IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 1, 2020, accepted April 25, 2020, date of publication May 4, 2020, date of current version June 4, 2020.

Digital Object Identifier 10.1109/ACCESS.2020.2992330

A Robust User Association, Backhaul Routing, and Switching Off Model for a 5G Network With Variable Traffic Demands

ENRICA ZOLA AND ISRAEL MARTIN-ESCALONA
Department of Network Engineering, Universitat Politècnica de Catalunya (UPC-BarcelonaTECH), ES-08034 Barcelona, Spain
Corresponding author: Enrica Zola (enrica.zola@upc.edu)

This work was supported by the Spanish Government and European Regional Development Fund (ERDF) through Comisión Interministerial de Ciencia y Tecnología (CICYT) under Project PGC2018-099945-B-I00.

ABSTRACT With the expected increase in data traffic (e.g., video) and in the user devices generating new traffic (e.g., device-to-device communication, Internet of Things, etc.), the evolution of next-generation mobile networks (e.g., 5G networks) has gone towards heterogeneous deployments where multiple small cells coexist in the same area covered by a macro base station. To reduce the capital expenses in the network, a wireless mesh can be used, which is made of millimeter-wave links that route the data traffic of the mobile users inside the backhaul network. Such an increase in the number of deployed base stations inevitably increases the power consumption; hence, the operating expenses and the CO_2 consumption also increase. To achieve greener mobile communications, sleep-mode strategies have been considered in order to switch off the unused network components. However, the switching on/off should be made according to the traffic demanded by the users and with the aim of guaranteeing the demanded service at any time. Given that the traffic demand and networking traffic fluctuate over time at each location, we propose a robust mixed integer linear problem that jointly solves the user association, the backhaul routing paths in the wireless mesh and the switching off of the unused links with the aim of minimizing the power consumption. The robust strategy is based on the Γ-robust approach and is able to guarantee the user demand while taking into account its intrinsic variability. A thorough evaluation has been performed in order to analyze the impact of the robust strategy on the network performance.

INDEX TERMS 5G, energy efficiency, green networks, mesh backhaul, millimeter wave, robust optimization, routing, switching off, user association.

I. INTRODUCTION

The use of mobile and wireless communications has been growing in the last decades, making it easier to develop several new services including the Internet of Things (IoT), e-health, smart cities, autonomous driving, etc. Currently, our society is continuously and increasingly making use of such technologies for everyday life activities, including both work and entertainment related. The trend for many years now has been an increase in the traffic demand of mobile users, which has shaped the evolution from 2G to 4G systems and has paved the way to the definition of a disruptive new technology; the fifth-generation (5G) paradigm will bring millimeter-wave (mmWave) communications, heterogeneous technologies and new devices [1] that contribute to an enormous growth in the data traffic that must be handled. For example, CISCO foresees that the number of devices connected to IP networks will be more than three times the global population by 2022 [2]. This increase is mainly due to the wide spread of the IoT in industry and the consequent increase in machine to machine (M2M) communications, as well as an increase in consumer video use: nearly 79% of the world's mobile data traffic is expected to be video by 2022 [2].

As these massive volumes of data need to be transferred throughout the network among a growing number of greedy users, the whole infrastructure should be reconsidered. A denser deployment of base stations (BSs) is expected in 5G networks, where several small cells (SCs) will be installed under the coverage of a macro eNodeB (eNB) in order to meet the increasing demand [3]. With this new paradigm, heterogeneous networks (HetNets) are expected to achieve

The associate editor coordinating the review of this manuscript and approving it for publication was Muhammad Omer Farooq.

energy efficient communications [4], as the SCs can be placed in strategic areas (e.g., hotspots) to enhance the network performance. While the energy consumption is expected to increase with the number of BSs [5], it becomes necessary to study new solutions that contribute to reducing both the capital (CAPEX) and operating expenses (OPEX). In this context, the SCs should communicate with each other and with the eNB through a mmWave mesh backhaul network instead of fixed optical links, as suggested in the 5GMiEdge project [6]. In fact, mmWave links are identified as one of the key enablers of 5G because of their very high data rates, effectiveness at handling interference and their wide bandwidth [7].

To further reduce the OPEX, a key challenge is determining which network nodes have to be used (according on the network traffic) and which can be switched off, thus decreasing the overall power consumption [4]. However, sleep-mode techniques should take into account the effects of traffic load-dependent factors on the energy consumption [8]. A mixed integer linear program (MILP) was proposed in [9] that minimizes the energy consumption of a meshed 5G network by switching off the unused resources (e.g., SCs, BSs, and mmWave links) and arranging the user associations and the routing paths in the backhaul (BH) network according to the traffic patterns, thus adapting the network topology to the user needs. More recently, the model has been improved, the network energy efficiency (EE) has been evaluated at different times of the day [10], and several heuristic algorithms have been considered that can provide a faster resolution of the given problem.

It is well known that the user distribution and traffic demand vary for a given location depending on the time of day and the day of the week [11]. Thus, the traditional trend of dimensioning the network resources based on the peak traffic may lead to the underutilization of the infrastructure, to unnecessary OPEX for the network operator and to unnecessary CO_2 emissions for our planet. On the other hand, deploying a less dense infrastructure than the user requirements may lead to congestion, especially in peak hours. The nondeterministic nature of wireless networks makes the dimensioning of network resources increasingly more challenging; some clear examples of uncertain parameters are the wireless channel conditions, the users' fluctuating bit rate requirements and users' movements [12]. Such fluctuations make the dimensioning of the network resources increasingly more challenging; not only should the EE of future communications be optimized, but also a valid topology guaranteeing that the user equipments (UEs) are given the demanded services should be provided. Since the UEs demand cannot be exactly predicted, a robust strategy [13] should be applied when optimizing the power consumption in a 5G network. The variability in the input data (e.g., user demand) may lead to infeasible solutions in the MILP proposed in [9] and in [10]. That is, higher demand from one or more users may make the current topology unable to satisfy all the user demands.

This work provides a robust formulation for the joint optimization of the user association, backhaul routing and on/off strategies, aiming to reduce the total power consumption in a 5G network while guaranteeing that the user needs are met. To this end, the robust approach from [13] is applied to the MILP presented in [10], and the solution is then protected against variability in the demand of a given number of users. The main contributions of this article are as follows:

1) identifying the budget of uncertainty and the uncertainty set in our problem according to the Γ-robust approach in [13];
2) providing a mathematical model for the robust counterpart of the MILP presented in [10];
3) developing a robust MILP for the abovementioned problem;
4) providing a thorough evaluation of the power saving, network performance and impact of the robustness level in our model.

Section II summarizes the main contributions of the switching on/off strategies and optimization approaches on the EE for 5G networks and provides the background of our research. The problem formulation is described in Section III. First, t he 5G system model underlying the optimization problem is described in Section III-A; then, the MILP for the nominal case (i.e., when demand fluctuations are not taken into account) is presented in Section III-B, which is based on the formulation in [10]. Section IV presents the uncertainty sets needed for the robust formulation that is described in Section IV-C; then, the robust counterpart for the robust MILP is detailed in Section IV-D. The evaluation assessment is detailed in Section V; the results of the nominal case are first presented and then used as a reference for the evaluation of the robust model in Section V-B. The final remarks and open issues are discussed in Section VI.

II. RELATED WORK

In the last decade, HetNets have been foreseen as the potential solution to achieve EE in 5G networks and beyond [4]. In a HetNet, eNBs are deployed in order to guarantee a minimum coverage for a large area, and several relays (e.g., SCs) are deployed in the same area for coverage extension, throughput enhancement, load balancing, etc., thus achieving overall lower energy consumption for the network. In this context, the authors in [4] investigate the optimal energy efficient deployment strategies for the SCs in 5G HetNets. Another crucial factor for the EE in HetNets is the user association problem. The high EE potential of a cognitive algorithm is studied in [14]; however, a tree topology is considered in the BH, and there are corresponding problems related to the centralized topologies (e.g., unreliability in case of link failure, bottlenecks in case of highly loaded links, inflexibility to adapt to changing demands, etc.). A meshed topology is thus introduced in [15], where the benefits of load balancing on the spectral efficiency and on the EE are shown.

Although EE has been largely studied in the literature, the authors in [16] argue that none of the previous works have taken the EE as the objective function; thus, the authors consider an EE maximization algorithm that jointly assigns optimal flows on BH links, minimizes the power consumption in the access network (AN) and BH, and maximizes the AN throughput. In addition, the work in [9] considers the possibility of switching off the unused network components, thus further minimizing the power consumed in a HetNet. However, the meshed topology used in [16] and [9] with a single aggregator (e.g., the macro eNB with a fixed connection to the core network) may represent a bottleneck. This issue was solved in [10], which improves the system model of [9] and proposes a fast online solution policy. Such a policy makes the implementation possible in realistic deployments, where an SDN controller triggers the switching on/off of the network components based on the results from the optimization policy. This paper also focuses on a meshed HetNet where multiple BSs may act as aggregators and route the data towards the core network; different from the previous works, the users' demand is allowed to deviate from the nominal value and the proposed robust MILP still guarantees feasible solutions to the optimal user association, backhaul routing and switching off in 5G HetNets.

Sleep-mode techniques have been widely proposed and analyzed in the literature [8] and have been found to be a promising solution for green networking; they may take advantage of changing traffic patterns to switch off the lightly loaded BSs. However, the authors in [8] argue that simplifying assumptions, such as ignoring the effects of traffic load-dependent factors on energy consumption, introduce inaccuracies, thus affecting the benefits of the sleep-mode technique. The authors conclude that major benefits are obtained in areas with a very dense deployment of BSs and where the average traffic is low but has a high deviation. Additionally, better EE can be obtained when turning off those BSs with larger fixed proportions of energy consumption; that is, a model where the macro eNBs may be switched off as in [10] can bring further benefits. Moreover, highly connected networks with high randomness seem to save more power by using sleep modes more efficiently [4]. More recently, the impacts of different paradigms for 5G networks, such as the mmWave network and ultradense HetNets, on the EE of radio access networks are discussed in [17].

The access network selection problem is studied in [18], where multi-radio terminals are considered and their download rate is maximized for a given amount of download time. The download rate is subject to uncertainty due to the possibility of selecting different access networks. An integer linear program (ILP) for the energy efficient planning of wireless networks is presented in [12], where the total power consumed in the network and the total number of unsatisfied users are minimized. The authors also propose applying cutting planes to reduce the complexity of the model. Similarly, the authors in [19] present a robust optimization approach to energy savings in wireless local area networks (WLANs), which also incorporates user mobility under the Γ-robust optimization paradigm presented in [13]. They argue that while the capacity of the wireless link can be assumed to be stable over long time periods, there are deviations from the average at specific time intervals. Multiband robust optimization (MRO) is used in [20] to model the user mobility uncertainty in a WLAN, while the channel fluctuations are modeled through the Γ-robust approach. The MRO forms a more accurate model of the user mobility and, thus, achieves stronger overall energy saving.

The approach followed in this work is in line with the Γ-robust approach of the abovementioned publications. Similar to our work, the authors applied robust optimization in order to deal with the uncertainty in the users' demand. However, their focus is limited to the access links only, disregarding the complexity of a joint optimization of the AN and BH links in heterogeneous 5G networks. The present work covers this gap by applying Γ-robust optimization in order to deal with the uncertainty in the data rates demanded by the users. The uncertainty affects not only the access links of the 5G network but also the wireless BH meshed network formed by the SCs where the data have to flow—and thus be routed—before being finally delivered to the user through the AN link of one of the available BSs. This is the first attempt to provide a robust energy saving solution in a 5G HetNet that copes with user demand fluctuations; to the best of the authors' knowledge, none of the works in the literature has addressed this issue before.

III. USER ASSOCIATION, BACKHAUL ROUTING AND SWITCHING OFF MODEL FOR A 5G NETWORK

A. SYSTEM MODEL

The focus of our study is a 5G network composed of a set $\mathcal{B}$ of BSs, which can be either eNBs or SCs. The SCs are located inside the area of each eNB and are interconnected to each other and with the eNB through a set of line-of-sight (LOS) mmWave BH links, denoted as $\mathcal{L}_{BH}$. Each eNB and a given number of SCs per eNB play the role of the aggregators for the eNB area traffic since they have a fixed fiber connection to the core network. The set of aggregators for each eNB's area traffic is indicated by $\mathcal{A}$, where $\mathcal{A} \subseteq \mathcal{B}$. Fig. 1 depicts the system model for the case with three eNBs, where each eNB and four SCs are the aggregators.

We consider a set $\mathcal{U}$ of 5G UEs that wants to download data from the core network, thus focusing on the downlink case as in [10]. Each traffic flow has to be routed from the core network through the aggregators over some specific BH links in the mesh in order to reach the UE. Each UE must be connected with only one BS using one access link from the set $\mathcal{L}_{AN}$ of available microwave links. A specific guaranteed bit rate (GBR) demand for each UE is considered, which is based on its service [21] and is denoted by d^u. Depending on the radio conditions and the user demand, a given number of physical resource blocks (PRBs) will be required to satisfy the demand of each UE for the access link. The maximum

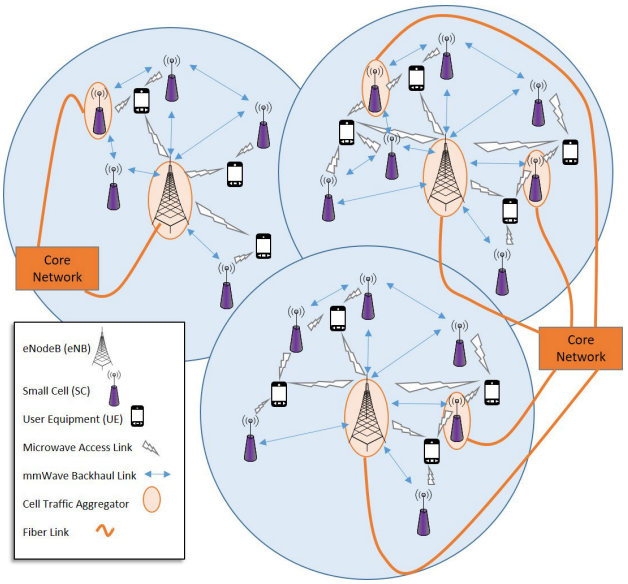

FIGURE 1. System model.

transmitted power of each BS is divided equally among the available PRBs at the BS (i.e., flat slow fading channels are assumed).

The proposed model provides 1) the set of UE associations, 2) the routing path for each UE's traffic flow in the BH mesh, and 3) the set of BSs and of BH links that may be switched off since they do not carry the users' traffic. For this, a MILP was formulated in [10] that was aimed at minimizing the power consumed in the network when providing the services demanded by a given number of users. The model is summarized in Section III-B. To easily introduce the reader to the robust approach presented in Section IV, a variable demand from each user will be considered and the proposed robust model will again be able to provide the set of UE associations, routing paths and ON/OFF settings under variable users' demand.

Without loss of generality and for the sake of simplicity, we assume that the fiber links from the core network to the aggregators are characterized by very high capacities (i.e., bottlenecks are avoided on those fiber links) and negligible power consumption (i.e., the power consumed by the fiber links has no impact on the total power consumption) [10].

B. MILP FORMULATION FOR THE NOMINAL CASE

The minimization of the power consumed in the proposed scenario was formulated as a MILP in [10] and is summarized in this section. Given the GBR demand d^u for each user $u \in \mathcal{U}$, the objective is to minimize the total power of all the active BSs in the network, which is given by

$$\begin{aligned} &\underset{x^u_{(i,j)}, s^{\text{AN}}_i, s^{\text{BH}}_{(i,j)}}{argmin} \sum_{i\in\mathcal{B}} p_i(x^u_{(i,j)}, s^{\text{AN}}_i, s^{\text{BH}}_{(i,j)}), \\ &\text{s.t. } a)\ x^u_{(i,j)} \in \{0,1\}, \quad \forall u \in \mathcal{U}, \\ &\qquad \forall (i,j) \in \mathcal{L}_{\text{BH}} \cup \mathcal{L}_{AN} \\ &\quad b)\ s^{\text{AN}}_i, s^{\text{BH}}_{(i,j)} \in \{0,1\}, \quad \forall i \in \mathcal{B}, \\ &\qquad \forall (i,j) \in \mathcal{L}_{\text{BH}} \\ &\quad c)\ \sum_{u\in\mathcal{U}} \sum_{(i,j)\in\mathcal{L}_{AN}} x^u_{(i,j)} c_{(i,j)} \le c_{i_{\max}}, \\ &\qquad \forall\, i \in \mathcal{B} \\ &\quad d)\ \text{Power constraints} \\ &\quad e)\ \text{Switch ON/OFF constraints} \\ &\quad f)\ \text{Path conservation constraints} \end{aligned} \tag{1}$$

$p_i(x^u_{(i,j)}, s^{\text{AN}}_i, s^{\text{BH}}_{(i,j)})$ is the power of a BS i considering its access and BH links (in the following, we will refer to it as p_i); $x^u_{(i,j)}$ is a binary variable that indicates whether the link (i,j) is used (1) or not (0) by user u (i.e., an access link is characterized by $x^u_{(i,u)}$); and s^{AN}_i and $s^{\text{BH}}_{(i,j)}$ are binary variables that indicate whether the BS i or BH link (i,j), respectively, is switched on (1) or off (0).

Equation (1c) is the **capacity constraint**, where $c_{i_{\max}}$ is the number of PRBs that are available at BS i, and $c_{(i,j)}$ is the number of PRBs that user j needs for association with BS i.

$$c_{(i,j)} = \left\lceil \frac{d^u}{se_{(i,j)}} \right\rceil, \quad \forall (i,j) \in \mathcal{L}_{AN} \tag{2}$$

$se_{(i,j)}$ is the spectrum efficiency of the access link (i,j) and is calculated as in [10].

$$se_{(i,j)} = BW_{PRB}\, log_2(1 + SINR_{(i,j)}), \tag{3}$$

where BW_{PRB} is the bandwidth (BW) of a PRB.

Regarding the **power constraints** in (1d), the power p_i of BS i can be computed as the sum of the power consumed on the microwave links that are used by each UE u once it associates with i [22] plus the power consumed by each BH link that is activated from BS i to the mesh.

$$\begin{aligned} p_i = {} & NTX^{\text{AN}}_i \left(s^{\text{AN}}_i p^{\text{AN}}_{0_i} + \Delta^{\text{AN}}_p\, p^{\text{AN}}_{out_i} \right) \\ & + \sum_{(i,j)\in\mathcal{L}_{\text{BH}}} NTX^{\text{BH}}_{(i,j)} \left(s^{\text{BH}}_{(i,j)} p^{BH}_{0_{(i,j)}} + \Delta^{\text{BH}}_p p^{BH}_{out_{(i,j)}} \right), \\ & \forall i \in \mathcal{B} \end{aligned} \tag{4}$$

NTX^{AN}_i and $NTX^{\text{BH}}_{(i,j)}$ represent the numbers of transceiver chains of BS i and the BH link (i,j), respectively; Δ_p is the factor for the load-dependent power consumption whose value can change for different types of antennas [22].

The output transmitted power of BS i is calculated as

$$p^{\text{AN}}_{out_i} = \frac{p^{\text{AN}}_{\max_i}}{c_{i_{\max}}} \sum_{u\in\mathcal{U}} \sum_{(i,j)\in\mathcal{L}_{\mathcal{AN}}} (x^u_{(i,j)}\, c_{(i,j)}), \quad \forall i \in \mathcal{B} \tag{5}$$

and, similarly, the output transmitted power $p^{BH}_{out_{(i,j)}}$ of the BH transceiver of link (i,j) can be computed in (6), as shown at the bottom of the next page since it spreads over two columns. Please recall that linear interpolation is applied in order to keep the model linear (e.g., here, the number of breakpoints is $nbkp$).

The load on each BH link is evaluated as

$$load^{\text{BH}}_{(i,j)} = \frac{\sum_{u\in\mathcal{U}} x^u_{(i,j)} d^u}{BW_{(i,j)}}, \quad \forall (i,j) \in \mathcal{L}_{\text{BH}}. \tag{7}$$

Moreover, $p^{BH}_{out_{(i,j)}}$ cannot exceed the maximum transmission power $p^{\mathrm{BH}}_{max_{(i,j)}}$ that is allowed on a BH link

$$0 \leq p^{\mathrm{BH}}_{out_{(i,j)}} \leq p^{\mathrm{BH}}_{max_{(i,j)}}. \tag{8}$$

The **switch ON/OFF constraints** (1e) allow a BS i or a BH link (i,j) to be on if and only if there is traffic of at least one user on that AN or BH link, respectively. As mentioned in [10], the linearized switching ON/OFF of BS i in the access is given by

$$\begin{cases} \sum_{u\in\mathcal{U}} \sum_{(i,j)\in\mathcal{L}_{\mathrm{AN}}} x^u_{(i,j)} + BIG\, y^{\mathrm{AN}}_i \geq 1 \\ s^{\mathrm{AN}}_i + BIG\, y^{\mathrm{AN}}_i \geq 1 \qquad\qquad \forall\, i \in \mathcal{B}, \\ \sum_{u\in\mathcal{U}} \sum_{(i,j)\in\mathcal{L}_{\mathrm{AN}}} x^u_{(i,j)} \leq BIG\,(1 - y^{\mathrm{AN}}_i) \end{cases} \tag{9}$$

and the linearized switching ON/OFF of BH link (i,j) is written as

$$\begin{cases} \sum_{u\in\mathcal{U}} x^u_{(i,j)} + BIG\, y^{\mathrm{BH}}_{(i,j)} \geq 1 \\ s^{\mathrm{BH}}_{(i,j)} + BIG\, y^{\mathrm{BH}}_{(i,j)} \geq 1 \qquad\qquad \forall\, (i,j) \in \mathcal{L}_{\mathrm{BH}}, \\ \sum_{u\in\mathcal{U}} x^u_{(i,j)} \leq BIG\,(1 - y^{\mathrm{BH}}_{(i,j)}) \end{cases} \tag{10}$$

where BIG is a large positive number, and y^{AN}_i and $y^{\mathrm{BH}}_{(i,j)}$ are binary variables needed for the linear transformation of the constraint.

According to our assumption of considering the downlink (i.e., the traffic flow from the core network towards a UE), the **path conservation constraints** (1f) must ensure that the traffic entering a node must exit it unless the node is the source of the traffic or its destination.

$$\sum_{(i,j)\in\mathcal{L}} x^u_{(i,j)} - \sum_{(j,i)\in\mathcal{L}} x^u_{(j,i)} = \begin{cases} 1, & \text{if } i = \text{ source}, \\ -1, & \text{if } i = u \text{ (sink)}, \\ 0, & \text{otherwise}, \end{cases} \tag{11}$$

$\forall u \in \mathcal{U}$, $\forall i$ and $j \in \mathcal{B} \cup \mathcal{U}$, and where $\mathcal{L} = \mathcal{L}_{\mathrm{BH}} \cup \mathcal{L}_{\mathrm{AN}}$.

Moreover, the traffic flow of one user cannot be split on multiple routes when exiting a node, that is,

$$\sum_{(i,j)\in\mathcal{L}_{\mathrm{BH}}\cup\mathcal{L}_{\mathrm{AN}}} x^u_{(i,j)} \leq 1, \quad \forall u \in \mathcal{U},\ \forall i \in \mathcal{B}. \tag{12}$$

Finally, a UE u cannot connect to more than one BS at a time; thus,

$$\sum_{(i,j)\in\mathcal{L}_{\mathrm{AN}}} x^u_{(i,j)} = 1, \quad \forall u \in \mathcal{U}. \tag{13}$$

TABLE 1. Notation used in the MILP.

Input parameters	
$\mathcal{A}$	Set of aggregators
$\mathcal{B}$	Set of BSs
$\mathcal{U}$	Set of UEs
$\mathcal{L}_{\mathcal{AN}}$	Set of access links
$\mathcal{L}_{\mathcal{BH}}$	Set of BH links
BW_{PRB}	Bandwidth of a PRB
$BW_{(i,j)}$	Bandwidth allocated to BH link (i,j)
$\alpha_{(i,j)}$	Total losses minus the gains on BH link (i,j)
$c_{(i,j)}$	Number of PRBs needed for association on AN link (i,j)
c^{max}_i	Max. number of PRBs at site i
d^u	GBR demand for user u
Δ_p	Slope of the load-dependent power consumption
NTX	Number of transceiver chains
p_0	Minimum nonzero output power
p_{max}	Maximum transmission power
Decision variables	
p_i	Total power consumption at BS i
p_{out}	Output transmitted power
s^{AN}_i	Binary indicator of the activation of BS i
$s^{BH}_{(i,j)}$	Binary indicator of the activation of BH link (i,j)
$x^u_{(i,j)}$	Binary indicator of the use of link (i,j) by UE u
y	Binary variable for the switch ON/OFF constraint

The input parameters and the output variables of our MILP are summarized in Table 1.

IV. ROBUST APPROACH

The motivation behind this work is understanding the impacts of variable traffic demands on the distribution of resources and on the energy consumed in a 5G network. To this end, we will consider that our MILP in (1) is affected by uncertainty in the GBR demand parameter d^u. As the demanded GBR from user u travels in the mesh network through some BH links and is then delivered to user u through one of the available access links of a given BS i, the uncertainty will affect the following: 1) the PRB allocation in the access link (2) and thus the capacity constraint (1.c) and the output transmitted power of BS i (5); and 2) the bandwidth required on the BH links, which in turn directly affects the load (7) and, consequently, the output transmitted power on those BH links (6).

A robust optimization approach is then required in order to alleviate the side effects of data uncertainty. We apply the Γ-robust optimization approach presented in [13] to the

$$p^{BH}_{out_{(i,j)}} = \alpha_{(i,j)} \cdot \begin{cases} sl_1 \cdot load^{\mathrm{BH}}_{(i,j)}, & \text{if } load^{\mathrm{BH}}_{(i,j)} \leq bkp_1 \\ sl_1 \cdot bkp_1 + sl_2\,(load^{\mathrm{BH}}_{(i,j)} - bkp_1), & \text{if } bkp_1 \leq load^{\mathrm{BH}}_{(i,j)} \leq bkp_2 \\ sl_1 \cdot bkp_1 + sl_2\,(bkp_2 - bkp_1) + sl_3\,(load^{\mathrm{BH}}_{(i,j)} - bkp_2), & \text{if } bkp_2 \leq load^{\mathrm{BH}}_{(i,j)} \leq bkp_3 \\ \vdots & \\ sl_1\, bkp_1 + \sum_{n=2}^{nbkp} sl_n\,(bkp_n - bkp_{n-1}) + sl_{nbkp+1}\,(load^{\mathrm{BH}}_{(i,j)} - bkp_{nbkp}), & \text{if } load^{\mathrm{BH}}_{(i,j)} \geq bkp_{nbkp} \end{cases}$$
$$\forall (i,j) \in \mathcal{L}_{\mathrm{BH}} \tag{6}$$

formulation in (1), where d^u is defined as a random variable. As the GBR demand of a user u remains constant over the used links (i, j), in our formulation, we have

$$d^u = d^u_{(i,j)} \quad \forall (i,j) \in \mathcal{L}_{BH} \cup \mathcal{L}_{AN} \tag{14}$$

For the uncertain matrix $D = (d^u)$, which represents the amount of data demanded by UE u, we can assume that each coefficient d^u has a nominal value $\bar{d}^u$ and a potential symmetric maximum deviation $\hat{d}^u \geq 0$, and thus, it lies in the interval $[\bar{d}^u - \hat{d}^u, \bar{d}^u + \hat{d}^u]$. A crucial issue in robust optimization is how the robust uncertainty set is defined. According to our approach, the uncertainty budget will affect 1) the PRB allocation in the access link and 2) the bandwidth required on each BH link, which in turn impacts the power required at both the access and the backhaul networks.

A. UNCERTAINTY SET FOR THE ACCESS LINKS

We assume that at most Γ_i coefficients in row i are allowed to deviate from their nominal value, meaning that in the worst case only Γ_i users connected with BS i will demand at most $\bar{d}^u + \hat{d}^u$ resources, instead of the nominal $\bar{d}^u$ [13]. As Γ_i represents the *budget of uncertainty* for BS i, all the values for which the sum of the relative deviations from their nominal values is at most Γ_i represent the robust uncertainty set for our formulation. More formally, a scaled variation ϕ^u of parameter d^u from its nominal value is defined as

$$\begin{aligned} &|\phi^u| \leq 1, \quad \forall (i,j) \in \mathcal{L}_{AN} \\ &\phi^u = \frac{d^u - \bar{d}^u}{\hat{d}^u}, \\ &\sum_{(i,j)\in\mathcal{L}_{AN}} |\phi^u| \leq \Gamma_i, \quad \forall i \in \mathcal{B} \end{aligned} \tag{15}$$

B. UNCERTAINTY SET FOR THE BACKHAUL LINKS

As the traffic demanded by the users may also flow in the mesh network, we assume that at most $\Xi_{(i,j)}$ coefficients in each BH link (i, j) are allowed to deviate from their nominal value. We define the *budget of uncertainty* $\Xi_{(i,j)}$ for which

$$\begin{aligned} &\sum_{u\in\mathcal{U}} |\sigma^u_{(i,j)}| \leq \Xi_{(i,j)}, \\ &|\sigma^u| \leq 1, \quad \forall (i,j) \in \mathcal{L}_{BH} \\ &\sigma^u = \frac{d^u - \bar{d}^u}{\hat{d}^u}. \end{aligned} \tag{16}$$

meaning that in the worst case only $\Xi_{(i,j)}$ UEs whose GBR demand is traveling on BH link (i, j) will demand at most $\bar{d}^u + \hat{d}^u$ resources, instead of the nominal $\bar{d}^u$.

Note that since it is not known for which user the variability effectively occurs (i.e., which user effectively demands a different amount of data compared to the nominal value), the budget of uncertainty on the AN links and the budget of uncertainty on the BH links are independently defined. That is, we protect against a maximum deviation that may occur for any of the users associated with a given BS i, and at the same time, we protect against a maximum deviation that may occur for any of the users whose traffic is flowing on a given BH link (i, j). Under these conditions, the protection is guaranteed when taking into account, independently at each BS and on each BH link, those users whose demand variability has the worst effect on the total power consumption.

C. ROBUST FORMULATION

Once the uncertainty of the traffic demand parameter is set, it will directly affect equations (2) and (7), where now, d^u is a random variable instead of a deterministic value; consequently, equations (1c), (5) and (6) need to be updated.

According to the capacity constraint (1c), the number of PRBs used at each BS cannot exceed the maximum number $c_{i_{\max}}$ of PRBs that are available at BS i. In the robust formulation, this can be rewritten as

$$\begin{aligned} &\left(\sum_{u\in\mathcal{U}} \sum_{(i,j)\in\mathcal{L}_{AN}} x^u_{(i,j)} \left\lceil \frac{\bar{d}^u}{se_{(i,j)}} \right\rceil \right. \\ &\qquad \left. + \max_{\mathcal{U}'\subseteq\mathcal{U}, |\mathcal{U}'|\leq\Gamma_i} \sum_{u\in\mathcal{U}'} \sum_{(i,j)\in\mathcal{L}_{AN}} x^u_{(i,j)} \left\lceil \frac{\hat{d}^u}{se_{(i,j)}} \right\rceil \right) \\ &\leq c_{i_{\max}}, \quad \forall i \in \mathcal{B}. \end{aligned} \tag{17}$$

Moreover, the power constraint in the access links is also affected by the uncertainty of the traffic demand. Thus, (5) can be rewritten as

$$\begin{aligned} p^{\mathrm{AN}}_{out_i} = \frac{p^{\mathrm{AN}}_{\max_i}}{c_{i_{\max}}} &\left(\sum_{u\in\mathcal{U}} \sum_{(i,j)\in\mathcal{L}_{AN}} x^u_{(i,j)} \left\lceil \frac{\bar{d}^u}{se_{(i,j)}} \right\rceil \right. \\ &\left. + \max_{\mathcal{U}'\subseteq\mathcal{U}, |\mathcal{U}'|\leq\Gamma_i} \sum_{u\in\mathcal{U}'} \sum_{(i,j)\in\mathcal{L}_{AN}} x^u_{(i,j)} \left\lceil \frac{\hat{d}^u}{se_{(i,j)}} \right\rceil \right), \quad \forall i \in \mathcal{B}. \end{aligned} \tag{18}$$

Similarly, the load on each BH link (7) should be reformulated as

$$\begin{aligned} load^{\mathrm{BH}}_{(i,j)} = &\left(\sum_{u\in\mathcal{U}} x^u_{(i,j)} \bar{d}^u + \max_{\mathcal{U}'\subseteq\mathcal{U}, |\mathcal{U}'|\leq\Xi} \sum_{u\in\mathcal{U}'} x^u_{(i,j)} \hat{d}^u \right) \\ &\cdot \frac{1}{BW_{(i,j)}}, \quad \forall (i,j) \in \mathcal{L}_{BH}, \end{aligned} \tag{19}$$

which in turn affects (6) and (4).

The new robust problem formulated in this section is not linear any more due to the *max* functions in (17), (18), and (19). Section IV-D shows how the problem can be transformed into a MILP.

D. ROBUST COUNTERPART OF THE MILP

In this section, by exploiting LP duality [13], the *max* functions in (17), (18), and (19) are transformed into linear functions. To this end, two dual variables are needed for each robust uncertainty set (i.e., one on the access links $\mathcal{L}_{AN}$ and one on the backhaul links $\mathcal{L}_{BH}$). This section provides the full description of the robust counterpart of the model presented in [10] and summarized in Section III-B.

The robust counterpart of the MILP in (1) is given by

$$\begin{aligned} &\underset{x^u_{(i,j)}, s^{\mathrm{AN}}_i, s^{\mathrm{BH}}_{(i,j)}}{argmin} \sum_{i\in\mathcal{B}} P_i(x^u_{(i,j)}, s^{\mathrm{AN}}_i, s^{\mathrm{BH}}_{(i,j)}),\\ &\text{s.t. } a)\ x^u_{(i,j)} \in \{0,1\},\ \forall u \in \mathcal{U},\\ &\qquad \forall (i,j) \in \mathcal{L}_{\mathrm{BH}} \cup \mathcal{L}_{AN}\\ &\qquad b)\ s^{\mathrm{AN}}_i, s^{\mathrm{BH}}_{(i,j)} \in \{0,1\},\ \forall i \in \mathcal{B},\\ &\qquad \forall (i,j) \in \mathcal{L}_{\mathrm{BH}}\\ &\qquad c)\ \text{Robust capacity constraint}\\ &\qquad d)\ \text{Robust power constraints}\\ &\qquad e)\ \text{Switch ON/OFF constraints}\\ &\qquad f)\ \text{Path conservation constraints} \end{aligned} \tag{20}$$

Two dual variables μ_i and $\nu_{(i,j)}$ need to be defined for each access link (i,j) such that

$$\begin{aligned} \mu_i + \nu_{(i,j)} &\geq x^u_{(i,j)} \left\lceil \frac{\hat{d}^u}{se_{(i,j)}} \right\rceil, \quad \forall\ (i,j) \in \mathcal{L}_{AN},\\ \mu_i &\geq 0, \quad \nu_{(i,j)} \geq 0. \end{aligned} \tag{21}$$

Thus, the **robust capacity constraint** in (20c) is obtained by linearizing (17) as

$$\begin{aligned} \left(\sum_{u\in\mathcal{U}} \sum_{(i,j)\in\mathcal{L}_{AN}} \left(x^u_{(i,j)} \left\lceil \frac{\bar{d}^u}{se_{(i,j)}} \right\rceil + \nu_{(i,j)} \right) + \Gamma_i \mu_i \right)\\ \leq c_{i_{\max}} \quad \forall\ i \in \mathcal{B} \end{aligned} \tag{22}$$

Regarding the **robust power constraints** in (20d), the power P_i of BS i can be computed as the sum of the power consumed on the microwave links that are used by each UE u once it associates with BS i plus the power consumed by each BH link that is activated from BS i to the mesh, that is,

$$\begin{aligned} P_i = {} & NTX^{\mathrm{AN}}_i (s^{\mathrm{AN}}_i p^{\mathrm{AN}}_{0_i} + \Delta^{\mathrm{AN}}_p P^{\mathrm{AN}}_{out_i})\\ & + \sum_{(i,j)\in\mathcal{L}_{\mathrm{BH}}} NTX^{\mathrm{BH}}_{(i,j)} \left(s^{\mathrm{BH}}_{(i,j)} p^{BH}_{0_{(i,j)}} + \Delta^{\mathrm{BH}}_p P^{BH}_{out_{(i,j)}} \right),\\ & \forall i \in \mathcal{B} \end{aligned} \tag{23}$$

where $P^{\mathrm{AN}}_{out_i}$ is obtained by linearizing (18), that is,

$$\begin{aligned} P^{\mathrm{AN}}_{out_i} = \frac{p^{\mathrm{AN}}_{\max_i}}{c_{i_{\max}}} \Bigg(& \sum_{u\in\mathcal{U}} \sum_{(i,j)\in\mathcal{L}_{AN}} x^u_{(i,j)} \left\lceil \frac{\bar{d}^u}{se_{(i,j)}} \right\rceil\\ & + \Gamma_i \mu_i + \sum_{u\in\mathcal{U}} \sum_{(i,j)\in\mathcal{L}_{AN}} \nu_{(i,j)} \Bigg), \quad \forall i \in \mathcal{B}. \end{aligned} \tag{24}$$

Two other dual variables $\lambda_{(i,j)}$ and $\kappa^u_{(i,j)}$ need to be introduced on the backhaul links such that

$$\begin{aligned} \lambda_{(i,j)} + \kappa^u_{(i,j)} &\geq x^u_{(i,j)} \hat{d}^u, \quad \forall\ (i,j) \in \mathcal{L}_{BH}, \quad \forall\ u \in \mathcal{U}\\ \lambda_{(i,j)} &\geq 0, \quad \kappa^u_{(i,j)} \geq 0. \end{aligned} \tag{25}$$

Then, the parameter $load^{\mathrm{BH}}_{(i,j)}$ in (19) can be written as

$$L^{\mathrm{BH}}_{(i,j)} = \frac{\left(\sum_{u\in\mathcal{U}} x^u_{(i,j)} \bar{d}^u + \Xi_{(i,j)} \lambda_{(i,j)} + \sum_{u\in\mathcal{U}} \kappa^u_{(i,j)} \right)}{BW_{(i,j)}} \quad \forall (i,j) \in \mathcal{L}_{BH} \tag{26}$$

TABLE 2. Notation of the robust counterpart MILP.

Input parameters	
$\mathcal{A}$	Set of aggregators
$\mathcal{B}$	Set of BSs
$\mathcal{U}$	Set of UEs
$\mathcal{L}_{\mathcal{AN}}$	Set of access links
$\mathcal{L}_{\mathcal{BH}}$	Set of BH links
$BW_{(i,j)}$	Bandwidth allocated to BH link (i,j)
BW_{PRB}	Bandwidth of a PRB
$\alpha_{(i,j)}$	Total losses minus the gains on BH link (i,j)
c^{max}_i	Max. number of PRBs at site i
$\bar{d}^u$	Nominal value of the GBR demand for user u
$\hat{d}^u$	Maximum deviation of the GBR demand for user u
Δ_p	Slope of the load-dependent power consumption
NTX	Number of transceiver chains
Γ_i	Budget of uncertainty for BS i
$\Xi_{(i,j)}$	Budget of uncertainty for the set of $\mathcal{L}_{BH}$
p_0	Minimum nonzero output power
p_{max}	Maximum transmission power
Decision variables	
P_i	Total robust power consumption at BS i
P_{out}	Robust output transmitted power
s^{AN}_i	Binary indicator of the activation of BS i
$s^{BH}_{(i,j)}$	Binary indicator of the activation of BH link (i,j)
$x^u_{(i,j)}$	Binary indicator of the use of link (i,j) by UE u
y	Binary variable for the switch ON/OFF constraint
μ_i and $\nu_{(i,j)}$	Dual variables for each AN link (see (21))
$\lambda_{(i,j)}$ and $\kappa^u_{(i,j)}$	Dual variables for each BH link (see (25))

and $P^{BH}_{out_{(i,j)}}$ can be obtained by substituting $load^{\mathrm{BH}}_{(i,j)}$ with $L^{\mathrm{BH}}_{(i,j)}$ in (6).

Finally, the **switch ON/OFF constraints** (20e) and the **path conservation constraints** (20f) remain unchanged and are given by 9-10 and by 11-13, respectively.

The notation for the robust counterpart is summarized in Table 2. The robust MILP guarantees a feasible optimal solution for a given amount of allowed variability of the user demand d^u; the allowed variability is defined through the budget of uncertainty Γ_i of the access link of each BS i, the dual variables μ_i and $\nu_{(i,j)}$ defined in (21), the budget of uncertainty $\Xi_{(i,j)}$ for each BH link (i,j), and the dual variables $\lambda_{(i,j)}$ and $\kappa^u_{(i,j)}$ defined in (25). The objective is to minimize the total power P_i of all the active BSs in the network. The solution provides the set of access links through which the users associate with the network, the set of active BH links through which the users' traffic flows in the mesh, and the set of BSs and BH links that are turned off in order to save power while they are not in use.

V. EVALUATION

Our model has been tested using a scenario with 17 BSs, which are composed of 1 eNB and 16 small cells, as depicted in Fig. 2; the SCs are grouped in two clusters inside the coverage area of the eNB (500 m radius),[1] according to the specifications by 3GPP [23]. The 16 SCs are randomly

[1] We have considered the same scenario used in [10].

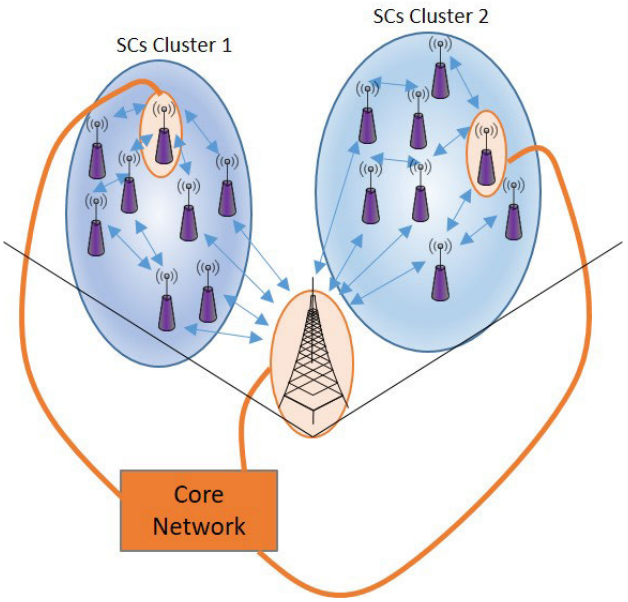

FIGURE 2. Simulation scenario with a single eNB sector and two clusters of SCs as in [10]. The eNB and one SC in each cluster act as the traffic aggregators.

TABLE 3. Values used for the evaluation [10], [23].

Parameter	Value
$BW_{(i,j)}$	200 MHz
BW_{PRB}	0.2 MHz
c_i^{max}	100
Δ_p^{BH}	10^5
NTX	8
p_0^{BH}	3.9 W
p_{max}^{BH}	0.0631 W

	eNB	SC
Δ_p^{AN}	4.7	4
p_0^{BH}	130 W	6.8 W
p_{max}^{BH}	39.8107 W	1 W

dropped in the eNB sector. We have considered 5 random drops (called random scenarios) in order to calculate the statistics. Only three BSs have a fixed connection with the core network (as depicted in Fig. 2). In each random scenario, two SCs are randomly selected to have this fixed/high capacity/zero-power link, together with the eNB. In each random scenario, the number of available mmWave links for the meshed backhaul network among the small cells and between an SC and the eNB may vary from 100 to 130. The BH links operate at 60 GHz with a channel BW of 200 MHz; the other parameters used in the evaluation are summarized in Table 3.

As in [10], eight different user densities (i.e., minimum of 13 UEs and maximum of 62 UEs) have been considered in the evaluation, which correspond to different times of the day (i.e., from midnight to 7 am [24]). The users are randomly dropped inside the coverage area of the eNB, and according

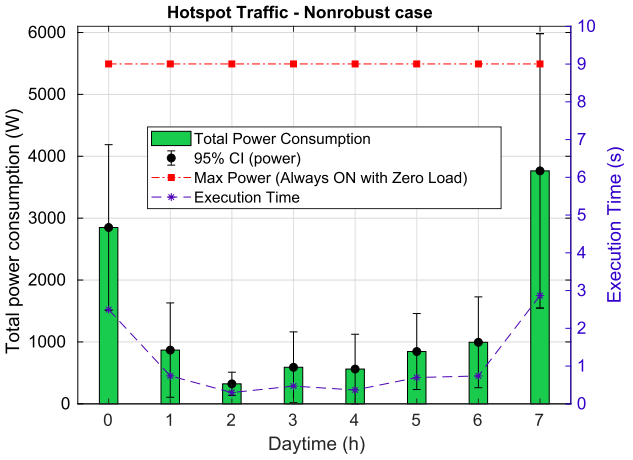

FIGURE 3. Average total power consumption (green bars) with the 95% CI (in black) and average execution time (blue asterisks) based on the time of day in the nonrobust case.

to their locations and the signal quality from the different BSs in the scenario, several microwave access links are eventually available for association. Each user has a predefined demand (i.e., bits per second that the UE requests from the core network), thus requiring a given number of PRBs. According to the LTE specification [25], we assume that the number of available PRBs at each BS is 100 when the subcarrier spacing is 15 kHz and the transmission bandwidth is 20 MHz.

Unless stated, the values presented in this section refer to the **average over the 5 random scenarios** at each time of the day. We will first present our results for the nominal case (i.e., when Γ and Ξ are set to zero), meaning that the demand fluctuations are not taken into account. Those results are taken as a reference for later comparison in Section V-B, where the impact of an increasing variation of the users' demand on the energy efficiency of the network is studied.

A. RESULTS IN THE NONROBUST CASE

When setting Γ and Ξ to zero, the robust problem will be equivalent to the original problem presented in [10]. That is, since the users' demand d^u is considered to be known and set to a fixed value (i.e., the nominal value $\bar{d}^u$ for each user u), the Robust MILP in (20) is then equivalent to the original MILP in (1).

Figure 3 presents the total power consumed in the network every hour from midnight to 7 am. The average is represented in green. It goes from 323 W when the lowest number of users are using the network (i.e., at 2 am) up to 3764 W for the scenario with the highest user density (i.e., at 7 am). The 95% confidence intervals (CIs) are shown in black and display a high variability among the five random scenarios. There may be an increase of up to 81% with respect to the mean (4 am), while it appears to more stable in higher density scenarios (up to 47% at 7 am). We can conclude that the higher the UE density is, the lower the CI (i.e., the SC distribution has an impact on the results).

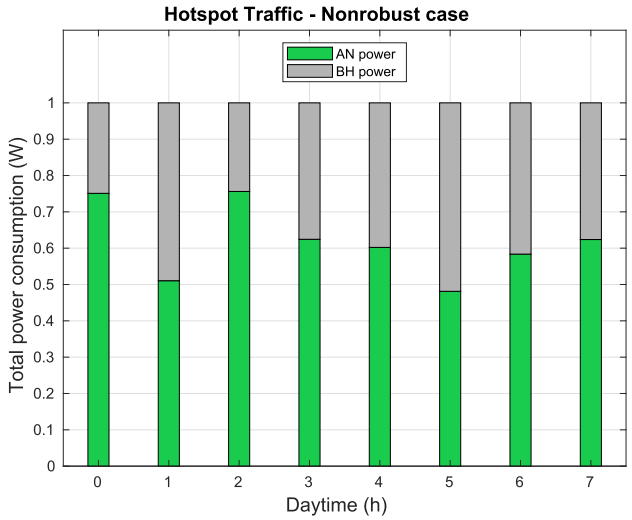

FIGURE 4. AN (green bars) versus BH (gray bars) power normalized to the total power consumption based on the time of day for the nonrobust case. The error bars for the access network are depicted in black.

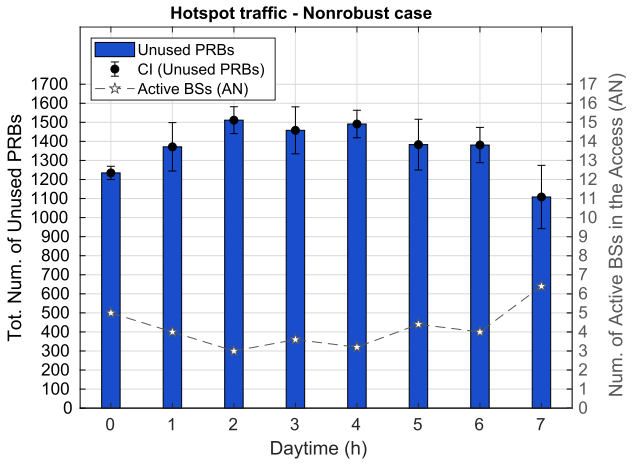

FIGURE 5. Average number of PRBs left unused in the network (blue bars) with the 95% CI (in black) and the average number of BSs that are switched ON (gray stars) in the nonrobust case.

The red line in Fig. 3 represents the reference power consumption when all the BSs in the network are switched on and the output transmitted power components P_{out}, which depend on the users density, are set to zero; we refer to this condition as "Zero Load". The red line thus represents the minimum power that would be consumed during the day in our scenario if the switching off strategy was not applied. Notice that the power consumption is actually higher once the output transmitted power components are also taken into account. As an example, the reference "Zero Load" power consumption is 5492.16 W, and it would increase to 6713.70 W at 7 am when considering the corresponding user demand and load at that hour of the day; that is, P_{out} is 1221.54 W at 7 am, and it varies at each time of the day depending on the user density. This also explains why the CI at 7 am intersects the reference line.

The blue asterisks in Fig. 3 represent the average execution time; again, it increases with the user density (i.e., from 0.3 seconds up to 2.9 seconds). The 95% CIs for the execution times for the five random scenarios are always lower than 40%. Thus, for the simulated scenarios, we can conclude that, in the nonrobust case, the optimal solution can be found in real time.

The total power consumption is the sum of the power consumed on the access links established from each BS to each UE in the network (i.e., AN power) plus the power consumed by each BH link that is activated from each BS to the mesh (i.e., BH power), as in (23). The AN and BH powers, which are normalized to the total power consumption, are represented in Fig. 4 with green and gray bars, respectively. In general, most of the power is consumed in the access, independent of the user density. In the two scenarios with the highest user densities, most of the power is consumed in the AN links (i.e., 75% at midnight and 62% at 7 am); the same trend exists in the three scenarios with lower densities (i.e., 76% at 2 am, 62% at 3 am and 60% at 4 am). The 95% CIs for the AN power normalized to the total power consumption stay from 0.2 to 0.4, showing relatively high variability in the 5 random scenarios that have been used in the evaluation. This result is expected since the SCs random drop has an impact on the number of PRBs that are required to satisfy the UE demand and, hence, the P_{out}. This high variability confirms the representativeness of the 5 random scenarios, which turn out to be quite heterogeneous.

When the number of BSs that are switched on in the AN links increases (represented by the gray stars in Fig. 5), the number of unused PRBs decreases (represented by the blue bars). The total number of available PRBs in our scenario is 1700 (i.e., 17 BSs, each one with 100 PRBs). Higher density scenarios (i.e., midnight and 7 am) need more BSs to serve the users and are thus switched on in the AN links (5.0 and 6.4, respectively), which also correspond to fewer PRBs left unused (1234.6 and 1108.2, respectively). It is worth noting that, according to the 95% CIs, at least 940 PRBs remain unused in all the evaluated scenarios, which is more than half of the available PRBs per BS.

The utilization of the BSs in the network is not uniform, as depicted in Fig. 6. The average number of PRBs that are used based on the time of day is shown in different subplots for different BSs. The eNB (Fig. 6(a)) can be switched off unless there is a high user density in the network (i.e., midnight and 7 am). Similarly, there are BSs that are barely used (e.g., SC8, SC11 and SC14) and that can be switched off most of the time (Fig. 6(b)), while the PRBs utilization is almost constant and low for others (e.g., from 8 to 20% for SC2, SC4, SC12 and SC15, as depicted in Fig. 6(c)). On the other hand, some BSs are constantly used (e.g., from 16 to 60% of PRBs utilization for SC3, SC5, SC9 and SC10), while others can be highly utilized at given hours or barely used and switched off during the day (Fig. 6(e)-(f)). It is worth noting that, on average, all the BSs need to be switched on in order to satisfy the user demand at 7 am.

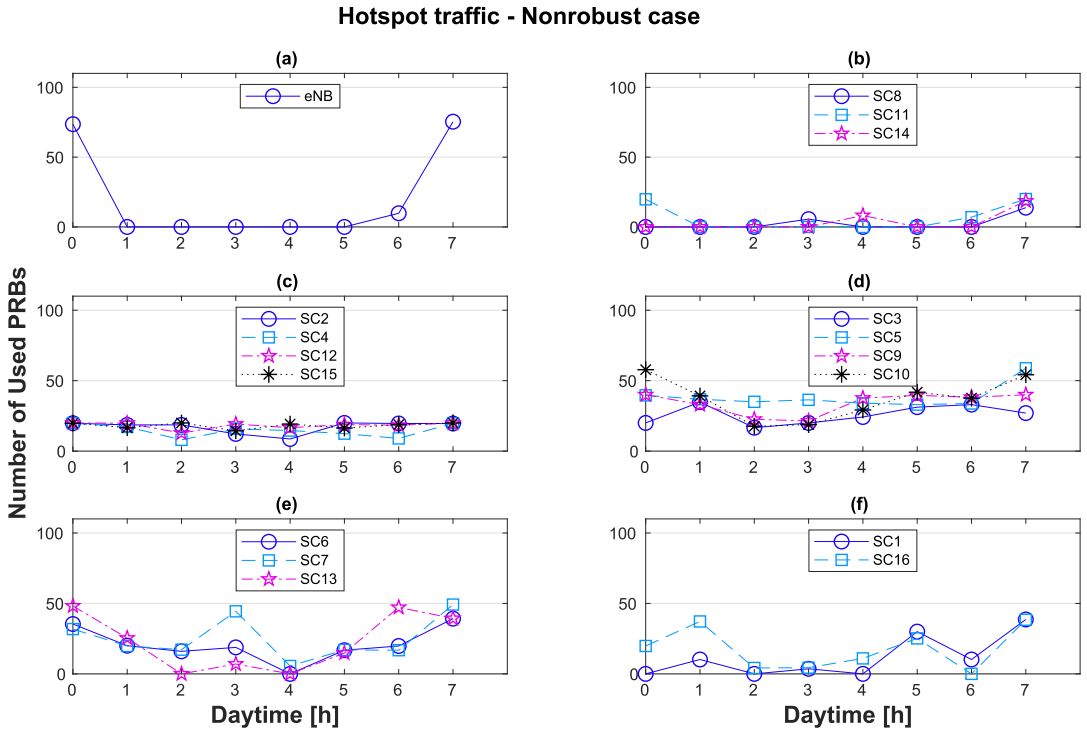

FIGURE 6. Average number of PRBs that are used at each BS based on the time of day in the nonrobust case.

B. IMPACT OF INCREASING THE ROBUSTNESS ON THE ENERGY EFFICIENCY OF THE NETWORK

This section is devoted to analyzing how the optimal solution changes when the user demand is assumed to be variable, according to the theory shown in Section IV. The user demand d^u that was used in the evaluation in Section V-A is now set as the nominal value of the GBR demand for each user u ($\bar{d}^u$). Several simulations have been run considering three different values for the maximum deviation $\hat{d}^u$ of the GBR demand of each user u (i.e., 0.1, 0.2 and 0.4) and considering two different values for the budgets of uncertainty (i.e., Γ_i and $\Xi_{(i,j)}$ are both set to 1 or both set to 5). As described in Sec. IV, the budget of uncertainty sets the maximum number of users for which the user demand is allowed to simultaneously deviate from the nominal value.

Thus, six robustness scenarios are depicted with increasing levels of variability: three where Γ_i and $\Xi_{(i,j)}$ are set to 1 and $\hat{d}^u$ is set to 0.1 (rob-scen-Γ1-dev0.1), 0.2 (rob-scen-Γ1-dev0.2) or 0.4 (rob-scen-Γ1-dev0.4); and three others where Γ_i and $\Xi_{(i,j)}$ are set to 5 and $\hat{d}^u$ is set to 0.1 (rob-scen-Γ5-dev0.1), 0.2 (rob-scen-Γ5-dev0.2) or 0.4 (rob-scen-Γ5-dev0.4). For instance, in the robustness scenario with Γ_i and $\Xi_{(i,j)}$ set to 1 and the maximum deviation $\hat{d}^u$ set to 0.2 (i.e., rob-scen-Γ1-dev0.2), only one user demand is allowed to deviate from the nominal value on each BS and on each BH link, with a maximum deviation of 20% from the nominal user demand. The six robustness scenarios are compared with respect to the nonrobust scenario summarized in Section V-A.

Regarding the total power consumption in the robustness scenarios, two values can be shown. First, we have the expected power consumption, which is the average power that will be needed since the deviations from the nominal users demand are symmetric. Second, the risk-adjusted power represents the maximum potential consumption when the total allowed increase in the user demand actually occurs (i.e., worst-case scenario). It is important to stress, however, that in both cases, an increase in the power consumption is expected when there is higher variability in the uncertainty set (i.e., in comparison with the results in Sec. V-A). The reason for this expectation is due to the spare resources that need to be left unallocated at each BS and link (even if they will not be used) in order to guarantee that they can be used in case that the maximum variability allowed in the uncertainty set actually occurs. Please recall here that Γ-Robustness theory is based on the assumption that it is highly unlikely that all the user demands vary at their maximum deviation for all the users at the same time. Thus, those spare resources may cause more BSs to be switched on in order to guarantee a feasible solution, even in the event of an increase in the demanded rate.

Fig. 7 depicts the total risk-adjusted power consumed in the network for each robustness scenario, which is normalized with respect to the power consumed in the nonrobust scenario (represented in black). Comparing the results with a given set of Γ_i and $\Xi_{(i,j)}$, we observe that the average normalized risk-adjusted power consumption increases when the maximum deviation grows; similarly, when we set a given maximum

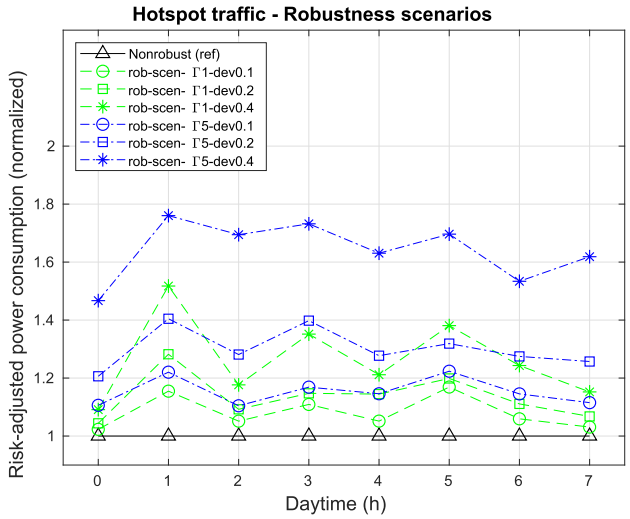

FIGURE 7. Average risk-adjusted power consumption for different Γ based on the time of day, which is normalized to the power in the nonrobust case.

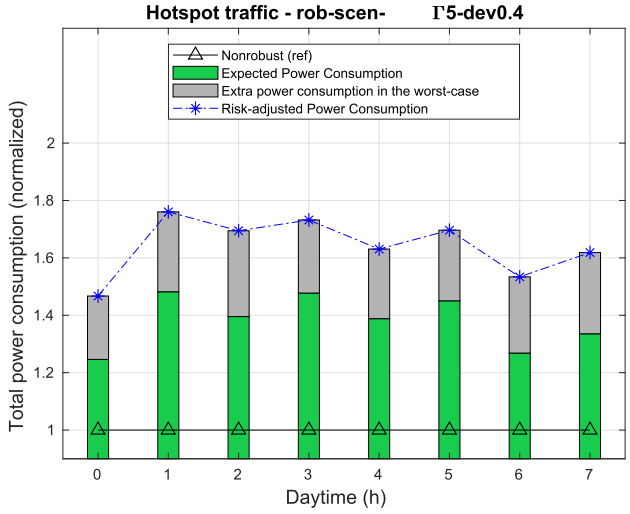

FIGURE 8. Average expected power consumption (green bars) plus the extra power needed in the worst-case scenario (gray bars, risk-adjusted power consumption), which is normalized to the nonrobust case for rob-scen-Γ5-dev0.4.

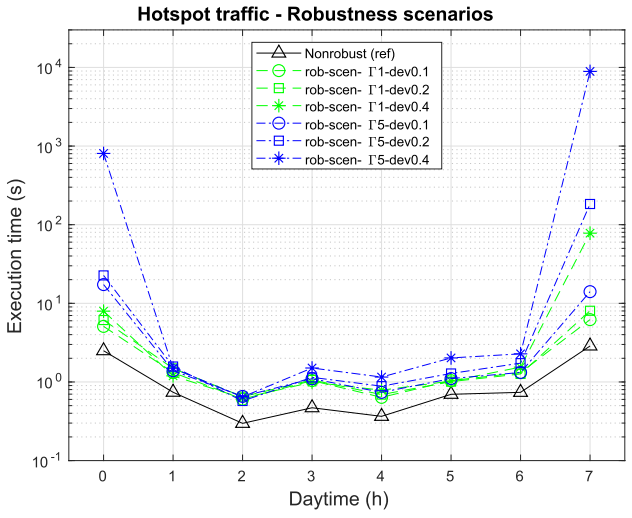

FIGURE 9. Average execution time for different Γ.

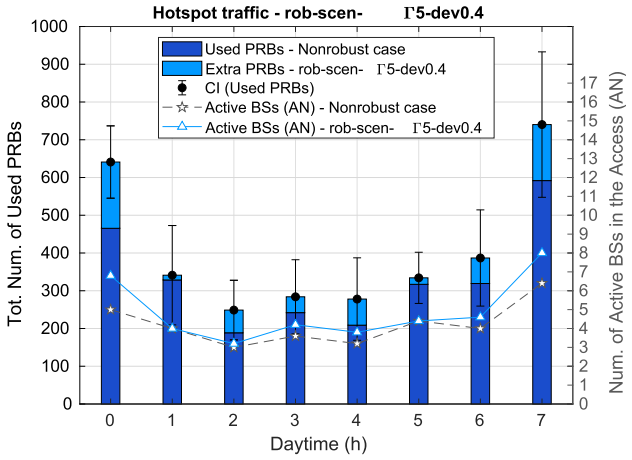

FIGURE 10. Average number of PRBs employed in the network versus the average number of BSs that are switched ON. The extra resources comparison of the nonrobust case and the rob-scen-Γ5-dev0.4 is depicted.

deviation, the average normalized risk-adjusted power consumption increases as the budget of uncertainty increases. In the most conservative of the six robustness scenarios (i.e., rob-scen-Γ5-dev0.4, which is represented with blue asterisks), the increase in the risk-adjusted power consumption remains from 45% to 75% (at midnight and 1 am, respectively).

However, it is worth noting that this increase only occurs in the worst case, while on average, less power is actually consumed in the robustness scenarios. For instance, the expected power consumption in the rob-scen-Γ5-dev0.4 is depicted in Fig. 8 and represented with green bars. On average, it increases from 25% to 48% compared to the nonrobust case (again represented with black triangles). The extra amount of power depicted in gray in Fig. 8 will be consumed in the unlikely event that all the allowed deviations occur simultaneously; also recall that the power consumption may be less than expected, as the deviations are symmetric [13]).

As a conclusion, a less risky solution (i.e., one that takes into account a higher variability in the user demand) will result in higher power consumption, which is needed in order to protect the feasibility of the solution (i.e., leaving extra resources free in case they are needed for the extra demand).

The execution time for each robustness scenario is depicted in Fig. 9 and increases as the budget of uncertainty increases. In general, the execution time is more than an order of magnitude longer at midnight and at 7 am (i.e., when the user density is higher). Especially critical is the most conservative of the six robustness scenarios (i.e., the one represented with blue asterisks), as the time to find the optimal solution may be up to two days. Of course, the proposed algorithm cannot

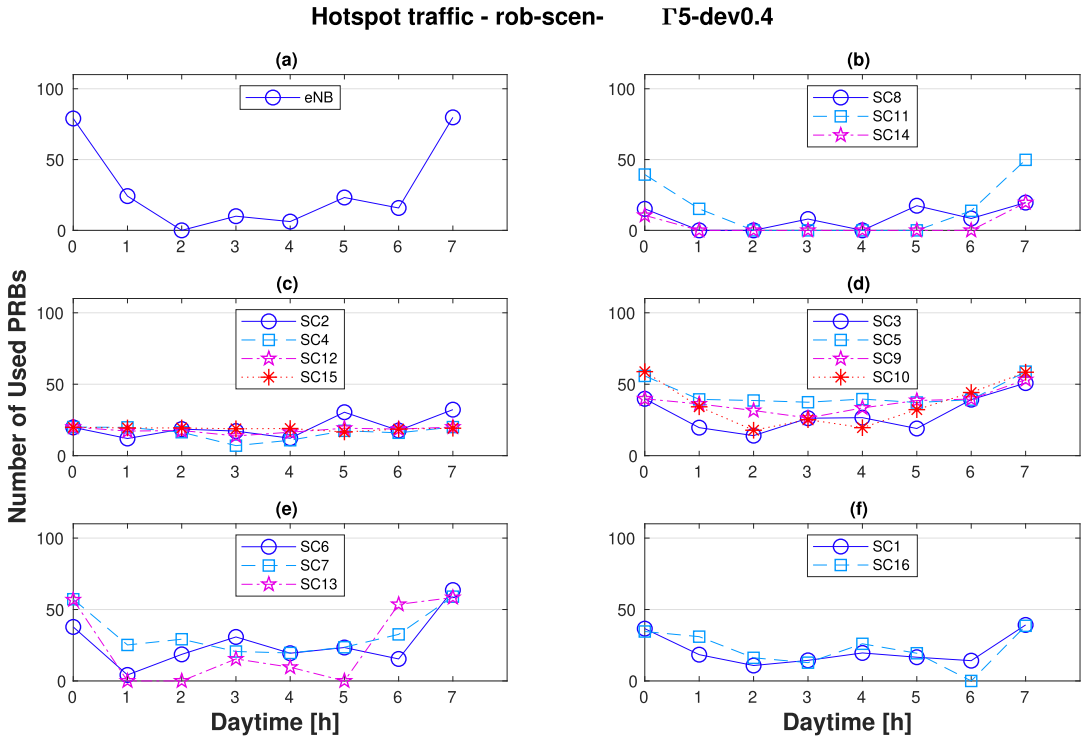

FIGURE 11. Average number of PRBs that are used at each BS based on the time of day for rob-scen-Γ5-dev0.4.

be implemented in real scenarios, but it can be used as a comparison threshold for any heuristic approach that will be developed in the future.

Again, when the number of BSs that are switched on in the AN links increases, more PRBs are utilized. As an example, Fig. 10 shows the results in the most conservative robustness scenario (i.e., with Γ and Ξ equal to 5 and dev $= 0.4$). The extra amounts needed in the robust case and the nonrobust case are depicted in light-blue and dark blue, respectively. Additionally, as a reference, the number of BSs that are switched on in the robust case is represented by light-blue triangles, and that for the nonrobust case is represented by gray stars. The two scenarios (nonrobust and robust case) present the same trend, where higher density scenarios require more PRBs and more BSs to serve the users. Additionally, more PRBs are used on average in the robustness scenario than the nonrobust scenario (i.e., from an increase of 12.6 PRBs on average at 1 am to 175.6 at midnight).

Again, the utilization of the BSs in the network is not uniform, as depicted in Fig. 11, which shows the number of PRBs that are used at each BS based on the time of day in the most conservative of the robustness scenarios (i.e., rob-scen-Γ5-dev0.4). As stated before, there are more PRBs allocated than in Fig. 6 as they may be used in case the deviation effectively occurs. For instance, the eNB (i.e., subplot (a)) can now be switched off only at 2 am when the user density is very low. Again, there are BSs that are barely used most of the time (e.g., SC8 and SC14) and that can be switched off (i.e., subplot (b)). There are others whose PRBs utilization is almost constant and low (e.g., from 7 to 20% for SC4, SC12 and SC15 and from 12 to 32% for SC2). It is worth noting that, on average, all the BSs need to be switched on at midnight and 7 am when the user demand is higher; however, even in this case, the power consumed by the whole network is reduced when the robust optimization approach is used.

VI. CONCLUSION

A robust MILP has been proposed for the joint optimization of the user association, backhaul routing and on/off strategies, with the objective of reducing the total power consumed in a 5G network while guaranteeing that the user needs are met. Because the user demand fluctuates over time and is mostly difficult to predict, the model takes into account this variability in order to protect the solution against a given number of simultaneously happening fluctuations. The variability not only affects the access links of the 5G network, thus requiring extra resources (i.e., PRBs) at the BS, but also impacts the BH link, where extra bandwidth should be allocated in case the demand increases with respect to the nominal value. The robust solution provides an association pattern and BH routing that may require extra resources to be left unallocated and available in the case of an increase in the user demand, thus turning on more network resources (e.g., more active BSs are expected). As expected, when accounting for higher deviations, the obtained solution is less risky and, thus, more energy consuming. However, since the Γ-robust approach from [13] considers symmetrical deviations from the nominal value, it has been shown that the

increase in the expected power consumption is smaller than in the risk-adjusted power consumption, which represents the worst-case scenario where all the deviations occur.

Regarding the execution time, a less risky solution requires more time for the algorithm to be solved. It has been shown that the proposed algorithm cannot be implemented in real scenarios since in some occasions, it may take more than two days to find the optimal solution. A hybrid solution is needed for realistic implementations, where a heuristic algorithm can provide a faster solution for those scenarios where the real optimum takes too long. The solution proposed in this paper should help to tune the parameters in the heuristic algorithm and should be used as a comparison threshold for such a heuristic hybrid approach.

ACKNOWLEDGMENT

The authors would like to thank Dr. Agapi Mesodiakaki for sharing with them the 5G scenarios used in the evaluation.

REFERENCES

[1] M. L. Attiah, A. A. M. Isa, Z. Zakaria, M. K. Abdulhameed, M. K. Mohsen, and I. Ali, "A survey of mmWave user association mechanisms and spectrum sharing approaches: An overview, open issues and challenges, future research trends," *Wireless Netw.*, vol. 26, no. 4, pp. 2487–2514, Mar. 2019, doi: 10.1007/s11276-019-01976-x.

[2] (Mar. 2019). *Cisco Visual Networking Index: Forecast and Trends, 2017–2022 White Paper*. [Online]. Available: https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html

[3] K. Sakaguchi *et al.*, "Where, when, and how mmWave is used in 5G and beyond," *IEICE Trans. Electron.*, vol. E100.C, no. 10, pp. 790–808, 2017.

[4] H. Y. Lateef, M. Dohler, A. Mohammed, M. M. Guizani, and C. F. Chiasserini, "Towards energy-aware 5G heterogeneous networks," in *Energy Management in Wireless Cellular and Ad-Hoc Networks*. Cham, Switzerland: Springer, 2016, pp. 31–44.

[5] M. Usama and M. Erol-Kantarci, "A survey on recent trends and open issues in energy efficiency of 5G," *Sensors*, vol. 19, no. 14, p. 3126, Jul. 2019.

[6] (Aug. 2019). *Millimeter-Wave Edge Cloud as an Enabler for 5G Ecosystem, 5G-MiEdge project, Final Annual Report*. [Online]. Available: https://5g-miedge.eu/deliverables/

[7] L. Wei, R. Hu, Y. Qian, and G. Wu, "Key elements to enable millimeter wave communications for 5G wireless systems," *IEEE Wireless Commun.*, vol. 21, no. 6, pp. 136–143, Dec. 2014.

[8] J. Wu, Y. Zhang, M. Zukerman, and E. K.-N. Yung, "Energy-efficient base-stations sleep-mode techniques in green cellular networks: A survey," *IEEE Commun. Surveys Tuts.*, vol. 17, no. 2, pp. 803–826, 2nd Quart., 2015.

[9] E. Zola, A. J. Kassler, and W. Kim, "Joint user association and energy aware routing for green small cell mmWave backhaul networks," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Mar. 2017, pp. 1–6.

[10] A. Mesodiakaki, E. Zola, R. Santos, and A. Kassler, "Optimal user association, backhaul routing and switching off in 5G heterogeneous networks with mesh millimeter wave backhaul links," *Ad Hoc Netw.*, vol. 78, pp. 99–114, Sep. 2018. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0360544218311319

[11] M. S. Dahal, J. N. Shrestha, and S. R. Shakya, "Energy saving technique and measurement in green wireless communication," *Energy*, vol. 159, pp. 21–31, Sep. 2018.

[12] G. Claßen, A. M. C. A. Koster, and A. Schmeink, "A robust optimisation model and cutting planes for the planning of energy-efficient wireless networks," *Comput. Oper. Res.*, vol. 40, no. 1, pp. 80–90, Jan. 2013.

[13] D. Bertsimas and M. Sim, "The price of robustness," *Oper. Res.*, vol. 52, no. 1, pp. 35–53, Feb. 2004.

[14] A. Mesodiakaki, F. Adelantado, L. Alonso, and C. Verikoukis, "Energy-efficient user association in cognitive heterogeneous networks," *IEEE Commun. Mag.*, vol. 52, no. 7, pp. 22–29, Jul. 2014.

[15] A. Mesodiakaki, E. Zola, and A. Kassler, "User association in 5G heterogeneous networks with mesh millimeter wave backhaul links," in *Proc. IEEE 18th Int. Symp. A World Wireless, Mobile Multimedia Netw. (WoWMoM)*, Jun. 2017, pp. 1–6.

[16] S. B. Aboagye, A. Ibrahim, and T. M. N. Ngatched, "Energy efficient power and flow control in millimeter wave backhaul heterogeneous networks," in *Proc. IEEE Global Commun. Conf. (GLOBECOM)*, Dec. 2018, pp. 1–7.

[17] J. Lorincz, A. Capone, and J. Wu, "Greener, energy-efficient and sustainable networks: State-of-the-art and new trends," *Sensors*, vol. 19, no. 22, p. 4864, Nov. 2019.

[18] E. Zola, P. Dely, A. J. Kassler, and F. Barcelo-Arroyo, "Robust association for multi-radio devices under coverage of multiple networks," in *Wired/Wireless Internet Communication*, V. Tsaoussidis, A. J. Kassler, Y. Koucheryavy, and A. Mellouk, Eds. Berlin, Germany: Springer, 2013, pp. 70–82.

[19] R. G. Garroppo, G. Nencioni, M. G. Scutellà, and L. Tavanti, "Robust optimisation of green wireless LANs under rate uncertainty and user mobility," *Electron. Notes Discrete Math.*, vol. 52, pp. 221–228, Jun. 2016.

[20] F. D'Andreagiovanni, R. G. Garroppo, and M. G. Scutellà, "Green design of wireless local area networks by multiband robust optimization," *Electron. Notes Discrete Math.*, vol. 64, pp. 225–234, Feb. 2018.

[21] *Study on Small Cell Enhancements for E-UTRA and E-UTRAN; Higher Layer Aspects, Version 12.0.0, Rel.12*, document 3GPP TR 36.842, Dec. 2013.

[22] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson, M. Imran, D. Sabella, M. Gonzalez, O. Blume, and A. Fehske, "How much energy is needed to run a wireless network?" *IEEE Wireless Commun.*, vol. 18, no. 5, pp. 40–49, Oct. 2011.

[23] *Small Cell Enhancements for E-UTRA & E-UTRAN-Physical Layer Aspects, Version 1.0.0, Rel. 12*, document 3GPP TR 36.872, Aug. 2013.

[24] G. K. Tran, H. Shimodaira, R. E. Rezagah, K. Sakaguchi, and K. Araki, "Practical evaluation of on-demand smallcell ON/OFF based on traffic model for 5G cellular networks," in *Proc. IEEE Wireless Commun. Netw. Conf. Workshops (WCNCW)*, Doha, Qatar, Apr. 2016, pp. 1–7.

[25] J. Zyren, "Overview of the 3 GPP long term evolution physical layer," Freescale Semicond. Inc., Austin, TX, USA, White Paper 3GPPEVOLUTIONWP, 2007. [Online]. Available: https://www.nxp.com/docs/en/white-paper/3GPPEVOLUTIONWP.pdf

ENRICA ZOLA received the double M.Sc. degree in telecommunications engineering from Politecnico di Torino, Italy, and the Universitat Politècnica de Catalunya (UPC-BarcelonaTECH), Spain, in 2003, and the Ph.D. degree from UPC, in 2011. From 2003 to 2006, she worked as a full-time Lecturer with UPC. Since May 2017, she has been serving as an Associate Professor with the Department of Network Engineering, UPC. She has been involved in several research projects on the performance modeling of wireless systems and networks. Her research interests include wireless networking, mobility management, radio resource management, performance optimization modeling, robust optimization techniques, and the design of 5G networks.

ISRAEL MARTIN-ESCALONA received the degree in telecommunications engineering from the Universitat Politècnica de Catalunya (UPC), in 2001, the Ph.D. degree from UPC, in 2010, and the degree in computer science from the Computer Science Faculty of Barcelona, UPC, in 2013. In 2003, he joined the School of Telecommunications Engineering, UPC, where he has been teaching networking, teletraffic, and simulation. He currently leads the Research Group on Cellular Communications Networks, UPC. His research interests include data analysis, on self-managed networks, and particularly on indoor location solutions with and without network infrastructure.