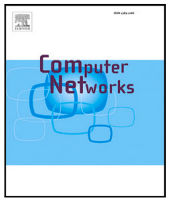
Base Station switching and edge caching optimisation in high energy-efficiency wireless access network

Computer Networks 192 (2021) 108100

Contents lists available at ScienceDirect

Computer Networks

journal homepage: www.elsevier.com/locate/comnet



Base Station switching and edge caching optimisation in high energy-efficiency wireless access network

Greta Vallero [a,*], Margot Deruyck [b], Michela Meo [a], Wout Joseph [b]

[a] Politecnico di Torino, Department of Electronics and Telecommunications, C.so Duca degli Abruzzi 24, 10129 Torino, Italy
[b] Ghent University, IMEC-WAVES, Technologiepark-Zwijnaarde 126, 9052 Ghent, Belgium

ARTICLE INFO

Keywords:
Radio Access Network
Multi-access Edge Caching
Multi-access Edge Computing
Renewable energy
Energy efficiency

ABSTRACT

The improvement of the energy efficiency and the reduction of the latency are two of the main goals of the next generation of Radio Access Networks (RANs). In order to achieve the latter, Multi-access Edge Computing (MEC) is emerging as a promising solution: it consists of the placement of computing and storage servers, directly at each Base Station (BS) of these networks. For the RAN energy efficiency, the dynamic activation of the BSs is considered an effective approach. In this paper, the caching feature of the MEC paradigm is considered in a portion of an heterogeneous RAN, powered by a renewable energy generator system, energy batteries and the power grid, where micro cell BSs are deactivated in case of renewable energy shortage. The performance of the caching in the RAN is analysed through simulations for different traffic characteristics, as well as for different capacity of the caches and different spread of it. New user association policies are proposed, in order to totally exploit the MEC technology and reduce the network energy consumption. Simulation results reveal that, thanks to this technology and the proposed methodologies, the experienced delay and the energy consumption drop, respectively, up to 60% and 40%.

1. Introduction

The Multi-access Edge Computing (MEC) technology, also known as Mobile Edge Computing, has been introduced to push computing and storage resources in physical proximity of end users, placing servers on the edges of the network [1]. In this way, the execution of applications, the pre-processing of data and the caching of popular contents are performed in proximity of end users. In Radio Access Networks (RANs), these servers are co-located on Base Stations (BSs). Therefore, these infrastructures are able to provide storage and computation services, in addition to access services [2]. Several benefits are derived from the introduction of this technology [3,4]. First, the backhaul traffic load is reduced, since the access to the cloud is unneeded. Second, the quality of the multimedia content can be adapted to the user's channel. Finally, as proved in [5–7], the latency is reduced, which is one of the objectives of the 5G. Because of these advantages, this technology has been largely investigated in literature. Many works focus on the optimisation file placement in caches located on BSs of RANs, with the goal of minimising the content delivery delay [8–10].

Another key goal of the 5th generation of networks, is the improvement of the energy efficiency [11]. Indeed, 5G systems aim at consuming a fraction of the energy consumption of 4G mobile networks, even if the amount of traffic which 5G networks are supposed to manage is much larger than in 4G ones, as highlighted in [12]. According to [12], the mobile IP traffic will reach 77.5 exabyte (EB) per month by 2022, which is an enormous increase compared to 11.5 EB per month in 2017. For this reason, the capacity of the 5G networks, is expected to increase by a factor 1000 more than 4G networks, supporting up to 9 billion of mobile devices, and an heterogeneous range of applications, services and devices [13]. The network energy efficiency has been recognised as a fundamental and urgent aspect of the communication community, since 80% of the total mobile network is consumed by mobile access equipment. As reported in [13], from this 80%, 90% is consumed by the BSs of these networks, whose energy consumption is an important actor of the Operational Expenditure (OPEX), which would grow because of the RAN densification, planned with the 5G RAN deployment [14,15]. In addition to this, the increase of the RAN energy consumption will contribute to the increase of the carbon emissions, generated during the energy production, which significantly contribute to the climate changes. For this reason, the design of energy efficient RANs has been receiving a lot of attention for many years. The European Commission, in [16], under the need for actions to improve the energy efficiency in communications, formalises a policy that regulates the energy consumption and carbon emissions for Broadband Communication Equipment. Meanwhile, in literature,

* Corresponding author.
E-mail address: greta.vallero@polito.it (G. Vallero).

https://doi.org/10.1016/j.comnet.2021.108100
Received 19 November 2020; Received in revised form 10 March 2021; Accepted 6 April 2021
Available online 20 April 2021

many works address this issue through the dynamic allocation of the network resources [17–21]. Indeed, the typical behaviour of the daily traffic demand presents short peaks and long valleys, during which the capacity of the RAN is under-utilised, since the traffic demand is very low. Therefore, during these periods, the unneeded capacity is deactivated, allowing energy saving [17]. Another trend proposes local Renewable Energy Sources (RES) as power supply of RANs, e.g. a wind turbine and/or a Photovoltaic (PV) panel system. This makes these networks self-sufficient and more sustainable, since the amount of energy that is produced by burning fossil fuels reduces [22]. Recently, these two approaches have been combined, so that the BSs of the RAN, supplied by RES, are dynamically switched to sleep mode, when the traffic demand is low, as in [18], or when the amount of renewable energy that is generated by RES is not enough to power the RAN [19].

The energy efficiency in RAN and the employment of the MEC technology in these networks have been largely investigated in the literature, but the impact of the MEC technology employment on the network energy efficiency is usually neglected. Meanwhile, the effect of the BSs switching on the MEC technology performance is ignored, as well. Indeed, these two topics are typically considered separately and their coexistence has not been investigated yet. For this reason, in our previous work, presented in [23], the simultaneous employment of the MEC technology and BSs switching is considered, providing an overview of their mutual effects. The growth of the energy consumption due to the supply of the MEC servers installed on each BS is analysed, as well as the effects of the BSs switching, and consequently of the MEC server deactivation, on the experienced delay. In this paper, we deepen the analysis of the energy saving strategy and we use association procedures, which aim at further improving the network energy efficiency, as well as the latency reduction. The MEC switching is also introduced to guarantee the load balancing among BSs. To do this, we considered a portion of an heterogeneous RAN in the city centre of Ghent, in Belgium, which dynamically adapts its capacity to the traffic demand. It is composed of a set of macro cell BSs, each supported by 4 micro cell BSs and powered by a PV panel system, energy batteries and the power grid. Each BS of the considered RAN is equipped with a caching server, where the most popular contents are stored. In case the renewable energy generation is not sufficient for the network supply, the micro BSs of the network are deactivated. The contributions of this work are:

- Using a simulation-based approach, we quantify the gain as well as the cost, which derive from the usage of caching servers, placed at each BS. The gain is measured in experienced delay drop, and the cost is expressed in growth of the energy consumption of the network. These quantities are also evaluated when an energy reducing strategy is used, which deactivates each micro cell BS in case of local renewable energy shortage.
- By simulations, we derive the impact of the different traffic characteristics and of the different capacities of each cache.
- Different spread of the cache capacity among the BSs of the network is investigated and we observe that caching on the macro BSs is always needed to significantly reduce delays, while caching also on the micro cells relieves the effort on the macro cell. This allows the micro cells to often respond without any involvement of the macro cell, providing a slight delay reduction.
- Different association policies are proposed, which aim at minimising the RAN energy consumption and/or the experienced delay, in order to maximise the benefits provided by the MEC technology usage, ensuring also the achievement of energy efficiency.

The paper is organised as follows. Related works are revised in Section 2. In Section 3, the scenario and the methodology of our work are presented, while the used Key Performance Indicators (KPIs) are described in Section 4. Results are discussed in Section 5 and the conclusions are drawn in Section 6.

2. State of the art

2.1. Energy efficiency in wireless access networks

The employment of the RES for the power supply of the BSs of RANs has been receiving much attention because it reduces the carbon emissions and of the electricity bill [24–26]. Various papers address the critical issue of properly dimensioning RE generation systems to power mobile networks [24–26]. The sizing process brings to a trade off among self-sustainability, cost and feasibility constraints due to the installation of a RE generation system. The RE system sizing problem for powering BSs of an heterogeneous RAN is formalised in [24], aiming at the minimisation of costs, and it is solved through an heuristic algorithm. A Markov reward process model is used in [25], to study a BS power system. This work demonstrates the deployability of green BSs in an urban environment, since even a small area solar panel and a battery with a connection to the power grid, is enough for a green BS operation. In [26], the optimal power supply to achieve the minimum annual OPEX and carbon emissions is treated, considering a diesel generator, wind turbines, PV panel and the grid as energy sources.

The Resource on Demand (RoD) strategies dynamically adapt the available radio resources to the current traffic demand, activating/deactivating the BSs according to the traffic demand, in order to make mobile network consumption more load proportional, thus allowing to reduce the capacity of the required RE generators and to increase the feasibility of the RE powering system. Overviews of the RoD approaches can be found in [27–29]. Depending on the objectives, various sleep mode based algorithms can be applied for managing the radio resources [18,30–34]. In [18], the aim of the RoD strategy is to adapt the energy consumption to the actual traffic load, to reduce the grid energy demand and to limit the operational cost. Similarly, authors of [30] propose a framework, whose goal is the minimisation of the network power consumption, to efficiently allocate spectrum resources to users, switching off unneeded BSs. In [31] and [32] an ON/OFF switching, based on reinforcement learning and on a time-varied probabilistic algorithm, respectively, is deployed to optimise the self-sustainability of the RAN, powered by RE. With the purpose of reducing the electricity bill and to provide ancillary services, authors in [33] and [34] apply RoD strategies in a green RAN to improve the interaction with the smart grid in a demand response framework. The work, discussed in [35], dynamically allocates resources of a RAN, aiming at the minimisation of the network power consumption, taking into consideration the delivery deadline of the user requests. The optimisation problem is formulated and then, because of its exponential complexity, solved with a greedy algorithm.

Closely related to these papers are our previous works presented in [36–40]. In [36], a capacity-based deployment tool for the design of energy-efficient wireless access networks is designed, which is used in [37] to investigate a wireless access network powered by a PV panel system, where additional sleep mode periods for the BSs are introduced, to cope with the renewable energy shortages. In [38], analytical models are deployed to predict the green network performance as a function of PV panel size and storage capacity. In [39], in order to reduce the operational cost of the network, micro BSs of an heterogeneous RAN enter the sleep mode when the energy is expensive. BSs switching is performed in [40], according to Neural Network traffic predictions, in order to minimise the RAN energy consumption.

2.2. Multi access Edge Computing

The MEC technology that uses computing and caching power at the edges of the network, has received much attention in the literature. It provides several benefits, such as the reduction of the experienced latency and of the load in the core network [3,4]. In [41,42], surveys of the MEC technology utilisation are provided. The overview presented in [41] distinguishes the MEC mechanisms into mechanisms

Fig. 1. Considered portion of RAN of the city centre of Ghent (Belgium), composed by 8 macro cell BSs, each supported by 4 micro cell BSs.

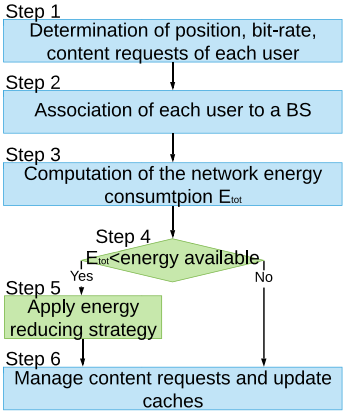


Fig. 2. Different steps of the simulations.

for computation offloading and mechanisms for caching, while [42] mainly focuses on caching features, discussing in details the caching optimisation and the content insertion/expulsion policies. The reduction of the latency given by the adoption of the MEC paradigm, used for content caching, in wireless networks is demonstrated in [5–7]. In [5], different use cases show the drop of the delay and backhaul links utilisation, with the employment of the MEC paradigm, whereas the work presented in [6] highlights that the spectral and energy efficiency are improved with caching at the wireless edges. Authors in [7] propose the Dynamic Programming based Adaptive Caching Algorithm, which, besides the minimisation of the experienced delay, optimise also the provided video quality. Many works formulate an optimisation problem to select the contents to cache in order to improve the network performance, when the MEC technology is used, [3,9,10,43–47]. Authors in [43] aim at minimising the experienced delay in an heterogeneous RAN, where caching servers are placed on each BS. In [44], the optimisation problem maximises the local hit, proposing a reduction of the problem in order to treat it analytically. With the same objective, in [3], the optimal content placement problem is given. The work discussed in [45] formulates through an integer programming problem the most appropriate content placement in order to optimise the system towards the hit occurrences and the power consumption, considering users mobility. Also in [46], the experienced delay and power consumption are jointly optimised and results are obtained through simulations. In [47], the allocation of femtocells and WiFi offloading, used as helper where some contents are cached, is optimised in order to minimise the time needed for each download. [9] proposes a content caching replacement algorithm, which uses predictions of the future request references, derived from a polynomial fit algorithm. Authors in [10] formulate an optimisation problem, which minimise the network cost and content delivery delay, in order to determine which chunks of a content should be cached and how they should be transcoded. The problem is solved applying the relaxation to the constraints and proposing an heuristic.

3. Methodology

In this work, the heterogeneous RAN portion considered in [22], covering an area of 0.3 km^2 of the city centre of Ghent, in Belgium, is considered (orange rectangle in Fig. 1).

The RAN that covers this area is composed of 8 macro cell BSs, marked by the blue points in Fig. 1. In order to provide additional capacity during high traffic demand periods, each macro cell BS is supported by 4 micro cell BSs. These are indicated with the brown points in the figure and their radio coverage overlaps with the macro cell. The considered wireless technology is the Long Term Evolution-Advanced (LTE-A), whose frequency and channel bandwidth are 2.6 GHz and 5 MHz, respectively. Single Input Single Output (SISO) antennas for both the micro and macro cell BSs are considered. The link budget is reported in Table 1, taken from [48].

Each BS of the network is equipped with a caching server, to push contents closer to the users. The capacity of the caching server installed on each macro is double the one installed in each micro BS. Similar to [49], the hardware technology of each cache is DRAM (Dynamic Random Access Memory). As in [46] and [47], these servers update their contents according to the Least Frequently Used (LFU) cache algorithm, which aims at storing the most popular contents. We assume that the file library is composed of 1000 files, with size equal to 100 Mbit. Similarly to [22,50], the considered portion of RAN is supplied by a centralised PV panel system, an energy battery, and the power grid. We consider a poly- or mono-crystalline silicon PV-panel system, whose capacity and energy efficiency are, respectively, 100 kWp and 20%, as in [51]. Its energy production is retrieved from PVWATT [52], which provides the hourly energy production, considering realistic solar irradiation patterns, during the typical meteorological year in the considered area, accounting for 14% the main typical losses occurring in a real PV panel system during the process of solar radiation conversion into electricity. In this study, the energy production is typical for the winter season is considered, in order to analyse the worst case scenario in terms of produced energy. In particular, the data collected during the week from 3 January to 9 January, in Turin (Italy) are considered in our simulations. The battery has effective size equal to 50 kWh. Notice that the actual energy storage capacity is 71–100 kWh, since we consider a maximum Depth of Discharge of 70% or 50%, respectively, to ensure the maximum battery life. The energy generated by the PV panel is used to power the BSs and their cache and when the production exceeds the BSs energy consumption, the unused energy is conserved into the battery. In case no renewable energy is available for the network supply, the required energy is taken from the power grid.

In our simulations, the steps reported in Fig. 2 are performed in each time slot, with a duration of 1 h. First, the traffic is generated: the users, their position, the required bit rate and the requested contents are determined (Step 1 in Fig. 2). Once the traffic has been generated, each user is associated with a BS (Step 2 in Fig. 2), if possible. Then, for each hour, the network energy consumption is computed (Step 3 in Fig. 2). If the hourly needed energy exceeds the available renewable energy (Step 4 in Fig. 2), a strategy is applied during that hour to reduce the energy consumption of the network (Step 5 in Fig. 2). Finally, the requested contents are delivered and caches are updated (Step 6 in Fig. 2). Details of each step are given in the following sections.

3.1. Generation of the traffic

In order to determine the number of active users for each time slot, lasting 1 h, we employ the user distribution used in [19,22,50],

Table 1
Link budget parameters for the LTE-A macro cell and micro cell BS.

Parameters	Macro cell BS	Micro cell BS
Frequency	2.6 GHz	
Maximum input power antenna	43 dBm	33 dBm
Antenna gain base station	18 dBi	4 dBi
Antenna gain mobile station	2 dBi	
Soft hand over gain	0 dB	
Feeder loss base station	0 dB	
Feeder loss mobile station	0 dB	
Fade margin	10 dB	
Yearly availability	99.995%	
Cell interference margin	2 dB	
Bandwidth	5 MHz	
Receiver SNR	1/3 QPSK = −1.5 dB, 1/2 QPSK = 3 dB, 2/3 QPSK = 10.5 dB, 1/2 16-QAM = 14 dB, 2/3 16-QAM = 19 dB, 4/5 16-QAM = 23 dB, 2/3 64-QAM = 29.4 dB	
Used sub-carriers	301	
Total sub-carriers	512	
Noise figure mobile station	8 dB	
Implementation loss mobile station	0 dB	
Height mobile station	1.5 m	
Coverage requirements	90%	
Shadowing margin	13.2 dB	
Building penetration loss	8.1 dB	

which varies according to the hour of the day to reflect the typical behaviour of the daily traffic demand. The position of each user is defined according to a uniform distribution and the bit rate is 1 Mbit/s. Similarly to [5], a Poisson distribution, with parameter λ of 1 request/minute, provides the number of generated content requests per user. The contents that are requested by each user are determined according to their popularity, which is defined by a distribution. As in [43,45–47,49,53], this popularity distribution is a Zipf's distribution, characterised by the parameter α. This parameter affects the difference among contents in terms of popularity. In case the value chosen for α is large, the most popular contents are significantly more popular than the other contents, and by decreasing α the popularity of contents behaves more similarly to the uniform distribution. In order to introduce slight differences among the files popularity at different locations (i.e., at different BSs), as in [5], the level of popularity of each content on each macro cell BSs is determined starting from a reference popularity and performing random shuffles on it. In particular, sorting the files of the library from the most popular to the least popular (according to the reference popularity), the popularity of 30% of content is randomly swapped to generate the popularity on each macro cell BS. A similar procedure is performed to determine the popularity at each micro cell BS. In this case, starting from the popularity of the corresponding macro cell BS, the popularity of 15% of the contents is shuffled. To generate the contents requested by each user, the popularity distribution associated to the BS from which the considered user experiences the lowest path loss is used; the computation is described in the following.

3.2. Creation of the network

The RAN adapts its capacity to the instantaneous bit rate requested by the active users. This means that, as in [19], the capacity of the network is not always totally used, but it dynamically responds to the instantaneous traffic demand. In each time slot, once the traffic has been generated, the association process starts (Step 2 in Fig. 2) and each user is associated with a BS. Each time slot starts with all deactivated BSs. First, a list of possible BSs is created, to which the user can be connected. A BS is inserted in the list, if it can provide the requested bit-rate and if the experienced path loss is lower than an allowable maximum. To determine the experienced path loss, the direct line between the user and each considered BS is determined and, taking into account the presence of existing buildings, whose 3D data are provided by a shape file of the city of Ghent, we determine if the user is in a LoS (Line-of-Sight) or NLoS (Non-Line-of-Sight). According to this, the appropriate Walfish Ikegami (WI) propagation model is used [54]. This path loss has to be lower than the maximum allowable, in order to guarantee to still have a sufficient quality at the receiver side. If the experienced path loss is larger than the allowable maximum, the input power of the BS is increased until the path loss becomes acceptable. In case the input power reaches the maximum allowable input power, but the path loss is still larger than the maximum, that BS is not inserted in the list. If the list results to be empty, the user remains uncovered.

To determine the BS of the list the user should be associated with, a fitness function f is used. It is computed for each BS of the list and gives a measure of the power consumption and of the delay experienced by users, assuming that the considered user is associated with that BS. The fitness function is defined as follows:

$$f = w_1 \cdot (1 - \frac{E}{E_{MAX}}) + w_2 \cdot (1 - \frac{D}{D_{MAX}}) \quad (1)$$

where w_1 and w_2 are weight factors between 0.0 and 1.0. The value of E, in watt, is the power consumption of the current solution, E_{MAX}, in watt, is the maximum power, consumed by the network when all BSs are active and consuming the maximum power. D is the global experienced delay, in milliseconds, if the current network is used and D_{MAX} is the global experienced delay if each requested content is retrieved in the cloud. Higher values of f mean that the network performs better in terms of power consumption and/or experienced delay. Once this function has been computed for each BS, the one that maximises it is picked. In case an off BS is activated, the simulator checks if it is possible to transfer users already covered by other BSs to this "new" BS, if the available bit rate is enough, the experienced path loss is acceptable and the value of the fitness function increases, possibly increasing the transmitting power of the BS, if possible and necessary. If all users of a certain BS are moved to this "new" BS, that BS is switched off.

The way the weights are set determines which aspect of the network is optimised while the association procedure is performed:

- *PC Opt*: the network power consumption is optimised; to realise this, w_1 is 1 and w_2 is 0;
- *Delay Opt*: the network is optimised towards the experienced delay, thus w_1 and w_2 are 0 and 1, respectively;
- *Delay-PC Opt*: in this case the optimisation is performed with respect to both the network power consumption and the experienced delay, w_1 and w_2 are 0.5.

- *PL Opt*: the goal is to select the solution that makes the users suffer the lowest path loss, and w_1 and w_2 are equal to 0. This is the typical association procedure and we use it as benchmark.

The usage of the fitness function for the association procedure to minimise the network energy consumption and/or the experienced delay, provides a greedy solution. Indeed, it takes optimal local decisions for each user and the association of a "new" user does not consider the association of the already associated users, unless a new BS is activated. This reflects the actual temporal succession of user arrival and the possible handover when a BS is activated, making our approach realistic for the real RAN environment. Moreover, the greedy approach is necessary since the optimisation of the network energy consumption and/or the experienced delay, controlling the BS emitted power and the user association, is an NP-Hard problem, as illustrated in [55]. Nevertheless, the association procedure is performed while the system is operating that means that a solution is needed on the fly. This makes the optimisation approach a not feasible solution.

Finally, we introduce the *MEC switching* variant. In this case, the MEC servers, which are installed on each macro BS, are deactivated during the day, from 5 a.m. to 11.00 p.m. and the cache capacity of each macro BS is distributed equally to its 4 micro cell BSs. The described association policies are named as *PC Opt-MEC Switching*, *Delay Opt-MEC Switching*, *Delay-PC Opt-MEC Switching* and *PL Opt-MEC Switching*. A summary of the used association strategies is given in Table 2.

3.3. Energy consumption of the network

Once the network is created, i.e., each user is associated with a BS if possible, the energy consumption of the network during each time slot is computed (Step 3 in Fig. 2). The hourly energy consumption of the network, in watthour, at time t, is given by:

$$E_{tot}^{(t)} = \sum_{b=1}^{N_{BS}} E_{b,comm}^{(t)} + \sum_{b=1}^{N_{BS}} E_{b,server}^{(t)} \tag{2}$$

where N_{BS} is the number of the active BSs, $E_{b,comm}^{(t)}$ is the energy consumption of the BS b at time t, due to the communication features and $E_{b,server}^{(t)}$ is the energy consumption of the BS b at time t, due to the supply of the cache located on that BS. The $E_{b,comm}^{(t)}$ component is computed according to the models for the macro cell and micro cell BS proposed in [48], which depend on the input power of the BS and on its hardware components. According to [46] and [49], $E_{b,server}$, in watthour, is given by:

$$E_{b,server}^{(t)} = \omega_{MEC} \cdot C_{server} \cdot t \tag{3}$$

where ω_{MEC} is in W/bit, C_{server} is the capacity of the server and t is the time (in hour, 1 in our case). If a BS is in sleep mode, its energy consumption is assumed to be negligible.

3.4. Energy reduction strategy

During each time slot, once the network energy consumption is computed, an energy reduction strategy is applied (Steps 4–5 in Fig. 2):

1. *No action*: in this case, no action is taken during that time slot.
2. *Deactivate all micro cell BSs*: all micro cell BSs are deactivated during that hour, in case the energy consumption is larger than the renewable energy that is available during that time interval. This is given by the energy produced by the PV panel system and stored in the battery. The users who have been connected to each deactivated micro cell BS are reconnected to a macro cell BS, if possible. They are reconnected to the macro BS that maximises the fitness function, see (1), and that has enough available capacity, if the experienced path loss is lower than the allowed maximum.
3. *Deactivate some micro cell BSs*: micro cell BSs are put in sleep mode in case of energy shortage, i.e. in case the available renewable energy is lower than the energy consumption of the network for the considered hour, but are switched off gradually. To do this, the network energy consumption is computed, when switching off 1, 2, 3 or 4 micro cell BSs per macro cell BSs. As soon as the network consumption becomes smaller than the amount of available renewable energy, that number of microcell BS per macro cell BS is deactivated. The order in which the microcell BSs are turned off follows the number of users served by each micro cell BS. Similar to *Deactivate all micro cell BSs* strategy, the users who were connected to a micro cell BS, which has been switched to sleep mode, are reconnected to an active BS, if possible.
4. *Deactivate some macro cell BSs*: once the network energy consumption is computed, in case the available renewable energy is smaller than the energy consumption of the network for the considered hour, macro cell BSs and their corresponding micro cell BSs are gradually switched in sleep mode. We do this computing the energy consumption of the network, when switching off 1, 2,.., 8 macro cell BSs and the corresponding micro cell BSs. As soon as the available renewable energy becomes sufficient for the network supply, that number of BSs is deactivated. The order in which the macro BSs are selected to be turned off, follows the number of users served by each macro BS. After the deactivation, the users who were connected to a sleeping BSs are reconnected to an active BS, if possible.

3.5. Content delivery

In each time interval, once the energy reduction strategy is applied, each content requested by each user is delivered (Step 7 in Fig. 2). When the requested content is cached in the server of the serving BS, the content is transmitted directly to the user. Notice that the content transfer to the user, from the BS, which he/she is associated with, always incurs [56]. We assume that this access latency $T_{bs,u}$ for the cells is identical and equal to 30 ms, as in [43]. If the requested content is not cached by the serving micro cell BS but by the macro cell BS, the macro cell BS transmits the content to that micro. The link between a micro cell BS and its corresponding macro is wired, through optic fibres. Nevertheless, due to the insufficient capacity of the BH links, which causes a bottleneck in this segment of the network, its latency contribution, $T_{BS,bs}$, is significant [57], but lower than the one in the wireless link, between the user and the micro BS. We assume that it is equal to 20 ms [43]. The resulting latency is given by $T_{BS,bs}$+$T_{bs,u}$ [43]. If the content is not present not even on the macro cell BS, the request is forwarded to the content provider. In this case, the experienced latency is given by $T_{cp,BS} + T_{BS,bs} + T_{bs,u}$, with $T_{cp,BS}$=50 ms, giving a total latency equal to 100 ms, as in [43,56]. In case a user is associated with a macro cell BS, which is caching the requested content, that content is received with latency $T_{bs,u}$. If that content is not stored in the server of that macro cell BS, it is retrieved on the content provider and the user receives it with delay $T_{cp,BS} + T_{bs,u}$. After each content delivery, the cache is updated (Step 7 in Fig. 2), according to Least Frequently Used (LFU) cache algorithm, so as to always cache the most popular contents.

3.6. Initial state of caches

We determine which contents are stored in each cache at the beginning of each simulation, through a preliminary phase. For each value of α, the scenario is simulated, assuming that the considered RAN operates for 100 weeks. Each simulation begins with empty caches and the occurrences of a content request are updated, as well as the stored files in each cache, at each time slot. When the number of variations in each cache stabilises, the transient phase for each cache filling is assumed to be over. As a result, the files in each cache at that time interval are the initial state of the caches in simulations.

Table 2
Summary of the different user association strategies.

Name	Objective	Energy reducing strategy	MEC switching
PL Opt	Min. path loss	No	No
Delay Opt No Switch	Min. delay	No	No
PC Opt No Switch	Min. power consumption	No	No
Delay-PC Opt No Switch	Min. delay and power consumption	No	No
PL Opt With Switch	Min. path loss	No	Yes
Delay Opt With Switch	Min. delay	No	Yes
PC Opt With Switch	Min. power consumption	No	Yes
Delay-PC Opt With Switch	Min. delay and power consumption	No	Yes
PL Opt Deactivate All micro cell BSs	Min. path loss	Yes	No
PL Opt Deactivate Some Macro cell BSs	Min. path loss	Yes	No

4. Key performance indicators

Energy consumption

The energy consumption of the network during the simulation, in Wh, is given by

$$E = \sum_{t=1}^{T} E_{tot}^{(t)} \tag{4}$$

where $E_{tot}^{(t)}$ is the energy consumption of the network at time t and it is computed as reported in (2) and T is the duration of the simulation.

Green energy

The green energy, in Wh, accounts for the amount of used energy which has been produced by the PV-panel system, which is locally installed.

Brown energy

The brown energy, in Wh, indicates the energy bought from the power grid, for the network supply. As already mentioned, in case the available renewable energy is not sufficient to power the RAN, even after the application of an energy reducing strategy, if employed, the missing energy is taken from the grid.

User coverage

The user coverage is the percentage of served users, considering that a user can be associated to a BS if he/she experiences a path loss lower than the allowable maximum and if that BS has enough capacity to provide the required bit rate.

Average delay

This is the average delay experienced by users to receive content. It is measured in ms and is given by:

$$D = \frac{1}{N_r} \sum_{r=1}^{N_r} d_r \tag{5}$$

where d_r is the delay which is experienced for the content request r and N_r is the total number of requests during the simulation.

Hit - 1 hop probability

This is the probability that the requested content is stored locally on the BS which the considered user is associated to.

Hit - 2 hops probability

This is the probability that the content requested by a user associated to a micro cell BS is not cached on that micro cell BS but on the corresponding macro cell.

Miss probability

This is the probability that the requested content is fetched from the content provider, since it is not stored in the cache of the BS, which the user is associated with nor in the one of the corresponding macro cell BS, if the considered user is associated with a micro cell BS.

Table 3
Values of parameter used in simulations.

$T_{bs,u}$	30 ms
$T_{BS,bs}$	20 ms
$T_{cp,BS}$	50 ms
ω_{MEC}	2.5 $\cdot 10^{-9}$ W/bit

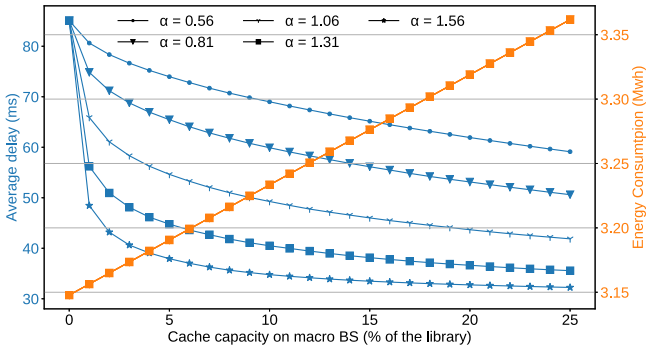


Fig. 3. Average delay (in blue) and energy consumption (in orange) varying the dimension of each cache, for different values of the parameter α.

5. Performance evaluation

In this section, we discuss the results obtained, when the considered RAN is simulated, assuming that it operates for 1 week. The values of latency, as well as the value of the parameter ω_{MEC} of (3) are reported in Table 3. They are taken from [43] and [49], respectively.

5.1. Impact of the cache size and the popularity

In the first part of our work, we analyse the effects of the parameters that affect the performance of the local caching. To do this, we simulate the scenario described in Section 3, using *No action* as energy reduction strategy. Fig. 3 shows on the left y-axis the average experienced latency, in blue, and on the right y-axis the energy consumption, in orange, versus the size of the cache on each micro and macro BS (the cache on the macro BS is double the one on the micro BSs), given in percentage of stored library. Each curve of the figure corresponds to different values of the parameter α, characterising the Zipf's distribution. When the percentage of the stored library is zero, no local caching is performed. The growth of the size of the cache generates a reduction of the experienced delay, since more contents can be stored locally. This reduction strictly depends on the characteristics of the popularity, e.g., on the parameter α. Indeed, as already mentioned, a large value of α means that there is a small part of the library which is very popular. If this is the case, even a small cache drastically reduces the experienced delay. When α is larger than 1, the experienced delay is reduced up to 40%, if 1% of the library is locally stored. Conversely, a small value of α indicates that the files have a similar popularity. In this scenario, larger caches are needed to achieve significant delay reduction: if

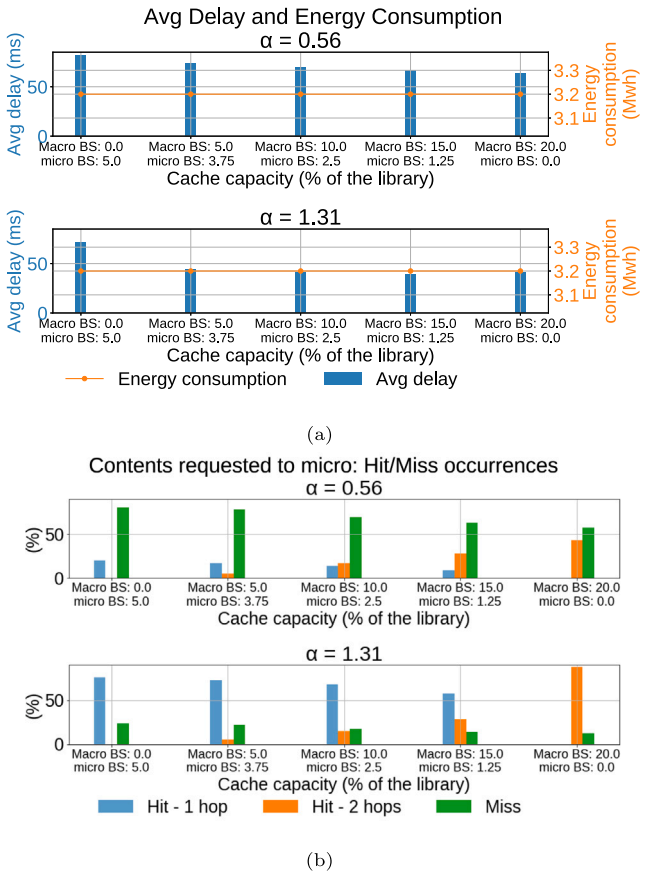


Fig. 4. Given a fixed caching capacity equal to 20% of the total library, change of its distribution among BSs: (a) Avg delay and energy consumption, (b) hit/miss occurrences probability on micro cell BS.

the popularity distribution is described by the Zipf's function with parameter α equal to 0.56, 10% of the library should be stored to reduce the experienced delay up to 30%. The energy consumption increases linearly with the cache size, see (3). Nevertheless, this growth is limited to 7%, when the cache stores 25% of the library and lower than 1.5%, if 5% of the library is cached.

Besides the impact of the cache size on the user experience, we also investigate the impact of its distribution among BSs. In particular, for each macro cell BS and its 4 micro cell BSs, a total capacity equal to 20% of the library is considered, and we vary its distribution among the BSs. We consider the case in which the cache on the macro cell BSs stores 0%, 5%, 10%, 15% and 20% of the library and, correspondingly, each micro cell BSs stores 5%, 3.75%, 2.5%, 1.25% and 0%. In Fig. 4(a), the average delay (blue bars) and the energy consumption (orange line) are shown, for these values of cache capacity, for α equal to 0.56 and 1.31, when *No action* strategy is used. Fig. 4(b) shows the probability of the possible events that a user might experience when he/she requires a content. Blue bins indicate the probability to experience a *hit - 1 hop*, i.e., the needed content is cached on the BS, with which that user has been associated, the orange bins show the *hit - 2 hops* probability and the green bins report the *miss* probability. As shown by the orange lines in Fig. 4(a), the energy consumption is constant since the total capacity does not change. Moreover, the plot reveals that the delay reduction not only depends strictly on the popularity, i.e., on parameter α, but also depends on the cache distribution among BSs. Indeed, as reported in Fig. 4(b), if α is 0.56 and the micro cell BSs can store up to 5% of the library, no more than 20% of the requests on a micro cell BSs can be satisfied locally (on that BS), while this number grows to 76% if α is 1.31. Similarly, when all the considered caching capacity is put on the macro cell BS, that BS satisfies 43% and 88% of the requests, respectively, as can be observed in the 4(b). Furthermore, even if the hit with a single hop is less frequent due to the reduction of the cache capacity installed on each micro cell BS (see Fig. 4(b)), from Fig. 4(a), it is evident that putting more cache on each macro cell BS generates the drop of the experienced delay. This is because the cache on each macro is reachable by the users connected to it, as well as by users connected to each corresponding micro cell BS. Therefore, the growth of the capacity on macro cell BSs corresponds to the growth of the cache capacity, which is reachable by all the users. For the same reason, the miss probability decreases, when the capacity on each macro cell BS increases (Fig. 4(b)). The resources on the macro cell are precious, but it is convenient to install some capacity on micro cell BSs too to relieve the effort on the macro cells and achieve some local hits(1 hop from users), especially if α is large. Indeed, when α is equal to 1.31 and all the cache capacity is located on the macro cell BSs, the average delay is larger than the case where 15% and 1.25% of the library are stored on the macro and on the micro cell BSs, respectively.

5.2. Impact of energy reduction strategy

In this section we first discuss the impact of the energy reducing strategies, when the MEC technology is not used, i.e., when each MEC server stores 0% of the library. In Figs. 5(a), 5(b), 5(c) and 5(d), the network energy consumption, the amount of green and brown energy, the number of active BSs and the user coverage are reported, respectively, for each time slot of the simulation. The curves marked by circles and triangles in the figures correspond to, respectively, *No Action* and *Deactivate all micro cell BSs*, which deactivates each micro cell BSs as soon as the renewable energy is not sufficient for the network supply. The lines with stars and squares in Fig. 5 are the cases in which *Deactivate some micro cell BSs* and *Deactivate some macro cell BSs* are employed, so as to deactivate gradually the micro and the macro BSs, respectively, in case the renewable energy is smaller than the network consumption. From these figures, we notice that, while the energy consumption and number of active BSs are almost the same during the light hours for each used strategy, see Figs. 5(a) and 5(b), the situation is different from 14:00 to 8:00. Indeed, in this time interval, since the renewable energy results insufficient to power the network because of lack of solar production, when *Deactivate all micro cell BSs*, *Deactivate some micro cell BSs* and *Deactivate some macro cell BSs* are used, BSs are deactivated, reducing the RAN energy consumption, see curves with triangles, stars and squares in Figs. 5(a) and 5(b). *Deactivate all micro cell BSs* and *Deactivate some micro cell BSs* behave very similarly in terms of number of switched micro cell BSs and energy consumption. This is because, even if the latter puts in sleep mode only the number of micro cell BSs which makes the renewable energy sufficient for the network supply, actually it often needs to deactivate all micro BSs (see Fig. 5(b)). As a consequence, it switches only a slightly lower number of BSs than the former strategy. This results in a very similar energy consumption, which is, respectively for *Deactivate all micro cell BSs* and *Deactivate some micro cell BSs*, 1900 and 1940 kWh, 40% lower than *No Action* case, whose consumption accounts for 3148 kWh. Besides the reduction of the network energy consumption, *Deactivate all micro cell BSs* and *Deactivate some micro cell BSs* allow to buy from the grid a lower amount of brown energy. Fig. 5(c) reports, for each hour of the simulation, in green the amount of used renewable energy and in brown the amount of brown energy, which is bought from the grid, for each energy reducing strategy. The amount of green energy is almost the same for each strategy, but with *Deactivate all micro cell BSs* and *Deactivate some micro cell BSs*, between 46% and 55% less energy is bought from the grid than when *No Action* is used, which needs to buy from the grid 86% of its energy consumption. This shows that making the RAN more sustainable through the usage of renewable energy sources has to be coupled with the micro cell BS switching, in order to be effective. Indeed, if this is the case, the energy consumption

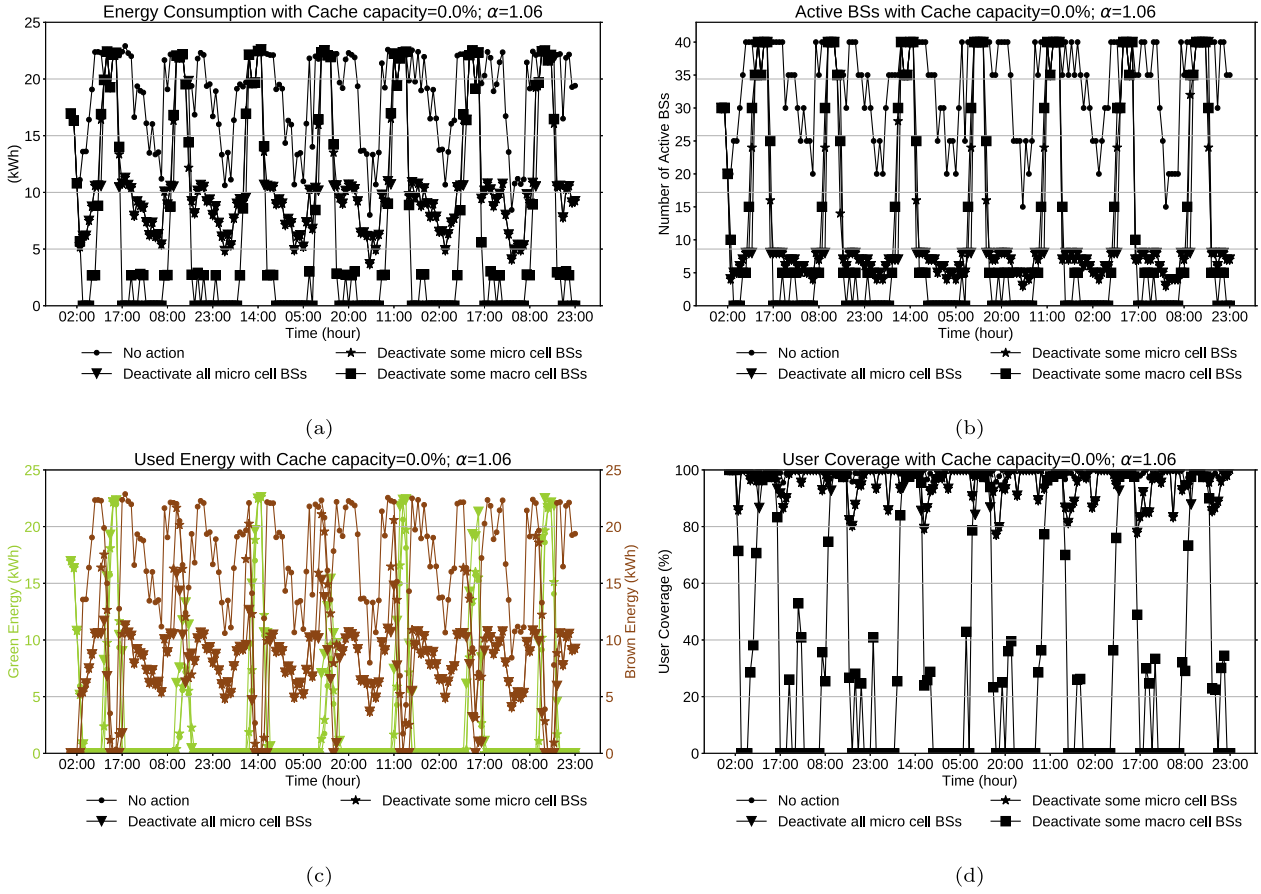


Fig. 5. Usage of the energy reducing strategies, when the MEC technology is not used: (a) Energy Consumption, (b) Number of active BSs, (c) Amount of Used Green and Brown Energy and (d) User Coverage.

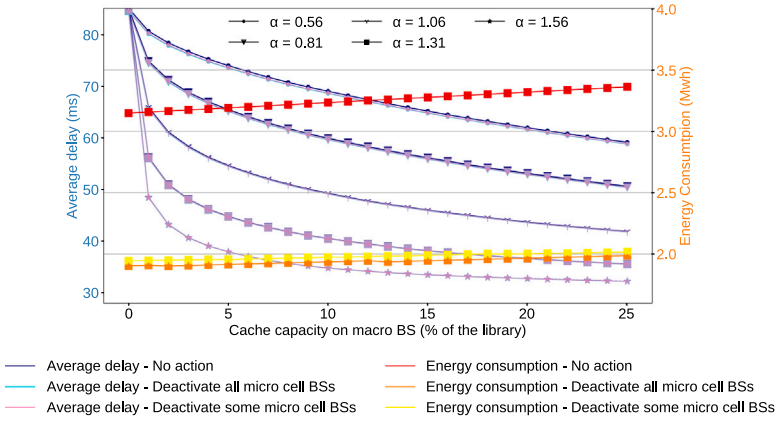


Fig. 6. Delay and energy consumption varying the size of each cache, for different values of the parameter α, when *No action* and *Deactivate all micro cell BSs* are used.

and the bought energy are strongly decreased, reducing the Opex expenditure and improving the RAN self-sufficiency. The trend of the energy consumption and of active BSs presented by *Deactivate all micro cell BSs* and *Deactivate some micro cell BSs* strategies also occurs for the user coverage, which results lower than when *No Action* is employed, because of micro cell BSs deactivation, but rarely lower than 95%. This drop of the user coverage underlines the key impact of the micro cell BSs on the QoS: they are necessary in order to always provide an optimal user coverage, i.e. 99%, which is provided when the micro cell BSs are not deactivated, as when *No Action* strategy is used.

The situation is different for *Deactivate some macro cell BSs*. Indeed, it often deactivates all BSs during the night, generating interruption of the service, as shown by the curve marked by triangles of Fig. 5(b). As a consequence, even if more than 64% of energy is saved, its nightly user coverage falls to 0%, resulting in unacceptable (see green curve in Fig. 5(d)) user coverage. This highlights the fundamental role of the macro cell BSs in hierarchical RANs, in order to provide an adequate QoS, revealing that the deactivation of the macro cell BSs to save energy is not a feasible solution for our scenario. For this reason, this strategy is not considered in the following discussions.

We now analyse the effect of the energy reduction strategy on the caching paradigm and vice versa, when *Deactivate all micro cell BSs* and *Deactivate some micro cell BSs* are used.

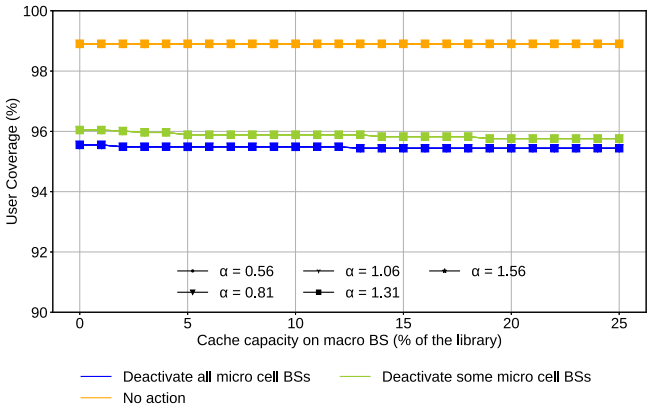


Fig. 7. User Coverage varying the size of each cache, for different values of the parameter α, when *No action*, *Deactivate all micro cell BSs* and *Deactivate some micro cell BSs* are used.

In Fig. 6, the impact of the variation of the size of each cache is shown, in terms of experienced delay (left y-axis) and of energy consumption (right y-axis) in blue, light blue, pink and in red, orange, yellow, when *No action*, *Deactivate all micro cell BSs* and *Deactivate some micro cell BSs* strategies are used, respectively. As already discussed, the usage of *Deactivate all micro cell BSs* and *Deactivate some micro cell BSs* significantly reduces the energy consumption of the network: when they are employed, the system drops its consumption by up to 41% and 40%, respectively. With *Deactivate all micro cell BSs* and *Deactivate some micro cell BSs*, the energy consumption does not grow as much as in the case of *No action*, with the increase of the cache capacity. Indeed, when each macro cell BS stores up to 25% of the library, the network consumes 4% more than the case with no caching servers. This is because with a larger cache, it is more likely that the available renewable energy is insufficient to power the RAN. As a result the switching of micro cell BSs occurs more often and above a given storage size (up to when 5% of the library is stored), the system stabilises: micro cell BSs are deactivated in the same period, since the energy is not sufficient for the network supply in the same instant. Moreover, with *Deactivate all micro cell BSs* and *Deactivate some micro cell BSs*, the experienced delay is slightly reduced. This is due to the fact that when the micro cell BSs are deactivated, the users are closer to the content provider, since they are always at 2 hops distance. This is more evident for low values of α, since in these cases the content needs to be taken from the content provider more often. Therefore, the impact of this reduction of distance is higher. The employment of *Deactivate all micro cell BSs* and *Deactivate some micro cell BSs* strategies reduce the user coverage, as noticed above. As can be noticed in Fig. 7, by varying the size of each cache, the user coverage drops from 99% given with *No action*, in orange, to, respectively, 95.5% and 96%, in blue and green, which is acceptable. When *Deactivate all micro cell BSs* and *Deactivate some micro cell BSs* are used, the growth of the cache capacity does not significantly impact the user coverage, since the number of time slots during which micro cell BSs are deactivated slightly grows by 4%, if 25% of the library is cached on macro BSs, with respect to the scenario that does not employ MEC server.

5.3. Impact of the association strategy

Now we analyse the effects of the user association policies which we propose and present in Section 3. Figs. 8(a) and 8(b) show the delay and the energy consumption with the different user association policies, with α equal to 0.56 and 1.31, respectively, increasing the cache capacity from 0% to 25% of the library. First, since the characteristic of the traffic demand, i.e. of the parameter α, has no impact on the network energy consumption, when *PC Opt* is used, variations on the

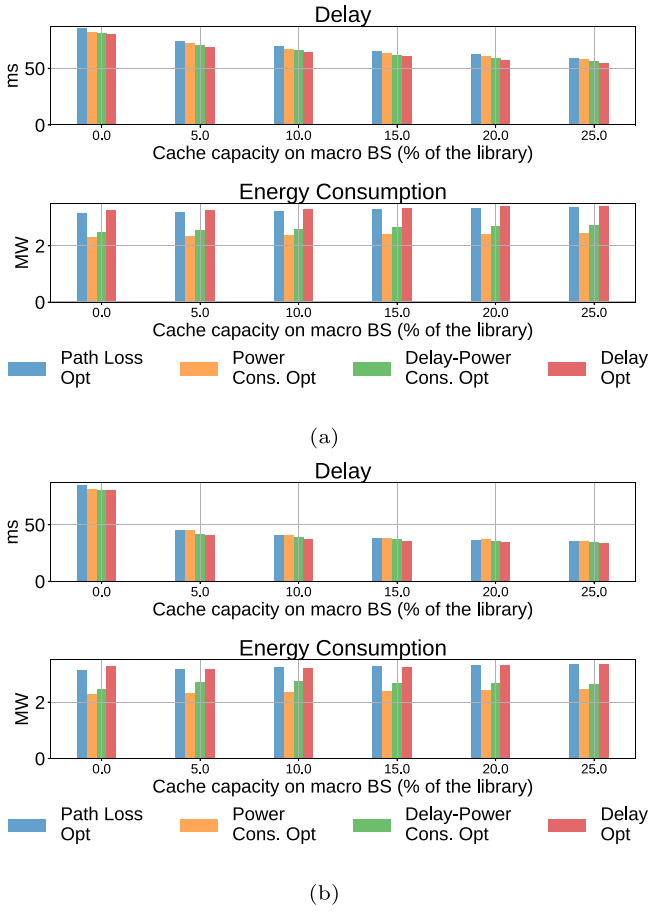


Fig. 8. Energy consumption and delay achieved with the proposed users association policies, with (a) α=0.56 and (b) α=1.31.

characteristic of the traffic do not impact the associations and, as a consequence, the network energy consumption, as can be noticed by the orange bars in the two figures. Moreover, when compared to the case in which the experienced path loss is minimised, between 22% and 27% of network energy consumption reduction is achieved. Using this users association procedure, users are typically associated with macro BSs, meaning that having a few but heavily loaded macro BSs is more efficient than having many active BSs under utilised. This also shortens the time needed to reach the cloud and because of this, when retrieving contents in the cloud occurs often, i.e. α is lower than 1 and/or the server size is small, a slight delay reduction, between 2.5% and 4%, with respect to *PL Opt* policy, is achieved.

When users are associated while minimising the experienced delay, i.e., *Delay Opt* is used, the delay drops up to 10% (see red bars in Fig. 8) with respect to the case with the same cache capacity, but where *PL Opt* is used as the association approach. This drop is more significant when misses are more likely, i.e. for decreasing cache size and α parameter, because performance can be improved significantly. Moreover, in order to make the access to the cloud faster, this policy forces users to be associated with the macro BS, which maximises the local hit. This is also because if a user is associated with a macro BS, the cloud is reached with 2 hops, while if associated with a micro BS, 3 hops are necessary to reach it. Thus, being associated with a macro makes the access to the cloud faster. Because of this trend, to achieve a path loss low enough to receive an acceptable quality of the signal, the output power of a BS results higher than when *PL Opt* is used. This determines the growth of the network energy consumption by 4%. When α is larger than 1, this policy is not particularly effective, as already mentioned, making the energy consumption and the delay almost unchanged. In

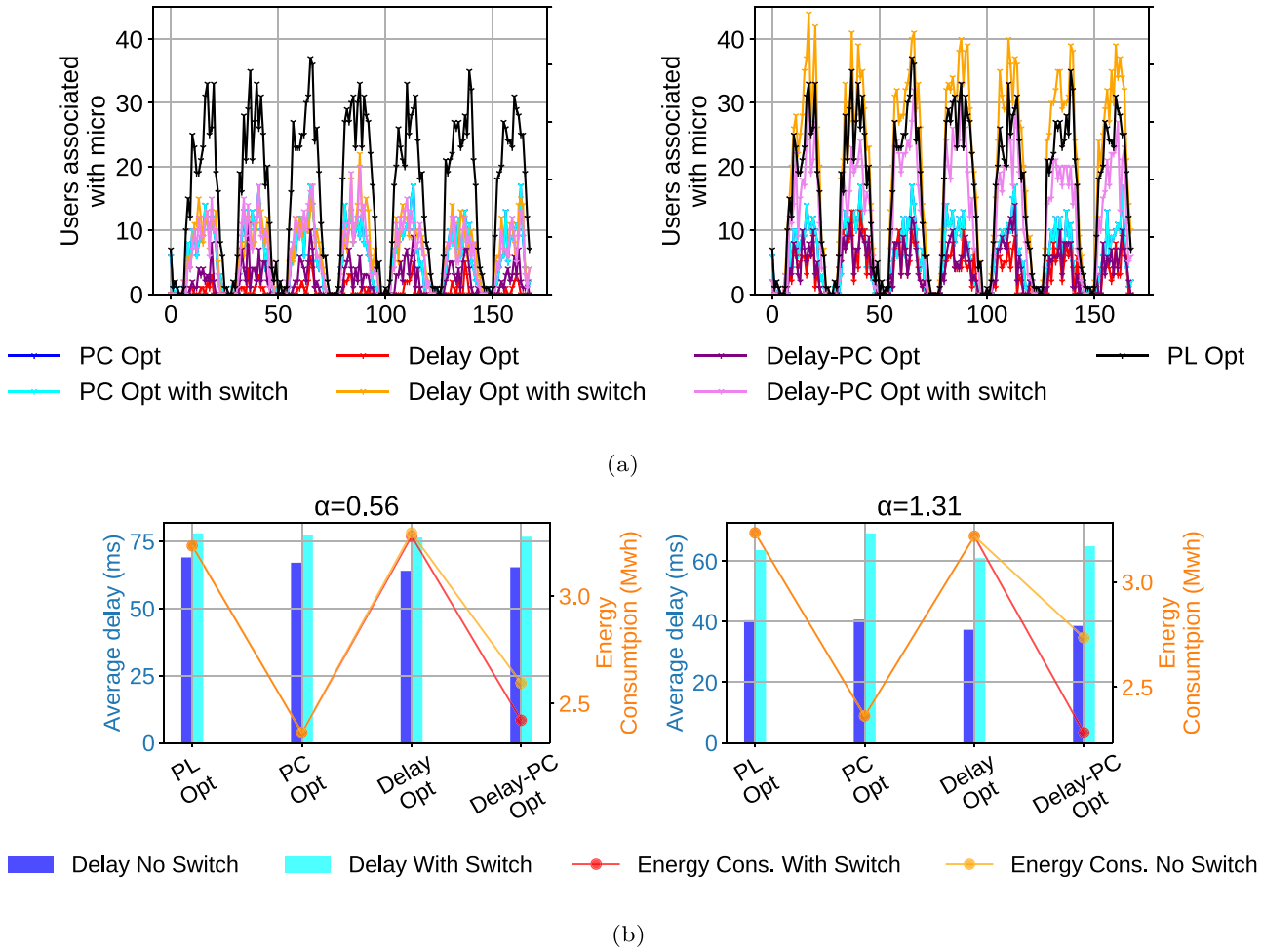


Fig. 9. Effects of the dynamic MEC server switching: (a) Energy Consumption and Delay and (b) number of users who are associated to micro BSs, the cache capacity on each macro is 10% of the library.

case the association procedure aims at the joint minimisation of delay and energy consumption, both of them are reduced by 5% and 20%, respectively.

5.4. Effects of the MEC switching

We have observed that the proposed association procedures tend to associate users to macro BSs, but this violates the load balancing. Fig. 9(a) shows the number of users, that are associated with a micro cell BS, in each time slot of the simulation. This value is shown for α equal to 0.56, on the left and to 1.31, on the right, assuming that each cache, which is installed on each macro cell BS, stores up to 10% of the library. We consider all the association policies, with and without *MEC switching* variant. As can be observed from the figure, if users are associated according to *Delay Opt*, in blue, *Delay-PC Opt*, in purple, and *PC Opt*, in red, users associated with micro cell BSs are 0, 1 and 6, respectively, against 19, obtained when *PL Opt* is used. As shown in [50], this increases the electromagnetic exposure of human beings, besides worsening the QoS, as claimed in [58,59]. To solve this issue, we introduce the *MEC switching*, which deactivates the MEC server of each macro BSs, between 5 a.m. and 11 p.m., distributing equally that amount of capacity from each macro cell BSs to its 4 micro cell BSs (see Section 3). Fig. 9(b) shows the delay, for each association policies and for α equal to 0.56, on the left and to 1.31, on the right, assuming that each cache on each macro stores up to 10% of the library, with the light blue and blue bins, when *MEC switching* is used or not, respectively. The network energy consumption is given by the red and orange lines, respectively.

First, we focus our attention to the cases that associate users according to *Delay Opt* or *Delay-PC Opt* policies. In these cases, users tend to be associated with micro cell BSs during the day (from 5.00 a.m. to 11.00 p.m.). As a consequence, when α is 0.56, the *MEC switching* variant slightly increases the number of users served by the micro BSs from 0 with *Delay Opt* to 11 with *Delay Opt-MEC Switching* and from 1 with *Delay-PC Opt* to 7 with *Delay-PC Opt-MEC Switching*. With α equal to 1.31, this growth is more evident: from 4 to 22 and from 3 to 9, if *Delay Opt*, *Delay Opt-MEC Switching* and *Delay-PC Opt* and *Delay-PC Opt-MEC Switching* are used, respectively. With low values of α this improvement is less evident. This is because in these cases the access to the cloud is frequently needed. For this reason, users continue to be associated to macro BSs to reduce the time to retrieve contents, despite the *MEC Switching* employment. With large values of α, several local hits occur. Hence, the *MEC switching* variant induces the association of users with micro BSs and this increases the network energy consumption, which grows up to 16% (see Fig. 9(b)). Moreover, since cache capacity is moved from each macro BS towards its micro cell BSs, less cache is reachable by users. This results in a higher experienced delay, which grows by up to 68%. When *PC Opt* is employed as association policy, the usage of the *MEC switching* variant does not improve the load-balancing issue, since it is independent of the MEC server presence and capacity. As a consequence, the *MEC switching* does not impact the energy consumption of the *PL Opt*, as well as of the *PC Opt*. Nevertheless, the delay grows by up to 70%, in case the *MEC*

switching is used. As already mentioned, removing cache capacity on the macro and spreading it among its micro cell BSs, deteriorates the delay, as less capacity is reachable from users.

6. Conclusion

In this paper, mechanisms to move towards two of the objectives of 5G networks, the delay reduction and the network energy efficiency, are revised and proposed. A portion of a RAN, composed by 8 macro cell BSs, each supported by 4 micro cell BSs is considered, where the MEC technology is employed, to push the most popular contents closer to users so as to reduce latency. The considered RAN is powered by a PV panel system and an energy battery and is connected to the power grid. Different users association policies are proposed in order to further improve the experienced delay and the energy consumption of the network. We notice that, even if strictly dependent on the characteristic of the traffic and on the server capacity, the MEC technology reduces the experienced delay up to 60%, without generating significant growth of the network energy consumption, limited to 7%. In addition, the employment of an energy reduction strategy, applied in case of renewable energy shortage, reduces the energy consumption but does not impact the experienced delay. The proposed users association policies effectively provide reductions of delay and power consumption.

Our results lead to three conclusions. First, caching at the edge and dynamic activation of the BSs, can be very effective in reducing latency and reducing the network power consumption, respectively, without deteriorating their performances because of their coexistence. Second, caching on the macro BSs is always needed to significantly reduce delays, while caching also on the micro cells relieves the effort on the macro cell. Finally, association procedures which minimise the delay and/or the energy consumption tend to associate users with macro BSs. In this way, when the network energy consumption is optimised, also the experienced delay is slightly reduced. Meanwhile, in case the association procedure is based on the minimisation of the delay, the energy consumption increases, since users can be associated to BS which are far, and the emitted power has to be increased so as to receive the signal with the adequate quality.

CRediT authorship contribution statement

Greta Vallero: Conceptualization, Methodology, Software, Writing - original draft. **Margot Deruyck:** Data curation, Conceptualization, Supervision. **Michela Meo:** Data curation, Conceptualization, Supervision. **Wout Joseph:** Data curation, Conceptualization, Supervision.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgement

M. Deruyck is a Post-Doctoral Fellow of the FWO-V (Research Foundation - Flanders, Belgium, no. 12Z5621N).

References

[1] X. Yang, X. Yu, H. Huang, H. Zhu, Energy efficiency based joint computation offloading and resource allocation in multi-access MEC systems, IEEE Access 7 (2019) 117054–117062.
[2] M.A. Habibi, M. Nasimi, B. Han, H.D. Schotten, A comprehensive survey of RAN architectures toward 5G mobile communication system, IEEE Access 7 (2019) 70371–70421.
[3] B. Blaszczyszyn, A. Giovanidis, Optimal geographic caching in cellular networks, in: 2015 IEEE International Conference on Communications, ICC, IEEE, 2015, pp. 3358–3363.
[4] K. Poularakis, G. Iosifidis, A. Argyriou, L. Tassiulas, Video delivery over heterogeneous cellular networks: Optimizing cost and performance, in: IEEE INFOCOM 2014-IEEE Conference on Computer Communications, IEEE, 2014, pp. 1078–1086.
[5] T.X. Tran, A. Hajisami, P. Pandey, D. Pompili, Collaborative mobile edge computing in 5G networks: New paradigms, scenarios, and challenges, IEEE Commun. Mag. 55 (4) (2017) 54–61.
[6] D. Liu, B. Chen, C. Yang, A.F. Molisch, Caching at the wireless edge: design aspects, challenges, and future directions, IEEE Commun. Mag. 54 (9) (2016) 22–28.
[7] R. Deng, DASH based video caching in MEC-assisted heterogeneous networks, Multimedia Tools Appl. (2020).
[8] Y. Sun, Z. Chen, H. Liu, Delay analysis and optimization in cache-enabled multi-cell cooperative networks, in: 2016 IEEE Global Communications Conference, GLOBECOM, IEEE, 2016, pp. 1–7.
[9] E.E. Ugwuanyi, S. Ghosh, M. Iqbal, T. Dagiuklas, S. Mumtaz, A. Al-Dulaimi, Co-operative and hybrid replacement caching for multi-access mobile edge computing, in: 2019 European Conference on Networks and Communications, EuCNC, IEEE, 2019, pp. 394–399.
[10] E. Baccour, A. Erbad, K. Bilal, A. Mohamed, M. Guizani, Pccp: Proactive video chunks caching and processing in edge networks, Future Gener. Comput. Syst. 105 (2020) 44–60.
[11] A. Osseiran, F. Boccardi, V. Braun, K. Kusume, P. Marsch, M. Maternia, O. Queseth, M. Schellmann, H. Schotten, H. Taoka, et al., Scenarios for 5G mobile and wireless communications: the vision of the METIS project, IEEE Commun. Mag. 52 (5) (2014) 26–35.
[12] C.V. Forecast, Cisco Visual Networking Index: Forecast and Trends, 2017–2022, White paper, Cisco Public Information, 2019.
[13] A. Gati, F.E. Salem, A.M.G. Serrano, D. Marquet, S.L. Masson, T. Rivera, D.-T. Phan-Huy, Z. Altman, J.-B. Landre, O. Simon, et al., Key technologies to accelerate the ICT Green evolution–An operator's point of view, 2019, arXiv preprint arXiv:1903.09627.
[14] D. Pompili, A. Hajisami, T.X. Tran, Elastic resource utilization framework for high capacity and energy efficiency in cloud RAN, IEEE Commun. Mag. 54 (1) (2016) 26–32.
[15] D. Sabella, A. De Domenico, E. Katranaras, M.A. Imran, M. Di Girolamo, U. Salim, M. Lalam, K. Samdanis, A. Maeder, Energy efficiency benefits of RAN-as-a-service concept for a cloud-based 5G mobile network infrastructure, IEEE Access 2 (2014) 1586–1597.
[16] P. Bertoldi, EU Code of Conduct on Energy Consumption of Broadband Equipment, 2017.
[17] Ł. Budzisz, F. Ganji, G. Rizzo, M.A. Marsan, M. Meo, Y. Zhang, G. Koutitas, L. Tassiulas, S. Lambert, B. Lannoo, et al., Dynamic resource provisioning for energy efficiency in wireless access networks: A survey and an outlook, IEEE Commun. Surv. Tutor. 16 (4) (2014) 2259–2285.
[18] M. Dalmasso, M. Meo, D. Renga, Radio resource management for improving energy self-sufficiency of green mobile networks, ACM SIGMETRICS Perform. Eval. Rev. 44 (2) (2016) 82–87.
[19] M. Deruyck, W. Joseph, E. Tanghe, L. Martens, Reducing the power consumption in LTE-advanced wireless access networks by a capacity based deployment tool, Radio Sci. 49 (9) (2014) 777–787.
[20] T. Shankar, et al., A survey on techniques related to base station sleeping in green communication and CoMP analysis, in: 2016 IEEE International Conference on Engineering and Technology, ICETECH, IEEE, 2016, pp. 1059–1067.
[21] S. Buzzi, I. Chih-Lin, T.E. Klein, H.V. Poor, C. Yang, A. Zappone, A survey of energy-efficient techniques for 5G networks and challenges ahead, IEEE J. Sel. Areas Commun. 34 (4) (2016) 697–709.
[22] M. Deruyck, D. Renga, M. Meo, L. Martens, W. Joseph, Accounting for the varying supply of solar energy when designing wireless access networks, IEEE Trans. Green Commun. Netw. 2 (1) (2017) 275–290.
[23] G. Vallero, M. Deruyck, W. Joseph, M. Meo, Caching at the edge in high energy-efficient wireless access networks, in: ICC 2020-2020 IEEE International Conference on Communications, ICC, IEEE, 2020, pp. 1–7.
[24] T. Pamuklu, C. Ersoy, Reducing the total cost of ownership in radio access networks by using renewable energy resources, Wirel. Netw. 26 (3) (2020) 1667–1684.
[25] A.P.C. da Silva, D. Renga, M. Meo, M.A. Marsan, Small solar panels can drastically reduce the carbon footprint of radio access networks, in: 2019 31st International Teletraffic Congress, ITC 31, IEEE, 2019, pp. 64–65.
[26] R. Mina, G. Sakr, Design and optimization of a renewable-energy fully-hybrid power supply system in mobile radio access networks, Int. J. Renew. Energy Res. 9 (3) (2019) 1339–1350.
[27] Ł. Budzisz, F. Ganji, G. Rizzo, M.A. Marsan, M. Meo, Y. Zhang, G. Koutitas, L. Tassiulas, S. Lambert, B. Lannoo, et al., Dynamic resource provisioning for energy efficiency in wireless access networks: A survey and an outlook, IEEE Commun. Surv. Tutor. 16 (4) (2014) 2259–2285.
[28] T. Shankar, et al., A survey on techniques related to base station sleeping in green communication and CoMP analysis, in: 2016 IEEE International Conference on Engineering and Technology, ICETECH, IEEE, 2016, pp. 1059–1067.
[29] S. Buzzi, I. Chih-Lin, T.E. Klein, H.V. Poor, C. Yang, A. Zappone, A survey of energy-efficient techniques for 5G networks and challenges ahead, IEEE J. Sel. Areas Commun. 34 (4) (2016) 697–709.
[30] H. Ghazzai, M.J. Farooq, A. Alsharoa, E. Yaacoub, A. Kadri, M.-S. Alouini, Green networking in cellular hetnets: A unified radio resource management framework with base station on/off switching, IEEE Trans. Veh. Technol. 66 (7) (2017) 5879–5893.

[31] M. Miozzo, L. Giupponi, M. Rossi, 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, IEEE, 2017, pp. 1–6.
[32] N.B. Rached, H. Ghazzai, A. Kadri, M.-S. Alouini, A time-varied probabilistic ON/OFF switching algorithm for cellular networks, IEEE Commun. Lett. 22 (3) (2018) 634–637.
[33] D. Renga, H.A.H. Hassan, M. Meo, L. Nuaymi, Energy management and base station on/off switching in green mobile networks for offering ancillary services, IEEE Trans. Green Commun. Netw. 2 (3) (2018) 868–880.
[34] M. Ali, M. Meo, D. Renga, Cost saving and ancillary service provisioning in green mobile networks, in: The Internet of Things for Smart Urban Ecosystems, Springer, 2019, pp. 201–224.
[35] S. D'Oro, M.A. Marotta, C.B. Both, L. DaSilva, S. Palazzo, Power-efficient resource allocation in C-RANs with SINR constraints and deadlines, IEEE Trans. Veh. Technol. 68 (6) (2019) 6099–6113.
[36] M. Deruyck, W. Joseph, E. Tanghe, L. Martens, Reducing the power consumption in LTE-advanced wireless access networks by a capacity based deployment tool, Radio Sci. 49 (9) (2014) 777–787.
[37] M. Deruyck, D. Renga, M. Meo, L. Martens, W. Joseph, Reducing the impact of solar energy shortages on the wireless access network powered by a PV panel system and the power grid, in: 2016 IEEE 27th Annual International Symposium on Personal, Indoor, and Mobile Radio Communications, PIMRC, IEEE, 2016, pp. 1–6.
[38] M. Deruyck, D. Renga, M. Meo, L. Martens, W. Joseph, Accounting for the varying supply of solar energy when designing wireless access networks, IEEE Trans. Green Commun. Netw. 2 (1) (2018) 275–290.
[39] G. Vallero, M. Deruyck, M. Meo, W. Joseph, Accounting for energy cost when designing energy-efficient wireless access networks, Energies 11 (3) (2018) 617.
[40] G. Vallero, D. Renga, M. Meo, M.A. Marsan, Greener RAN operation through machine learning, IEEE Trans. Netw. Serv. Manag. 16 (3) (2019) 896–908.
[41] M. Mehrabi, D. You, V. Latzko, H. Salah, M. Reisslein, F.H. Fitzek, Device-enhanced MEC: Multi-access edge computing (MEC) aided by end device computation and caching: A survey, IEEE Access 7 (2019) 166079–166108.
[42] S. Safavat, N.N. Sapavath, D.B. Rawat, Recent advances in mobile edge computing and content caching, Digit. Commun. Netw. 6 (2) (2020) 189–194.
[43] M. Chen, Y. Qian, Y. Hao, Y. Li, J. Song, Data-driven computing and caching in 5G networks: Architecture and delay analysis, IEEE Wirel. Commun. 25 (1) (2018) 70–75.
[44] K. Poularakis, G. Iosifidis, L. Tassiulas, Approximation algorithms for mobile data caching in small cell networks, IEEE Trans. Commun. 62 (10) (2014) 3665–3677.
[45] M. Chen, Y. Hao, L. Hu, K. Huang, V.K. Lau, Green and mobility-aware caching in 5G networks, IEEE Trans. Wireless Commun. 16 (12) (2017) 8347–8361.
[46] Z. Luo, M. LiWang, Z. Lin, L. Huang, X. Du, M. Guizani, Energy-efficient caching for mobile edge computing in 5G networks, Appl. Sci. 7 (6) (2017) 557.
[47] K. Shanmugam, N. Golrezaei, A.G. Dimakis, A.F. Molisch, G. Caire, Femtocaching: Wireless content delivery through distributed caching helpers, IEEE Trans. Inform. Theory 59 (12) (2013) 8402–8413.
[48] M. Deruyck, W. Joseph, L. Martens, Power consumption model for macrocell and microcell base stations, Trans. Emerg. Telecommun. Technol. 25 (3) (2014) 320–333.
[49] N. Choi, K. Guan, D.C. Kilper, G. Atkinson, In-network caching effect on optimal energy consumption in content-centric networking, in: 2012 IEEE International Conference on Communications, ICC, IEEE, 2012, pp. 2889–2894.
[50] M. Deruyck, E. Tanghe, D. Plets, L. Martens, W. Joseph, Optimizing LTE wireless access networks towards power consumption and electromagnetic exposure of human beings, Comput. Netw. 94 (2016) 29–40.
[51] S. Hua, Q. Zhou, D. Kong, J. Ma, Application of valve-regulated lead-acid batteries for storage of solar electricity in stand-alone photovoltaic systems in the northwest areas of China, J. Power Sources 158 (2) (2006) 1178–1185.
[52] A.P. Dobos, PVWatts Version 5 Manual, Tech. rep., National Renewable Energy Lab.(NREL), Golden, CO (United States), 2014.
[53] Y.-C. Wang, K.-C. Chien, A load-aware small-cell management mechanism to support green communications in 5G networks, in: 2018 27th Wireless and Optical Communication Conference, WOCC, IEEE, 2018, pp. 1–5.
[54] P.E. Mogensen, J. Wigard, COST Action 231: Digital mobile radio towards future generation system, final report, in: Section 5.2: On Antenna and Frequency Diversity in GSM. Section 5.3: Capacity Study of Frequency Hopping GSM Network, 1999.
[55] H. Wu, H. Lu, F. Wu, C.W. Chen, Energy and delay optimization for cache-enabled dense small cell networks, IEEE Trans. Veh. Technol. 69 (7) (2020) 7663–7678.
[56] T.X. Tran, A. Hajisami, D. Pompili, Cooperative hierarchical caching in 5G cloud radio access networks, IEEE Netw. 31 (4) (2017) 35–41.
[57] I. Parvez, A. Rahmati, I. Guvenc, A.I. Sarwat, H. Dai, A survey on low latency towards 5G: RAN, core network and caching solutions, IEEE Commun. Surv. Tutor. 20 (4) (2018) 3098–3130.
[58] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, J.G. Andrews, User association for load balancing in heterogeneous cellular networks, IEEE Trans. Wireless Commun. 12 (6) (2013) 2706–2716.
[59] T. Han, N. Ansari, A traffic load balancing framework for software-defined radio access networks powered by hybrid energy sources, IEEE/ACM Trans. Netw. 24 (2) (2015) 1038–1051.

Greta Vallero was born in Torino (Italy) on September 21st, 1993. She got her bachelor degree in Computer Engineering, in Politecnico di Torino, in 2015; she then obtained the master degree with *summa cum laude*, in ICT for Smart Societies (Telecommunication Engineering), in October 2017, in Politecnico di Torino. From March 2017 to August 2017, she was hosted by Ghent University, in Belgium, to work on her master thesis, supervised by Prof. Michela Meo, in collaboration with Prof. Wout Joseph (Ghent University) and Dr. Margot Deruyck (Ghent University). In 2018, she starts officially her Ph.D., under the supervision of Professor Michela Meo, at Politecnico di Torino. Her main research interests are Multi-Access Edge Computing, as well as the energy efficiency in Radio Access Networks, using the support of Machine Learning algorithms, through radio resource management and network renewable energy supply.

Margot Deruyck received the M.Sc. degree in Computer Science Engineering and the Ph.D. degree from Ghent University, Ghent, Belgium, in 2009 and 2015, respectively. From September 2009 to January 2015, she was a Research Assistant with Ghent University — IMEC – WAVES (Wireless, Acoustics, Environment & Expert Systems – Department of Information Technology). Her scientific work is focused on green wireless access networks with minimal power consumption and minimal exposure from human beings. This work led to the Ph.D. degree. Since January 2015, she has been a Postdoctoral Researcher at the same institution where she continues her work in green wireless access network.

Michela Meo received the Laurea degree in electronic engineering and the Ph.D. degree in electronic and telecommunications engineering from the Politecnico di Torino, Italy, in 1993 and 1997, respectively, where she has been a Professor since 2006. She has co-authored about 200 papers and edited a book with Wiley and special issues of international journals, including ACM Monet, Performance Evaluation, and Computer Networks. Her research interests include performance evaluation and modelling, green networking, and traffic classification and characterisation. She is an Associate Editor of the IEEE Communications Surveys & Tutorials and an Area Editor of the IEEE Transactions on Green Communications and Networking. He was an Associate Editor of the IEEE Transactions of Networking. She chairs the Steering Committee of IEEE OnlineGreenComm and the International Advisory Council of ITC. She was the Program Co-Chair of several conferences, including ACM MSWiM, IEEE Online GreenComm, IEEE ISCC, IEEE Infocom Miniconference, and ITC.

Wout Joseph received the M.Sc. degree in electrical engineering from Ghent University (Belgium) in July 2000. From September 2000 to March 2005, he was a research assistant at the Department of Information Technology (INTEC) of the same university. During this period, his scientific work was focused on electromagnetic exposure assessment. His research work dealt with measuring and modelling of electromagnetic fields around base stations for mobile communications related to the health effects of the exposure to electromagnetic radiation. This work led to a Ph.D. degree in March 2005. Since April 2005, he has been a postdoctoral researcher for IBBT-Ugent/INTEC (Interdisciplinary institute for Broadband Technology). Since October 2007, he has been a Post-Doctoral Fellow of the FWO-V (Research Foundation – Flanders). Since October 2009 he has been a professor in the domain of "Experimental Characterisation of wireless communication systems." His professional interests are electromagnetic field exposure assessment, propagation for wireless communication systems, antennas and calibration. Furthermore, he specialises in wireless performance analysis and Quality of Experience.