Caching at the edge in high energy-efficient wireless access networks

Caching at the edge in high energy-efficient wireless access networks

Greta Vallero*, Margot Deruyck**, Wout Joseph** and Michela Meo*

*Politecnico di Torino, Italy
**Ghent University-IMEC, Belgium

***Abstract*—In the next generation of Radio Access Networks (RANs), Multi-access Edge Computing (MEC) is considered a promising solution to reduce the latency and the traffic load of backhaul links. It consists of the placement of servers, which provide computing platforms and storage, directly at each Base Station (BS) of these networks. In this paper, the caching feature of this paradigm is considered in a portion of a RAN, powered by a renewable energy generator system, energy batteries and the power grid. The performance of the caching in the RAN is analysed for different traffic characteristics, as well as for different capacity of the caches and different spread of it. Finally, we verify that the usage of a strategy that aims at reducing the energy consumption does not impact the benefits provided by the mobile edge caching.**

***Index Terms*—Radio Access Network, Multi-access Edge Caching, Multi-access Edge Computing, renewable energy, energy efficiency**

I. Introduction

The mobile IP traffic will reach 77.5 exabyte (EB) per month by 2022, enormous increase compared to 11.5 EB per month in 2017 [1]. To meet this growth, the capacity of the new generation of networks, the 5G networks, is expected to increase by a factor 1000 more than 4G networks, and it will support up to 9 billion of mobile devices, and an heterogeneous range of applications, services and devices [2]. Beside this, the network energy efficiency is supposed to improve: given the same transmitted amount of traffic, 5G systems aim at consuming a fraction of the energy consumption of 4G mobile networks. To reach these goals, the latency should be reduced, the communication reliability improved and the bit rate increased [3]. Regarding the reduction of the latency, one of the most exploited solutions in literature is the MEC paradigm [4]. It consists of the implementation of computing platforms and storage on servers, which are placed at the edge of the network. 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 RANs, these servers are placed directly on BSs, in order to provide storage and computation services, in addition to access services.

As claimed in [5] and [6], several benefits are achieved when this approach is used, in addition to the reduction of the latency. First, the backhaul traffic load is reduced, since the content is stored locally and not taken from the content provider. Second, the quality of the multimedia content can be adapted to the user's channel. For these reasons, many works in literature focus on this topic. In [7], different use cases highlight the impact given by the usage of the MEC paradigm. In [5], the content placement problem is revisited and an effective solution, which aims at maximising the hit probability, is proposed. Authors in [4] and [8] treat the same objective, in an heterogeneous access network. The optimal size of the cache for a multi cell scenario is discussed in [9].

Meanwhile, the problem of the network energy efficiency has become more and more urgent. This is because 80% of the total mobile network is consumed by mobile access equipment. As reported in [2], from this 80%, 90% is consumed by the BSs of these networks. For this reason, the design of energy efficient RANs has been receiving lot of attention for many years. Many works address this issue through the activation of the sleep mode [10]–[14]. With its usage, the typical behaviour of the daily traffic demand is exploited: it usually presents short peaks and long valleys, during which the capacity of the RAN is under-utilised. Therefore, during these periods, characterised by a low traffic demand, the unneeded capacity is deactivated, allowing energy saving [10]. In addition to this, a trend is to consider local Renewable Energy Sources (RES), e.g., a wind turbine and/or a Photovoltaic (PV) panel system, for the power supply of the RAN, to limit the amount of energy which is produced by burning fossil fuels [15]. 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 [11], or when the amount of renewable energy that is generated by RES is not enough to power the RAN [12]. In the literature, an overview of the dynamic resource activation, according to the current traffic demand is presented in [10], [13], [14]. The proper sizing of the RES used for the RAN supply is discussed in [16] by simulation and in [17], [18] by analytical modeling.

Typically, the MEC paradigm and the usage of the strategy for the reduction of the energy consumption of a RAN are considered separately. For this reason, the novelty of this paper is to consider their simultaneous employment and to provide an overview of their mutual effects. In particular, in this work, the heterogeneous RAN in the city of Ghent, Belgium, proposed in [15], is considered. It is supplied by a PV panel system and energy batteries. Each BS of the considered RAN is equipped with a caching server, where the most popular contents are stored. The capacity of the network is dynamically adapted to the traffic demand and in case the

renewable energy generation is not sufficient for the network supply, an energy reducing strategy is applied. In the first part of the work, the benefits provided by the caching on the BSs are analysed. Then, the impact of the energy reducing strategy on the caching is discussed.

The paper is organised as follows. In section II, the scenario and the methodology of our work are presented. Results are discussed in section IV and the conclusions are drawn in section V.

II. Methodology

In this work, a similar scenario to the one considered in [15] is used. A part of heterogeneous RAN, covering an area of 0.3 km^2 of the city centre of Ghent, in Belgium, is considered (orange rectangle in Fig. 1).

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.

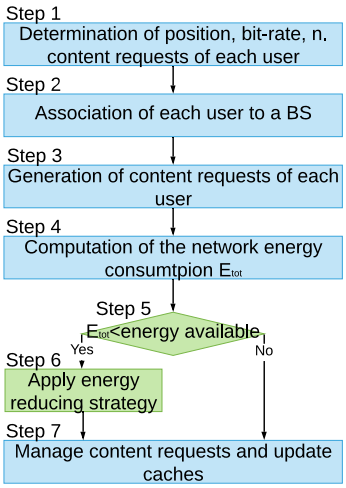


Fig. 2. Different steps of the simulations.

The RAN that covers this area is composed of 8 macro cell BSs, marked by the blue points in Fig. 1, each supported by 4 micro cell BSs, indicated with the brown points in the figure, whose radio coverage overlaps with the macro cell. Thus, micro cell BSs are deployed to provide additional capacity during high traffic demand periods. Long Term Evolution-Advanced (LTE-A) is the wireless technology considered with a frequency of 2.6 GHz and a channel bandwidth of 5 MHz with a single transmitting and receiving antenna, i.e., Single Input Single Output (SISO) for both the micro and macro cell BSs. The link budget assumed in this work is the same of [19].

Each BS of the cluster is equipped with a caching server, to push contents closer to the users. Similarly to [20], the hardware technology of each cache is DRAM (Dynamic Random Access Memory). As in [21] and [22], these servers update their contents according to Least Frequently Used (LFU) cache algorithm, to store the most popular contents of a file library composed by 1000 files, of 100 Mbit size each. With respect to the energy supply of the cluster, a centralised PV panel system, an energy battery, and the power grid are considered. As in [15], the capacity of the PV panel is 100 kWp, while the effective battery size is 50 kWh, with actually 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 in case of additional production, is conserved into the battery. If no renewable energy is available, the BSs take the required energy from the power grid. The data of the produced energy are taken from PVWATT [23], which reports the hourly generated energy. The data collected during the week from 3 January to 9 January, in Turin (Italy) are used in our simulations. The use of winter data leads to a worst case scenario in terms of produced energy.

Our simulations consist of different steps, performed in each time slot, lasting 1 hour, as shown in Fig. 2. First, the traffic during each hour is generated: the users, as well their position, required bit rate and requested content are determined (Step 1 in Fig. 2). Then, once each user has been associated to a BS (Step 2 in Fig. 2), if possible, the content requested by each user is determined (Step 3 in Fig. 2). At this point, for each hour, the energy consumption of the network is computed (Step 4 in Fig. 2). If the hourly available renewable energy is not enough for the network supply, a strategy is applied during that hour to reduce the energy consumption of the network (Steps 5, 6 in Fig. 2). Finally, the requested content are delivered and caches are updated (Step 7 in Fig. 2). Details of each step are given in the following sections.

A. Generation of the traffic

As in [12], [15] and [24], a user distribution is used to determine the number of active users for each time slot, lasting 1 hour. This distribution varies according to the hour of the day, to reflect the typical behaviour of the daily traffic demand. Then, for each user of each time slot, his/her position, requested bit rate and number of generated content requests are determined (Step 1 in Fig. 2). The position is defined according to a uniform distribution, while we assume that the required bit rate is 1 Mb/s. As in [7], the number of generated content requests is determined by a Poisson distribution, whose parameter λ is 1 request/minute.

B. Creation of the network

The capacity of the considered RAN responds to the instantaneous bit rate requested by the active users, in order to optimise the network with respect to the power consumption. This means that the input power of each BS is reduced as much as possible, to guarantee the user coverage [12]. In each time slot, which lasts for 1 hour, once the users are generated as described above, each of them is associated to the BS from which he/she experiences the lowest path loss (Step 2 in Fig. 2), provided that the BS has enough capacity to serve that user. This experienced path loss has to be lower than the maximum allowable path loss to receive the signal with the sufficient quality. When a user is associated to a BS, the requested files are determined, according to the popularity distribution of that BS (Step 3 in Fig. 2). According to [4], [20]–[22], [25], [26], this popularity distribution is a Zipf distribution, characterised by the parameter α. This parameter impacts the difference among contents in terms of popularity. A large α means that the most popular contents are significantly more popular than the other contents, and decreasing α, the popularity of content behaves more similarly to the uniform distribution. 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 this reference popularity, the popularity of 30% of content is randomly swapped to generate the popularity on each macro cell BS, so that slight differences among the files popularity at different locations (i.e., at different BSs) are introduced [7]. 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.

C. Energy consumption of the network

Once each user is associated to a BS and the content requests are determined, the energy consumption of the network during each time slot is computed (Step 4 in Fig. 2). The hourly energy consumption of the network, in watthour, is given by:

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

where N_{BS} is the number of the active BSs, $E_{b,comm}$ and $E_{b,server}$ are the energy consumption of the BS *b* due to the communication features and to the supply of the cache located on that BS, respectively. The $E_{b,comm}$ component is computed according to the model for the macro cell and micro cell BS proposed in [19]. According to [21] and [20], $E_{b,server}$, in watthour, is given by:

$$E_{b,server} = \omega_{MEC} \cdot C_{server} \cdot t \tag{2}$$

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.

D. Energy reduction strategy

At the beginning of each time slot, once the energy consumption is computed, if it is larger than the renewable energy, which is available during that time interval, given by the energy produced by the PV panel system and stored in the battery, an energy reduction strategy is applied (Steps 5-6 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 generated and the stored renewable energy are not enough to power the network. The users who have been connected to each deactivated micro cell BS are reconnected to a macro cell BS, if possible, e.g., if there is a macro cell BS that has enough available capacity and if the experienced path loss is lower than the maximum possible.

E. Content delivery

During each time interval, once the energy reduction strategy is applied, if needed, 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, with latency $T_{bs,u}$. 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 latency is given by $T_{BS,bs} + T_{bs,u}$. 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 $T_{cp,BS}+T_{BS,bs}+T_{bs,u}$. In case a user is associated to 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 LFU cache algorithm, so as to always cache the most popular contents.

F. Initial state of caches

To determine which contents are stored in each cache at the beginning of a simulation, a preliminary phase is required. For each value of α, a very long simulation, lasting 100 weeks, is performed, starting with empty caches. At each time slot, the occurrences of a content request are updated, as well as the stored files in each cache. We assume that when the number of variations in each cache stabilises, the transient phase for each cache filling is over. For this reason, the files in each cache at that time interval are the contents which are cached at the beginning of the simulation.

III. Key Performance Indicators

1) Energy consumption: The energy consumption of the network during the simulation is given by

$$E = \sum_{t=1}^{T} E_{tot,t} \tag{3}$$

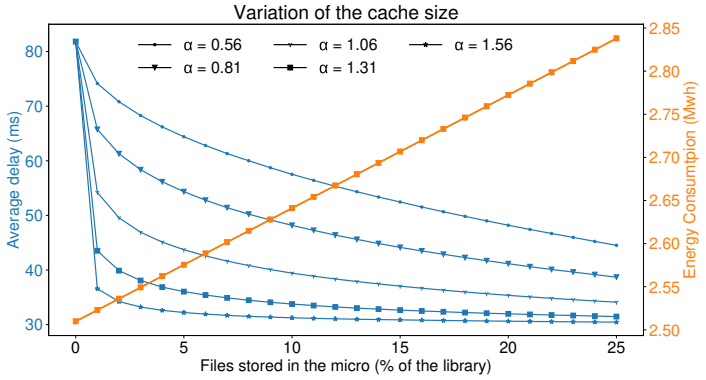


Fig. 3. Average delay (in blue) and energy consumption (in orange) varying the dimension of each cache, for different values of the parameter α.

TABLE I
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

where $E_{tot,t}$ is the energy consumption of the network at time t and it is computed as reported in (1) and T is the duration of the simulation.

2) 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 a given threshold and if that BS has enough capacity to provide the required bit rate.

3) Average Delay: This is the average delay experienced by users and it is given by:

$$D = \frac{1}{\sum_{t=1}^{T} U_t \sum_{u=1}^{U_t} R_u} \sum_{t=1}^{T} \sum_{u=1}^{U_t} \sum_{r=1}^{R_u} d_{u,r} \qquad (4)$$

where $d_{u,r}$ is the delay which is experienced by the user u, for the content request r. R_u, U_t and T correspond to the number of requests required by the user u, the number of served users at time t and the duration of the simulation, respectively.

4) Hit - 1 hop probability: This is the probability that the requested content is stored locally on the BS to which the considered user is connected.

5) 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.

6) Miss probability: This is the probability that the requested content is taken from the content provider, since it is not stored in the cache of the BS, to which the user is associated nor in the one of the corresponding macro cell BS, if the considered user is associated to a micro cell BS.

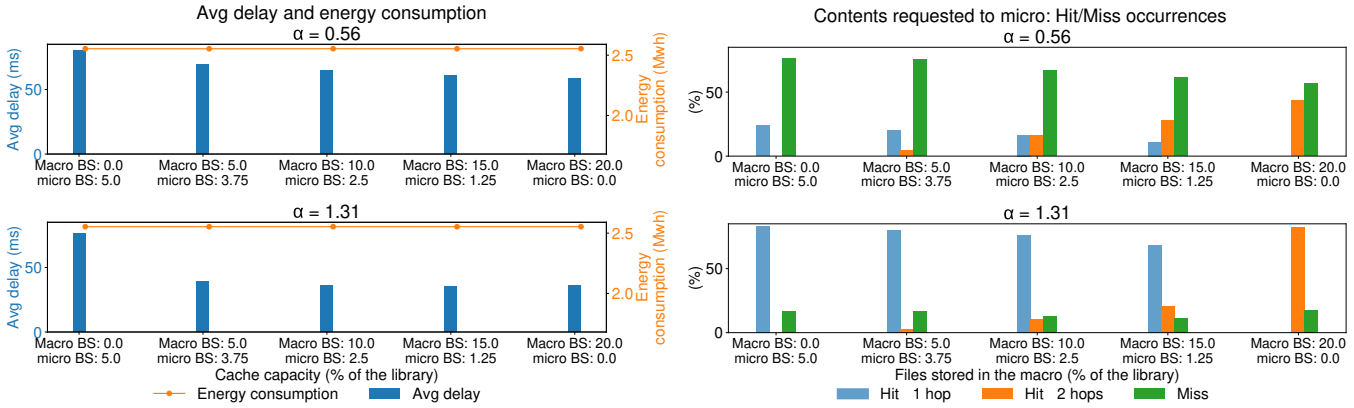
IV. PERFORMANCE EVALUATION

In this section, we discuss the results obtained by simulations, lasting 1 week. The values of latency, as well as the value of the parameter ω_{MEC} of (2) are reported in Table I. They are taken from [4] and [20], respectively.

A. Impact of the cache size and the popularity

In the first part of our work, we analyse the effects of the parameters which affect the performance of the local caching. To do this, we simulate the scenario described in Section II, 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, varying the size of the cache on each micro and on each macro (the cache on the macro 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, which characterises the Zipf's distribution. When the percentage of stored library is zero, we are in the case in which no local caching is performed. The growth of the size of the cache generates a reduction of the experienced delay, since more content 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 50%, if 1% of the library is locally stored. Conversely, a small value of α indicates that the files have similar popularity. In this scenario, larger caches are needed to achieve significant delay reduction: if 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 by 30%. The energy consumption increases linearly with the cache size, see (2). Nevertheless, this growth is limited to 13%, when the cache stores 25% of the library and lower than 3% if 5% of the library is cached.

Besides the impact of the size of the cache on the user experience, we also investigate the impact of its distribution. 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



(a) (b)

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.

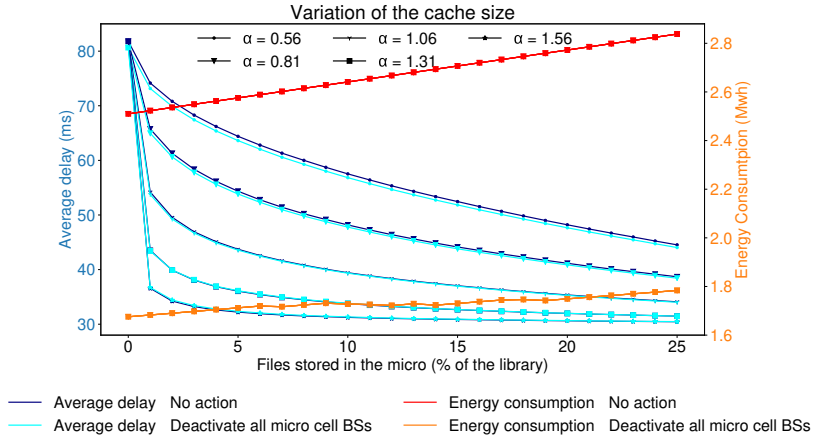


Fig. 5. Delay and energy consumption varying the dimension of each cache, for different values of the parameter α, when *No action* and *Deactivate all micro cell BSs* are used.

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. 4a, the average delay (blue bars) and the energy consumption (orange line) are reported, for these values of cache capacity, for α equal to 0.56 and 1.31, when *No action* strategy is used. Fig. 4b shows the probability of the possible events that a user might experience when served by a micro cell BS. Blue bins indicate the probability to experience a *hit - 1 hop*, i.e., is the content is cached on that micro cell BS, 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. 4a, the energy consumption is constant since the total capacity does not change. Moreover, the plot reveals again that the delay reduction strictly depends on the popularity, i.e., on the parameter α and marginally depends on the cache distribution among BSs. Indeed, as reported in Fig. 4b, if α is 0.56 and the micro cell BSs can store up to 5% of the library, no more than 24% of the requests on a micro cell BSs can be satisfied locally (on that BS), while this number grows to 83% if α is 1.31. Similarly, when all the considered caching capacity is put on the macro cell BS, that BS satisfies 43% and 83% of the requests, respectively. 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. 4b), from Fig. 4a, 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. 4b). Nevertheless, the resources on the macrocell are precious and it is convenient to install some capacity on micro cell BSs too: this relieves the effort on the macrocells and allows to achieve some local (1 hop from users) hits, 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.

B. Impact of using energy reduction strategy

We now analyse the effect of the usage of the energy reduction strategy that we called *Deactivate all micro cell BSs* on the caching paradigm. If this strategy is employed, the micro cell BSs are switched off in case the locally renewable produced and stored energy is not sufficient for the network supply. In Fig. 5, the impact of the variation of the dimension of each cache is shown, in terms of experienced delay (left y-axis) and of energy consumption (right y-axis) in blue and light blue and in red and in orange, when *No action* and *Deactivate all micro cell BSs* strategies are used, respectively. The usage of *Deactivate all micro cell BSs* significantly reduces the energy consumption of the network: when it is employed, the system drops its consumption by 33% to 37%, according to the dimension of the cache. With *Deactivate all micro cell BSs*, the energy consumption does not grow constantly with the increase of the cache capacity. This is because above a given storage dimension (up to when 5% of the library is stored), the system stabilises, e.g. all 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* 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 with low values of α, since in these cases the content needs to be taken from the content provider more often, so the impact of this reduction of distance is higher. The employment of *Deactivate all micro cell BSs* strategy reduces the user coverage, from 97% given with *No action* to 94%, which is acceptable anyway.

V. Conclusion

In this paper, the caching paradigm is considered in a portion of a RAN, composed by 8 macro cell BSs, each supported by 4 micro cell BSs. The caching paradigm consists of the installation of caching servers on the BSs, to push the most popular contents closer to users so as to reduce latency and traffic in the backhaul network. The considered RAN is powered by a PV panel system, an energy battery and the grid. First, we verify the benefits given by the employment of these local caching servers and we analyse the effects of the cache capacity, as well as the impact of the popularity on the delay experienced by users and the energy consumption of the network. We notice that these metrics strictly depend on the characteristics of the popularity, on the capacity of the caching servers and on the distribution of this capacity among the BSs. Then, we check that 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.

Our results lead to two conclusions. First, caching can be very effective in reducing latency also when dynamic activation of the BSs is implemented as in the case of green RANs that are powered by renewable energy sources and that activate the BSs based on the availability of locally produced energy. 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 allowing the micro cells to often respond without any involvement of the macro cell. As next steps of our work we will consider wind turbines as local renewable energy generator, and strategies for dense RANs, which implement the MEC paradigm.

References

[1] C. V. Forecast, "Cisco visual networking index: Forecast and trends, 2017–2022," *White paper, Cisco Public Information*, 2019.

[2] 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," *arXiv preprint arXiv:1903.09627*, 2019.

[3] 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 communications magazine*, vol. 52, no. 5, pp. 26–35, 2014.

[4] M. Chen, Y. Qian, Y. Hao, Y. Li, and J. Song, "Data-driven computing and caching in 5g networks: Architecture and delay analysis," *IEEE Wireless Communications*, vol. 25, no. 1, pp. 70–75, 2018.

[5] B. Blaszczyszyn and A. Giovanidis, "Optimal geographic caching in cellular networks," in *2015 IEEE international conference on communications (ICC)*. IEEE, 2015, pp. 3358–3363.

[6] K. Poularakis, G. Iosifidis, A. Argyriou, and 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.

[7] T. X. Tran, A. Hajisami, P. Pandey, and D. Pompili, "Collaborative mobile edge computing in 5g networks: New paradigms, scenarios, and challenges," *IEEE Communications Magazine*, vol. 55, no. 4, pp. 54–61, 2017.

[8] K. Poularakis, G. Iosifidis, and L. Tassiulas, "Approximation algorithms for mobile data caching in small cell networks," *IEEE Transactions on Communications*, vol. 62, no. 10, pp. 3665–3677, 2014.

[9] X. Peng, J. Zhang, S. Song, and K. B. Letaief, "Cache size allocation in backhaul limited wireless networks," in *2016 IEEE International Conference on Communications (ICC)*. IEEE, 2016, pp. 1–6.

[10] 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 Communications Surveys & Tutorials*, vol. 16, no. 4, pp. 2259–2285, 2014.

[11] M. Dalmasso, M. Meo, and D. Renga, "Radio resource management for improving energy self-sufficiency of green mobile networks," *ACM SIGMETRICS Performance Evaluation Review*, vol. 44, no. 2, pp. 82–87, 2016.

[12] M. Deruyck, W. Joseph, E. Tanghe, and L. Martens, "Reducing the power consumption in lte-advanced wireless access networks by a capacity based deployment tool," *Radio Science*, vol. 49, no. 9, pp. 777–787, 2014.

[13] 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.

[14] 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*, vol. 34, no. 4, pp. 697–709, 2016.

[15] M. Deruyck, D. Renga, M. Meo, L. Martens, and W. Joseph, "Accounting for the varying supply of solar energy when designing wireless access networks," *IEEE Transactions on Green Communications and Networking*, vol. 2, no. 1, pp. 275–290, 2017.

[16] M. Meo, Y. Zhang, R. Gerboni, and M. A. Marsan, "Dimensioning the power supply of a lte macro bs connected to a pv panel and the power grid," in *2015 IEEE International Conference on Communications (ICC)*. IEEE, 2015, pp. 178–184.

[17] V. Chamola and B. Sikdar, "Resource provisioning and dimensioning for solar powered cellular base stations," in *2014 IEEE Global Communications Conference*. IEEE, 2014, pp. 2498–2503.
[18] ——, "A multistate markov model for dimensioning solar powered cellular base stations," *IEEE Transactions on Sustainable Energy*, vol. 6, no. 4, pp. 1650–1652, 2015.
[19] M. Deruyck, W. Joseph, and L. Martens, "Power consumption model for macrocell and microcell base stations," *Transactions on Emerging Telecommunications Technologies*, vol. 25, no. 3, pp. 320–333, 2014.
[20] N. Choi, K. Guan, D. C. Kilper, and 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.
[21] Z. Luo, M. LiWang, Z. Lin, L. Huang, X. Du, and M. Guizani, "Energy-efficient caching for mobile edge computing in 5g networks," *Applied sciences*, vol. 7, no. 6, p. 557, 2017.
[22] K. Shanmugam, N. Golrezaei, A. G. Dimakis, A. F. Molisch, and G. Caire, "Femtocaching: Wireless content delivery through distributed caching helpers," *IEEE Transactions on Information Theory*, vol. 59, no. 12, pp. 8402–8413, 2013.
[23] A. P. Dobos, "Pvwatts version 5 manual," National Renewable Energy Lab.(NREL), Golden, CO (United States), Tech. Rep., 2014.
[24] M. Deruyck, E. Tanghe, D. Plets, L. Martens, and W. Joseph, "Optimizing lte wireless access networks towards power consumption and electromagnetic exposure of human beings," *Computer Networks*, vol. 94, pp. 29–40, 2016.
[25] M. Chen, Y. Hao, L. Hu, K. Huang, and V. K. Lau, "Green and mobility-aware caching in 5g networks," *IEEE Transactions on Wireless Communications*, vol. 16, no. 12, pp. 8347–8361, 2017.
[26] Y.-C. Wang and 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.