Location-Aware Sleep Strategy for Energy-Delay Tradeoffs in 5G with Reinforcement Learning

Ali El Amine, Hussein Al Haj Hassan, Mauricio Iturralde, Loutfi Nuaymi

Location-Aware Sleep Strategy for Energy-Delay Tradeoffs in 5G with Reinforcement Learning

Ali El-Amine*, Hussein Al Haj Hassan†, Mauricio Iturralde* and Loutfi Nuaymi*
*IMT Atlantique, IRISA, UMR CNRS 6074, F-35700 Rennes, France
†American University of Science and Technology, Beirut, Lebanon
e-mail: *name.lastname@imt-atlantique.fr, †hhajhassan@aust.edu.lb

***Abstract*—In this paper, we propose a sleep strategy for energy-efficient 5G Base Stations (BSs) with multiple Sleep Mode (SM) levels to bring down energy consumption. Such management of energy savings is coupled with managing the Quality of Service (QoS) resulting from waking up sleeping BSs. As a result, a tradeoff exists between energy savings and delay. Unlike prior work that studies this problem for binary state BS (ON and OFF), this work focuses on multi-level SM environment, where the BS can switch to several SM levels. We propose a Q-Learning algorithm that controls the state of the BS depending on the geographical location and moving velocity of neighboring users in order to learn the best policy that maximizes the tradeoff between energy savings and delay. We evaluate the performance of our proposed algorithm with an online suboptimal algorithm that we introduce as well. Results show that the Q-Learning algorithm performs better with energy savings up to 92% as well as better delay performance than the heuristic scheme.**

***Index Terms*—Energy consumption, service delay, 5G, sleep modes, LTE positioning, Q-learning**

I. Introduction

In order to support future traffic requirements (high-speed mobile data services and better coverage), the next generation cellular networks are expected to deploy denser mobile access networks with different Base Stations (BSs) sizes (large, small and pico). Although this improves the capacity and coverage of the network, it also brings new challenges such as pushing the limits of energy consumption and CO_2 emissions. In order to meet these challenges, green cellular networks with improved energy efficiency have become a key factor in future 5G cellular networks. Since more than 50% of the network power is consumed by the Radio Access Network (RAN) and in particular the BS [1], it has become a high-priority objective for network management to reduce the energy consumption of this part of the network. Under these circumstances, BS sleep strategies are drawing increasing attentions recently [2], [3].

The introduction of BS Sleep Mode (SM) is consistent with the objective of reducing the energy consumption of the network. For instance in [4], the authors studied the energy saving problem by switching-off some macro BSs under coverage constraints. Using tools from stochastic geometry to model the location of the BSs, the authors achieved an energy efficiency gain up to 60% compared to when all the BSs are active. BS sleep strategies also play an important role in renewable energy equipped BSs to better manage the energy harvested and the energy stored in the battery [5], [6], [7]. However, such management of energy saving has to be carefully coupled with managing the Quality of Service (QoS) to ensure the user satisfaction since it can bring additional delay. In [8], the authors studied the tradeoff between energy savings and delay for different wake-up schemes. Similarly, in [9] this tradeoff is studied in a renewable energy environment where the BSs switch to sleep mode in a cooperative manner to further reduce the grid energy consumption.

In contrast to the above binary sleep scheme models (ON and OFF), [10] proposed a multi-level sleep model for future 5G BSs. The model categorizes the power consumption for different types of BSs (e.g., Large, small, signal, data). This model provides different stages of SM levels. Each stage is characterized by an activation/deactivation period and a power consumption. In 4G networks (LTE), implementing these SM levels is challenging due to backwards compatibility and reference signaling requirements. However, in 5G networks the Cell Reference Signals (CRS) are removed [11] making it possible for a BS to explore these different SM levels paving the road to the ultimate goal of achieving almost zero power consumption at zero load. The work in [12], [13] explored the potential of these SM levels. Using Reinforcement Learning (RL), the authors achieved an energy saving gain up to 90% in a low load traffic.

On the other hand, Location Based Services (LBS) are increasing in both commercial applications and emergency services. LBS involves the process of determining the geographical position of a device, such as a mobile phone. Due to the increase demands on positioning, the Federal Communications Commission (FCC) set several location accuracy and reliability requirements, especially for emergency services [14]. Unlike previous radio-access standards, LTE incorporates positioning capabilities to support higher level of application-adaptive requirements in order to meet the requested positioning QoS. LTE adopts three independent positioning techniques: Assisted Global Navigation Satellite Systems (A-GNSS), Enhanced Cell ID (E-CID) and Observed Time Difference of Arrival (OTDOA) [15].

In this paper, we study the joint problem of energy savings and delay in a multi-level SM environment, where the BS can switch to several SM levels to save energy while maintaining the QoS of the users. We propose a methodology that allows the operator to freely manage the tradeoff between energy

consumption and service delay according to the different requirements of the 5G use cases, such as Ultra-Reliable Low Latency Communications (URLLC). Different from the above mentioned work, in this paper we propose a Q-Learning-based algorithm that controls the state of the BS depending on the geographical location of the user, and his/her moving velocity in order to learn the best policy that maximizes the tradeoff between energy savings and delay. In order to evaluate the performance of the Q-Learning algorithm, we propose a heuristic policy that maximizes the energy savings while minimizing the delay of the network. We show that Q-Learning outperforms the heuristic scheme. To the best of our knowledge, the joint study of energy-delay and device positioning has not been well investigated in the literature. The closest work to ours can be found in [16] where the authors studied the state of small cells depending on the position of the users. However, the work focuses on maximizing the energy efficiency and is based on an offline algorithm using tools from stochastic geometry that cannot be implemented in an online manner.

This paper is organized as follows. In Section II, we detail the system model along with the 5G sleep mode model. In Section III, we start by presenting an overview of the positioning techniques for LTE before proposing our Q-Learning and heuristic algorithms. Finally, we present the simulation results in Section IV before concluding in Section V.

II. System model

We consider a 5G network composed of M gNBs serving k users. A gNB is a part of the Next Generation Radion Access Network (NG-RAN) that is part of the 3GPP 5G NextGen System [17]. Each user requests a real-time service, e.g., VoIP call. We further assume the BS has active mode and different SM levels with different activation times. When a user requests a service from a gNB in SM, it triggers the activation mode and the user is buffered until the BS wakes up. This wake-up delay could have an impact on the latency added to the system. The deeper the SM is, the more time the user will have to wait until the gNB reactivates.

A. Sleep modes in 5G networks

In [10], GreenTouch Project identified four distinct SM levels by grouping sub-components with similar transition latency when being activated or deactivated. The presented model enables to quantify the power consumption of the BS in each of the four SMs. These are:

- SM 1: It considers the shortest time unit of one OFDM symbol (i.e. $71\mu s$) comprising both deactivation and reactivation times. In this mode only the power amplifier and some processing components are deactivated.
- SM 2: It corresponds to the case of sub-frame or Transmission Time Interval (TTI) (i.e. 1 ms). In this SM, more components enter the sleep state.
- SM 3: It corresponds to the frame unit of 10 ms. Most of the components are deactivated in this mode.
- SM 4: This is the deepest sleep level. Its unit corresponds to the whole radio frame of 1s. It is the standby mode where the BS is out of operation but retains wake-up functionality.

Higher energy savings can be achieved when switching BSs to a deeper SM, since more components will be deactivated. However, this will be associated with longer transition latency which may impact the QoS for the users. In Table I, we present the SM levels characteristics.

Along with SM and users' dynamics, the BS has to wake up periodically to send signaling bursts of Synchronization Signaling (SS) and Physical Broadcast CHannel (PBCH). It has been agreed in Third Generation Partnership Project (3GPP) [11] that the transmission periodicity of the SS/PBCH block can be set to any value among [5, 10, 20, 40, 80, 160 ms]. With these values, SM 4 cannot be used. Hence, we limit our work to the first three SM levels.

TABLE I: BS Sleep Modes Characteristics[10].

Sleep level	Deactivation duration	Activation duration	Power consumption
SM 1	35.5 μs	35.5 μs	48%
SM 2	0.5 ms	0.5 ms	13%
SM 3	5 ms	5 ms	9%
SM 4	0.5 s	0.5 s	7.5%

B. Dynamic user model

We consider two cases for users service requests from a BS:

1) Camping in a gNB: in this case, the user is in any of the three 5G Radio Resource Control (RRC) modes: connected, idle or inactive [17].
2) Handover arrival: in this case, a user already performing a service transmission/reception from a BS is moving towards the coverage of a neighboring BS. The user is handed over the other BS for service continuation.

In the first case, when a service is being called for the first time, the delay impact on the user is tolerable since the transmission has not yet started. However, in the case of a handover (case 2), the service might be interrupted or degraded if the BS hosting the user is inactive (i.e., in sleep mode). This might have a severe impact on the QoS of the transmission. It is important to note that if the requested service is of type non real-time (e.g., web browsing), then this delay will have a less impact on the QoS since non real-time services are delay tolerant.

In this work, we consider the case of real-time services, and in particular the handovers of users between BSs. The objective is to jointly minimize the delay associated with handovers, and the energy consumption of the network. As shown in Fig. 1, we consider a user being served by BS A moving towards BS B with a velocity v (Km/h). Since BS B is not serving any users, it decides to switch to a specific SM level. This decision depends on the position of the user from BS B, i.e., the closer the user moves towards BS B, the lighter its SM level becomes. Since positioning with LTE is not always accurate and is prone to measurement errors, we divide the geographic region of interest (the region of neighboring sites) into several

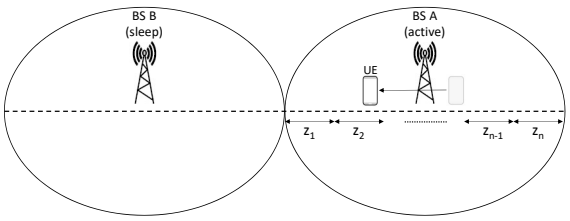

Fig. 1: The position of the user moving towards a neighboring gNB can be measured by the LTE positioning system.

zones ($z_1, z_2, \ldots, z_n$). Each zone corresponds to a region that can be detected by the BS. For example, if a user is in zone j (z_j), the BS senses the presence of the user in z_j, but it does not know its exact location. The size of these regions depends on the accuracy of the positioning method used, i.e., the smaller the zone size, the more accurate the user position is measured.

III. Location-aware sleep mode strategy

A. Positioning methods with LTE

In wireless networks, positioning is very challenging due to the user mobility and the dynamic nature of both the environment and radio signals. On the other hand, the FCC set some stringent requirements on the accuracy of positioning especially for emergency services [14]. As a result, 3GPP described an architecture, standards and methods for LTE to support different positioning techniques that reach a high level of accuracy as requested by the regulations.

LTE supports three positioning techniques: A-GNSS, E-CID, and OTDOA.

1) A-GNSS: To overcome the line of sight and low signal level drawbacks problems of GNSS, the cellular network assists the GNSS receiver to increase its positioning operation. A-GNSS works best in outside conditions with clear view of the sky (line of sight).
2) E-CID: This method is based on Cell of Origin (COO) that estimates the position of the device in the geographical area of its serving BS. Since this method is not accurate enough (linked to the cell size), E-CID performs measurements on radio signals such as Reference Signal Received Power (RSRP) and Time Different Of Arrival (TDOA) to measure the Round Trip Time (RTT) and Angle-of-Arrival (AoA) for more accuracy. With E-CID, it is also possible to have information on the direction of the device.
3) OTDOA: According to the LTE specification [18], this method relies on the downlink radio signals from multiple BSs to compute the user position. In [18], LTE standard specifies Positioning Reference Signal (PRS) that is dedicated for positioning purposes. By properly configuring the PRS and with interference management techniques, the error in position can be significantly improved to 1 meter, however, at the expense of decreasing the spectral efficiency [15].

Each of the above proposed technique has its advantages and limitations. To improve positioning in challenging radio environment, hybrid positioning using the different above methods is also supported in Release 10 [19]. In Table II, we summarize the differences between these methods. For more details on LTE positioning, the reader is referred to [20].

TABLE II: Comparison of LTE positioning techniques.

Technique	Accuracy	Limitations
A-GNSS	High (10-15 m)	Requires line of sight
E-CID	Low (80-800 m)	Variable accuracy depending on the environment
OTDOA	Medium-High (10-40 m but can go down to 1 m at the expense of SE)	Requires synchronized network and operator dependency

B. Sleep mode policy: Q-Learning approach

Distributed Q-Learning is an online optimization technique that aims at finding the best policy of an agent (e.g., BS) via real-time interactions with the environment. Q-Learning obtains the optimal policy by maximizing the expected value of the total reward (Q-value) over all successive episodes.

In Q-learning, an agent m takes an action a_m^t from an action set $\mathcal{A}$, then moves to a new state s_m^{t+1} while receiving a reward r_m^t. The Q-value is then updated locally indicating the level of convenience of selecting action a_m^t when in state s_m^t. The Q-value is updated as follows:

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

where α is the learning rate that represents the speed of convergence, $\gamma \in [0, 1]$ is the discount factor that determines the current value of the future state costs, and t is the time at which the action has been taken.

During the learning phase, the agent selects the corresponding action based on the ϵ-greedy policy, i.e., it selects with probability $1 - \epsilon$ the action associated with the maximum Q-value, and with probability ϵ selects a random action:

$$a_m^t = \begin{cases} \underset{a \in \mathcal{A}}{\operatorname{argmax}}\, Q(s_m^t, a_m^t), & \text{if } y > \epsilon \\ rand(\mathcal{A}), & \text{otherwise} \end{cases} \quad m = 1, \ldots, M. \quad (2)$$

By implementing the ϵ-greedy policy, the agent would have explored all possible actions and avoided local minima. For more details on RL and Q-learning the reader is referred to [21].

In this work, we define the set of possible actions $\mathcal{A}$ the state of the BS, i.e., active or in sleep mode (SM_1, SM_2 or SM_3). The BS chooses the appropriate action based on the geographical zone the user lies in. Thus, we define the state space and action space as follows:

$$\mathcal{S} = \{z_1, z_2, \ldots, z_n\}$$

$$\mathcal{A} = \{\mathrm{Active}, \mathrm{SM}_1, \mathrm{SM}_2, \mathrm{SM}_3\}$$

The number of zones depends on the accuracy of the positioning method used, and it spans the area of the closest neighboring BSs.

An episode starts when a user asks for a service, moves towards a neighboring cell, and finishes when the handover is complete. During each episode, the neighboring BS that the user is approaching to, changes its state by taking an action following Eq. (2). Then, it stores a quality-value linking the states $s \in \mathcal{S}$ to the chosen action $a_m \in \mathcal{A}$ following Eq. (1). The optimal policy consists of choosing the best sleep mode level having the highest Q-value.

The goal is to find the best policy for each state (zone) along which the user is moving in order to maximize the reward r_m^t. We define the reward as the weighted-sum of the energy gain G and the added delay D, both resulting from the sleep mode level chosen during an episode.

$$r = (1 - \eta)G - \eta D \tag{3}$$

where $\eta \in [0, 1]$ is a parameter which controls the trade-off between energy gain and the delay performance. We note that in the delay-tolerant services case (e.g., web browsing), $\eta \to 0$, thus emphasizing on saving energy. In the other case where the service is delay-sensitive (e.g., VoIP), $\eta \to 1$. We note that the weight parameter (η) is freely chosen by the operator depending on the 5G use cases.

Algorithm 1 : Q-Learning Algorithm

Initialize $q(s, a) = 0$, $\forall s \in \mathcal{S}$ and $\forall a \in \mathcal{A}$.
Set the weight η, and the average user velocity v.
procedure Training ($Q(s, a)$)
while Learning **do**
    Visit state s.
    Select an action a using ϵ-greedy rule in (2).
    Receive a reward r.
    Observe next state s'.
    Update the Q-value $q(s, a)$ from (1).
end while
end procedure

procedure Online
From $Q(s, a)$, store best action in Q-table $\forall s \in \mathcal{S}$ and $\forall a \in \mathcal{A}$.
Run Q-Learning.
end procedure

In details, the training phase consists of running the BS (agent) with different instances of user arrivals in an offline fashion. After this step is completed, the system goes online. During this phase, a Q-table stores the best policy using the trained Q-values of the previous offline step. The algorithm is described in Algorithm 1.

C. Delay-Sensitive Sleep Mode (DS-SM) algorithm: Heuristic approach

In contrast to the Q-Learning-based algorithm that includes an offline training phase with a high degree of system information knowledge (e.g., exact user position, d, and velocity, v), we propose a Delay-Sensitive Sleep Mode (DS-SM) online algorithm that takes decisions on which sleep mode level to switch to based on the estimated measured position ($\tilde{d}$) and velocity ($\tilde{v}$). We use this algorithm as a benchmark to study the performance of the Q-Learning algorithm described in the previous section.

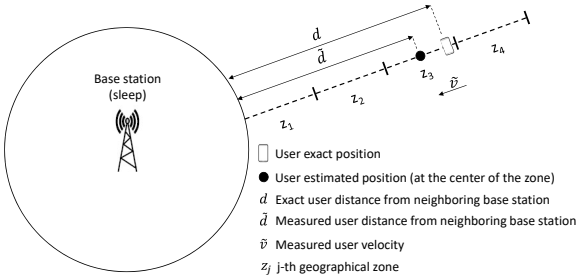

Fig. 2: User positioning measurement following DS-SM algorithm.

When a user reaches a geographical zone (z_j), the estimated distance from the neighboring cell ($\tilde{d}$) is calculated from the center of the zone it is located, as shown in Fig. 2. The estimated velocity on the other hand is related to the actual velocity by the following expression: $\tilde{v} = v + \alpha$, where α is the error in the measurement. The BS takes the decision to switch to SM_i in a geographic zone (z_j) if after the sleep duration T_{SM_i} has elapsed, the user did not enter the neighboring cell, thus minimizing the delay. If the user enters the neighboring cell while it is inactive, his/her QoS will degrade due to the added delay resulting from the waking up time from SM_i level. In order to maximize the energy savings, the algorithm starts with the deepest sleep mode level allowed (i.e., SM_3). The algorithm is summarized in Algorithm 2.

Algorithm 2 : Delay-Sensitive Sleep Mode (DS-SM) Algorithm

Measure user location $\tilde{d}$ and velocity $\tilde{v}$.
Repeat until handover is complete.
for $i = 3..1$ **do**
    if $T_{SM_i} \times \tilde{v} < \tilde{d}$ **then**
        Switch the BS to SM_i.
        Update $\tilde{d}$ after T_{SM_i} is elapsed.
        Set $i = 3$.
    else
        if $i = 1$ **then**
            Switch the BS to active mode until handover is complete.
        end if
    end if
end for
Output: Energy consumption and network added delay.

IV. Simulation results

A. Simulation parameters

In this section, we demonstrate the performance of the proposed approach for deciding SM levels for neighboring BSs. We consider a low traffic period where a BS serves a user with a continuous real-time service. We focus on the delay as the performance metric for QoS. The reason is twofold. First, we consider a low traffic period. Thus, bandwidth is not a problem. Second, we focus on real-time services, such as VoIP call, that do not require high throughput, but are delay sensitive. The user is randomly generated in a geographical zone and moving towards a neighboring BS with a constant velocity. We further consider that the inter-site distance is $800m$.

In order to point the impact of the different SM levels on the energy savings, we consider a low load traffic. From [22], we found the power figures for the different states of the BS. Then we define the sleep duration times for each SM level as: $T_{SM_1} = 0.5s$, $T_{SM_2} = 10s$ and $T_{SM_3} = 30s$. The reason behind these values is to minimize the measurements done by the BS for user positioning. Table III summarizes these values.

TABLE III: Power Consumption of a 2×2 MIMO BS.

State	SM 1	SM 2	SM 3	Idle
Power consumption (W)	52.3	14.3	9.51	109
Sleep duration (T_{SM_i}) (s)	0.5	10	30	-

B. Convergence analysis

First, we analyze the convergence of the proposed Q-Learning algorithm for $\eta = 0.5$. In Fig. 3, we evaluate the impact of the system positioning accuracy on the performance of the Q-Learning training phase. It is clear to observe that as the system accuracy decreases (e.g., from $10m$ to $80m$), the learning delay decreases as well. For instance, after 50 episodes the Q-Learning algorithm converges for a positioning accuracy of $80m$, whereas, 400 episodes are required for a sharper accuracy of $10m$. This is related to the number of states that increases with higher accuracy. Thus, requiring more time to converge. Once the algorithm converges, we stop the training phase and exploit the obtained policies. Then, we store the best policy in a look-up table that will be used during the exploitation phase.

C. Energy consumption and service delay trade-off

In Fig. 4, we present the tradeoff between the average network energy consumption and added delay for different values of η and with a positioning accuracy of $10m$. In other words, the difference between two geographical zones is $10m$. We normalize the energy consumption with the case when the BS is always in idle mode. First, we observe that significant energy savings up to 92% (corresponding to the normalized energy consumption of 0.098) can be achieved using the multi-level sleep modes. The proposed Q-Learning algorithm allows a wide range of control over the energy consumption. For instance, with $\eta = 0$, the energy consumption is the lowest. However, this is at the expense of added service delay that is at its highest ($5ms$). This is because for this case, the SM policy is chosen solely considering energy savings. The resulting delay corresponds to the waking up time when the BS is in the deepest sleep state (SM_3). For $\eta = 1$, we observe that the energy consumption is higher, but the delay is zero. So when the user arrives to the neighboring cell, the BS is already active. It is then important to carefully choose η in order to satisfy the requirements of the different 5G use cases.

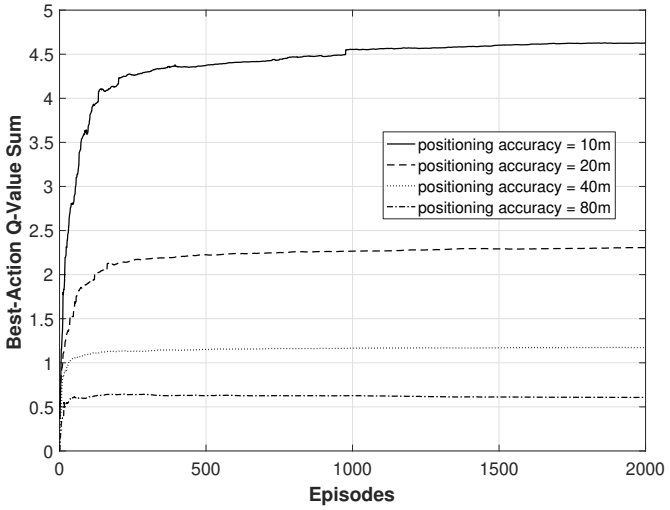

Fig. 3: Convergence of the Q-Learning algorithm for $\eta = 0.5$.

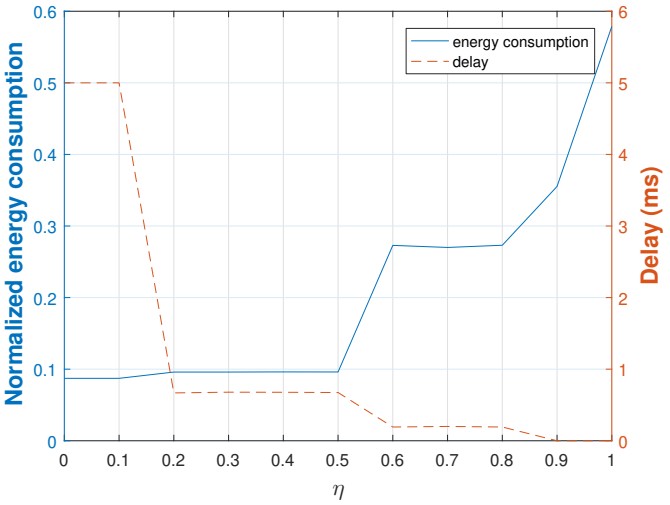

Fig. 4: Performance assessment of the selected policies during the exploitation phase.

D. Performance evaluation

As a benchmark for comparison, we consider the DS-SM online scheme presented in Section III-C. This algorithm does not depend on η. However, its goal is to minimize the service delay while reducing the energy consumption. Hence, the corresponding η in the Q-Learning algorithm can be found for the case where the delay is minimized. Since DS-SM also takes into account the energy consumption, we look for the smallest value of η that minimizes the delay. For example in

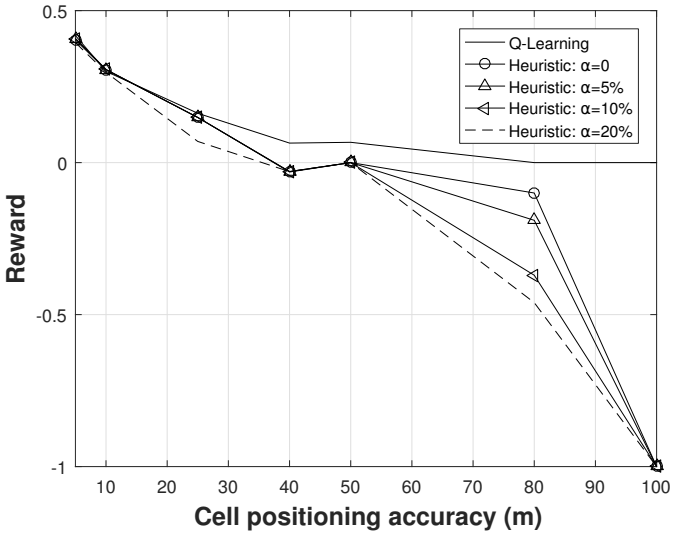

Fig. 5: Performance comparison between Q-Learning and DS-SM algorithms with different user velocity measurement accuracies.

Fig. 4, this corresponds to $\eta = 0.9$. We note that the value of η for the designated condition changes for different user velocity and position accuracy. In the following, we compare the reward expression in Eq. (3) that takes into account both energy consumption and service delay. For a fair comparison, this weighted parameter will be also used when computing the reward resulting from the DS-SM algorithm the same way it is computed for the Q-Learning algorithm. We note that this reward can be regarded as a utility function of η.

From Fig. 5, we show that the Q-Learning algorithm leads to a higher reward than DS-SM and in particular for low positioning accuracy. The poor performance in DS-SM results from choosing the policy based on the inaccurate positioning techniques (always in the center of the zone), since it does not include a training phase to tune these decisions to optimize the reward. For low position accuracy (e.g., $90m$), the gap between the user's real and measured positions is increased. This results in a performance degradation for both algorithms. However, DS-SM adds high delay when $R(\eta) < 0$.

V. Conclusion

In this work, we investigated the tradeoff between energy consumption and service delay associated with sleep strategies in 5G networks. While BS sleeping significantly reduce the energy consumption of the BS (up to 92%), it is coupled with QoS degradation by bringing additional delay to the users. We proposed a methodology for reducing the energy consumption of the BS while ensuring a good QoS. This methodology also permits the operator to freely manage the tradeoff between energy consumption and service delay. This objective was achieved by switching the neighboring BS to different SM levels depending on the location of the user in the network, and his/her moving velocity towards another cell. In order to choose the best SM policy, we proposed a Q-Learning algorithm. Compared with the heuristic DS-SM algorithm, the proposed Q-Learning algorithm outperforms the benchmark scheme in terms of bringing more energy savings to the network while maintaining a good QoS, and it allows the operator to manage the tradeoff factor (η) according to the 5G use cases.

References

[1] L. Suarez, L. Nuaymi, and J. Bonnin. An overview and classification of research approaches in green wireless networks. *EURASIP Journal on Wireless Communications and Networking*, 2012.

[2] F. Han, S. Zhao, L. Zhang, and J. Wu. Survey of strategies for switching off base stations in heterogeneous networks for greener 5G systems. *IEEE Access*, 2016.

[3] M. Feng, S. Mao, and Tao. Base station ON-OFF switching in 5G wireless networks: Approaches and challenges. *IEEE Wireless Communications*, Aug. 2017.

[4] J. Peng, P. Hong, and K. Xue. Stochastic analysis of optimal base station energy saving in cellular networks with sleep mode. *IEEE Communications Letters*, Apr. 2014.

[5] A. El-Amine, H. A. H. Hassan, and L. Nuaymi. Analysis of energy and cost savings in hybrid base stations power configurations. In *2018 IEEE 87th Vehicular Technology Conference (VTC Spring)*, June 2018.

[6] A. El-Amine, H. A. H. Hassan, and L. Nuaymi. Services kpi-based energy management strategies for green wireless networks. In *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2018.

[7] A. El-Amine, H. A. H. Hassan, and L. Nuaymi. Battery aging-aware green cellular networks with hybrid energy supplies. In *2018 IEEE 29th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Sep. 2018.

[8] X. Guo, Z. Niu, S. Zhou, and P. R. Kumar. Delay-constrained energy-optimal base station sleeping control. *IEEE Journal on Selected Areas in Communications*, May 2016.

[9] V. Chamola, B. Sikdar, and B. Krishnamachari. Delay aware resource management for grid energy savings in green cellular base stations with hybrid power supplies. *IEEE Transactions on Communications*, Mar. 2017.

[10] B. Debaillie, C. Desset, and F. Louagie. A flexible and future-proof power model for cellular base stations. In *2015 IEEE 81st Vehicular Technology Conference (VTC Spring)*, May 2015.

[11] 3GPP TS 38.331. 5G; NR; Radio Resource Control (RRC); Protocol specification (Rel. 15). 2017.

[12] Fatma Ezzahra Salem et al. Reinforcement learning approach for advanced sleep modes management in 5G networks. In *2018 IEEE Vehicular Technology Conference (VTC-Fall)*, July 2018.

[13] 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 WCNC 2019)*, Apr. 2019.

[14] Edwin L. Baker. Wireless enhanced 911 working group: report of proceedings. honolulu, HI: Legislative reference bureau. Jan 2004.

[15] J. A. del Peral-Rosado, J. A. López-Salcedo, G. Seco-Granados, F. Zanier, and M. Crisci. Achievable localization accuracy of the positioning reference signal of 3gpp lte. In *2012 International Conference on Localization and GNSS*, June 2012.

[16] C. Liu, B. Natarajan, and H. Xia. Small cell base station sleep strategies for energy efficiency. *IEEE Transactions on Vehicular Technology*, 65, Mar. 2016.

[17] 3GPP TS 23.501. System architecture for the 5G system (Rel. 15). Sep. 2018.

[18] 3GPP TS 36.305. Stage 2 functional specification of User Equipment (UE) positioning in E-UTRAN (Rel. 9), std. Jan 2011.

[19] 3GPP TS 136 305. LTE; Evolved Universal Terrestrial Radio Access Network (E-UTRAN); Stage 2 functional specification of User Equipment (UE) positioning in E-UTRAN (Release 10). Jan. 2011.

[20] Positioning with LTE, Ericsson White Paper. Sep. 2011.

[21] Richard S. Sutton and Andrew G. Barto. *Introduction to Reinforcement Learning*. MIT Press, Cambridge, MA, USA, 1st edition, 1998.

[22] IMEC power model tool. [Online]. Available: https://www.imec-int.com/powermodel.