

RESEARCH ARTICLE

Using Reinforcement Learning to Reduce Energy Consumption of Ultra-Dense Networks With 5G Use Cases Requirements

SILVESTRE MALTA^{1,4}, PEDRO PINTO^{1,2,3}, (Member, IEEE),
AND MANUEL FERNÁNDEZ-VEIGA⁴, (Senior Member, IEEE)

¹ Applied Digital Transformation Laboratory (ADiT-LAB), Instituto Politécnico de Viana do Castelo, 4900-347 Viana do Castelo, Portugal

² Institute for Systems and Computer Engineering, Technology and Science (INESC TEC), 4200-465 Porto, Portugal

³ Departamento de Ciências da Comunicação e Tecnologias da Informação, Universidade da Maia, 4475-690 Maia, Portugal

⁴ AtlanTTic Research Center, University of Vigo, 36310 Vigo, Spain

Corresponding author: Silvestre Malta (smalta@estg.ipvc.pt)

This work was supported by the Spanish Government through “Enhancing Communication Protocols with Machine Learning while Protecting Sensitive Data (COMPROMISE)” funded by MCIN/AEI/10.13039/501100011033 under Grant PID2020-113795RB-C33.


ABSTRACT In mobile networks, 5G Ultra-Dense Networks (UDNs) have emerged as they effectively increase the network capacity due to cell splitting and densification. A Base Station (BS) is a fixed transceiver that is the main communication point for one or more wireless mobile client devices. As UDNs are densely deployed, the number of BSs and communication links is dense, raising concerns about resource management with regard to energy efficiency, since BSs consume much of the total cost of energy in a cellular network. It is expected that 6G next-generation mobile networks will include technologies such as artificial intelligence as a service and focus on energy efficiency. Using machine learning it is possible to optimize energy consumption with cognitive management of dormant, inactive and active states of network elements. Reinforcement learning enables policies that allow sleep mode techniques to gradually deactivate or activate components of BSs and decrease BS energy consumption. In this work, a sleep mode management based on State Action Reward State Action (SARSA) is proposed, which allows the use of specific metrics to find the best tradeoff between energy reduction and Quality of Service (QoS) constraints. The results of the simulations show that, depending on the target of the 5G use case, in low traffic load scenarios and when a reduction in energy consumption is preferred over QoS, it is possible to achieve energy savings up to 80% with 50 ms latency, 75% with 20 ms and 10 ms latencies and 20% with 1 ms latency. If the QoS is preferred, then the energy savings reach a maximum of 5% with minimal impact in terms of latency.

INDEX TERMS 5G, energy efficiency, sleep mode, reinforcement learning.

I. INTRODUCTION

Telecom as an industry is continuously growing to meet customers' requirements. Mobile applications generate a high traffic volume with multiple connections and a high throughput density with QoS concerns. According to [1], UDNs have been considered one of the advanced technologies in 5th Generation (5G) and could be the key to meeting user expectations. Compared to existing non-dense deployment in

heterogeneous networks based on the Long-Term Evolution (LTE) standard, with UDNs, small, femto and pico cells can be densely deployed by network operators and or even by users. In UDNs, the access nodes and/or the number of communication links per unit area are dense. To consider a network as ultra-dense, two measures can be considered, the number of cells and the number of users. The density of a network is mainly defined by the number of cells or users in a given area. Quantitatively, the definition of UDN varies from literature to literature, according to [1] there must exist more than 10^3 cells/ km^2 , or more than 600 active users/ km^2

The associate editor coordinating the review of this manuscript and approving it for publication was Adamu Murtala Zungeru .

to consider a network as UDN. An UDN is defined in [2], as a network in which the density of the BSs or access points is potentially as high as or even higher than the density of the users. In [3], an UDN is characterized by the fact that the distance between the sites in the network is not greater than a few meters. As indicated in [4] the key patterns of a UDN are the User End (UE) density and the UE mobility. By 2027, 5G networks are expected to carry 62% of the total mobile data traffic [5]. According to [6], between 2020 and 2030, the compound annual growth rate will increase by 55% per year, reaching 607 EBytes in 2025 and 5.016 EBytes in 2030.

The implementation of 5G wireless networks must provide uninterrupted service to connected clients with strict QoS requirements but, on the other hand, the network must be energy efficient. Due to the demands of 5G services and applications and the characteristics of UDN, the promotion of energy efficiency is a constant challenge for network operators.

Studies in [7], [8], and [9] highlight that BSs consume about 60% to 80% of the energy used in cellular networks. Next Generation Mobile Networks (NGMN) have documented 6G Use Cases and Analysis [10], where a variety of usage scenarios have been forecasted in the 6th Generation (6G) time horizon. One of them is related to the evolution of the network, which describes aspects related to the evolution of core technologies, including Artificial Intelligence (AI) as a service, energy efficiency, and the delivery of ubiquitous coverage. It is stated that by engaging the capabilities of native AI/Machine Learning (ML) models in the intelligent allocation of networking, computing and storage resources, the optimization of energy consumption can be accomplished in the network and devices. NGMN consider that optimization of energy consumption can be achieved by cognitive management of the dormant, inactive, and active states of a network element or device and that the associated state durations could be modulated elastically. To realize a 5G network-wide energy efficiency, 3rd Generation Partnership Project (3GPP) New Radio (NR) has redesigned operation features [11] to support technologies that reduce energy consumption, namely Massive Multiple-Input Multiple-Output (mMIMO), Lean carrier design, Advanced Sleep Modes (ASMs), and AI.

ML plays a very important role in the field of communications. Reinforcement Learning (RL), a subset of ML, features an agent that learns using a trial-and-error approach to map situations into actions. A reward or penalty is received after each action, allowing action-based learning to improve rewards. The agent will then pursue an objective by capturing the most important aspects of the real problem, interacting over time with the environment [12]. Thus, RL can be deployed to explore the balance between energy consumption and performance in 5G networks.

This paper presents a 5G BS sleep mode management based on SARSA that balances a set of metrics to find the best tradeoff between energy reduction and QoS constraints. This tradeoff between energy consumption and end-to-end (E2E)

user latency is explored for three 5G main use cases requirements, respectively Enhanced Mobile Broadband (eMBB), Massive Machine Type Communications (mMTC) and Ultra-Reliable Low Latency Communications (URLLC).

The remainder of this document is organized as follows. Section II presents related work in the area. Section III presents the background of the technologies and concepts addressed in the system model designed in this work. Section IV introduces the system model. Section V presents a discussion relative to the results obtained; and in Section VI conclusions are made.

II. RELATED WORK

Energy efficiency is a commonly studied subject in several technological fields. Several approaches to green cellular techniques have been proposed, such as hardware enhancements, sleep mode techniques, optimization in radio transmission, network planning and deployment, and adoption of renewable energy resources. To address such problems and challenges, RL has recently been used in the networking and communication areas. RL has been also used in other real life applications such as healthcare, robotics, gaming, image processing, and manufacturing. As the level of complexity of future networks increases, traditional approaches to network planning and deployment, and operation will no longer be adequate. With 5G, mobile network operators can enable new services and experiences for enterprises and consumers. Those services can be mapped with 5G use cases that have different requirements to meet, thus different radio access network architectures are also needed. Network slicing is a potential solution to simplify network architectures and operations, as it enables customization of specific demands for specific services or customers, using the same physical network infrastructure. In [13], RL is used to enable a scheduling solution for data traffic management. The scheduling framework allows different scheduling rules to be selected as the RL agent maps the rules to each state and learns when to apply each. This framework allows minimizing packet delays and packet drop rates for strict QoS requirement applications. In [14] the authors propose a slice admission policy based on RL for a 5G flexible Radio Access Network (RAN), that splits the RAN functions to support diverse service requirements. The authors considered a central office and a remote regional data center to divide the processing tasks. High-priority services with strict latency restrictions are placed in the central office and low-priority services with non-strict latency restrictions can be placed in both centers. The authors concluded that the proposed policy outperforms the benchmark heuristics, as the system manages the slice requirements with different latency requirements. This allows the maximization of accepted slices while minimizing the overall penalty paid by the infrastructure provider. The authors of [15] refer that, to provide services with different QoS, mobile edge computing is a technology that enables lower latencies with more flexible services. They used Software-Defined Networks (SDNs) as a solution to

decouple the data forwarding from the centralized control and RL to define a QoS-aware adaptive routing in a multi-layer hierarchical SDN. With this distributed hierarchical control plane, used to minimize signalling delay, the authors highlight that this framework enables adaptive, time-efficient, and QoS-aware packet forwarding. Reducing 5G energy consumption is still a challenge in cellular networks, and RL has also been explored to improve energy efficiency. To address the energy consumption problem of densely and overlapped BSs used by vehicular social networks, in [16] a joint mode-selection and power-control algorithm using RL is proposed. The authors have used the SARSA algorithm to obtain a RL policy that learns how to maximize energy efficiency by adjusting the minimum signal-to-interference plus noise ratio to guarantee the outage probability. In [17] authors studied the joint access control and battery prediction problems in a small-cell Internet of Things (IoT) system that includes multiple energy harvesting user equipment and one BS with limited uplink access channels. The authors concluded that the battery prediction problem has been solved using a fixed round-robin access control policy. The RL-based algorithm allowed to minimize prediction loss error without any model knowledge about the energy source and the energy arrival process.

Sleep mode techniques and RL have received important attention and are being used to achieve energy efficiency in 5G systems. In [18] the authors used the SARSA RL algorithm to decide which sleep mode to choose at a given time considering the BS load, the system dynamically hops from active state to any sleep mode based on the instantaneous traffic load.

In [19] the authors have studied scenarios with different traffic profiles and periodicity of signalling bursts to quantify the outcomes in terms of energy consumption reduction and QoS serving elastic data traffic in the downlink direction. The authors concluded that in scenarios with low traffic load and with increased signalling periodicity, a considerable energy consumption reduction is observed. In [20] the authors design a RL system that aims to find the optimal duration for each sleep mode level according to the requirements of the network operator in terms of energy consumption reduction and delay constraints. The Q-learning algorithm uses solely local information to implement the energy savings policy. Reference [21] is an extension of [19] and [20] where the authors implement a Q-learning codebook to map the traffic load to the possible actions to be performed by the RL agent, using the same approach to maximize the number of times that each sleep mode level is used, starting with the deepest sleep level first. Besides sleep mode traffic-aware strategies, sleep mode location awareness has also been studied. In [22], a Q-Learning algorithm is proposed to control the state of the BS depending on the geographical location and moving velocity of neighbouring users. The objective is to learn the best policy that maximizes the tradeoff between energy savings and delay.

To the best of our knowledge, no studies have integrated the wake-up delay of the sleep mode level with E2E user packet latency for the uplink traffic, being aware of the type of traffic and not only of the mobile user request to activate the BS.

III. REQUIREMENTS

This section presents the requirements and criteria related to energy consumption and delay that have been evaluated in this work. In a 5G network, the BSs have two power consumption components: node power consumption and communication power consumption. The node power consumption includes the power consumption of signal processing, cooling, and backup batteries. Communication power includes the power consumption to transmit a signal with a certain coverage, which depends on the distance of each UE: a distant UE consumes more energy in data transmission than a closer one [1].

The sleep mode strategy assumes the deactivation/activation of the BS hardware components. These components can be grouped by similar deactivation/activation time and assigned to different sleep mode levels. As stated in [23], the following sleep mode levels can be assumed:

- Sleep Mode (SM)1 level: the power amplifier and some processing components of the digital baseband and analogue front-end are disabled. This is the fastest level with 0.071 ms of (Orthogonal Frequency Division Multiplexing (OFDM) symbol) deactivation/activation time.
- SM2 level: needs 1 ms (1 sub-frame of Transmission Time Interval (TTI)) to deactivate/activate additional components of the analogue front-end.
- SM3 level: the power amplifier, all the components of the digital baseband, and almost all the parts of the analogue front-end (except the clock generator) are switched off. The deactivation/activation time is 10 ms (a frame).
- SM4 level: is the standby mode where a large part of the components of BS is deactivated. The wake-up functionality will take 1s, as this is the minimal sleep duration.

Different sets of hardware can be grouped by sleep mode, being an example of the work in [23]. Table 1 presents the power consumption measurements for 4 different types of BSs where the authors have grouped the power consumption and transition times by sleep mode. Both 2×2 and 4×4 macro BSs, radiate 49 dBm over 3 sectors with a bandwidth of 20 MHz, being the difference between them the amount of Multiple-Input Multiple-Output (MIMO) streams. The pico BS radiates 1 W over a bandwidth of 20 MHz. The femtocell BS radiates 250 mW over a bandwidth of 10 MHz and the LSAS BS radiate 41 dBm with 200 antennas. As depicted, regardless of BS, in SM2, SM3 and SM4 the energy consumption is below 10% compared to load mode zero where all hardware is activated.

Table 2 presents the energy consumption, duration of activation/deactivation, percentage of reduction in energy consumption and energy savings, grouped by SMs for the

TABLE 1. Base station power consumption by base station type as defined in [23].

BS Type	BS Power Consumption [W]					
	Load Mode		Sleep Mode			
	Full	Zero	1	2	3	4
2×2 macro	702.6	114.5	76.5	5.6	8.6	5.3
4×4 macro	742.2	138.9	86.3	12.4	7.3	6.2
pico	6.9	2.3	1.5	0.4	0.3	0.2
femto	2.2	1.0	0.6	0.2	0.2	0.1
LSAS	40.5	32.2	21.0	4.1	2.4	1.6

TABLE 2. 2 × 2 macro BS energy consumption, deactivation/activation duration, energy consumption percentage reduction and energy saving in watts, grouped by SMs.

Sleep Mode	Energy Consumption	Act./Deac. Duration	Energy Consumption Reduction	Energy Saving
Awake	114.5 W	n/a	0%	n/a
SM1	76.5 W	0.035 ms	33.19%	38 W
SM2	8.6 W	0.5 ms	92.49%	105.9 W
SM3	6.0 W	5 ms	94.76%	108.5 W
SM4	5.3 W	500 ms	95.37%	109.2 W

2 × 2 macro BS defined in [23]. As the system enters a deeper sleep mode, less energy is consumed, however, the transition latency to activate the BS takes longer, which degrades the end user QoS. In each sleep mode level, the activation and deactivation times are equal. The minimum time of each sleep mode is the sum of the deactivation and activation times. When the BS is awake in idle mode or serving users, it is not possible to reduce energy consumption by deactivating some hardware components, and when BS is idle, the energy consumption ranges from 5.3 W to 114.5 W with activation times from 0.035 ms to 500 ms.

According to the International Telecommunication Union (ITU) nomenclature for IMT-2020 [24], 5G networks target the following three main use case families with distinct connectivity requirements: respectively eMBB, mMTC and URLLC.

- eMBB: enables the transfer of large volumes of data at extreme data rates. It is a human-centric use case, with typical usage on mobile phones and mobile PC's/tablets.
- mMTC: is a machine-centric use case that provides access to a massive number of Low-Power Wide-Area (LPWA) devices that occasionally send or receive small volumes of data. Typical usage includes wearable, low-cost sensors, actuators, meters, and trackers.
- URLLC: is a machine-centric use case with rigorous requirements for reliability and latency. With typical usage on AR/VR, autonomous vehicles, cloud robotics, real-time coordination and control of machines and processes, advanced wearables, and real-time human-machine collaboration.

Table 3 summarizes the 5G use cases presenting for each of the application categories, the general characteristics, and the standard user plane latency requirements [24]. Services within each use case may not have the same constraints, and

thus in certain scenarios, it is possible to relax latency requirements towards the long term. Table 3 also summarizes some sub-use cases that have more relaxed E2E latencies [25], [26]. eMBB applications such as online gaming or 4k downlink video streaming are compatible with E2E latencies of 10 ms and 20 ms respectively. In mMTC there are applications such as autonomous vehicles sensor and video dynamic latency of 5 ms and autonomous vehicles video fixed with 50 ms latency. In URLLC some use cases are suitable with latencies beyond 5 ms, 20 ms or even 50 ms. The automotive 5G URLLC use cases can be divided into *assisted*, *co-operative* and *tele-operated* driving. Their user plane requirements in terms of E2E latency for both uplink and downlink are 5 ms for assisted, 10 ms for cooperative and 20 ms for teleoperated. In the context of Industry 4.0, the use cases *motion control*, *factory automation* and *process automation* can be considered where the monitoring and control of industrial processes are critical in terms of latency and resilience requirements. Their E2E latency requirements are, respectively, of 1 ms, 10 ms and 50 ms.

This opens the opportunity to have systems built with a tradeoff between acceptable latency and energy consumption reduction. This work has not considered the SM4 level due to its minimum sleep duration of 1 s, which is not compatible with latency demands of eMBB, mMTC and URLLC 5G use cases.

IV. SYSTEM MODEL

The design of the system model considers the tradeoff between the energy consumption and the E2E user traffic latency. When in a given sleep mode, the BS does not transmit or receive traffic from the end user but listens to incoming traffic from the core network intended for the end user. In case the BS is sleeping, incoming traffic from the core network is stored on a packet buffer and therefore latency increases, however, this enables a reduction in energy consumption. Figure 1 presents the system model.

The system model (see Fig 1) is defined with respect to (1) traffic modulation, (2) energy consumption, and (3) sleep mode policies. Each of these definitions is detailed in the following subsections. The system model was coded in Python to simulate the integration of the RL agent within the BS in order to obtain a proper sleep mode policy. During simulations, the agent is trained to interact with the environment by making observations, taking actions, and earning rewards. The observations depend on the incoming traffic received by the BS that feeds the packet buffer, and on the energy consumption that depends on the sleep mode level that has been chosen by the agent.

A. TRAFFIC MODULATION

The incoming traffic from the core network was modulated using a stochastic process to match the behaviour of physical quantities of data traffic to be used during simulations. A Poisson process was defined to obtain different traffic models with a different arrival rate λ over time $\lambda(t)$. The rate

TABLE 3. 5G application categories and user plane latency requirements.

Use Case	Application Categories	General Characteristics	Standard user plane latency [24]	Sub-use cases user plane latencies [25] [26]
eMBB	- Broadcasting - Media delivery - Online gaming	- Extreme data rates - Large data volumes - Low latency (best effort)	4 ms, one way for both downlink and uplink	- Online gaming, 10 ms - Downlink video streaming 4k, 20 ms
mMTC	- Actuators - Sensors - Trackers - Wearables	- Low cost devices - Extreme coverage - Long device battery life	1 ms, one way for both downlink and uplink	- Autonomous vehicles: sensor, 5 ms - Autonomous vehicles: video dynamic, 5 ms - Autonomous vehicles: video fixed, 50 ms
URLLC	- Augmented reality - Mobile robots - Motion control - Remote control	- High reliability - Ultra-low latency - High availability	1 ms, one way for both downlink and uplink	- Automotive: Assisted, 5 ms - Automotive: Co-operative, 10 ms - Automotive: Tele-Operated, 20 ms - Industry 4.0: Motion control, 1 ms - Industry 4.0: Factory automation, 10 ms - Industry 4.0: Process automation, 50 ms

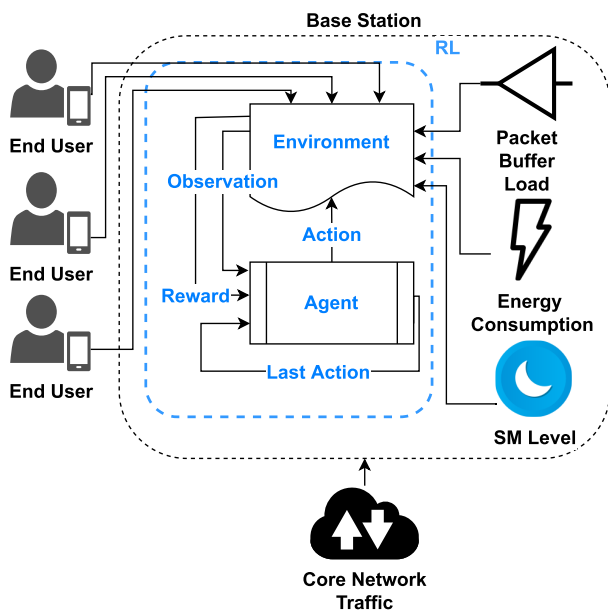


FIGURE 1. System model.

parameter λ is defined as the average number of arrivals per time unit, as in (1).

$$\lambda = \frac{\#events}{interval} \times interval \text{ length.} \quad (1)$$

The value of interval value has been set to 20 ms and #events represents the number of expected events at each 20 ms interval, the interval length represents the total time used in each simulation that was defined with 1000 ms. The system was simulated with loads ranging from 5% to 95% in steps of 15% (i.e. 5%, 20%, 35%, 50%, 65%, 80% and 95%). The number of #events per 20 ms interval varied as the load increased. In the case of 5% traffic load, 1 #event is expected at each 20 ms, so 50 events will be received by the BS during 1000 ms. Table 4 presents the variation of #events used per traffic load.

TABLE 4. Variation of #events used per traffic load.

#events	Traffic Load
1	5%
4	20%
7	35%
10	50%
13	65%
16	80%
19	95%

Each event is composed of one or more packets. Figure 2 presents on the left y-axis the number of events in each traffic load. Each event consists of 1 up to 7 packets, and this number of packets varies depending on the modulated traffic load. This variation in the number of packets per event can be observed on the x-axis. The line graph, with the caption on the y-axis on the right, represents the total packets received in each modulated traffic load. As the traffic load increases, the number of events with more than one packet also increases. To generate the number of packets received in each event, a random function is used to return a random element chosen from a population (number of packets). The population ranges from 1 to 7 depending on the traffic load. Relative weight is given to each element in order to influence the random choice, weights were decreasingly assigned to each element of the population.

B. ENERGY CONSUMPTION MODEL

The energy consumption model was defined with an approximate energy saving estimation at the BS node level, which can be set as the fraction of time that the BS spends in each sleep mode, or awake mode over a certain period of time. The energy consumption during a given state is given by (2) where the T_s is the time in ms spent in the state s , and EC_s is the energy consumed per ms for the given state.

$$EC_s = T_s \times EC_s, \quad \forall s \in \{awake, SM_1, SM_2, SM_3\}. \quad (2)$$

The transition between the states awake, SM1 and SM2 takes less than 1 ms, where the specific energy consumption for that transition period is not considered. When the

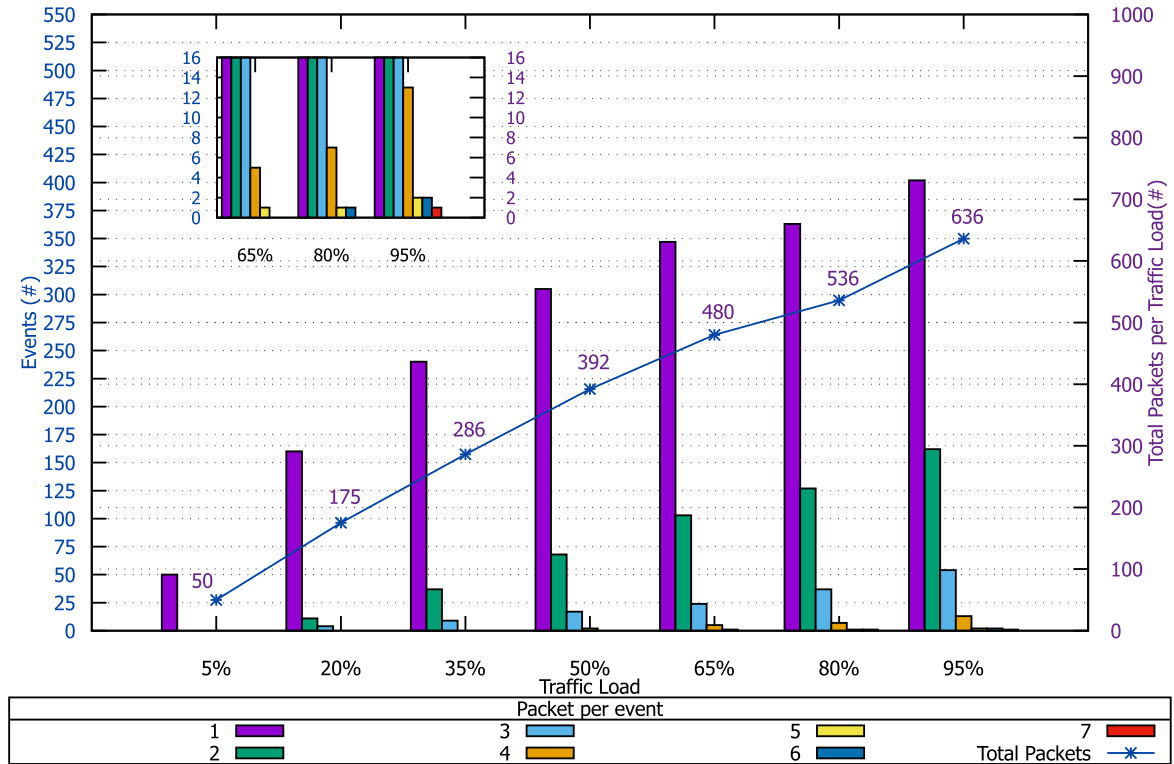


FIGURE 2. Number of events in each traffic load.

transition is between SM_3 and SM_2 , SM_3 and SM_1 or vice-versa, a transaction duration of 4 ms, and a value of 5 ms between SM_3 and awake or vice-versa is considered. For these transition periods, the average energy consumption during the transition period with $EC_{tr1} = 0.0303 W$ and $EC_{tr2} = 0.0514 W$ being defined for the two scenarios. The energy consumption calculation during a transition process between states is calculated as in (3) where T_{tr_x} is the time in ms spent in the transition x (i.e., 4 ms for transition 1, and 5 ms for transition 2), and EC_{tr_x} is the energy consumed per ms for the given state transition.

$$EC_{tr} = T_{tr_x} \times EC_{tr_x}, \quad x = 1, 2. \quad (3)$$

The total energy consumption for an episode of 1000 ms is then given by (4).

$$EC_{total} = \sum_{i=1}^n EC_s(i) + EC_{tr}(i). \quad (4)$$

C. SLEEP MODE POLICY

The sleep mode policy is the outcome of the RL algorithm that enables the system to learn the best policy to use in each simulated environment.

This system model considers transitions from the state-action pair to state-action pair, and learns the values of state-action pairs using Markov decision process policy with a rewarding process. The chosen algorithm is SARSA an On-Policy Temporal-Difference Learning algorithm used

to find the best sleep mode policy. This method estimates the state-value function denoted by $q_\pi(s, a)$ for the current behavior policy and for all states s and actions a . This update is done after every transition from a non-terminal state S_t as presented in Algorithm 1. A sleep mode strategy is proposed that relies on the tradeoff between the energy consumption and the wake-up delay added to the packet buffer latency, where α is the learning rate, γ is the discount factor, S is the state, S' is the state after the action A is set in state S , and R is the reward received after the action is taken.

In order for the RL agent to find the best sleep mode policy, actions are tested in the environment, and observations and rewards are obtained.

1) STATES AND ACTIONS

At each timestep, t the state space of the BS is represented by s_t and can take a value from the set S . The state s is part of the environment and indicates the current sleep mode level that is set in the BS and the status of the packet buffer load if it is *low* or *high* as follows:

$$S = \{\text{awake}_{low}, \text{awake}_{high}, SM_{1,low}, SM_{1,high}, SM_{2,low}, SM_{2,high}, SM_{3,low}, SM_{3,high}\}.$$

The action space enables the possible decisions denoted by a that the agent can set in the BS. The set A represents all possible actions, as follows.

$$A = \{\text{awake}, SM_1, SM_2, SM_3\}.$$

Algorithm 1 SARSA (On-Policy TD Control) for Estimating

$$Q \approx q_*$$

Require:

Algorithm parameters: step size $\alpha \in (0, 1]$, $\varepsilon > 0$
 Initialize $Q(s, a) \forall s \in S^+, a \in A(s)$, arbitrarily except that
 $Q(\text{terminal}, \cdot) = 0$

Loop for each episode do

Initialize S

Choose A from S using policy derived from $\varepsilon - greedy$

Loop for each step of episode do

Take action A , observe R, S'

Choose A' from S' using policy derived from $\varepsilon - greedy$

$$Q(S, A) \leftarrow Q(S, A) + \alpha[R + \gamma Q(S', A') - Q(S, A)]$$

$S \leftarrow S'$

$A \leftarrow A'$

end for until S is terminal

end for

2) REWARD/PENALTY FUNCTION

The rewards and penalties depend on two main variables: SM_{weight} and Buf_{lim} . SM_{weight} can be set between $[0, 1]$ where this normalized weight can prioritize QoS over energy consumption reduction, or vice-versa. Buf_{lim} penalizes the reward when the packet buffer load is high. The RL environment allows the agent to constantly monitor the binary status of the buffer (low or high) and the variable Buf_{lim} defines the threshold in terms of the latency of the packets that are in the buffer while the BS is in sleep mode. When the BS is in the awake state and not serving traffic, there is an opportunity to save energy, thus there is no reward for saving energy, but also no penalty for the delay introduced by a deeper sleep mode level. The system is rewarded when energy is saved and penalized in the following two situations:

- when a delay is introduced to wake up the BS;
- when the packet buffer load is high, as this increases the E2E latency of packets.

The power-saving reward is calculated using (5), whereas the deeper the sleep mode level, the greater the reward.

$$RPS = \frac{EC_{\text{awake}} - EC_{SM_x}}{EC_{\text{awake}}}, \quad x \in \{\text{awake}, SM_1, SM_2, SM_3\}. \quad (5)$$

The delay to wakeup penalty, whereas the deeper the sleep mode level the greater the penalty, is obtained using (6).

$$PD = -\frac{\text{Delay}_x}{\text{Delay}_{sm_3}} \quad x \in \{\text{awake}, SM_1, SM_2, SM_3\}. \quad (6)$$

The reward function in (7) includes a term SM_{weight} that prioritizes the reduction of energy consumption or prioritizes QoS with less delay to wake up. If the packet buffer load is high and the action is different from sleep mode awake, then the reward or penalty represented by BL_{reward} or BL_{penalty} is

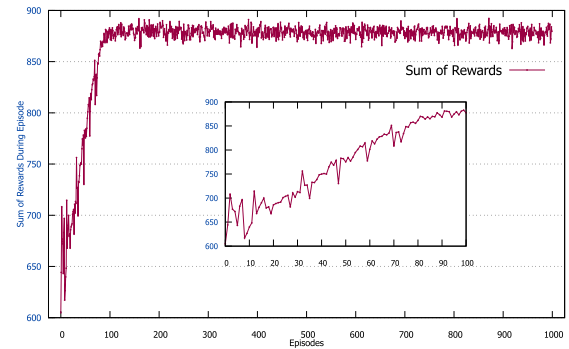


FIGURE 3. Convergence of the total reward function per episode.

used by the following reward function:

$$R = \begin{cases} (SM_{\text{weight}} \times RPS + (1 - SM_{\text{weight}}) \times PD), & \text{if BufferLoad} == 0 \\ BL_{\text{reward}}, & \text{if action} == 0 \\ BL_{\text{penalty}}, & \text{otherwise} \end{cases}, \quad \text{otherwise.} \quad (7)$$

V. RESULTS AND ANALYSIS

This section presents the performance of the event-based simulations implemented in Python. During the learning phase, the consumption and delays values from the 2×2 macro BS presented in Table 2, the different traffic loads presented in Section IV-A and the following SARSA parameters were considered: $\epsilon = 0.1$ (exploration parameter), $\gamma = 0.1$ (discount factor), $\alpha = 0.1$ (learning rate). To deduce the RL policies, combinations with the parameters $SM_{\text{weight}} \in [0, 1]$, $Buf_{\text{lim}} \in \{1 \text{ ms}, 5 \text{ ms}, 10 \text{ ms}, 20 \text{ ms}, 50 \text{ ms}\}$ and traffic loads ranging from 5% to 95% in steps of 15% were simulated. During the training process, the action selection is defined using the $\varepsilon - greedy$ policy. This mechanism aims to find a tradeoff between the *exploration-exploitation* where the agent performs random exploration occasionally with probability ϵ and takes the optimal action most of the time with probability $1 - \epsilon$. In each simulated combination of parameters, the training process took 1000 episodes with 1000 steps in each episode, each step representing a *ms*. At each step, the ε -greedy policy selects an action a_i from the estimated $q_i(s, a_i)$. The agent sets the sleep mode level in the BS accordingly with the chosen action, observes the immediate reward defined in (7) and the next state S' and updates the state-value function $q_i(s, a_i)$. Each episode ends when the maximum step of 1000 is reached and each total episode reward is obtained by the accumulation of instantaneous rewards in all steps. Figure 3 presents the convergence obtained with the following parameters: $SM_{\text{weight}} = 1$, $Buf_{\text{lim}} = 50 \text{ ms}$ and traffic load = 5%. It can be seen that as the number of episodes increases, the value of accumulated instantaneous rewards stabilizes after around 100 iterations. The same behaviour occurred for the remaining combinations of simulated parameters.

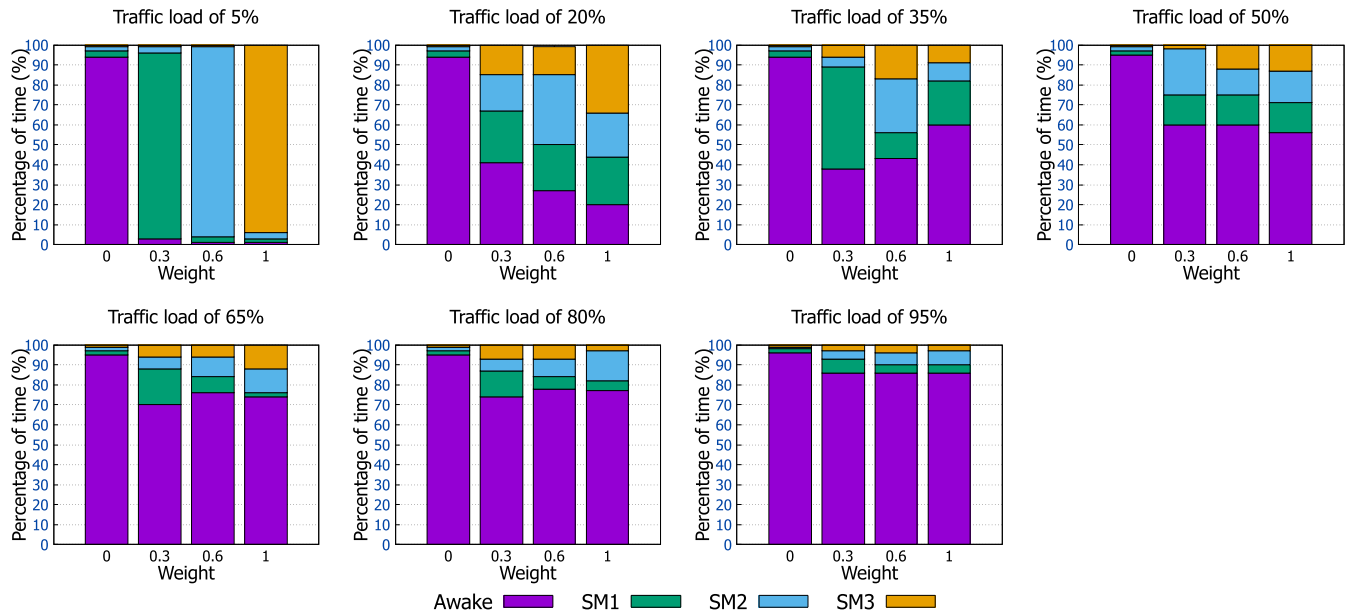


FIGURE 4. States of the BS with a threshold of 50 ms.

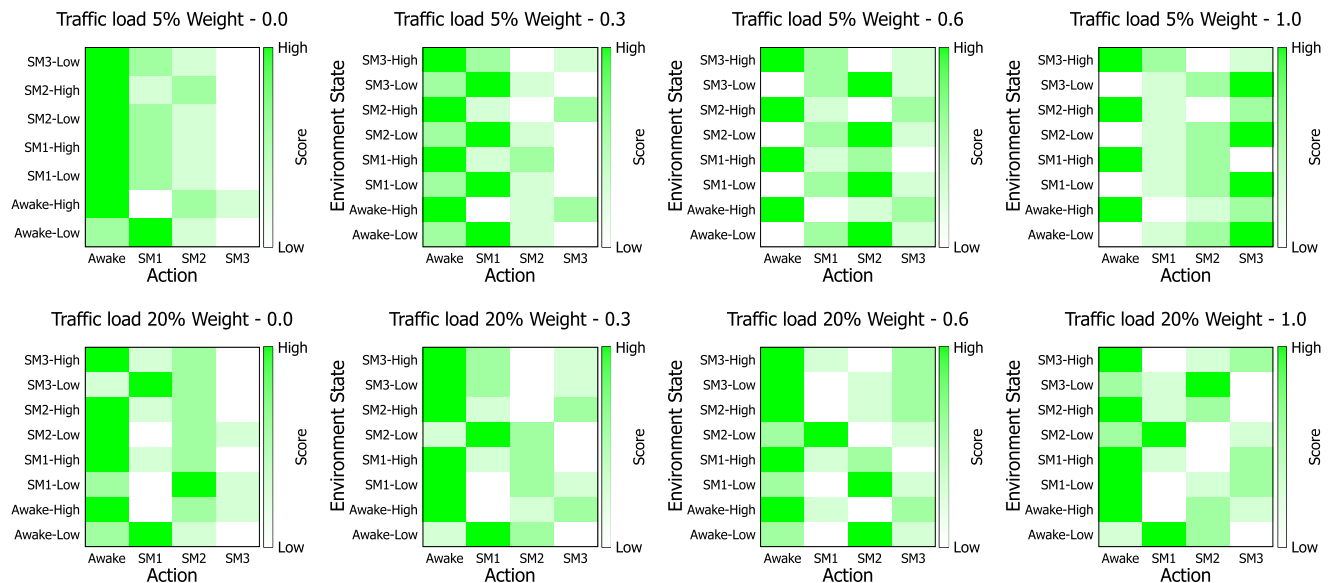


FIGURE 5. SARSA Q-Table HeatMap reward function for different $SM_{weights}$ and traffic loads.

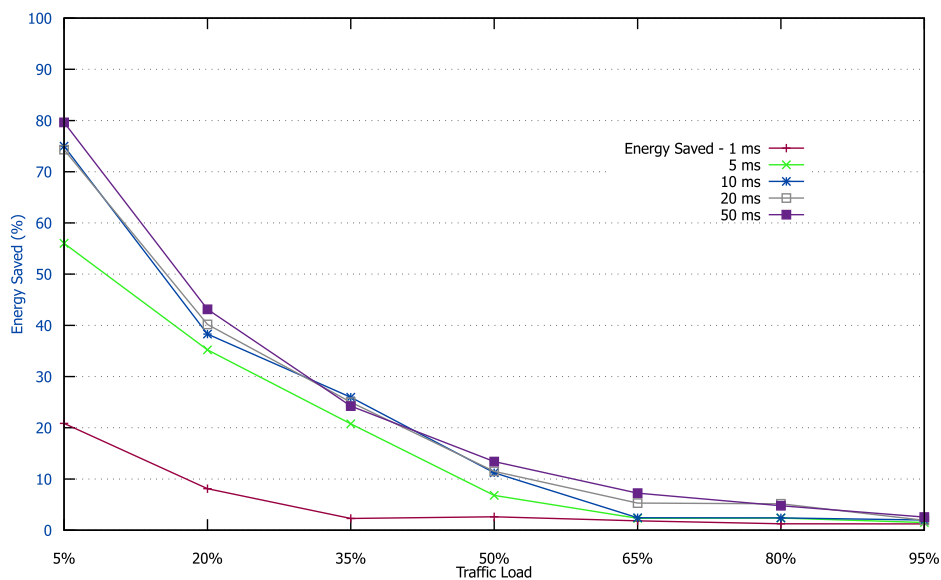
Figure 4 presents the distribution of the different sleep mode states as a function of SM_{weight} and the traffic load for a Bu_{lim} of 50 ms using the exploited policy from the training phase. This policy can accommodate the URLLC industry 4.0 use case for process automation that is compatible with 50 ms latencies.

Considering that with a $SM_{weight} \approx 1$, system maximizes the energy consumption reduction and with $SM_{weight} \approx 0$ it restricts latency. For a traffic load of 5%, the system adjusts the policy to the chosen weight, where it is verified that for a SM_{weight} of 0, the system stays in the awake state during

94% of the time, 93% in SM_1 with a SM_{weight} of 0.3, 95% in SM_2 with a SM_{weight} of 0.6 and 94% in SM_3 with a SM_{weight} of 1. With a very low traffic load, sleep mode policies can have a significant impact on energy reduction because the system can choose the sleep mode level in function of the SM_{weight} without receiving penalties due to the packet buffer load being high. As traffic load increases, it is verified that for a SM_{weight} of 0 the system continues to restrict the latency, predominating in approximately 95% of the time the awake state. For a traffic load range from 20% to 95% with a SM_{weight} greater than 0, there is a clear tendency to increase

TABLE 5. Energy saving percentage per use case for a traffic load of 5%.

Use Case	Sub Use Case	Buf _{lim} [ms]	Energy Saving per SM _{weight} [%]				
			0	0.25	0.5	0.75	1
eMBB	Online gaming	10	3.62	50.92	70.34	75.0	75.38
	Downlink video streaming	20	3.91	54.12	70.82	71.45	74.29
mMTC	Autonomous vehicles: sensor	5	3.23	33.33	40.53	48.28	56.02
	Autonomous vehicles: video dynamic	5	3.23	33.33	40.53	48.28	56.02
	Autonomous vehicles: video fixed	50	3.99	62.30	75.49	76.07	79.64
URLLC	Automotive: Assisted	5	3.23	33.33	40.53	48.28	56.02
	Automotive: Co-operative	10	3.62	50.92	70.34	75.0	75.38
	Automotive: Tele-Operated	20	3.91	54.12	70.82	71.45	74.29
	Industry 4.0: Motion control	1	3.04	16.43	17.44	19.38	20.84
	Industry 4.0: Factory automation	10	3.62	50.92	70.34	75.0	75.38
	Industry 4.0: Process automation	50	3.99	62.30	75.49	76.07	79.64

**FIGURE 6.** Energy saving percentage per Buf_{lim} and traffic load.

the use of the awake state as an alternative to the other states. Even so, for a traffic load of 20% and 35%, the percentage of use of the awake state is less than 50%, allowing a distribution of the remaining time between the states of SM₁, SM₂ and SM₃ according to SM_{weight} which enables the system to save energy. From 50% to 95% traffic load, the choice of awake state gradually increases between 55% and 85% of the time. This tendency is expected, as the system tends to wake up the BS so that it can remove packets from the buffer to guarantee maximum Buf_{lim} latency times.

Figure 5 shows the Q-Table heatmap comparison in a scenario with a traffic load of 5% and a slightly higher with 20%. Depending on the current state, the learned policy will be more likely to choose the greener action. In both scenarios, when the environment state has the packet buffer high, the chosen action will be to awake the BS. When the packet buffer is low, despite there being only a 15% traffic increase between the two scenarios, it is possible to verify that the Q-table allows an adaptation of the chosen actions through the traffic increment. Given an SM_{weight} of 1.0, in a scenario with 20%

of traffic load, the system tends to not choose the SM₃ action to avoid an increase in latency packets that are in the packet buffer, unlike the 5% of traffic load scenario where the load is more relaxed, the system tends to choose the SM₃ action.

Regarding performance metrics, the outcome of this system was quantified by analyzing the energy consumption and latency. To test the dynamics of the different incoming traffic loads with the 5G use cases presented in Section III, energy-critical and latency-critical simulations have been carried out.

Table 5 summarizes the energy saving percentage per SM_{weight} for each use case in a scenario with a traffic load of 5%. When the system is configured to prefer QoS over energy consumption reduction, the values obtained are around 3%. When the traffic load is greater than 5% and depending on the use case, it is possible to obtain energy-saving gains between 20% and 80%.

Figure 6 presents the energy saving percentage for each Buf_{lim} among the simulated traffic loads.

Figure 7 presents the energy consumption reduction and packet latency for each Buf_{lim} and the traffic load simulated.

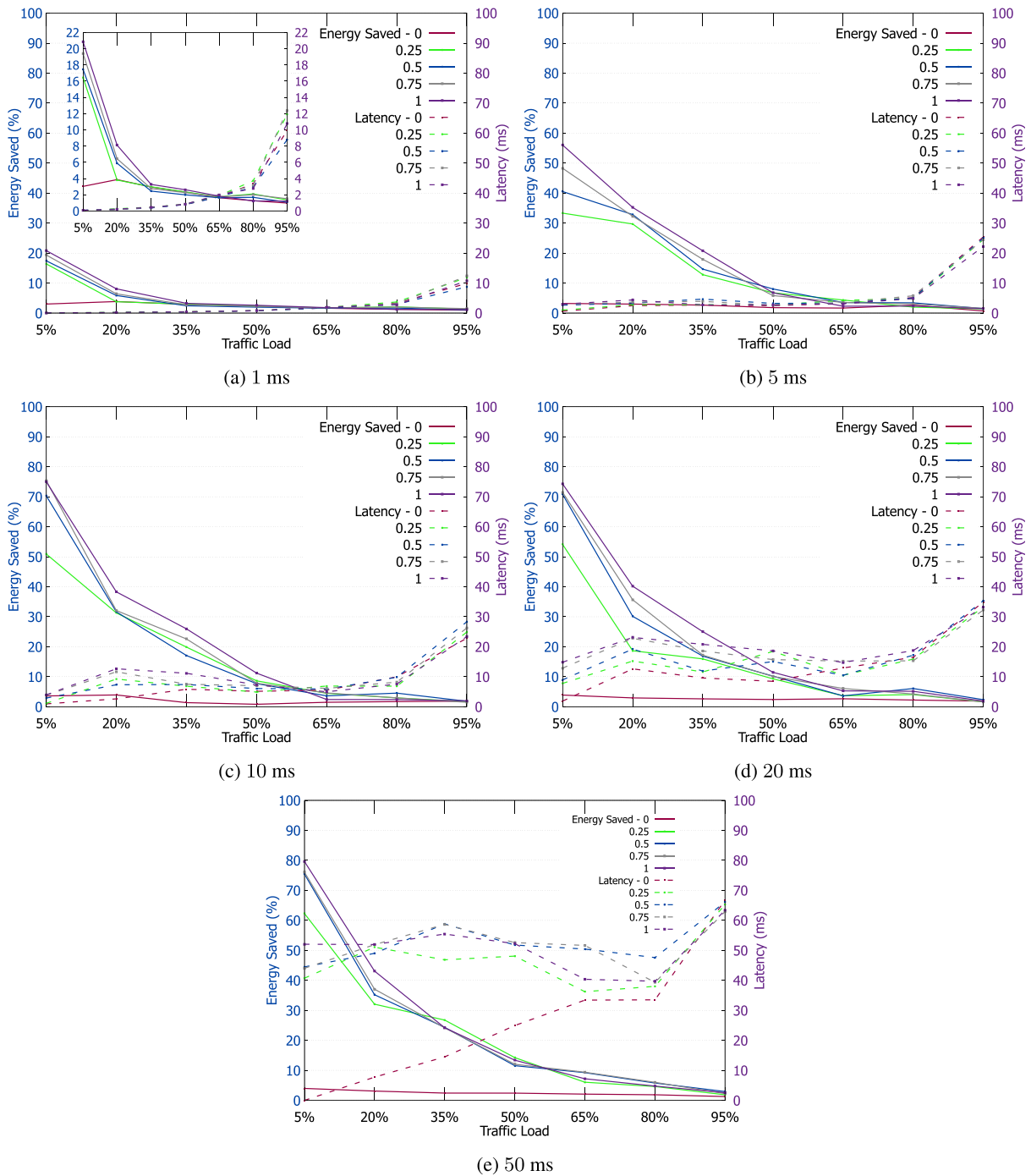


FIGURE 7. Energy consumption reduction and packet latency with thresholds of 1 ms, 5 ms, 10 ms, 20 ms, and 50 ms.

Figure 7 includes figures presenting the results for a Buf_{lim} of 1 ms, 5 ms, 10 ms, 20 ms, and 50 ms. Figure 7a presents the value of the energy saving percentage per Buf_{lim} for 1 ms which is suitable for more latency-restrictive use cases such as mMTC and URLLC. Despite restrictions in terms of latency, in environments with traffic load up to 50%, it is possible to obtain energy gains between 3% and 20% within the maximum limit of 1 ms latency accomplished. Figure 7b

presents a Buf_{lim} of 5 ms that meets the latency requirement for autonomous vehicles sensor and video dynamic from mMTC use case and assisted automotive driving from URLLC use case. Figure 7c presents a Buf_{lim} of 10 ms that meets the latency requirement for factory automation in industry 4.0 or cooperative automotive driving URLLC use case. Figure 7d presents a Buf_{lim} of 20 ms that meets the latency requirement for the 4K video streaming downlink

eMBB use case and the teleoperated automotive driving URLLC use case. Figure 7e presents a Bu_{lim} of 50 ms that meets the latency requirement for mMTC autonomous vehicles sensor and video fixed and URLLC process automation in industry 4.0 use cases.

The highest energy consumption reduction of 80% is obtained with a more relaxed latency threshold of 50 ms and a low traffic load of 5%. As the latency threshold is decreased, the energy consumption reduction also decreases, but still achieving values of 75% for 20 ms and 10 ms thresholds. With a threshold of 5 ms, a 56% of energy consumption reduction is achieved.

As the traffic load increases, the energy saving decreases, this is an expected behaviour considering that the packet buffer receives more packets to deliver and the latency-related constraints imply a more frequent BS awakening, thus reducing energy saving opportunities. Analyzing the packet buffer latency, it is concluded that the latency threshold is respected up to 80% traffic load. Above that, the latency value increments due to the throughput configured in the simulations that allow sending only one packet per millisecond to the end users.

VI. CONCLUSION

To enhance coverage and capacity, 5G networks can be deployed as UDNs. In these networks, as the number of BSs increases, challenges arise with respect to their power management and efficiency. As energy consumption increases with the number of BSs, strategic SMs policies can have a positive impact on reducing power consumption, thus increasing energy efficiency.

In this proposal, the reduction of energy consumption with latency restrictions in a 5G BS is presented. Unlike the revised works in chapter II, this paper introduces the latest proposals in 3GPP NR Release 18 [27]. According to such proposals, some information may be sent from the UE to the BS to assist in setting the sleep modes and transmission parameters. A latency requirement is one example of such information, but it is not limited to it [28]. In this proposal, several 5G use cases with different concerns in terms of maximum latency combined with different traffic loads were tested, and it is certain that the E2E maximum user latency of each chosen 5G use case has been respected. Therefore, power savings can be maximized without negatively affecting the quality of service of end-users. This work, proposes a sleep mode technique that monitors the traffic that is arriving at the BS, this allows the BS to fall asleep at times when there is no traffic to serve. The period when the BS is asleep is maximized by monitoring the latency of packets that arrive in the buffer. The proposal was tested using a combination of several factors, namely a weight that allows giving more importance to energy consumption reduction or having a high constraint on the latency, a buffer load state indicating whether the buffer is low or high in terms of packet latency, all combined with various traffic load scenarios. Testing various thresholds for the buffer latency allows obtaining policies that can be mapped with the

different 5G eMBB, mMTC and URLLC use cases where the constraints with latency can be more or less restricted. The combination of these factors defines multiple sleep mode policies, enabling the exploration of the tradeoff between energy consumption reduction and latency. The results show a significant gain in energy savings, particularly in low traffic load and less restricted latency scenarios. Even with more constraints in terms of latency control and higher traffic loads, energy savings are possible. As management of the different metrics is possible, the mobile operator can orchestrate a tradeoff between energy reduction consumption and latency constraints taking into account the different use case requirements of 5G.

ACRONYMS

3GPP	3rd Generation Partnership Project.
5G	5th Generation.
6G	6th Generation.
AI	Artificial Intelligence.
ASM	Advanced Sleep Mode.
BS	Base Station.
E2E	end-to-end.
eMBB	Enhanced Mobile Broadband.
IoT	Internet of Things.
ITU	International Telecommunication Union.
LPWA	Low-Power Wide-Area.
LTE	Long-Term Evolution.
MIMO	Multiple-Input Multiple-Output.
ML	Machine Learning.
mMIMO	Massive Multiple-Input Multiple-Output.
mMTC	Massive Machine Type Communications.
NGMN	Next Generation Mobile Networks.
NR	New Radio.
OFDM	Orthogonal Frequency Division Multiplexing.
QoS	Quality of Service.
RAN	Radio Access Network.
RL	Reinforcement Learning.
SARSA	State Action Reward State Action.
SDN	Software-Defined Network.
SM	Sleep Mode.
TTI	Transmission Time Interval.
UDN	Ultra-Dense Network.
UE	User End.
URLLC	Ultra-Reliable Low Latency Communications.

REFERENCES

- [1] M. Kamel, W. Hamouda, and A. Youssef, "Ultra-dense networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 4, pp. 2522–2545, 4th Quart., 2016. [Online]. Available: <http://ieeexplore.ieee.org/document/7476821/>
- [2] J. Park, S.-L. Kim, and J. Zander, "Asymptotic behavior of ultra-dense cellular networks and its economic impact," in *Proc. IEEE Global Commun. Conf.*, Dec. 2014, pp. 4941–4946.
- [3] R. Baldemair, T. Irnich, K. Balachandran, E. Dahlman, G. Mildh, Y. Selén, S. Parkvall, M. Meyer, and A. Osseiran, "Ultra-dense networks in millimeter-wave frequencies," *IEEE Commun. Mag.*, vol. 53, no. 1, pp. 202–208, Jan. 2015.
- [4] W. Yu, H. Xu, A. Hematian, D. Griffith, and N. Golmie, "Towards energy efficiency in ultra dense networks," in *Proc. IEEE 35th Int. Perform. Comput. Commun. Conf. (IPCCC)*, Dec. 2016, p. 8.

- [5] (2021). Ericsson. *Mobile Data Traffic Forecast—Mobility Report—Ericsson*. Accessed: Jun. 6, 2022. [Online]. Available: <https://www.ericsson.com/en/reports-and-papers/mobility-report/dataforecasts/mobile-traffic-forecast>
- [6] *IMT Traffic Estimates for the Years 2020 to 2030*, document ITU-R M.2370-0, 2015.
- [7] W. Yu, H. Xu, H. Zhang, D. Griffith, and N. Golmie, “Ultra-dense networks: Survey of state of the art and future directions,” in *Proc. 25th Int. Conf. Comput. Commun. Netw. (ICCCN)*, Aug. 2016, pp. 1–10. [Online]. Available: <http://ieeexplore.ieee.org/document/7568592/>
- [8] A. Fehske, G. Fettweis, J. Malmodin, and G. Biczok, “The global footprint of mobile communications: The ecological and economic perspective,” *IEEE Commun. Mag.*, vol. 49, no. 8, pp. 55–62, Aug. 2011. [Online]. Available: <http://ieeexplore.ieee.org/document/5978416/>
- [9] S. Herrería-Alonso, M. Rodríguez-Pérez, M. Fernández-Veiga, and C. López-García, “An optimal dynamic sleeping control policy for single base stations in green cellular networks,” *J. Netw. Comput. Appl.*, vol. 116, pp. 86–94, Aug. 2018. [Online]. Available: <https://www.sciencedirect.com/science/article/pii/S1084804518301760>
- [10] (2022). NGMN. *6G Use Cases and Analysis*. Accessed: Jun. 6, 2022. [Online]. Available: <https://www.ngmn.org/highlight/ngmn-identifies-6g-use-case.html>
- [11] (2022). 3GPP. *3GPP Specification Detail*. Accessed: Jul. 6, 2022. [Online]. Available: <https://www.3gpp.org/DynaReport/38series.htm>
- [12] R. S. Sutton and A. G. Barto, *Reinforcement Learning: An Introduction*. Cambridge, MA, USA: MIT Press, 2018.
- [13] I. S. Comsa, S. Zhang, M. Aydin, P. Kuonen, Y. Lu, R. Trestian, and G. Ghinea, “Towards 5G: A reinforcement learning-based scheduling solution for data traffic management,” *IEEE Trans. Netw. Service Manage.*, vol. 15, no. 4, pp. 1–14, Aug. 2018.
- [14] M. R. Raza, C. Natalino, P. Ohlen, L. Wosinska, and P. Monti, “A slice admission policy based on reinforcement learning for a 5G flexible RAN,” in *Proc. Eur. Conf. Opt. Commun. (ECOC)*, Sep. 2018, pp. 1–3.
- [15] S.-C. Lin, I. F. Akyildiz, P. Wang, and M. Luo, “QoS-aware adaptive routing in multi-layer hierarchical software defined networks: A reinforcement learning approach,” in *Proc. IEEE Int. Conf. Services Comput. (SCC)*, Jun. 2016, pp. 25–33.
- [16] H. Park and Y. Lim, “Reinforcement learning for energy optimization with 5G communications in vehicular social networks,” *Sensors*, vol. 20, no. 8, p. 2361, Apr. 2020.
- [17] M. Chu, H. Li, X. Liao, and S. Cui, “Reinforcement learning based multi-access control and battery prediction with energy harvesting in IoT systems,” 2018, *arXiv:1805.05929*.
- [18] M. Masoudi, M. G. Khafagy, E. Soroush, D. Giacomelli, S. Morosi, and C. Cavdar, “Reinforcement learning for traffic-adaptive sleep mode management in 5G networks,” in *Proc. IEEE 31st Annu. Int. Symp. Pers., Indoor Mobile Radio Commun.*, Aug. 2020, pp. 1–6.
- [19] F. E. Salem, A. Gati, Z. Altman, and T. Chahed, “Advanced sleep modes and their impact on flow-level performance of 5G networks,” in *Proc. IEEE Veh. Technol. Conf.*, Sep. 2018, pp. 1–7.
- [20] F. E. Salem, Z. Altman, A. Gati, T. Chahed, and E. Altman, “Reinforcement learning approach for advanced sleep modes management in 5G networks,” in *Proc. IEEE 88th Veh. Technol. Conf. (VTC-Fall)*, Aug. 2018, pp. 1–5.
- [21] F. E. Salem, T. Chahed, Z. Altman, and A. Gati, “Traffic-aware advanced sleep modes management in 5G networks,” in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Apr. 2019, pp. 1–6.
- [22] A. El-Amine, H. A. H. Hassan, M. Iturralde, and L. Nuaymi, “Location-aware sleep strategy for energy-delay tradeoffs in 5G with reinforcement learning,” in *Proc. IEEE 30th Annu. Int. Symp. Pers., Indoor Mobile Radio Commun. (PIMRC)*, Sep. 2019, pp. 1–6.
- [23] B. Debaillie, C. Desset, and F. Louagie, “A flexible and future-proof power model for cellular base stations,” in *Proc. IEEE 81st Veh. Technol. Conf. (VTC Spring)*, May 2015, pp. 1–7.
- [24] (2015). ITU-R. *IMT Vision—Framework and Overall Objectives of the Future Development of IMT for 2020 and Beyond*. [Online]. Available: https://www.itu.int/dms_pubrec/itu-r/rec/m/R-REC-M
- [25] E. A. J. Lorca. (2017). *Deliverable D2.1: Scenarios, Kpis, Use Cases and Baseline System Evaluation*. Accessed: Feb. 8, 2022. [Online]. Available: https://one5g.eu/wp-content/uploads/2017/12/ONE5G_D2.1_finalversion.pdf
- [26] S. R. Pokhrel, J. Ding, J. Park, O.-S. Park, and J. Choi, “Towards enabling critical mMTC: A review of URLLC within mMTC,” *IEEE Access*, vol. 8, pp. 131796–131813, 2020.
- [27] *3GPP NR Release 18*. Accessed: Jul. 1, 2022. [Online]. Available: <https://www.3gpp.org/specifications-technologies/releases/release-18>
- [28] D. Lopez-Perez, A. De Domenico, N. Piovesan, G. Xinli, H. Bao, S. Qitao, and M. Debbah, “A survey on 5G radio access network energy efficiency: Massive MIMO, lean carrier design, sleep modes, and machine learning,” *IEEE Commun. Surveys Tuts.*, vol. 24, no. 1, pp. 653–697, Jan. 2022.



SILVESTRE MALTA received the B.S. degree in computer engineering from the Instituto Politécnico de Leiria, Portugal, and the M.S. degree in information systems from the Instituto Politécnico de Viana do Castelo, Portugal. He is currently pursuing the Ph.D. degree in computer science with Vigo University, Spain. Since 2004, he has been a network engineer in several private companies, and since 2010, he has been an Invited Assistant Professor with the Instituto Politécnico de Viana do Castelo. His research interest includes IA/ML applied in the area of computer networks.



PEDRO PINTO (Member, IEEE) received the B.Eng. degree in electrical and computers engineering and the M.Sc. degree in communication networks and services from the University of Porto, Portugal, in 2002 and 2007, respectively, and the Ph.D. degree in telecommunications jointly from the Universities of Minho, Aveiro, and Porto, Portugal, in 2015. Currently, he is an Assistant Professor, the Director of the M.S. degree in cybersecurity, and the Data Protection Officer with the Instituto Politécnico de Viana do Castelo, Portugal. He is also with the INESC TEC Research Institution. His research interests include the areas of computer networks, data privacy, and cybersecurity.



MANUEL FERNÁNDEZ-VEIGA (Senior Member, IEEE) received the Ph.D. degree in telecommunications engineering from the University of Vigo, Spain, in 2001. He is currently an Associate Professor and a Researcher with the Atlantic Research Center within the Information and Computing Laboratory. He has authored or coauthored over 80 papers in peer-reviewed international conferences and journals. His main research interests include wireless communications, distributed algorithms, and quantum internet.

...