number of active transmission time slots T_{Active} , SM_1 time slots T_{S_1} , SM_2 time slots T_{S_2} and SM_3 time slots T_{S_3} , respectively. Let R be the target service rate promised to each user in one transmission frame of duration τ_{frame} in 5G networks; hence, $R \geq R_i, i = 1, 2, 3, \dots, k$. Have to minimize the power consumption, but QoS needs to be maximized, i.e., delay needs to be minimized in this model; hence, a tradeoff between the average power consumption, average delay, and service rate provided given the promised service rate is analyzed. The proposed model is examined using the Q-learning algorithm.

III. Q-LEARNING APPROACH

Using real-time interactions with the environment, distributed Q-learning is an online optimization technique that seeks to determine the best policy for an agent (e.g., BS). By

maximizing the expected value of the overall reward (Q-value) throughout all succeeding episodes, Q-learning is able to determine the best course of action. In order to govern multi-agent systems, or a system with BS that takes decisions (selects the right SM level) in an uncoordinated manner, distributed Q-learning is an online optimization technique [10]. Through real-time interactions with the environment, the BS must autonomously learn a policy (active, idle, SM_1, SM_2, SM_3). The action a_m^t is the mechanism by which the system transition between different states of the environment. In the sense that it maximizes the expected value of the overall reward (Q-value) over all succeeding episodes, Q-learning discovers the best policy. The agents, i.e., BSs, only have a partial view of the entire system, and because of the network's uneven distribution of users, their activities may vary. A BS's decision to select a SM level is mainly influenced by the number of users it must provide service and how patient these users are [11]. In Q-learning, an agent takes action a_m^t from an action set $\mathcal{A}$ while receiving a reward r_m^t , moves to a new state s_m^{t+1} . When the system is in state s_m^t , indicating the level of convenience of selecting action a_m^t , then Q-value is updated locally. The Q-value is updated as follows:

$$Q(s_m^t, a_m^t) \leftarrow Q(s_m^t, a_m^t) + l \left[r_m^t + \gamma \max_{a \in \mathcal{A}} Q(s_m^{t+1}, a_m^{t+1}) - Q(s_m^t, a_m^t) \right], \quad (3)$$

where l is the learning rate that symbolizes the convergence speed (i.e., it defines how quickly or slowly we will approach the optimal reward), $\gamma \in [0, 1]$ is the discount factor that establishes the current value of the future state reward, and t is the moment at which the action has been taken.

During the learning phase, the agent chooses the appropriate action based on the ϵ -greedy policy, choosing the action with probability $1 - \epsilon$ that corresponds to the highest Q value and choosing a random action with probability ϵ :

$$a_m^t(y) = \begin{cases} \underset{a \in \mathcal{A}}{\operatorname{argmax}} Q(s_m^t, a_m^t), & \text{if } y > \epsilon \\ \operatorname{rand}(\mathcal{A}) & \text{if } y \leq \epsilon \end{cases} \quad m \in S. \quad (4)$$

Here, y is the uniformly chosen random number between $[0, 1]$. The agent would have analyzed every option and avoided the local minima using the ϵ -greedy policy.

The transition between the various states of the BS, i.e., active, idle, or sleep mode, is defined as the set of possible actions $\mathcal{A}$, where (SM_1, SM_2 , or SM_3), i.e., the state to which the BS can switch is represented by the collection of possible actions $\mathcal{A}$. Based on the system's traffic, the BS chooses the most appropriate action to take. Thus, the state space is defined as $S = \{\text{Idle}, \text{Active}, SM_1, SM_2, SM_3\}$.

The goal is to find the best policy for each state along which the user moves to maximize the reward r_m^t at any time t and $m \in S$. Define the reward as the weighted sum of the average power consumption $P_{cons, m}$, delay D_m^t , and the service rate $R - R_k$, resulting from the sleep strategy chosen.

$$r_m^t = \alpha(R - R_k) - \beta D_m^t - (1 - \alpha - \beta) P_{cons, m}, m \in S, t \geq 0 \quad (5)$$

where $\alpha, \beta \in [0, 1]$ are the normalized weights that denote the importance given to the factors power consumption, delay, and rate. When the BS is in any of the sleep modes, then the service rate provided to users R_k is zero. Similarly, when the BS is active, then the delay is zero. Hence, equation (5) deduces to,

$$r_m^t(y) = \begin{cases} \alpha R - \beta D_m^t \\ -(1 - \alpha - \beta) P_{cons,m} & \text{if } s_m^t = Idle, SM_1, \\ & SM_2, SM_3. \\ \alpha(R - R_k) \\ -(1 - \alpha - \beta) P_{cons,m} & \text{if } s_m^t = Active. \end{cases} \quad (6)$$

Fig. 1 denotes the action state containing the transitions from one state to another. In this proposed model, the system can move from the active state to the idle state if there are no service requests, idle state to SM_3 , or active state based on the service requests. Then SM_3 to SM_2 and SM_2 to SM_1 if there is no service request and from SM_3 and SM_2 and SM_1 to active state, if there is a service request. This state transition diagram shows that the system is ergodic as it is a finite state model and irreducible. Hence, it is positive recurrent and aperiodic. Hence, this proves the convergence of this proposed model.

The delay D_m^t is a random variable defined by the difference between the total duration of the state and the time taken by the arrival of the first user request after the initiation of that state.

IV. NUMERICAL ILLUSTRATION

We assume low load traffic in this analysis to demonstrate the impact of SM levels on energy consumption. The arrival of the users following a log-normal distribution with mean λ and variance $\lambda/10$ is considered [10]. Consider a mean arrival rate of $\lambda = 1$ user/s/km², which corresponds to a load of about 1%, meaning that the BS is only used for 1% of the simulation, the rest of the time being spent inactive. A single BS is examined where the user requests are served following exponential distribution with mean 4. The end of the service corresponds to their departure from the active state. The ASMs can significantly impact EC and delay when there is a very low load. In Table I, the power values for the various BS states are provided. Table II provides the sleep periods of different states of the BS. Table III provides the simulation parameter values. Now, based on the sleep duration of different states, D_m^t in μs is defined as:

$$D_m^t = \begin{cases} 0 & \text{if } m = \text{Active} \\ 50 - t_{Idle} & \text{if } m = \text{Idle} \\ 71 - t_{SM_1} & \text{if } m = SM_1 \\ 1000 - t_{SM_2} & \text{if } m = SM_2 \\ 10000 - t_{SM_3} & \text{if } m = SM_3 \end{cases} \quad (7)$$

where t_m denotes the time taken by the arrival of the first user request after the initiation of the state m , $m \in S$.

TABLE I
POWER CONSUMPTION IN DIFFERENT STATES

State	Active	Idle	SM_1	SM_2	SM_3
Power Consumption	250 W	109W	52.3W	14.3W	9.51 W

TABLE II
BASE STATION SLEEP MODE CHARACTERISTICS

State	Idle	SM_1	SM_2	SM_3
Sleep Duration	50 μs	71 μs	1 ms	10 ms

TABLE III
SIMULATION PARAMETERS

Variable	Value
w	20 MHz
$\frac{\tau}{T_{frame}}$	10ms/1ms
T	10
$P_{0,MT}$	185W/260W
Δ_{PM}	4.7
P_{max}	46 dBm
N_0	4×10^{-21} W/Hz

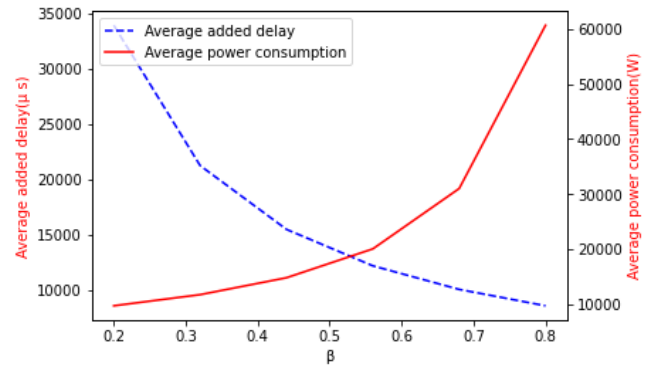


Fig. 2. Tradeoff between average added delay and power consumption varying values of β

Fig. 2 shows the trade-off between the average added delay and the average power consumption varying the various values of β with $\alpha = 0.1$. With the increase in values of β , the average delay decreases whereas power consumption increases for the optimal value of reward earned. The value of both the parameters, i.e., average delay and average power consumption is optimized for $\beta \in (0.5, 0.6)$.

Fig. 3 shows the trade-off between the average service rate provided to the users and the average power consumption varying the various values of α with $\beta = 0.1$. With the increase in values of α , the average service rate provided decreases whereas power consumption increases for the optimal value of reward earned. The value of both the parameters, i.e., average service rate provided and average

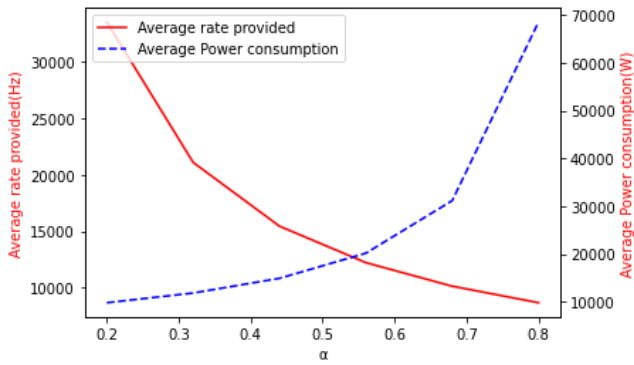


Fig. 3. Variation between rate and power consumption varying values of α

power consumption is optimized for $\alpha \in (0.5, 0.6)$.

V. CONCLUSION

In contrast to models that only use one type of SM, the suggested model uses ASMs to investigate the trade-off between the average power consumption, service rate, and average delay (i.e., the amount of time that the user requests must wait in the buffer until the BS reactivates). The reinforcement learning method known as the Q-learning algorithm is used to examine the presented trade-off. An unbalanced traffic cellular network where each BS chooses its optimal SM level to balance QoS and power consumption is considered. For a beneficial trade-off, optimal values of α and β are obtained.

ACKNOWLEDGMENTS

One of the authors (Dharmaraja Selvamuthu) thanks Bharti Airtel Limited, India, for financial support in this research work.

REFERENCES

- [1] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. A. Imran, D. Sabella, M. J. Gonzalez, O. Blume *et al.*, "How much energy is needed to run a wireless network?" *IEEE wireless communications*, vol. 18, no. 5, pp. 40–49, 2011.
- [2] M. Feng, S. Mao, and T. Jiang, "Base station on-off switching in 5g wireless networks: Approaches and challenges," *IEEE Wireless Communications*, vol. 24, no. 4, pp. 46–54, 2017.
- [3] 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, 2011.
- [4] Y.-F. Huang, T.-H. Tan, N.-C. Wang, Y.-L. Chen, and Y.-L. Li, "Resource allocation for d2d communications with a novel distributed q-learning algorithm in heterogeneous networks," in *2018 International Conference on Machine Learning and Cybernetics (ICMLC)*, vol. 2. IEEE, 2018, pp. 533–537.
- [5] K. Guo, C. Yang, and T. Liu, "Caching in base station with recommendation via q-learning," in *2017 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2017, pp. 1–6.
- [6] A. Galindo-Serrano and L. Giupponi, "Distributed q-learning for aggregated interference control in cognitive radio networks," *IEEE Transactions on Vehicular Technology*, vol. 59, no. 4, pp. 1823–1834, 2010.
- [7] M. Miozzo, L. Giupponi, M. Rossi, and P. Dini, "Distributed q-learning for energy harvesting heterogeneous networks," in *2015 IEEE International Conference on Communication Workshop (ICCW)*. IEEE, 2015, pp. 2006–2011.
- [8] T. Kudo and T. Ohtsuki, "Cell range expansion using distributed q-learning in heterogeneous networks," *Eurasip journal on wireless communications and networking*, vol. 2013, pp. 1–10, 2013.
- [9] H. Holtkamp, G. Auer, S. Bazzi, and H. Haas, "Minimizing base station power consumption," *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 2, pp. 297–306, 2013.
- [10] A. El-Amine, M. Iturralde, H. A. H. Hassan, and L. Nuaymi, "A distributed q-learning approach for adaptive sleep modes in 5g networks," in *2019 IEEE Wireless Communications and Networking Conference (WCNC)*. IEEE, 2019, pp. 1–6.
- [11] F. E. Salem, Z. Altman, A. Gati, T. Chahed, and E. Altman, "Reinforcement learning approach for advanced sleep modes management in 5g networks," in *2018 IEEE 88th Vehicular Technology Conference (VTC-Fall)*. IEEE, 2018, pp. 1–5.