



HAL
open science

Energy Optimization with Multi-Sleeping Control in 5G Heterogeneous Networks using Reinforcement Learning

Ali El Amine, Jean-Paul Chaiban, Hussein Al Haj Hassan, Paolo Dini, Loutfi Nuaymi, Roger Achkar

► **To cite this version:**

Ali El Amine, Jean-Paul Chaiban, Hussein Al Haj Hassan, Paolo Dini, Loutfi Nuaymi, et al.. Energy Optimization with Multi-Sleeping Control in 5G Heterogeneous Networks using Reinforcement Learning. IEEE Transactions on Network and Service Management, 2022, 19 (4), pp.4310 - 4322. 10.1109/TNSM.2022.3157650 . hal-03614844

HAL Id: hal-03614844

<https://laas.hal.science/hal-03614844v1>

Submitted on 21 Mar 2022

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Energy Optimization with Multi-Sleeping Control in 5G Heterogeneous Networks using Reinforcement Learning

Ali El Amine, Jean-Paul Chaiban, Hussein Al Haj Hassan, Paolo Dini, Loutfi Nuaymi and Roger Achkar

Abstract—The massive deployment of small cells in 5G networks represents an alternative to meet the ever increasing mobile data traffic and to provide very-high throughput by bringing the users closer to the Base Stations (BSs). This large increase in the number of network elements demands a significant increase in the energy consumption and carbon footprint followed by complex interference management. In order to address these challenges, we consider multi-level Sleep Mode (SM) where BS components with similar activation/deactivation times can be put to sleep. The deeper and higher energy efficient the SM is, the longer it will take the BS to activate, which might impose degradation in the Quality of Service (QoS). While this adds operational flexibility to the BS, it brings complex management to the operator. In this paper, we consider a heterogeneous network architecture where small cells can switch to different SM levels to save energy and reduce dropping rate. We propose a reinforcement learning algorithm for small cells that adapts their activities subject to service delay constraint. In this regard, the algorithm intelligently learns from the environment based on the co-channel interference, the cell buffer size and the expected cell throughput in order to decide the best SM policy. Numerical values show that important energy savings can be obtained with an acceptable dropping rate. Moreover, we show that while offloading users to the macro cell can significantly reduce their delay, dropping rate and the cluster energy consumption, it comes at a cost of decreasing the network energy efficiency up to 5 times compared with the case of no offload.

Index Terms—5G, cellular networks, multi-sleep mode levels, energy consumption, reinforcement learning.

I. INTRODUCTION

The world-wide mobile traffic is witnessing an unprecedented rise triggered by the high speed internet and the advent of smart phone technology. Based on Ericsson’s Mobility Report in November 2020, the mobile data traffic is expected to grow 7 times from 2020 to 2026 [1]. This remarkable growth in wireless data traffic is in an upward trend with no signs of slowing down. For instance, the number of 5G subscribers is expected to reach 2.8 billion by the year 2025 [1]. Addressing this expected increase in mobile data traffic

is one of the main challenges of future cellular networks. Today, the system design of 5G is envisioned to support 1000 times increase in capacity and 100 times more connected devices than today’s 4G networks [2]. In order to meet these challenges, next generation cellular networks are expected to massively deploy Small Base Stations (SBSs) as part of future 5G. These small cells are expected to be a promising solution to increase the area capacity in macro cell sites.

One of the main challenges of this architecture is the surge in the network energy consumption, placing network Energy Efficiency (EE) as one key pillar in future 5G networks [3]. According to the power consumption breakdown in [4], the Base Station (BS) consumes the major part of the total network energy consumption which is about 60% to 80%. As such, reducing the energy consumption at this level will amount to significant reduction in the network total energy consumption. As the mobile traffic pattern fluctuates over both time and space, the ideal energy consumption of the BS is to scale with the traffic load. This is however not the case, as the BS energy consumption at zero load is about 50 – 60% of its maximum energy consumption [5]. This results in energy wastage and low EE especially for underutilized BSs. Among the different approaches proposed to reduce the energy consumption in cellular networks and to scale the BS energy consumption to traffic variations, BS sleep scheme is considered as one of the best methods since it does not require changes to current network architecture, and it is easy to implement [6]. Many studies have focused on adapting the operation of the BS to save energy during off-peak periods [4]–[8]. However, completely shutting down or switching to sleep a BS for a long period of time can create coverage holes in the network, and thus impact the offered Quality of Service (QoS) to the served users (e.g., delay and packet dropping). Moreover, with the emergence of new services and the development of 5G, strict requirements are imposed on some services in terms of delay or dropped packets making the classical ON/OFF sleep unattractive to such services for the reasons mentioned above.

The recent finalized 5G New Radio (NR) 3GPP Rel-15 standard is designed to enable very low energy consumption. As such, new features for energy savings are introduced. Compared to 4G, NR supports Discontinuous Transmission (DTX) duration up to 160 ms as opposed to 1 ms in 4G networks [9]. This enables the implementation of new hardware deep sleep states in which the BS can lower its total energy consumption significantly. Embracing this framework, the work in [10] proposed four distinct Sleep Mode (SM) levels. Each SM level

A. El Amine is with LAAS-CNRS, 7 Av. du Colonel Roche, 31400, Toulouse France - Email: aelamine@laas.fr

J. Chaiban is with American University of Science and Technology (AUST), Beirut, Lebanon - Email: Jgc00002@students.aust.edu.lb

H. Al Haj Hassan is with Acklio, 1137A Av. des Champs Blancs, 35510 Cesson-Sevigne France - Email: hussein.alhajhassan@acklio

P. Dini is with Centre Tecnològic Telecomunicacions Catalunya (CTTC), Barcelona, Spain - Email: paolo.dini@cttc.es

L. Nuaymi is with IMT-Atlantique, 2 Rue de la Chataigneraie, 35576 Cesson Sevigne, France - Email: loutfi.nuaymi@imt-atlantique.fr

R. Achkar is with American University of Science and Technology (AUST), Beirut, Lebanon - Email: rachkar@aust.edu.lb

is characterized by an activation/deactivation time and a minimum sleep duration. Compared to the classical binary sleep scheme, multi-level SM adds network flexibility to cope with the traffic requirements and allows more energy to be saved. In this framework, understanding the impact of SMs on the network load distribution, the resulting inter-cell interference, the related radio resource management procedures and finally the end-users' QoS emerges as a new paradigm in future 5G and beyond cellular networks.

In this paper, we study the impact of SMs on the overall 5G network performance under end-users' QoS constraint while taking into account the inter-cell interference. In particular, we consider a Heterogeneous Network (HetNet) architecture where a Macro Base Station (MBS) is covering a cluster of small cells. While the MBS is always active to ensure coverage, the small cells can switch to different SM levels to mitigate QoS degradation while maximizing the network energy savings. In order to decide which SM to use, we leverage the off-policy Q-learning [11], a well known reinforcement learning algorithm, and propose a distributed Q-learning algorithm that controls the small cells activities based on their level of interference, their expected throughput and their buffer size. In contrast to previous works, we focus on the recent multi-level SM model proposed in [10]. We handle the traffic of users connected to SBSs in SM by a handover mechanism and study its impact on the overall network-level performance. We show that while offloading users to the active MBS reduces traffic drop rate, it comes at the expense of decreasing the network EE.

The rest of the paper is organized as follows. Section II reviews the existing work and presents the paper contributions. Section III describes the system architecture and the 5G multi-level SMs. In Section IV, we formulate our optimization problem and design the states, actions and cost function. We detail our proposed Q-learning algorithm in Section V. Finally, we discuss and conclude the results in Sections VI and VII.

II. RELATED WORK

The use of SM technique in Radio Access Networks (RANs) is among the most studied approaches to reduce their energy consumption [4], [6], [12]. We distinguish between two types of sleep schemes: binary SM level (ON and OFF) and multi-level SM. In binary SM, the BS shuts down completely when the traffic is low to save energy. Depending on the objectives, different sleep scheme based algorithms are proposed [13]–[20]. In [13], the authors investigated the potential of saving energy by switching-off macro BSs under coverage constraints using stochastic geometry. Compared with the baseline that keeps the BSs always active, their results achieved an EE gain of around 60%. In [14], the authors highlighted the energy consumption behavior of small cell deployments while emphasizing on sleep mode solutions to reduce the network energy consumption. The authors discussed three different sleep mode strategies, which lead to energy savings between 10 and 60%. The authors in [15] opted to maximize the EE of a HetNet by dynamically switching-off under-utilized SBSs along with user association and power control. As opposed to

switching-off SBSs, the authors in [16] proposed to switch off macro BSs and offload their users to low-power SBSs. In [17], the authors proposed an optimization mechanism based on delay-constrained energy-optimal BS sleeping policies. In [18], dynamic programming is applied to find the optimal BS ON-OFF policy given the on-grid energy price in order to minimize the on-grid energy cost purchased by the operator while guaranteeing the downlink transmission quality at the same time. In [19], a data-driven energy-saving framework is proposed. The authors adopted a heuristic switch-off policy based on the traffic conditions to maximize the network energy savings while satisfying the minimal service requirements for users. The authors in [20] proposed a heuristic load-based switching-off algorithm for SBSs to optimize the energy savings in an ultra dense network while minimizing the power costs related to switching-off.

In the literature of binary sleep schemes, several works considered combining SM with Machine Learning (ML) approaches for traffic prediction and action decisions. For example, [21] proved the effectiveness of binary sleep schemes and energy savings with ML to predict traffic in a two-tier HetNet powered by renewable energy and the smart grid. Instead of depending on the traffic forecast, conventional reinforcement learning techniques such as Q-learning have been applied in recent works for cell switching [7], [22]–[24]. By interacting with the environment, the agent (i.e., the BS) learns how to adapt to the dynamic network changes in order to decide the optimal ON-OFF policy. In [25], we proposed a Q-learning algorithm to decide the number of active resource blocks based on the traffic load in a green cellular network to save energy. Fuzzy Q-Learning (FQL), a stochastic optimization tool that combines fuzzy inference system and Q-learning, has also been investigated in the problem of cell switching. For example, [26] proposed a FQL algorithm to switch SBSs based on their buffer state, cell capacity and estimated spectral efficiency loss due to interference from nearby SBSs. For high order complexity networks with large state-action space, classical reinforcement learning are combined with value function approximation to derive the optimal policy. In this regard, linear function approximation and deep learning techniques are proposed to reduce the energy consumption by switching-off cells in ultra-dense networks [27]–[29].

Unlike the previous works with binary scheme, recent works have considered multi-level SM based on the model proposed by GreenTouch in [10] that introduces four distinct SM levels [30]–[34]. The authors in [30] proposed a Q-learning algorithm to determine the best duration of time a BS can spend in a particular SM level in order to minimize the network energy consumption subject to the BS (de)activation latency and users service requirements constraints. In [31], we proposed a distributed Q-learning algorithm to control the sleep depth of the BSs in the network to maximize the trade-off between energy savings and network latency. In [32], we leveraged the multi-SM levels to decide the SM depth of the BS given the location of the users in the network and their velocities towards neighboring cells. The authors in [33] proposed an online traffic-adaptive BS energy saving algorithm with Q-learning that utilizes the different SM levels subject to packets drop-

ping. In [34], we proposed a distributed Q-learning algorithm in a two-tier HetNet that controls the activity of the SBSs while considering the co-channel interference between neighboring cells in order to maximize the energy savings of the network subject to the service delay constraint. Different from the model of [10], others works such as [35] studied multi-level sleep under a different model. The authors proposed a heuristic to solve their Lyapunov optimization formulation problem with LSTM-based user predication.

Despite the exhaustive works on binary sleep scheme in [13], [15], [17]–[20], this scheme is undesirable for future 5G services with hard constraints. Compared to the existing works with multi-level SM in [30]–[35], this paper provides additional insights on understanding advanced SM levels of [10]. In contrast to [30], [33] that consider only one BS, we study a 3GPP scenario where we have a cluster of SBSs covered by a MBS. Different from our works in [31], [32] that consider a homogeneous network of BSs, in this paper we include the co-channel interference in the decision making process. Different from our previous work in [34], in this paper we further handle traffic offloading and investigate the system performance on a network-level scale. Finally, the work in [35] proposes a sub-optimal heuristic algorithm. We summarize in tables I and II the above literature for binary and multi-level sleep schemes, respectively. Moreover, we highlight the limitations that we addressed in multi-level SM.

A. Paper contributions

In this paper, we investigate the energy/delay trade-off problem in a HetNet architecture where a macro cell covers a cluster of small cells. We propose an intelligent cell switching mechanism using Q-learning where small cells can switch to different SM levels to save energy while maintaining a good QoS depending on the service use case. We leverage the recent multi-level SM model proposed in [10] that allows flexibility in controlling the state of the BS. Compared to existing work, this paper provides the following key insights:

- We propose an intelligent distributed controller based on reinforcement learning that chooses the best decision of SM level in order to minimize the energy consumption and data lost in the network. Due to the limited size of the defined state and action spaces, we rely on a Q-learning model that is proven to be effective and able to manage the computational requirements needed for our problem.
- In contrast to [30]–[33] that consider interference-limited models, we study a more realistic 3GPP scenario composed of a cluster of SBSs covered by a MBS where the co-channel interference between SBSs is regarded in the decision making process.
- In contrast to our previous work in [34], we handle the traffic of users when their connected SBS is turned off by a handover technique to the MBS. In this regard, we study the impact of this offloading on the overall network EE.
- We analyze the performance of the proposed algorithm for different use cases and network parameters through extensive simulations. In contrast to previous works that

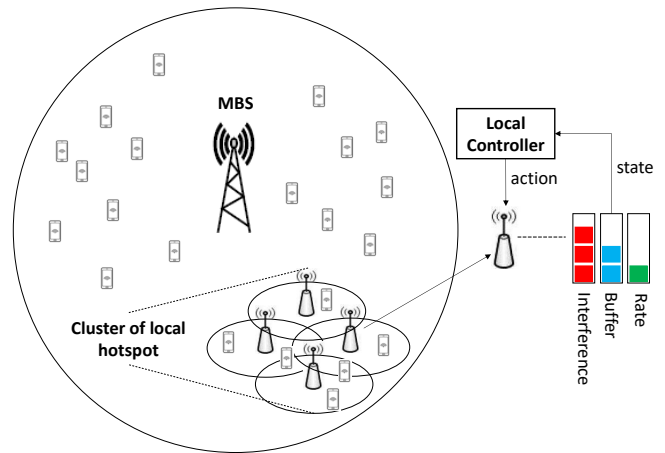


Fig. 1: A HetNet architecture with a macro cell and a cluster of small cells each with a distributed local controller.

focus on either network or BS level, we study the system performance on both levels. Our numerical results reveal a lower traffic drop rate when users offload from their inactive SBSs, however at the expense of decreasing the network EE.

III. SYSTEM MODEL

A. Network Model

Following the scenario described by 3GPP in [36], we consider a two-tier HetNet architecture where a MBS has four SBSs with overlapping coverage regions as shown in Fig. 1. The MBS serves as a basic constant coverage and it is responsible for control signaling and data services. The SBSs on the other hand are densely deployed in the coverage area of the macro cell to create a local hotspot to increase the capacity in the designated area, thus they serve as capacity cells. Moreover, the macro cell and the small cells are operating at different frequency bands, hence they do not interfere on each others.

In order to enhance the network EE, the SBSs can dynamically switch to different SM levels while the MBS stays active for maintaining coverage. This may not only save energy and limit the CO₂ footprint, but it may also enhance the system capacity by limiting the interference between neighboring cells. A distributed local network controller (e.g., Software Defined Network (SDN)) is in charge of managing the activity of these cells (see Fig. 1). We note that for binary SM levels (ON and OFF), finding the optimal set of active SBSs is a combinatorial problem with high computational complexity. In this work, we further add another layer of complexity by considering multi-level SMs. In order to solve this problem, a local controller manages a given SBS by means of interacting with the environment in order to map what to do (action) when faced with a current situation (state). This method is known as Q-Learning [11]. In particular, each SBS, at every iteration, exploits the expected average interference at its users from neighboring cells and the activation decision taken in

TABLE I: Overview of binary switch-off strategy.

Ref.	Architecture	Tool	Strategy	Contributions
[13]	HetNet: macro and micro cells	Stochastic geometry	Switch-off MBSs to save energy	Derive expressions of average coverage probability and users' power consumption. Determine the optimal proportion of sleep MBSs
[14]	HetNet: macro and small cells	Heuristic algorithms	Three control algorithms: (1) small cell driven; (2) core network driven and (3) UE driven	Investigate the network energy efficiency with sleep mode algorithms for SBSs
[15]	HetNet: macro and small cells	Convex Optimization	Switch-off underutilized SBSs and user association	Maximize the network energy efficiency
[16]	HetNet: macro and small cells	Heuristic algorithm	Switch-off MBSs and offload users to SBSs	Minimize the network energy consumption subject to QoS constraint
[17]	Single BS	Queueing theory	Hysteresis sleep mode scheme	Study the energy-delay tradeoff
[18]	Homogeneous network powered by green energy and power grid	Dynamic programming and Stochastic geometry	BS switching-off scheme	Obtain optimal BS on/off policy that reduces the on-grid energy cost
[19]	Large scale homogeneous network	Heuristic algorithm	Switch-off BSs according to traffic conditions	Maximize energy savings while guaranteeing minimum service requests
[20]	Ultra dense network	Heuristic algorithm	Load based switching-off algorithm	Optimize energy savings while minimizing the energy cost from switching users to other BSs
[21]	HetNet: macro and small cells	Supervised ML	Sleep action decisions based on traffic prediction	Investigate the effectiveness of sleep mode combined with ML approaches for traffic forecast
[7], [22]	HetNet: macro and small cells. Small cells are solely powered by green energy	Q-learning	Switch-off SBSs based on harvested energy, traffic load and battery level	Propose a distributed Q-learning approach that finds the optimal policy that improves the system performance (energy efficiency, drop rate, throughput)
[23]	HetNet: macro and small cells	Q-learning	Switch-off SBSs and offload users to the macro cell	Study the impact of offloading users to neighboring cells on the energy consumption
[24]	Large scale homogeneous network	Transfer learning with reinforcement learning	BS switching-off scheme	Propose a transfer actor-critic algorithm to speed up the learning process
[25]	Single BS powered by green energy and smart grid	Q-learning	Switching-off resource blocks	Decide the number of active resource blocks based on the traffic load to save energy
[26]	HetNet: macro and small cells	Fuzzy Q-learning (FQL)	Switch-off SBSs according to the cells' buffer, throughput and spectral efficiency loss	Propose a FQL approach to find the optimal policy that reduces energy consumption subject to QoS constraint
[27]–[29]	HetNet: macro and small cells	Deep Q-learning (DQL)	Switch-off SBSs	Propose a DQL approach to find the optimal policy that reduces energy consumption subject to QoS constraint

TABLE II: Overview of multi-level sleep strategy.

Ref.	Architecture	Strategy	Contributions	Limitations
[35]	Large scale homogeneous network	Adaptive sleep mode based on user prediction	Determine the optimal BS switching solution while ensuring a stable data queue state	Heuristic algorithm
[30]	Single BS	Advanced SM management	Determine the best duration of time a BS can spend in a particular SM level that minimizes the network energy consumption subject to QoS constraint	Interference-limited Only one BS
[31]	Large scale homogeneous network	Adaptive sleep mode	Control the sleep depth of the BSs that maximizes the tradeoff between energy savings and network latency	One type of BS Ignore users offloading
[32]	Large scale homogeneous network	Location-aware sleep strategy	Determine the BS sleep depth according to neighboring approaching users	Interference-limited One type of BS Focus only on handovers
[33]	Single BS	Traffic adaptive BS energy saving algorithm	Determine the BS sleep depth according to BS traffic	Interference-limited Only one BS
[34]	HetNet: macro and small cells	Switch-off SBSs according to the cells' buffer, throughput and spectral efficiency loss	Find the best sleep policy that maximize the energy savings of the network subject to QoS constraint	Ignore users offloading

the previous iteration. Then, the local controller uses this information along with the SBS buffer and traffic load levels to control its activity, as we detail later in this section.

B. Multi-Level Sleep Modes

In [10], GreenTouch 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 time unit corresponds to the whole radio frame of 1s. In this SM level, the BS enters a standby mode where it 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 that may impact the QoS of the system. In Table III, we present the SM levels characteristics for a 2×2 MIMO SBS. For simplicity and convenience, in this paper we consider a SISO system model. Even though a 2×2 MIMO system can enhance the achievable rate which in turn can improve the network performance, it consumes more energy. Hence the overall gain is relatively insignificant and out of the scope of this paper. For MIMO systems with a large number of antennas, key parameters such as spectral efficiency, interference and energy

consumption may be greatly affected. However, this case is not very realistic for small cells since there are likely not a lot of users in the studied hotspot.

Along with SM and users' dynamics, the BS has to wake up periodically to send signalling bursts. In contrast to Long Term Evolution (LTE) systems where each antenna must transmit every 0.2 ms a unique Cell Reference Signals (CRS) for channel quality estimates and mobility measurements among other Synchronization Signaling (SS), no CRS is required for 5G [37]. Instead, SS and Physical Broadcast CHannel (PBCH) are transmitted in SS/PBCH block periodically. It has been agreed in Third Generation Partnership Project (3GPP) [37] that this periodicity 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 III: 2x2 MIMO SBS Sleep Modes Characteristics [10].

Mode	Active		SM1	SM2	SM3	SM4
	100%	Idle				
Power (W)	20.7	13.2	8.22	3.55	3.36	3
(de)Activation duration			35.5 μ s	0.5 ms	5 ms	0.5 s
Minimum sleep duration			71 μ s	1 ms	10 ms	1 s

C. Traffic Model

In our model, we consider connected users continuously requesting packets at each time slot. The users have a minimum rate requirement (R^{min}). We use the term data block ($DB_{m,t}$) to describe the total packets requested by the users served by BS m (MBS or SBS) at time slot t . Given a cell bandwidth B_m and the noise spectral density N_0 , we can calculate the rate of user u served by cell m ($r_{m,u}$), the average cell throughput (r_m) and the number of bits served by cell m (y_m) as follows:

$$r_{m,u} = B_m \cdot \log_2 \left(1 + \frac{P_m \cdot h_{m,u}}{N_0 \cdot B_m + \sum_{m' \neq m} P_{m'} \cdot h_{m',u}} \right) \quad (1)$$

$$r_m = \frac{\sum_{u=1}^{K_m} r_{m,u}}{K_m} \quad (2)$$

$$y_m = T \cdot r_m \quad (3)$$

where P_m is the cell transmit power, $h_{m,u}$ is the user channel gain served by BS m , K_m is the total number of users served by cell m and T is the duration of a time slot. The average throughput of the macro cell does not include the interference term since we are considering only one macro BS.

At the beginning of each time slot, a data block is buffered and stacked in the designated cell buffer. The size of the data block (number of bits) is determined by the number of packets requested by the users served by this cell. We assume that the size of each packet (λ) is related to the minimum user rate requirement R^{min} as follows:

$$\lambda = T \cdot R^{min} \quad (4)$$

In this work, we consider the same rate requirement for all users. Thus, the size of each packet is the same. As a result, the size of the data block depends on the number of users

served by the corresponding cell. Furthermore, we characterize the data block by a Time-To-Live (TTL), L (ms), which is the lifetime of the block above which the remaining data in the block are lost. If a data block i is requested at time slot t_i , then its remaining lifetime at time slot t can be obtained as follows:

$$d_{m,t}^i = t_i + L - t \quad (5)$$

When a data block i is being served to the users at time slot t , its remaining data can be obtained as follows:

$$DB_{m,t}^i = \begin{cases} \max \left(0, \lambda \cdot K_{m,t} - \sum_{j=t_i}^t y_{m,j} \right) & \text{if } d_{m,t}^i < L \\ 0 & \text{otherwise} \end{cases} \quad (6)$$

where $y_{m,j}$ is the number of bits served by cell m at time slot j . Finally, we can express the buffer of a BS m at time slot t as the set of data blocks stacked in the buffer as follows:

$$B_{m,t} = \{DB_{m,t}^i | d_{m,t}^i > 0\} \quad (7)$$

To better understand the system traffic dynamics, we illustrate in Fig. 2 the evolution of a cell buffer with time as well as the lifetime of each arriving data block. Since the users continuously request packets, at each time slot a new DB_m^i arrives to the cell buffer m . We assume that all users have the same R^{min} , hence the DB_m^i 's have the same size. Furthermore, we consider in this example a small cell that is switching between active mode and any SM level. During the first time slot, the first data block (DB_m^1) arrives. Since the BS is in SM, this block stays in the buffer until the BS reactivates during the next time slot and part of DB_m^1 is served while another DB_m arrives (DB_m^2). At this time slot, the buffer of this cell following eq. (7) is expressed as $B_{m,t+T} = \{DB_m^1, DB_m^2\}$. At the end of the third time slot, the remaining of DB_m^1 is lost because it reached its lifetime before being served, and the buffer is updated to $B_{m,t+3T} = \{DB_m^2, DB_m^3, DB_m^4\}$. During each time slot, if the BS is active, the amount of served data depends on the cell throughput that is directly related to the level of interference with neighboring cells. Thus, the states of the SBSs have an impact on the latency added to the system. On the one hand, the deeper the SM level is, the more time the user will have to wait until the SBS reactivates. On the other hand, if all the SBSs are active, the level of interference is the highest resulting in lower throughput.

D. Users offloading

As previously mentioned, when an SBS switches into an SM level, its users experience service delay and possible data loss. In order to limit these degradations in QoS, users served by the cluster can offload to the MBS that is always active. In this regard, we propose to offload a user, u' , served by the cluster to the MBS if the latter has enough radio resources (i.e., bandwidth) to guarantee the user minimum rate requirement (R^{min}). Hence, the following user offloading condition:

$$r_{m,u'} = \frac{r'_m}{K_m + 1} \geq R^{min} \quad (8)$$

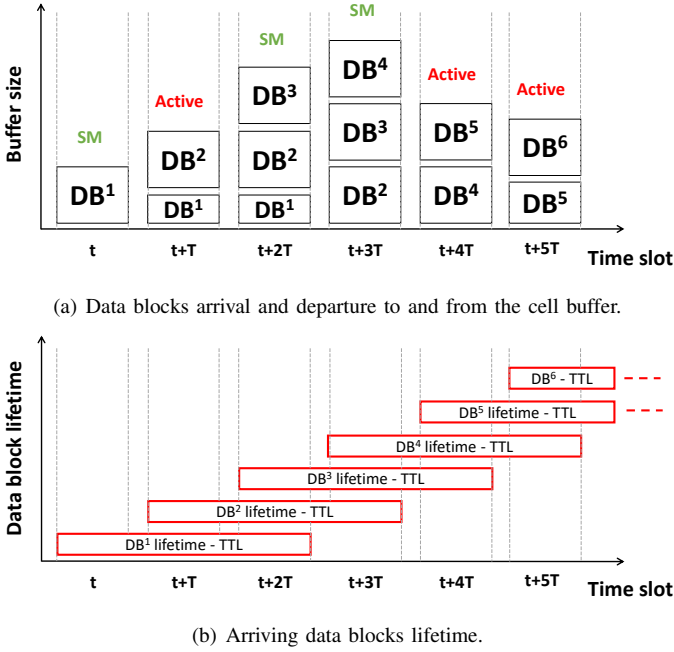


Fig. 2: An example of the flow of data blocks requested by the users served by a small cell and their life time.

where r'_m is the new average throughput of BS m (in this case $m = \text{MBS}$) that considers the offloaded user u' .

IV. PROBLEM FORMULATION

In this work, we consider the problem of minimizing the SBSs energy consumption and the data lost (dropped). The problem can be formulated as follows:

$$\min_{a \in \{\text{Act.}, \text{SM1}, \text{SM2}, \text{SM3}\}} \sum_{m=1}^M \sum_{t=1}^N \left((1-w) \cdot p_{m,t}(a) + w \cdot l_{m,t}(a) \right) \quad (9)$$

where M is the number of SBS and N is the total number of time slots. p and l represent the SBS power consumption and the amount of data lost, respectively after an action a has been taken. w is a weighting parameter that prioritizes between the power consumption and dropping rate.

In order to reduce the cluster's energy consumption and its data lost, we model the above problem as a discrete-state Markov Decision Process (MDP) problem, where a SBS m in state $s_m(t)$ takes an action $a_m(t)$ then moves to another state $s_m(t+1)$. We denote by $\mathcal{S} = \{R, B, I_l\}$ the state space of the SBS, where R , B and I_l are the sets related to the cell throughput, buffer state, and the estimated spectral efficiency loss (bit/s/Hz) due to the interference from nearby active SBSs, per SBS, respectively.

At each time slot, the local controller decides an action $a_m(t) \in \mathcal{A} = \{\text{Active}, \text{SM1}, \text{SM2}, \text{SM3}\}$ for a SBS after observing its state. In order to evaluate the cost associated with each action-pair, we use the cost function that takes into account the power consumption of the SBS $_m$, and the users' QoS:

$$c_m(s, a) = (1-w) \cdot p_m(s, a) + w \cdot l_m(s, a) \quad (10)$$

In order to solve the above MDP, we first need to derive the transition probabilities between the system states and the corresponding state-action cost. Then, a backward dynamic programming needs to be performed to evaluate the optimal value function, in order to derive the optimal policy. However, this requires a complete knowledge of the system information which is not practical for real-time applications. To alleviate the non-causal system information, we rely on Q-learning to find the optimal policy by exploring and interacting with the current state of the system. Due to the limited number of state-action pairs, we have a fairly small lookup table that traditional Q-learning algorithm can handle. Thus, in this paper, we do not need to use advanced reinforcement learning tools such as Deep Q-learning, and we focus on the classical Q-learning algorithm.

V. ONLINE INTERFERENCE-AWARE BS SLEEPING ALGORITHM

A. Preliminaries on Reinforcement Learning

Distributed Q-learning is an online optimization technique that aims at controlling multi-agent systems, i.e., a system featuring M SBSs that take decisions (select the appropriate SM level) in an uncoordinated fashion [11]. Each SBS has to learn independently a policy through real-time interactions with the environment. Q-learning finds the optimal policy in the sense that it maximizes the expected value of the total reward (Q-value) over all successive episodes. The agents (i.e., BSs) have a partial view of the overall system, and their actions may differ since the users are unevenly distributed over the network. In particular, the decision of a BS to choose a SM level is affected by how many users it has to serve, and on how delay-tolerant these users are.

In Q-learning, each agent 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 (in this work, the reward function is the cost function in eq. (10)). This reward is then used to update the Q-value locally, $Q(s_m^t, a_m^t)$, indicating the level of convenience of selecting action a_m^t when in state s_m^t . The Q-value is updated following the update rule:

$$Q(s_m^t, a_m^t) \leftarrow Q(s_m^t, a_m^t) + \alpha [r_m^t + \gamma \max_{a \in \mathcal{A}} Q(s_m^{t+1}, a_m^{t+1}) - Q(s_m^t, a_m^t)] \quad (11)$$

where α is the learning rate that represents the speed of convergence, and $\gamma \in [0, 1]$ is the discount factor that determines the current value of the future state costs.

During the learning phase, each 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:

$$a_m^t = \begin{cases} \operatorname{argmax}_{a \in \mathcal{A}} Q(s_m^t, a_m^t), & \text{if } y > \epsilon \\ \operatorname{rand}(\mathcal{A}), & \text{otherwise} \end{cases} \quad m = 1, \dots, M. \quad (12)$$

By implementing the ϵ -greedy policy, the BS would have explored all possible actions and avoided local minima. For more details on Reinforcement Learning (RL) and Q-learning the reader is referred to, e.g., [11].

B. Proposed Q-Learning Algorithm

At time slot t , the local state of the small cell m is $s_m(t) = \{r_m(t), b_m(t), i_{l,m}(t)\}$. where r_m , b_m and $i_{l,m}$ are the states of the throughput, buffer and interference of SBS m , respectively. Their values are quantized into discrete levels. Indeed, increasing the number of levels leads to more accurate modeling of the system but it also increases the system complexity. In our work and in order to balance between performance and complexity, we quantized R , B and I_l into 4, 3 and 3 levels, respectively. Due to the limited number of state-action pairs, the convergence of the cumulative reward towards the optimal value is not compromised. Therefore, there is no need for function approximations as that performed by Deep Q-Learning (DQL). In order to explore different combinations of states, we make the users move inside their serving cell at each time slot. We define an episode as a fixed simulation run where the users in motion request data continuously with a given minimum rate requirement. During this episode, the users move randomly in the cluster. However, they remain in the same SBS radius area to avoid offloading inside the cluster. Allowing users to freely move inside the SBS cells makes the environment more dynamic, thus an impact on the states of each cell. As a result, the algorithm explores different states to find the optimal policy that minimizes the cost function in (10). After the training phase is complete, a look-up table is generated mapping for every state the best corresponding action. Thus, when the system goes online to operate in real-time, the appropriate action to take for each state is fetched from that table. Since the total number of states is small, the time required to fetch this action is negligible. The Q-Learning algorithm is described in Algorithm 1.

VI. SIMULATION AND RESULTS

In this section, we demonstrate the performance of our proposed Q-learning SBS controller using MATLAB. Our simulator mimics a scenario of one macro BS serving 100 users (unless stated otherwise) covering a cluster of 4 SBSs with 30 users. The macro and cluster users are randomly distributed inside the macro cell and the cluster, respectively. We obtain the results over 200 independent runs, simulating 10 seconds of network activity. The network parameters for simulations were set according to 3GPP specifications [37]. We detail in Table IV the parameters used in our simulation.

A. Convergence analysis

We analyze the convergence of the proposed Q-learning algorithm for different values of w during the offline training phase in figures 3, 4 and 5. In this phase, the system is fed with a training episode that imitates the dynamics of the network. Furthermore, we disable users offloading in the training phase. During this episode, we consider that the users require a service with $R^{min} = 1$ Mbit/s and are moving inside the cluster in order gather enough experience by covering as much as visiting states as possible. Based on their partial view of the environment (cell buffer size, cell throughput and estimated spectral efficiency loss), each SBS takes an action from the

Algorithm 1 : Online Interference-Aware BS Sleeping

```

1: procedure Training ( $Q_m^T(s, a)$ )
2: Initialize the positions of the users and  $q_m(s, a) = 0, \forall m \in \mathcal{M}, \forall s \in \mathcal{S}$  and  $\forall a \in \mathcal{A}$ .
3: Set the weight  $w$ , and the average user velocity.
4: while Learning do
5:   for  $m \in \mathcal{M}$  do
6:     Visit state  $s_m$ .
7:     Select an action  $a_m$  using  $\epsilon$ -greedy rule in (12).
8:     Calculate the cost  $c_m$ .
9:     Observe next state  $s'_m$ .
10:    Update the Q-value  $q_m(s, a)$  from (11).
11:   end for
12: end while
13: end procedure

1: procedure Online
2: for  $u \in$  cluster of SBSs do
3:   if  $r_{MBS,u} \geq R^{min}$  then
4:     Offload  $u$  to the MBS.
5:   end if
6: end for
7: for  $m \in \mathcal{M}$  do
8:    $Q_m(s, a) = Q_m^T(s, a)$ 
9:   Run Q-Learning.
10: end for
11: end procedure

```

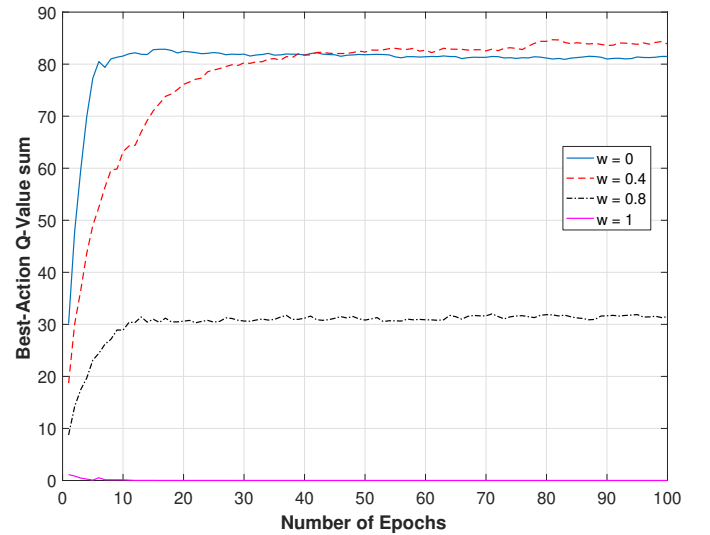


Fig. 3: Evolution of the Q-table with the number of epochs.

action set in \mathcal{A} to minimize the cost function in (10). By repeating this episode a number of epochs, the SBSs learn how to map their observable states to actions.

We present in Fig. 3 the convergence of the algorithm to a stable best action Q-Value sum that varies for different QoS trading factor w . This results from the complexity of finding the optimal trade-off for a given value of w when both energy consumption and data loss are considered. For example, when $w = 0$ or 1, the equation in (10) simplifies to minimizing the

TABLE IV: Simulation parameters

Parameter	Value
Macro/SBS TX power	43 dBm / 30 dBm
Macro/SBS cell radius	500 m / 40 m
Min. dist. Macro-UE / SBS-UE	35 m / 5 m
Min. dist. Macro-Hotspot center	105 m
Min. dist. SBS-SBS	20 m
Macro cell / SBS path loss	ITU UMa / Umi with LOS
Shadowing distribution	Log-normal

Parameter	Value	
Carrier frequency	2.0 GHz (Macro cell) 3.5 GHz (Small cell)	
Bandwidth (Macro and Small cell)	10 MHz	
Latency constraint (TTL)	100 ms	
Number of time slots	10^4	
Q-learning	ϵ (exploration parameter)	0.9
	α (learning rate)	0.9
	γ (discount factor)	0.1

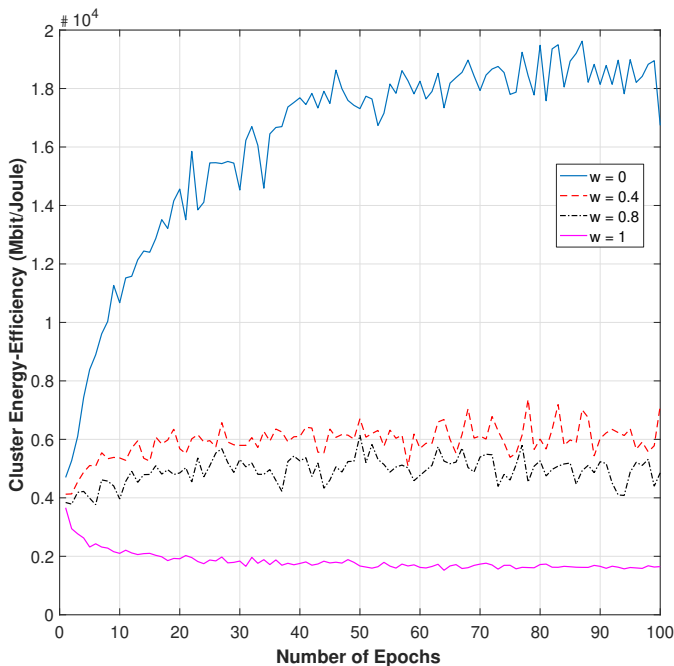


Fig. 4: Evolution of the cluster energy-efficiency with the number of epochs.

energy consumption or the amount of data lost, respectively. In the first case, the agent (i.e., small cell) ignores the co-channel interference between the cells, the size of the data in the buffer and the possible achieved throughput, and focuses only on reducing the energy consumption. In the second case, the SBS ignores the power consumption resulted from the actions chosen and focuses only on serving all the packets. In these special cases, the SBS is able to find the optimal policy faster compared to the other values of w where the agent requires more iterations to capture the trade-off that is influenced by different variables such as interference, buffer size, throughput and energy consumption. Regarding the different asymptotic values of the convergence, it depends on the reward function. For example, when $w = 1$, the reward function converges to zero for zero data loss. However, when $w = 0$ the energy consumption decreases to the lowest power saving value that is greater than zero.

In Fig. 4, we illustrate the evolution of the cluster's EE. The figure shows a decrease in the EE as the weighting parameter

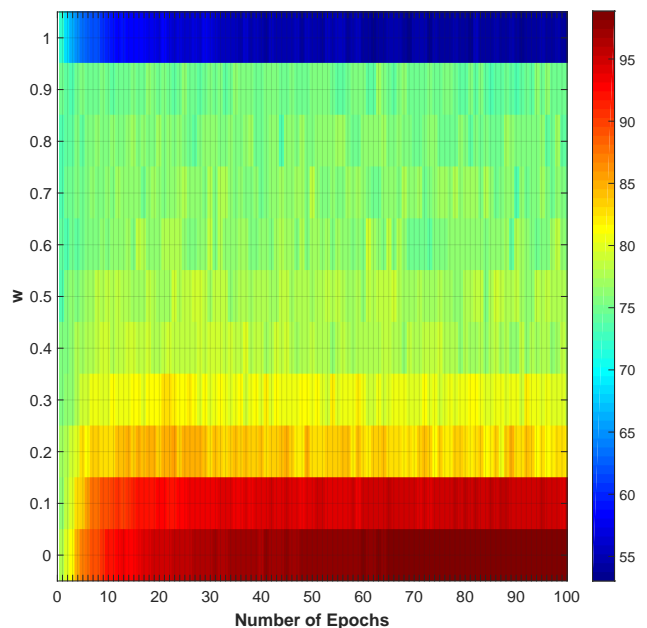
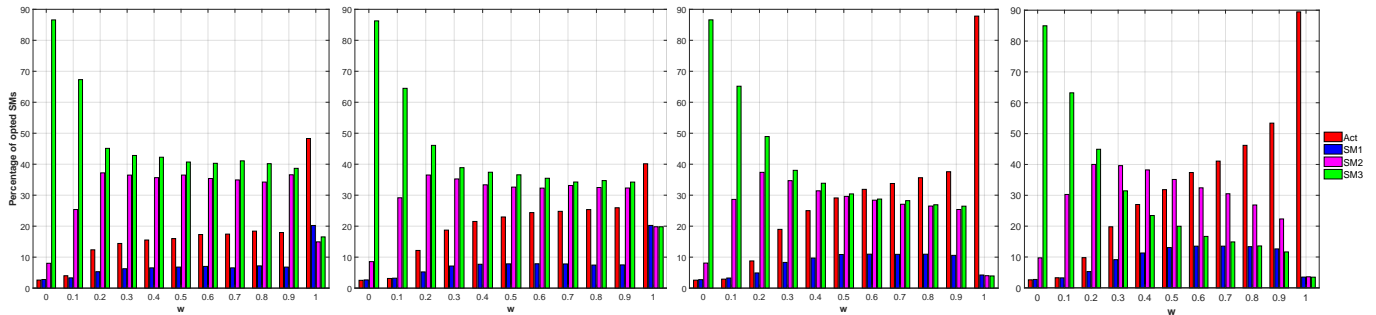


Fig. 5: Heatmap of the small base stations activities for different energy saving-data lost weights in the reward function with the number of epochs. The intensity of the colors represents the percentage of switch-off action (i.e., SM1, SM2 or SM3).

w increases. This is because the small cell tends to stay in SM for a longer duration of time when w is small resulting in lower energy consumption and higher throughput due to lower interference from neighboring cells in SM. This is opposed to the case of high values of w that not only increases the energy consumption, but also the interference between the small cells from staying active for a longer duration of time. The activity of these cells can be depicted from the heatmap in Fig. 5 showing the percentage of time the small cell is in SM. As expected, the percentage of time the SBS switches to SM decreases as w increases. For instance when w approaches zero, the SBSs are in SM for the longest duration of time (above 95%). As w approaches one, the states of the SBSs shifts to active mode to decrease the data lost. We note that in order to minimize the data lost for this setup (i.e., $R^{min} = 1$ Mbit/s), the SBSs can still switch to SM almost half of the time. This will be discussed in more details in the next section.



(a) Rate requirement $R = 0.5$ Mbit/s (b) Rate requirement $R = 1$ Mbit/s (c) Rate requirement $R = 1.5$ Mbit/s (d) Rate requirement $R = 2$ Mbit/s

Fig. 6: Percentage of SMs for different users' minimum rate requirement.

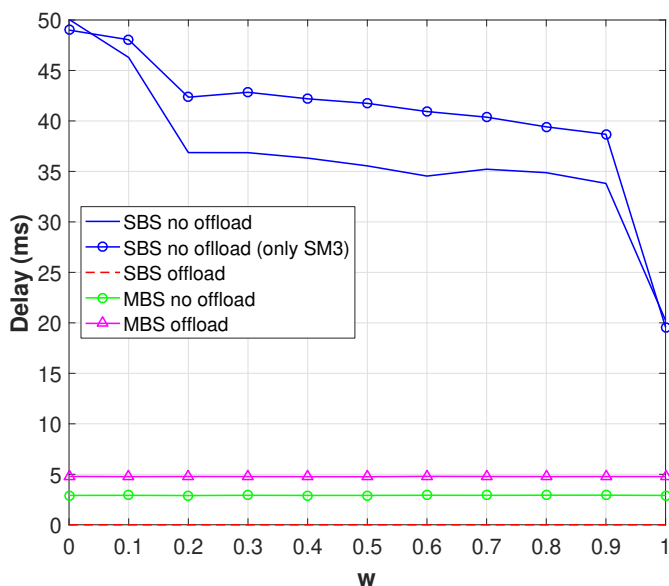


Fig. 7: Delay assessment as a function of w .

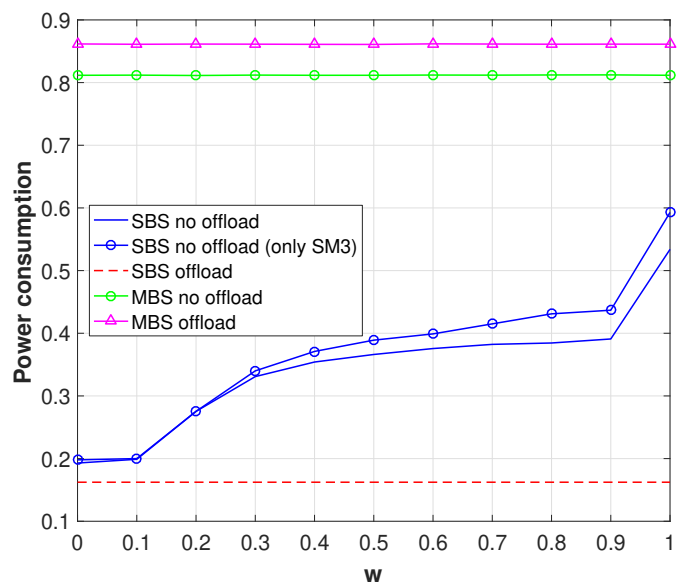


Fig. 8: Normalized power consumption assessment as a function of w .

B. Policy analysis

In this section, we discuss and evaluate the performance of the macro BS and the hotspot cluster. We compare two cases: 1) allowing users in the cluster to offload at the beginning of the first time slot, if the rate requirement condition in eq. (8) is met and 2) no users offloading. First, we report in Fig. 6 the different policies the SBSs take by changing the QoS trading factor (w) and for different users rate requirements without offloading. We observe that the cluster tends to save more energy when w approaches zero by switching to the deepest allowed SM level most of the time, while a gradual shift to active mode is observed when increasing w . On the other hand and by increasing the users rate requirement, the size of the data blocks requested in each time slot increases. This requires the SBS to stay in active mode for a longer period of time as can be depicted in the results in order to serve these blocks in time (i.e., within the delay constraint - TTL) and minimize the data lost. For the case when $w = 1$, each SBS is required to stay active for a minimum period of time to minimize the

data lost. This can be observed more clearly for $R = 0.5$ and $R = 1$ Mbit/s. Once the goal of reducing data lost is satisfied (i.e., $l = 0$), the SBS switches randomly to any of its defined actions. This explains the almost uniform distribution state modes for SM1, SM2 and SM3 after successfully minimizing the data lost. For instance, it is enough to stay active 40% of the time to minimize the data lost for $R = 1$ Mbit/s, while the SBS is required to stay active 90% of the time for $R = 2$ Mbit/s.

In figures 7 to 11, we evaluate the performance of the algorithm for different policies by varying w . We compare the performance between the macro cell and the cluster for the two cases (with and without users offload) and we set $R^{min} = 1$ Mbit/s. From Fig. 7, when cluster users are allowed to offload, the service delay from the SBSs to their users drops to zero while the macro users delay increase from 5.5 ms to almost 7.5 ms. This is because all 30 cluster users are offloaded as it is illustrated in Fig. 12 where 100 macro users correspond

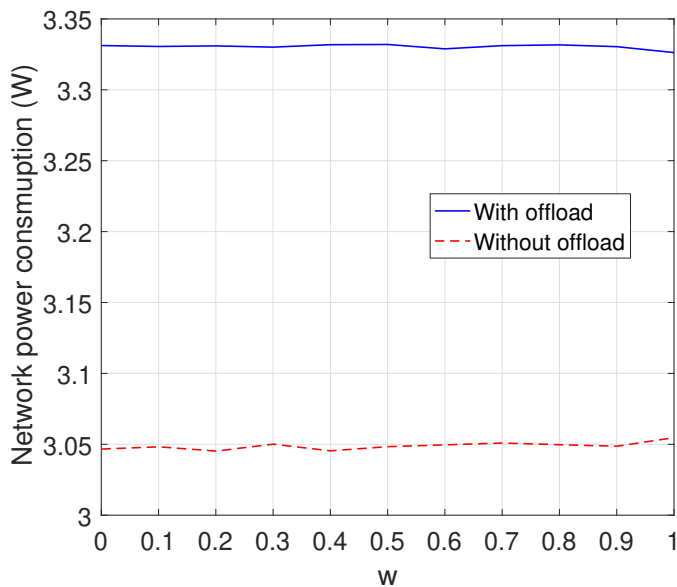


Fig. 9: Network overall power consumption assessment as a function of w .

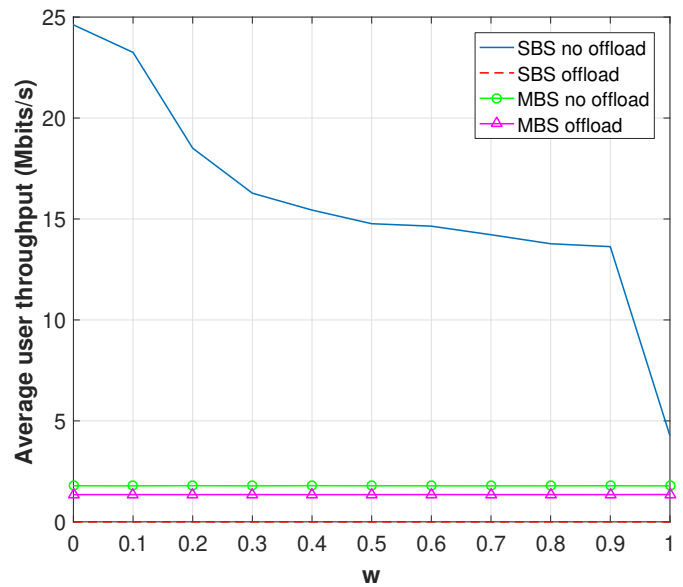


Fig. 11: Average user throughput assessment as a function of w .

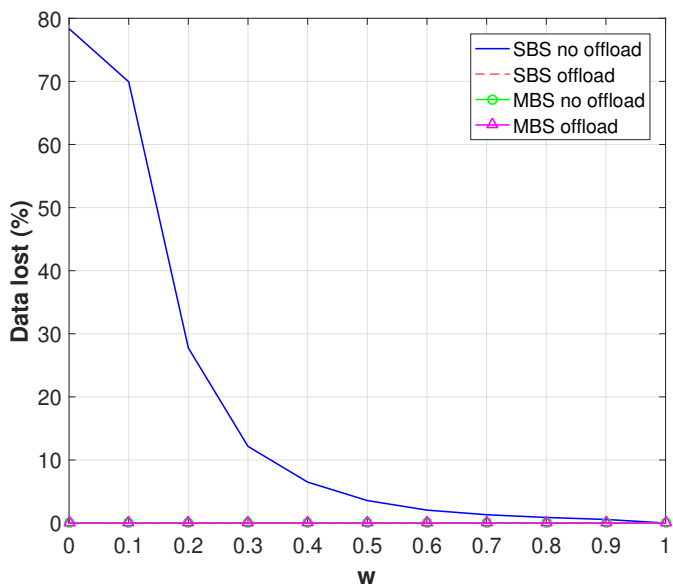


Fig. 10: Data lost assessment as a function of w .

to a macro traffic load of 63%. Similarly and as depicted in Fig. 8, the normalized power consumption of the cluster drops when all users are offloaded since the small cells switch to the deepest SM level when there are no users to serve. On the other hand, the power consumption of the macro cell increases as its traffic load rises from 100 users (63% traffic load) to 130 users (81% traffic load). We note that the small and macro cells power consumption are normalized independently to the case where the former is always on (that is the MBS without SM) and the latter (SBS) is at full load. Finally, we compare our proposed multi-level SM scheme with a baseline that considers only SM3. We will note this scheme as binary sleep scheme. We illustrate the case with no offload to highlight the impact of SM on the cluster delay and energy consumption. The results

show that our multi-level SM scheme outperforms the binary sleep scheme. This is due to the added flexibility with multi SM levels.

Even though the average delay of the requested data in Fig. 7 falls under the delay constraint limit of 100 ms for all the discussed cases, around 80% of these requested data are dropped (lost) for small values of w when the users stay connected to the cluster as shown in Fig. 10. This is because the average delay calculated does not take into account the dropped packets. As for the case of macro users or when the cluster users are offloaded, there are no dropped packets as the macro cell, that is always active, is able to serve the packets (data blocks) to its users under the TTL constraint.

In the above discussion, it is clear that offloading the users to the MBS improves the delay and minimizes the loss of packets for the users. However, this comes at a cost. First, even though the power consumption of the small cells is reduced when all of their users are offloaded, the increase in the macro cell power consumption due to the traffic increase is more significant than the energy savings from the small cells. This is because the macro cell consumes a lot more energy than the small cell, which can be verified in Fig. 9. Fig. 9 shows the overall network power consumption during the observation time. It is clear that when the users offload to the macro cell, the network power consumption increases by 9.3% compared to the no offload scenario. We also notice that when the users offload, the network power consumption is almost constant regardless of w . This is because our parameters setup (100 macro users and $R^{min} = 1$ Mbit/s) allows all cluster users to offload while guaranteeing enough radio resources from the MBS. Hence the SBSs are always in SM3 since they are not serving any users. On the other hand and for the other scenario (without offload), the network power consumption is slightly increased with w due to the increase in the cluster power consumption. This is because the power consumption

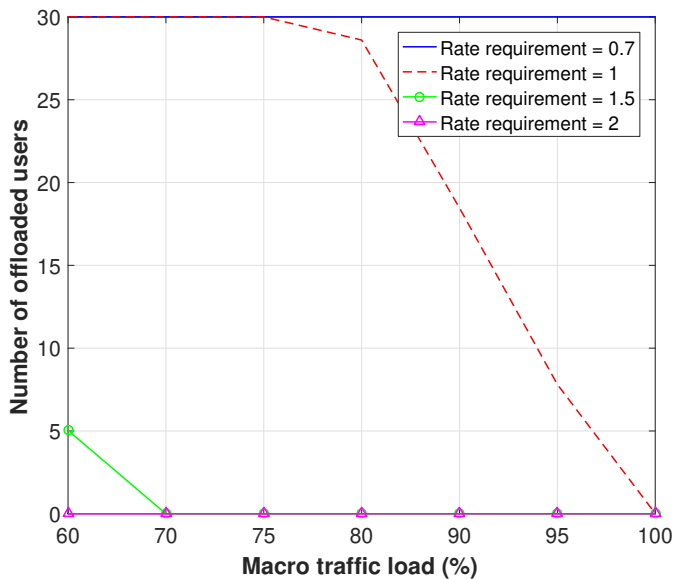


Fig. 12: Number of offloaded cluster users as a function of Macro traffic load.

of the macro cell dominates the network power consumption. Second, the idea of having a hotspot is to boost the capacity in a designated area, hence when the users are offloaded from the cluster their average rate decreases down to 10 times as shown in Fig. 11 because the macro cell that is initially serving 100 users spreads over an area of radius 500 meters providing less throughput than the cluster serving 30 users spread around an area of radius 40 meters per small cell. The reason behind having a high average throughput when w is small is due to the interference limitation from having more cells in SM which improves the SINR. This decrease in the user throughput along with the increase in the overall network energy consumption results in a decrease in the network EE as will be discussed in the next section.

C. Network performance assessment

In Fig. 12, we illustrate the number of offloaded users as a function of the macro cell traffic load and for different user rate requirement. Since, a user does not offload unless the macro cell guarantees his/her minimum rate requirement, less users are expected to be offloaded as the macro traffic load or the rate requirement increases. In particular, when $R = 2$ Mbit/s, no users are offloaded since the macro cell cannot guarantee such rate to the offloaded users.

In Fig. 13, we present the network EE and compare the two scenarios (with and without offload). As previously explained, the users in the cluster enjoy higher throughput. Moreover since the SBSs, alternating to different SM levels, consume less energy than the active MBS, the network EE for the no offload scenario is higher, up to 5 times when $w = 0$, compared with the offload scenario. As the macro traffic cell increases less cluster users are prone to offload until the macro cell is fully loaded and no users are offloaded. This explains why for high macro traffic load, the network EE for the two scenarios meet.

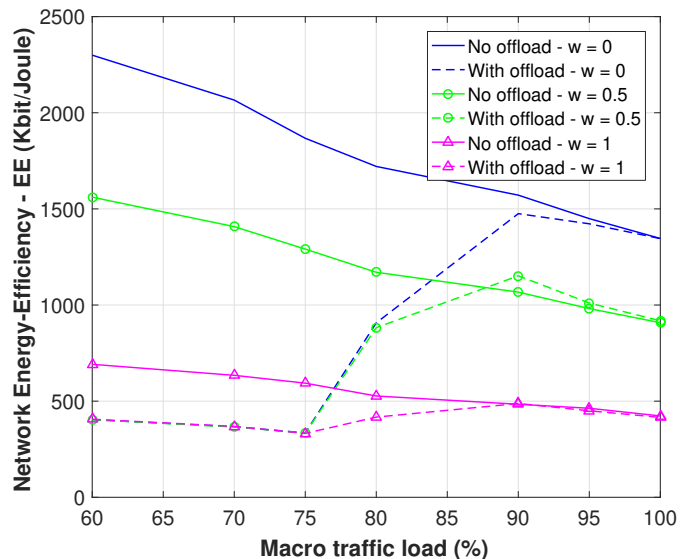


Fig. 13: Network energy efficiency assessment.

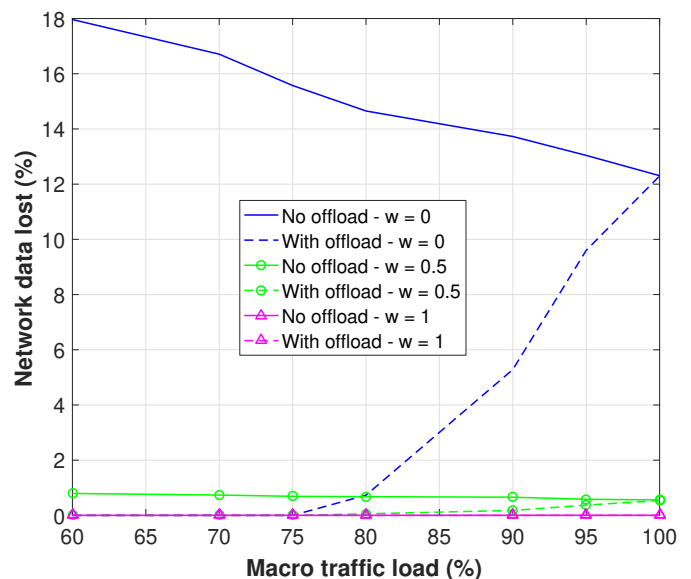


Fig. 14: Network data lost assessment.

In figures 14 and 15, we illustrate the network data lost and delay as a function of the macro traffic load, respectively by averaging over all network's users. As noted in the previous section, offloading users from the small cells to the macro cell enhances their experience in terms of reduced data lost and service delay for $R^{min} = 1$ Mbit/s. This can be clearly observed since the macro cell does not switch off as in the case of the small cells. Furthermore, a user does not offload to the macro cell unless enough radio resources are provided to him/her. Hence, the QoS is not affected.

VII. CONCLUSION

In this paper, we study the problem of energy saving optimization in 5G BSs by implementing multi SM levels. We consider a HetNet architecture where a MBS covers a cluster of SBSs. Given the distinct power and time consumption

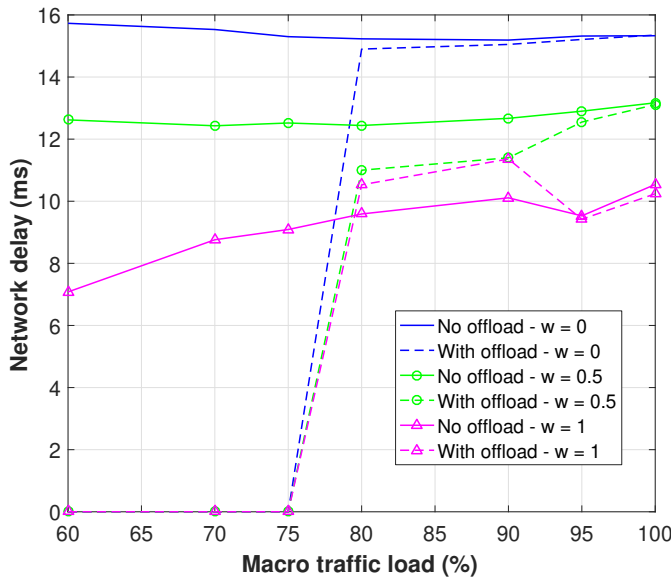


Fig. 15: Network delay assessment.

characteristics of these modes, we propose a distributed Q-learning algorithm that is able to learn the best policy to control the activity of the SBSs in order to minimize the energy consumption and data dropping. Thus, we investigate the trade-off between energy saving and QoS. Results show that the cluster energy consumption can be significantly reduced if the operator prioritizes energy savings at the cost of higher dropping rate. Moreover, we analyze the impact of offloading users from the small cells to the macro cell on the network performance. We show that under low traffic load, offloading users to the MBS reduces significantly their service delay and dropping rate. However, even though this reduces the energy consumption of the cluster, it comes with an increase in the network energy consumption and a decrease in the average user throughput resulting in a drop in the network EE. This could open the door to investigate load management algorithms for small cells solely equipped with renewable energy sources and batteries where EE could be sacrificed for a short duration of time to maintain continuous users' connectivity.

REFERENCES

- [1] Ericsson, "Ericsson Mobility Report," [Online]. Available: <https://www.ericsson.com/49da93/assets/local/mobility-report/documents/2020/june2020-ericsson-mobility-report.pdf>, Jun. 2020.
- [2] Q. C. Li, H. Niu, A. T. Papathanassiou, and G. Wu, "5G network capacity: Key elements and technologies," *IEEE Vehicular Technology Magazine*, Mar. 2014.
- [3] M. Masoudi, M. G. Khafagy, A. Conte *et al.*, "Green mobile networks for 5G and beyond," *IEEE Access*, Aug. 2019.
- [4] M. Feng, S. Mao, and T. Jiang, "Base station on-off switching in 5G wireless networks: Approaches and challenges," *IEEE Wireless Communications*, Aug. 2017.
- [5] S. Buzzi, I. Chih-Lin, T. E. Klein, H. V. Poor, C. Yang, and A. Zappone, "A survey of energy-efficient techniques for 5G networks and challenges ahead," *IEEE Journal on Selected Areas in Communications*, Apr. 2016.
- [6] J. Wu, Y. Zhang, M. Zukerman, and E. K. Yung, "Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey," *IEEE Communications Surveys Tutorials*, Feb. 2015.

- [7] M. Miozzo, L. Giupponi, M. Rossi, and P. Dini, "Switch-on/off policies for energy harvesting small cells through distributed Q-learning," in *2017 IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*, May 2017.
- [8] 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)*, Jul. 2018.
- [9] 3GPP TSG RAN, "TR 37.910, Study on new radio access technology: Radio access architecture and interfaces," Mar. 2017.
- [10] B. Debaille, 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] R. S. Sutton and A. G. Barto, *Introduction to Reinforcement Learning*, 1st ed. Cambridge, MA, USA: MIT Press, 1998.
- [12] F. Salahdine, J. Opadere, Q. Liu, T. Han, N. Zhang, and S. Wu, "A survey on sleep mode techniques for ultra-dense networks in 5G and beyond," *Computer Networks*, 2021.
- [13] J. Peng, P. Hong, and K. Xue, "Stochastic analysis of optimal base station energy saving in cellular networks with sleep mode," *IEEE Communications Letters*, Mar. 2014.
- [14] I. Ashraf, F. Boccardi, and L. Ho, "Sleep mode techniques for small cell deployments," *IEEE Communications Magazine*, vol. 49, no. 8, 2011.
- [15] M. Feng, S. Mao, and T. Jiang, "Boost: Base station on-off switching strategy for green massive MIMO hetnets," *IEEE Transactions on Wireless Communications*, Sep. 2017.
- [16] P. Ren and M. Tao, "A decentralized sleep mechanism in heterogeneous cellular networks with QoS constraints," *IEEE Wireless Communications Letters*, 2014.
- [17] X. Guo, Z. Niu, S. Zhou, and P. Kumar, "Delay-constrained energy-optimal base station sleeping control," *IEEE Journal on Selected Areas in Communications*, Jan. 2016.
- [18] Y. L. Che, L. Duan, and R. Zhang, "Dynamic base station operation in large-scale green cellular networks," *IEEE journal on selected areas in communications*, Aug. 2016.
- [19] H. Jiang, S. Yi, L. Wu, H. Leung, Y. Wang, X. Zhou, Y. Chen, and L. Yang, "Data-driven cell zooming for large-scale mobile networks," *IEEE Transactions on Network and Service Management*, Feb. 2018.
- [20] C. Luo and J. Liu, "Load based dynamic small cell on/off strategy in ultra-dense networks," in *2018 10th International Conference on Wireless Communications and Signal Processing (WCSP)*, Oct. 2018.
- [21] G. Vallero, D. Renga, M. Meo, and M. A. Marsan, "Greener RAN operation through machine learning," *IEEE Transactions on Network and Service Management*, Jun. 2019.
- [22] 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)*, Sep. 2015.
- [23] A. I. Abubakar, M. Ozturk, S. Hussain, and M. A. Imran, "Q-learning assisted energy-aware traffic offloading and cell switching in heterogeneous networks," in *2019 IEEE 24th International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD)*, Oct. 2019.
- [24] R. Li, Z. Zhao, X. Chen, J. Palicot, and H. Zhang, "Tact: A transfer actor-critic learning framework for energy saving in cellular radio access networks," *IEEE transactions on wireless communications*, 2014.
- [25] H. Sayed, A. El-Amine, H. A. H. Hassan, L. Nuaymi, and R. Achkar, "Reinforcement learning for radio resource management of hybrid-powered cellular networks," in *2019 International Conference on Wireless and Mobile Computing, Networking and Communications (WiMob)*, Dec. 2019.
- [26] A. De Domenico and D. Kténas, "Reinforcement learning for interference-aware cell dtx in heterogeneous networks," in *2018 IEEE Wireless Communications and Networking Conference (WCNC)*, Jun. 2018.
- [27] M. Ozturk, A. I. Abubakar, J. P. B. Nadas, R. N. B. Rais, S. Hussain, and M. A. Imran, "Energy optimization in ultra-dense radio access networks via traffic-aware cell switching," *arXiv preprint arXiv:2007.04133*, 2020.
- [28] J. Ye and Y.-J. A. Zhang, "Drag: Deep reinforcement learning based base station activation in heterogeneous networks," *IEEE Transactions on Mobile Computing*, 2019.
- [29] H. Li, H. Gao, T. Lv, and Y. Lu, "Deep Q-learning based dynamic resource allocation for self-powered ultra-dense networks," in *2018 IEEE International Conference on Communications Workshops (ICC Workshops)*, Jul. 2018.
- [30] 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)*, Apr. 2018.

- [31] 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)*, Oct. 2019.
- [32] 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 *2019 IEEE 30th Annual International Symposium on Personal, Indoor and Mobile Radio Communications (PIMRC)*, Nov. 2019.
- [33] 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 *2020 IEEE 31st Annual International Symposium on Personal, Indoor and Mobile Radio Communications*, Oct. 2020.
- [34] A. El-Amine, P. Dini, and L. Nuaymi, “Reinforcement learning for delay-constrained energy-aware small cells with multi-sleeping control,” in *2020 IEEE International Conference on Communications Workshops (ICC Workshops)*, Jul. 2020.
- [35] G. Jang, N. Kim, T. Ha, C. Lee, and S. Cho, “Base station switching and sleep mode optimization with lstm-based user prediction,” *IEEE Access*, vol. 8, 2020.
- [36] 3GPP TSG RAN, “3GPP TR 36.872, Technical Specification Group Radio Access Network; Small cell enhancements for E-UTRA and E-UTRAN- Physical layer aspects (Release 12),” Dec. 2013.
- [37] —, “3GPP TR 36.814, Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects (Release 9),” Mar. 2017.



Ali El Amine received his Ph.D. degree in computer and networking from IMT Atlantique (formerly, Telecom Bretagne), Rennes, France in 2019. He also received his B.E. degree in electrical engineering and M.S. degree in computer engineering from the Lebanese American University (LAU) in 2013 and 2016, respectively. He is currently a postdoctoral fellow in CNRS-LAAS, Toulouse, France. His research focuses on green wireless communication, cellular networks, energy efficiency, network slicing optimization, NFV and machine learning.



Jean-Paul Chaiban received his M.S. and B. S. degree in computer engineering from the American University of Science and Technology (AUST), Lebanon. He is currently continuing his studies in the American University of Beirut (AUB), Lebanon. His research focuses on wireless communications, cellular networks, cybersecurity and machine learning.



Hussein Al Haj Hassan received the Engineering degree in telecommunication and computer from the Lebanese University and the master degree in networks of telecommunications from the Lebanese University and Saint Joseph University, Beirut, Lebanon in 2011 and 2012 respectively. He received his Ph.D. degree in Networking from IMT Atlantique (Telecom Bretagne), Rennes, France in 2015. He was a faculty member at the American University of Science and Technology between 2016 and 2021. He is currently a standardization and

research engineer at Acklio, Rennes, France. His research focuses on wireless communication, cellular networks IoT, energy efficiency and machine learning.



Paolo Dini received MSc and PhD from Università di Roma La Sapienza, in 2001 and 2005, respectively. He served as a Post-Doc Researcher at the Research Centre on Software Technology (RCOST) - Università del Sannio, and contracted Professor at Università di Roma La Sapienza in 2005. He is now with the Centre Tecnològic de Telecomunicacions de Catalunya (CTTC) as a Senior Researcher. He received two awards from the Cisco Silicon Valley Foundation for his research on heterogeneous mobile networks in 2008 and 2011, respectively. He has

been involved in over 25 research projects related to network management, optimization and energy efficiency. He is currently the Coordinator of the EU H2020 MSCA SCAVENGE European Training Network on sustainable mobile networks with energy harvesting capabilities. His research interests include sustainable networking and computing, distributed optimization and optimal control, machine learning and data analytics.



Loutfi Nuaymi (Senior IEEE member) received the Engineering Diploma degree in telecommunication from the Saint Joseph University, Beirut, Lebanon in 1993. He got his PhD in Telecommunication from the Ecole Nationale Supérieure des Télécommunications (ENST), now Telecom Paris Tech, Paris, France in 2001. He is Full Professor at IMT Atlantique (previous names: Telecom Bretagne and ENST Bretagne), Rennes, France. His fields of interest are radio resource management, energy-efficiency and machines cellular coverage in wireless networks. He

is the author of ‘WiMAX’, published by Wiley (January 2007) and many journal and conference papers.



Roger Achkar is the Dean of the Faculty of Engineering at the American University of Science and Technology and an Associate Professor in Electrical Engineering. Dr. Achkar received a Ph.D degree in Energetic System and Information from the University of Technology of Compiègne, France, in 2008. A Master degree in Research (DEA) from the Lebanese University, Faculty of Engineering, in 2004 and a B.E. degree in Electrical Engineering from the Lebanese University, Faculty of Engineering, Branch II, Roumieh, Lebanon in 2002. Dr.

Achkar is a member of the Order of Engineers and Architects in Lebanon since 2005 and Treasurer of the IEEE Lebanon Section. His Research interests are in Electrical, Communications and Mechanical Engineering Education. He developed and applied, on an active magnetic bearing (AMB), an artificial neural network able to control non-linear systems. Currently, he is leading several research projects in the area of Artificial Intelligence, Machine Learning, IOT, Communications, Vision and Control. Dr. Achkar played a major role, as a Chairman of the Computer and Communications Engineering (CCE) Department, in receiving the ABET Accreditation for the Program at the American University of Science and Technology. He organized many conferences, workshops and trainings in the fields of Communication, Robotics, Programming and Artificial Intelligence. He was responsible of establishing many MOUs with different national and international companies, and co-operation agreements with competitive national and international programs in Computer and Communications, Petroleum, Electrical and Mechanical Engineering.