

RESEARCH

Open Access



User association strategies in HetNets leading to rate balancing under energy constraints

Javier Rubio, Antonio Pascual-Iserte , Jaume del Olmo*  and Josep Vidal 

Abstract

This paper deals with the development of several strategies for associating users to base stations (BSs) in heterogeneous networks. These strategies are able to balance the rate among users and BSs and increase the overall network utility. Constraints related to the energy availability at BSs are considered explicitly in the design, assuming that the BSs are equipped with batteries that are recharged through energy harvesting. We develop a general association strategy, and then we present several suboptimum but less complex solutions suitable for scenarios with high mobility or deployments of BSs with low computational capabilities. We also present an implementation that is to be executed in a distributed way among users and BSs without the need of having a central entity gathering all the information. The performance of the proposed strategies is evaluated through simulations in terms of rate balancing and the effect of the energy harvesting capabilities on the network throughput is shown. We also compare the proposed strategies with the traditional max-SINR user association approach.

Keywords: User association strategies, Heterogeneous networks, Energy harvesting, Rate balancing

1 Introduction

Traffic demands are currently increasing extremely fast [2, 3], which implies that technical solutions have to be provided to face this fact. Among these solutions, multi-tier Heterogeneous Networks (HetNets) [4] are being considered as a way to increase the system capacity and coverage. The main concept behind HetNets is the joint deployment of small base stations (BSs) of different sizes (e.g., picocells, femtocells, etc.) with already existing macro BSs. The main difference among these BSs relies on characteristics such as the transmit power, physical size, size of coverage areas, and cost, among others. HetNets enable a more flexible, easy to install, and economical deployment as new infrastructure can be deployed when and where necessary, thus, reducing the deployment cost.

In this paper, we are targeting scenarios where energy harvesting (based, for example, on solar panels) is the only

power source at the BSs. An example is a rural deployment in developing countries, where access to the electric power grid is too expensive or impossible. In such locations, some main challenges arise: cost of BSs, cost of solar panels and batteries, cost of terminals, and business models adapted to people with low incomes. The solution adopted in some real deployments consists of an access network based on 3G BSs empowered by solar panels of limited size in outdoor scenarios [5]. 3G is usually the only viable technology to be deployed thanks to the low costs of the equipment and the terminals.

In these scenarios where small cells and macro BSs have often overlapping coverage regions, users will have several possible BSs to connect with. This implies that user association strategies need to be developed. Classical solutions adopted by the 3GPP standard based on maximizing the received signal to interference plus noise ratio (SINR) may lead to situations where the macro BS will be heavily loaded whereas small cells will provide service only to a few users. This will imply that a load imbalance will happen, which in turn may produce a heavy reduction of the overall network throughput. This motivates the need for the development of dynamic user association strategies

*Correspondence: jaume.delolmo@tsc.upc.edu

Partial results of this paper have been presented previously by the same authors in conference IEEE GLOBECOM 2014 (©2014 IEEE: some materials are reprinted, with permission, from [1]).
 Department of Signal Theory and Communications, Universitat Politècnica de Catalunya (UPC), Jordi Girona 1-3, Building D5, 08034 Barcelona, Spain

that offload traffic from the macro BSs to the small cells in order to obtain a more fairly balanced rate among users and load among all BSs.

Another aspect to be taken into account is the fact that the BSs have batteries with limited capacity that can only be recharged through energy harvesting. This should be taken into account explicitly in the design of the user association strategies; otherwise, the batteries of the BSs providing larger data rates would get empty very fast.

Although the problem of user association under the objective of balancing the load in the network has been studied so far, most works have addressed only the case of networks composed of macro BSs (see examples [6–9]). The generalization to the case of HetNets has been considered only recently, such as in [10]. In this paper, some simple heuristic user association techniques for long-term evolution (LTE) setups are proposed taking into account the interference among cells. Another strategy for user association based on the dual coordinate descent method is derived in [11] in order to balance the network load by adopting a design criterion driven by the network utilization and proportional fairness. One of the drawbacks of this proposed solution is that the transmit powers of the BSs are considered fixed and not optimized. The optimization of the long-term data rate for user association is addressed in [12]. Unlike the previous papers, there are only a few works that present joint resource allocation and mechanisms for BS assignment. One of them is [13], where the authors consider the joint resource allocation and cell selection in code division multiple access (CDMA) networks. In this paper, a pricing-based distributed algorithm is developed taking into account the congestion of the BSs in addition to the channel states of the mobile terminals. The case of orthogonal frequency-division multiple access (OFDMA) networks is considered in [14], where the joint BS selection and the resource allocation is analyzed. The main drawback of the previous proposals is that the proposed user association techniques are run at the same time scale as the scheduling, whereas in practice this is not realistic since the mobility of the users is slower than the changes of the wireless channels.

Regarding user association strategies with BSs powered by energy harvesting sources, only a few works are available in the literature. In [15], authors propose a user association problem where the average traffic delay is modeled using queuing theory and optimized. In [16], authors propose a strategy that aims to maximize the number of accepted users and minimize the radio resource consumption where the available energy of BSs is dependent on the harvested energy in a certain period of time. A lexicographic minimization of on-grid energy consumption is proposed in [17], which involves the optimization in both the space and time dimensions, due to the temporal and spatial dynamics of mobile traffic and green

energy generation. Finally, the authors in [18] propose a user association based on the concept of topology potential which takes the traffic load of users and the available renewable energy of base stations into consideration for energy-load tradeoffs.

In this paper, we generalize our previous conference version [1] by, first of all, presenting a generalized formulation of the user association problem taking into account the temporal evolution of the scenario. Then, based on a previous strategy by ourselves (summarized and extended in Section 4.1.1) we derive several new user association algorithms that perform load balancing (based on user connections to BSs) among the different network tiers and that consider explicitly the battery status of the BSs as well as the energy that is being harvested. We also consider that the coverage areas depend explicitly on the available energies at the batteries. We propose a general centralized approach as well as several new suboptimum ones featuring reduced complexity solutions. We also design a new distributed strategy that is to be executed at each individual. As it will be discussed in the corresponding section, the decision to apply either the centralized or the distributed approach depends on several aspects, characteristics, and requirements of the concrete scenario at hand since each approach has its own pros and cons. Finally, we also study the asymptotic behavior of the evolution of the batteries.

The organization of the paper is as follows. Section 2 introduces the description of the system model and the assumptions considered. Then, in Section 3 we introduce the formulation of the most general user association scenario. Later in Section 4, we introduce some simplification and develop the proposed centralized and distributed user association techniques. Some results concerning the asymptotic behavior of the batteries are derived in Section 5. Simulation results are presented in Section 6, and finally, conclusions are drawn in Section 7.

2 System model

2.1 System description

Let us consider a Downlink (DL) cellular multi-tier Single-Input Single-Output (SISO) system composed of several BSs. Each of these BSs belongs to a particular BS class, each having different capabilities (transmission power, battery size, etc.), where each class is categorized as a tier. Each BS is indexed by i and the set of all BSs is denoted as \mathcal{B} . We consider that there are two types of users: users that demand a fixed service rate and users that request a flexible service rate. This could correspond to a cellular system comprised of voice users¹, that demand a fixed service rate, and data users, who are usually provided with a dynamic flexible rate depending on the system load and the channel conditions. Let us denote the set of voice users as \mathcal{U}_V and the set of data users as \mathcal{U}_D . $\mathcal{U}_V^{(i)}$ and $\mathcal{U}_D^{(i)}$

represent the set of voice and data users that are associated to BS i , respectively. The set that contains all users in the system is denoted as $\mathcal{U}_T = \mathcal{U}_V \cup \mathcal{U}_D$ and the set of all users connected to BS i as $\mathcal{U}_T^{(i)} = \mathcal{U}_V^{(i)} \cup \mathcal{U}_D^{(i)}$.

We assume a wide band code division multiple access (WCDMA) network [13, 19], which implies that users are multiplexed using codes. As we mentioned in the introduction, energy harvesting HetNets are usually applied in low cost rural scenarios, where network operators can only deploy 3G-based systems in order to have a successful business. However, the methodology presented in this paper could be easily extended to consider 4G systems by just considering the specific rate-power function and multiplexing users with carriers instead of codes. In any case, we assume that the network operator has already reserved a set of codes for the voice users and the remaining codes are to be allocated among the data users. Thus, the amount of available codes in each set is known and fixed at each BS, a strategy followed by network operators [20].

Each BS is solely powered with its own harvesting source and battery. The energy harvesting sources allow the BSs to collect energy from the environment and recharge the batteries. We consider that, at a given moment of time, only information of the past and current harvesting and battery dynamics of the BSs will be available to perform the user association, yielding to an *online approach*.

Let us denote the time instants (epochs) between association periods by index $\tau \in \Upsilon$ and the duration of these epochs as T_e . Let us also denote the scheduling periods by $t \in \mathcal{T}$ with a duration of T_f . In general, the user association procedure should be executed in a time scale longer than the scheduling periods. This makes the association problem very challenging since the decision has to be taken considering, amongst others, the current channel state information (CSI), but the channel may vary during the whole association period. As a consequence, the association decision should be implemented considering the expected throughput (over the channel realizations) within the duration of the association period.

2.2 Power consumption model and battery dynamics

The power consumption at any BS is modeled as the addition of the radiated power, which is divided into the power devoted to the common DL channels, P_{CPICH}^i , the power consumed by the user-specific physical channels, $P_{\text{BS}}^i(t)$, and a fixed power consumed by the electronics and cooling, P_c^i . Let $P_{\text{RAD}}^i(t) = P_{\text{BS}}^i(t) + P_{\text{CPICH}}^i$ be the overall radiated power by the i -th BS.

We consider that the amount of power that can be used for traffic services is limited in each BS and is denoted by $P_{\text{BS}_i}^{\text{max}}$, so $P_{\text{BS}}^i(t) \leq P_{\text{BS}_i}^{\text{max}} \forall i$. Let $C_i(t)$ be the energy stored at the battery of the i -th BS at the beginning of the

t -th scheduling period and let $E_i(t)$ be the overall energy consumed during this scheduling period:

$$E_i(t) = T_f \cdot (P_{\text{CPICH}}^i + P_{\text{BS}}^i(t) + P_c^i). \quad (1)$$

Then, at period $t + 1$, the battery level is updated in general as [21]

$$C_i(t+1) = (C_i(t) - E_i(t) + H_i(t))_0^{C_{\text{max}}^i}, \quad t \in \mathcal{T}, \quad (2)$$

where $H_i(t)$ is the energy harvested in Joules during the whole scheduling period t , $(x)_a^b$ is the projection of x onto the interval $[a, b]$, i.e., $(x)_a^b = \min\{\max\{a, x\}, b\}$, which accounts for possible battery overflows and assures that the battery levels are non-negative, and C_{max}^i is the battery capacity. Notice that the whole harvest collected during period t is assumed to be available in the battery at the end of the epoch for simplicity.

2.3 Energy harvesting model

In this paper we assume a discretized model for the energy arrivals as also did in [21–23]. In the formulation and algorithms proposed for the user association in Sections 3 and 4, we assume no concrete model for the energy arrivals. However, in Section 5, devoted to the analysis of the asymptotic behavior of the batteries and in the simulations (Section 6), we will assume that $H_i(t)$ follows an ergodic Bernoulli process. As a result, only two values of harvested energy are possible, i.e., $H_i(t) \in \{0, e_i\}$, where e_i is the amount of Joules contained in an energy packet. The probability of receiving an energy harvesting packet during one epoch depends on the actual harvesting intensity (in the case of solar energy, it depends on the particular hour of the day) and is denoted by $p_i(t)$. Note that a higher value of $p_i(t)$ will be obtained in epochs where the harvesting intensity is higher, e.g., during the day, and a lower value of $p_i(t)$ during the night or during cloudy days. Note that, in some cases, it is possible to predict or estimate partially the expected energy available to be harvested. This situation can be included in our formulation by adjusting our model parameters by knowing that $\mathbb{E}[H_i(t)] = p_i(t) \cdot e_i$. As a result, the majority of harvesting sources can be modeled with the previous stochastic model. Nevertheless, that analysis and also the simulations results could be extended to other statistics for the energy arrivals.

2.4 System assumptions

Let us define a set of indicator variables $x_{ji}(\tau) \in \{0, 1\}$ to denote whether a given user j is associated to a particular BS i during the τ -th epoch as follows

$$x_{ji}(\tau) = \begin{cases} 1, & \text{if user } j \in \mathcal{U}_T^{(i)}, \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

Let $\mathbf{x}(\tau) \triangleq \{x_{ji}(\tau), j \in \mathcal{U}_T, i \in \mathcal{B}\}$ be the set containing the indicator variables. We consider that a user can only be

connected to one BS at a given epoch. This is formulated through the following unique association constraint:

$$\sum_{i \in \mathcal{B}} x_{ji}(\tau) = 1, \quad \forall j \in \mathcal{U}_T, \forall \tau \in \Upsilon. \quad (4)$$

In order to be able to connect to a given BS, users need to receive its pilot signals above a minimum SINR threshold to estimate the channel. Thus, a particular user j will only be able to connect to the following set of BSs:

$$\mathcal{S}_j(t) = \left\{ i \in \mathcal{B} \mid \frac{P_{\text{CPICH}}^i}{P_{\text{BS}}^i(t) + A_{ji}(t)} \geq \gamma_{\text{CPICH}} \right\} \subseteq \mathcal{B}, \quad (5)$$

where $A_{ji}(t) = \frac{\sigma_j^2 + I_{ji}(t)}{\tilde{h}_{ji}(t)}$, being σ_j^2 the noise power at j -th receiver, $I_{ji}(t)$ the inter-cell interference, i.e., $I_{ji}(t) = \sum_{m \in \mathcal{B}, m \neq i} \tilde{h}_{jm}(t) P_{\text{RAD}}^m(t)$, and $\tilde{h}_{ji}(t)$ the instantaneous channel between BS i and user j . Given that, from (5), we obtain the following association constraints:

$$x_{ji}(\tau) = 0, \quad \forall i \notin \mathcal{S}_j(t), \quad (6)$$

for all scheduling periods t corresponding to the same epoch τ . The traffic power, $P_{\text{BS}}^i(t)$ can be split into power for voice and data connections as $P_{\text{BS}}^i(t) = \sum_{j \in \mathcal{U}_V^{(i)}} \tilde{p}_{ji}(t) + \sum_{k \in \mathcal{U}_D^{(i)}} \tilde{p}_{ki}(t)$, where $\tilde{p}_{ji}(t)$ and $\tilde{p}_{ki}(t)$ are the instantaneous powers corresponding to the transmission toward the j -th and k -th voice and data user connected to the i -th BS, respectively. The set of voice users request a fixed rate service and we assume that just one code is assigned to each of them. This is translated into a minimum SINR requirement (γ_j) as follows:

$$\sum_{i \in \mathcal{B}} \frac{x_{ji}(\tau) \tilde{p}_{ji}(t) M_V}{\theta \left(P_{\text{RAD}}^i(t) - \tilde{p}_{ji}(t) \right) + A_{ji}(t)} \geq \gamma_j, \quad \forall j \in \mathcal{U}_V, \quad (7)$$

where M_V is the spreading factor for voice codes and θ is the orthogonality factor [19]. The set of data users request a flexible rate service. As it was commented before, the user association procedure should be executed in a longer time scale compared to the coherence time of the channel. For this reason, the user association strategy should not take decisions based on the instantaneous channels, as they will vary over the entire epoch. As a consequence, we propose to take decisions based on the expected value of the throughputs (through the application of the rate function based on expected channels as will be shown next). Hence, the expected throughput (expressed in b/s/Hz) achieved during one particular scheduling period by the i -th user connected to the j -th BS is

$$R_{ji}(t) = \mathbb{E}_{\tilde{\mathbf{h}}} \left[\mathbb{E}_{\tilde{\mathbf{p}}, \tilde{\mathbf{n}} | \tilde{\mathbf{h}}} \left[\tilde{r}_{ji}(t) \log_2 \left(1 + \frac{M_D \tilde{p}_{ji}(t)}{\tilde{n}_{ji}(t) \left(\theta \left(P_{\text{RAD}}^i(t) - \tilde{p}_{ji}(t) \right) + A_{ji}(t) \right)} \right) \right] \right], \quad (8)$$

where M_D is the spreading factor for data codes, $\tilde{\mathbf{p}} \triangleq \{\tilde{p}_{ji} \mid j \in \mathcal{U}_D, i \in \mathcal{B}\}$ and $\tilde{\mathbf{n}} \triangleq \{\tilde{n}_{ji} \mid j \in \mathcal{U}_D, i \in \mathcal{B}\}$ are the power and code allocation variables (random variables), and where we have assumed that the same power is allocated to all the codes of the same user. The code allocation variable $\tilde{n}_{ji}(t)$ refers to the number of codes that the j -th BS assigns to the i -th user. Note that the instantaneous allocated powers and codes depend upon the concrete scheduling policy and, generally, they do not only depend on the instantaneous channels, but also on other factors. That is why we model $\tilde{\mathbf{p}}$ and $\tilde{\mathbf{n}}$ as random variables, and average w.r.t. them have to be applied in (8).

We use the power and code variables to measure the load that a given BS is experiencing so that a well load-balanced network can be obtained by a proper user association strategy.

3 General user association formulation

In systems where variables with temporal evolution affect the long-term system performance (such as, for example, the energy available in the battery that constrains the amount of power that can be allocated among the users), it is important to take decisions considering not only the current impact on the overall system performance, but also taking into account the future impact. It is evident that if we spend much power in the current epoch and the conditions of the wireless channels are poor, we may be missing the opportunity to use better this energy in the near future when the channel conditions improve. In this sense, the user association procedure should be coupled in time, and associations that are being carried out at present time should consider the past knowledge as well as the future impact. For example, if a given BS is harvesting a lot of energy, this BS should accept more users as more energy will be available in the future. If, on the other hand, the energy is decaying fast, then an intelligent strategy would be to allocate such users to the neighboring BSs gracefully (an aggressive strategy would not be smart either since if there was a massive reallocation of users, the neighboring BSs would need a lot of energy to satisfy the users demands).

Another point to consider is the fact that the association strategy should run in real time. This implies that the association strategy must be implemented considering only current information of the channels, the batteries, and the harvesting being collected. Contrary, offline solutions that require future knowledge, i.e., non-causal knowledge of the channel, the batteries, the harvesting, etc., have been proposed extensively in situations where network nodes are provided with energy harvesting sources [22, 24]. However, the assumption of knowing when the energy will be available in future instant times is usually hard to accept. Therefore, in this paper we focus only on

online approaches that are more suitable for a realistic implementation.

A formal way to model the dynamic behavior and consider only causal information consists in the use of Dynamic Programming (DP) techniques [25]. DP has been shown to be a good mathematical tool to solve time-coupled problems when only past and current information is available, i.e., an online approach, and we desire to have a control over how the resources used in the current time instant can affect the future performance. Unfortunately, in such DP problems, the optimum value of the variables are functionals instead of scalars as in classical optimization. In addition, the computation of such functionals is usually extremely difficult and computationally challenging. In any case, let us present now the most general user association strategy and then propose some simplifications that will make the problem tractable and solvable with a reasonable computational complexity. The system performance under consideration is the sum-utility of the overall expected throughput of all data users in the system. For the sake of generality, we consider a general utility function denoted by $U_j(\cdot)$ for the j -th user. Let us introduce the following definitions: $\tilde{\mathbf{r}} = \{\tilde{R}_j, j \in \mathcal{U}_D\}$, $\tilde{\mathbf{p}}(t) = \{\tilde{p}_{ji}(t), j \in \mathcal{U}_D, i \in \mathcal{B}\}$, $\tilde{\mathbf{p}}(t) = \{\tilde{p}_{ji}(t), j \in \mathcal{U}_V, i \in \mathcal{B}\}$, $\mathbf{n}(t) = \{n_{ji}(t), j \in \mathcal{U}_D, i \in \mathcal{B}\}$, $\mathbf{x}(\tau) = \{x_{ji}(\tau), j \in \mathcal{U}_T, i \in \mathcal{B}\}$, and $\mathbf{p}_{\text{RAD}}(t) \triangleq \{P_{\text{RAD}}^i(t), i \in \mathcal{B}\}$. The association procedure is formulated through the following optimization problem:

$$\begin{aligned}
 & \underset{\substack{\tilde{\mathbf{r}}, \tilde{\mathbf{p}}(t), \tilde{\mathbf{p}}(t), \\ \mathbf{n}(t), \mathbf{x}(\tau), \\ \mathbf{p}_{\text{RAD}}(t)}}}{\text{maximize}} & \sum_{j \in \mathcal{U}_D} U_j(\tilde{R}_j) \\
 & \text{subject to} & \\
 & \text{C1: } \tilde{R}_j \leq \frac{1}{|\mathcal{T}|} \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{B}} R_{ji}(t), \forall j & \\
 & \text{C2: } \sum_{i \in \mathcal{B}} \frac{M_V \tilde{p}_{ji}(t)}{\theta \left(P_{\text{RAD}}^i(t) - \tilde{p}_{ji}(t) \right) + A_{ji}(t)} \geq \gamma_i, \forall j \in \mathcal{U}_V, \forall t & \\
 & \text{C3: } \sum_{i \in \mathcal{B}} x_{ji}(\tau) = 1, \quad \forall j \in \mathcal{U}_T, \forall \tau \in \Upsilon & \\
 & \text{C4: } \sum_{j \in \mathcal{U}_D} \tilde{n}_{ji}(t) \leq n_D^{(i)}, \quad \forall i \in \mathcal{B}, \forall t & \\
 & \text{C5: } T_f \left(\sum_{j \in \mathcal{U}_D} \tilde{p}_{ji}(t) + \sum_{j \in \mathcal{U}_V} \tilde{p}_{ji}(t) + P_{\text{CPICH}}^i + P_c^i \right) \leq C_i(t) & \\
 & \text{C6: } C_i(t+1) = (C_i(t) - E_i(t) + H_i(t))_0^{C_i^{\max}}, \forall i, \forall t & \\
 & \text{C7: } P_{\text{RAD}}^i(t) = \sum_{j \in \mathcal{U}_D} \tilde{p}_{ji}(t) + \sum_{j \in \mathcal{U}_V} \tilde{p}_{ji}(t) + P_{\text{CPICH}}^i, \quad \forall i, \forall t & \\
 & \text{C8: } 0 \leq \tilde{n}_{ji}(t) \leq x_{ji}(\tau) n_D^{(i)}, \quad \forall i \in \mathcal{B}, \forall j \in \mathcal{U}_D, \forall t \in \mathcal{T} & \\
 & \text{C9: } 0 \leq T_f \tilde{p}_{ji}(t) \leq x_{ji}(\tau) C_i(t), \quad \forall i \in \mathcal{B}, \forall j \in \mathcal{U}_V, \forall t \in \mathcal{T} & \\
 & \text{C10: } x_{ji}(\tau) = 0, \quad \forall i \notin S_j(t), \forall j \in \mathcal{U}_T, \forall \tau \in \Upsilon & \\
 & \text{C11: } x_{ji}(\tau) \in \{0, 1\}, \quad \forall i \in \mathcal{B}, \forall j \in \mathcal{U}_T, \forall \tau \in \Upsilon & \\
 & \text{C12: } \tilde{p}_{ji}(t) \geq 0, \quad \forall i \in \mathcal{B}, \forall j \in \mathcal{U}_D, \forall t \in \mathcal{T}. &
 \end{aligned} \tag{9}$$

where $|\mathcal{T}|$ is the cardinality of set \mathcal{T} and $n_D^{(i)}$ is the number of available codes for data users at BS i . Notice that we have considered a general per-user utility function $U_j(\cdot)$ as long as it is an increasing, strictly concave, and continuously differentiable function [26]. The problem presented in (9) is extremely difficult to solve and very challenging for the following reasons: *i*) the expected rate expression that appears in constraint C1 is very difficult to handle; *ii*) it is time-coupled through constraints C1 and C6; *iii*) the interference term in the rate expression makes the whole expression not concave; *iv*) the coverage areas depend on the radiated powers and, hence, they should also be optimized jointly; and *v*) the association variables $x_{ji}(\tau)$ are integer variables, making the whole problem a combinatorial optimization problem. In this regard, all this complicating issues should be handled before presenting the association strategy. Since the user association algorithm must be executed regularly in real time, the solution of the previous problem based on DP will not be considered as the computational complexity is prohibitively high. Instead, in this paper, we develop simpler and less complex association strategies that could be closer to a real implementation. The basic idea of the proposed strategy is to decouple the problem in time and solve it for each time instant independently as will be shown later.

4 Proposed greedy-based user association strategies

In the previous section, we defined the most general user association strategy. However, the resolution of the problem was impractical. In the section, we are going to propose a simplified version of the problem 9 in which we simplify the time-coupling issues and solve the association at each time instant independently. Note that, even though we do not provide a method for solving problem 9, it is crucial to have it defined if one wants to understand the simplifications that we provide along this section. Having said that, the simplifications of the challenging issues presented in the previous section that we are considering now are the following:

- i*) The expected rate expression given in (8) is very difficult to compute as it depends on the probability density function (PDF) of the powers, the codes, and the channels. Instead, an approximation of the average rate is considered in this paper following the same criteria used in papers like [12, 27, 28]. In this approximation, we consider average channels instead of instantaneous ones so that the channels now change from epoch to epoch, i.e., $h_{ji}(\tau) = \mathbb{E}_{\mathbf{h}}[\tilde{h}_{ji}(t)]$, which includes shadowing, path loss, and antenna

gain. Now, all the variables change at the time scale given by τ and the expectations w.r.t. the powers and the codes are also removed since these new variables, $p_{ji}(\tau)$, $\check{p}_{ji}(\tau)$, and $n_{ji}(\tau)$, can be interpreted as average allocation variables, i.e., estimates of the average values of the instantaneous powers and codes allocated by the scheduler during one epoch. By carrying out this can be approximated as (again in b/s/Hz)

$$\begin{aligned} \bar{R}_{ji}(\tau) &= n_{ji}(\tau) \log_2 \left(1 + \frac{M_D p_{ji}(\tau)}{n_{ji}(\tau) (\theta (P_{\text{RAD}}^i(\tau) - p_{ji}(\tau)) + A_{ji}(\tau))} \right). \end{aligned} \quad (10)$$

Since all variables now depend on τ , there are no more short-term variables dependent on t and, hence, there is no need to compute the user scheduling together with user association as the output of the optimization problem.

Note that all other variables depending on the powers, such as $A_{ji}(\tau)$, will be considered dependent on the average channels $h_{ji}(\tau)$. However, the battery evolution will be updated with the “true” radiated powers that are the outcome of the scheduler running at time scale t . In any case, we will use the notation dependent on τ , which means that $C_i(\tau)$ is the battery level at the beginning of epoch τ once all powers allocated by the scheduler have been incorporated. Additionally, in order to guarantee that constraints in (5) are still fulfilled (with a certain outage probability since they cannot be strictly fulfilled for all possible channel realizations), we could increase the threshold γ_{CPICH} to make the constraints more restrictive.

ii) The most important simplification is described next. We are transforming the DP problem (9) into a standard optimization problem. Regarding constraint C1, we will not consider the summation term over all time instants. Then, instead of having constraint C6 that couples the battery evolution and, hence, couples all allocation variables, our online user association strategy is executed at each epoch independently by allowing each BS to spend just a given fraction of the energy available at the battery during that particular epoch. The amount of energy to be spent at each epoch can be further optimized offline (see [29]). This can be seen as a greedy policy, where at each epoch we execute the association strategy with just past and current information of the battery levels, channel states, etc.

In general, the total energy consumed by a given BS during one epoch, i.e., $E_i(\tau)$, is limited by a function of the current battery level as

$$T_e \cdot (P_{\text{CPICH}}^i + P_{\text{BS}}^i(\tau) + P_c^i) \leq g_i(C_i(\tau)), \quad \forall i, \tau, \quad (11)$$

where the function $g_i(\cdot)$ is defined as $g_i(C_i(\tau)) \triangleq \min \left\{ T_e \cdot (P_{\text{CPICH}}^i + P_{\text{BS}_i}^{\text{max}} + P_c^i), \alpha_i C_i(\tau) \right\}$, where $0 \leq \alpha_i \leq 1$. Note that α_i controls the fraction of energy that is allowed to be used in a given epoch.

iii) For simplicity in the notation, in the expression in (10), we will approximate $\theta (P_{\text{RAD}}^i(\tau) - p_{ji}(\tau)) + A_{ji}(\tau)$ by $\theta P_{\text{RAD}}^i(\tau) + A_{ji}(\tau)$ throughout the paper². Notice that now $\bar{R}_{ji}(\tau)$ is jointly concave in $n_{ji}(\tau)$ and $p_{ji}(\tau)$ and $\bar{R}_{ji}(\tau) = 0$ if $n_{ji}(\tau) = 0$, for any $p_{ji}(\tau) \geq 0$.

iv) Note that the power variables affect both the coverage (see (5)) and the data rates. However, it is extremely difficult to deal with both effects jointly. For this reason, we assume a worst-case approach when defining the coverage areas assuming that the interference is maximum³. The worst-case coverage sets are, therefore, defined as

$$\mathcal{S}_j(\tau) = \left\{ i \in \mathcal{B} \mid \frac{P_{\text{CPICH}}^i}{\bar{P}_{\text{BS}}^i(C_i(\tau)) + A_{ji}(\tau)} \geq \gamma_{\text{CPICH}} \right\}, \text{ where}$$

$$\bar{A}_{ji}(\tau) = \frac{\sigma_j^2 + \sum_{m \in \mathcal{B}, m \neq i} h_{jm}(\tau) (\bar{P}_{\text{BS}}^m(C_m(\tau)) + P_{\text{CPICH}}^m)}{h_{ji}(\tau)}.$$

With this, the coverage areas only depend on the current battery and channel state.

v) Due to the integrality of the variables $x_{ji}(\tau)$, it is a combinatorial problem, whose computational burden grows as $\mathcal{O}(|\mathcal{B}|^{|\mathcal{U}_T|})$. A typical approach to follow in this situation is to relax the association variables so that $x_{ji}(\tau) \in [0, 1]$, $\forall i, j$ [12]. Notice that this relaxation will result in another optimization problem with a different optimum objective value as the constraint set has been expanded. If the optimum solution of the relaxed problem is integer, this corresponds to the optimum solution of the original problem. Otherwise, we can, for example, round the relaxed association variables so that $x_{ji}(\tau) \in \{0, 1\}$.

Now, we present the different user association techniques developed assuming the previous simplifications.

4.1 General epoch-by-epoch user association formulation

In this section, we develop the most general epoch-by-epoch user association strategy. Let us introduce the following definitions: $\mathbf{p}(\tau) \triangleq \{p_{ji}(\tau), j \in \mathcal{U}_D, i \in \mathcal{B}\}$, $\check{\mathbf{p}}(\tau) \triangleq \{\check{p}_{ji}(\tau), j \in \mathcal{U}_V, i \in \mathcal{B}\}$, and $\mathbf{n}(\tau) \triangleq \{n_{ji}(\tau), j \in \mathcal{U}_D, i \in \mathcal{B}\}$. Given that, we formulate the optimization problem for the association strategy to be solved at the beginning of each epoch, which involves finding the indicators $\mathbf{x}(\tau)$ corresponding to the association as well as the average resource allocation variables, $\mathbf{p}(\tau)$, $\check{\mathbf{p}}(\tau)$, and $\mathbf{n}(\tau)$, that maximizes the aggregate utility function as follows:

$$\begin{aligned}
 & \underset{\mathbf{p}(\tau), \check{\mathbf{p}}(\tau), \mathbf{n}(\tau), \mathbf{x}(\tau), \mathbf{p}_{\text{RAD}}(\tau)}{\text{maximize}} && \sum_{j \in \mathcal{U}_D} U_j \left(\sum_{i \in \mathcal{B}} \bar{R}_{ji}(\tau) \right) \\
 & \text{subject to} && \\
 & \text{C1: } \sum_{i \in \mathcal{B}} \frac{\check{p}_{ji}(\tau) M_V}{\theta P_{\text{RAD}}^i(\tau) + A_{ji}(\tau)} \geq \gamma_j, \quad \forall j \in \mathcal{U}_V && \\
 & \text{C2: } \sum_{i \in \mathcal{B}} x_{ji}(\tau) = 1, \quad \forall j \in \mathcal{U}_T && \\
 & \text{C3: } \sum_{j \in \mathcal{U}_D} n_{ji}(\tau) \leq n_D^{(i)}, \quad \forall i \in \mathcal{B} && \\
 & \text{C4: } \sum_{j \in \mathcal{U}_D} p_{ji}(\tau) + \sum_{j \in \mathcal{U}_V} \check{p}_{ji}(\tau) \leq \bar{P}_{\text{BS}}^i(C_i(\tau)), \quad \forall i \in \mathcal{B} && \\
 & \text{C5: } P_{\text{RAD}}^i(\tau) = \sum_{j \in \mathcal{U}_D} p_{ji}(\tau) + \sum_{j \in \mathcal{U}_V} \check{p}_{ji}(\tau) + P_{\text{CPICH}}^i, \quad \forall i \in \mathcal{B} && \\
 & \text{C6: } 0 \leq n_{ji}(\tau) \leq x_{ji}(\tau) n_D^{(i)}, \quad \forall i \in \mathcal{B}, \forall j \in \mathcal{U}_D && \\
 & \text{C7: } 0 \leq \check{p}_{ji}(\tau) \leq x_{ji}(\tau) \bar{P}_{\text{BS}}^i(C_i(\tau)), \quad \forall i \in \mathcal{B}, \forall j \in \mathcal{U}_V && \\
 & \text{C8: } x_{ji}(\tau) = 0, \quad \forall i \notin \mathcal{S}_j(\tau), \forall j \in \mathcal{U}_T && \\
 & \text{C9: } x_{ji}(\tau) \geq 0, \quad \forall i \in \mathcal{B}, \forall j \in \mathcal{U}_T && \\
 & \text{C10: } p_{ji}(\tau) \geq 0, \quad \forall i \in \mathcal{B}, \forall j \in \mathcal{U}_D. &&
 \end{aligned} \tag{12}$$

It is important to note that, even though we are using average resource allocation to model the BS load, the previous problem only models the user association procedure, which needs to be run at time scale τ . After running the user association procedure, it is required to run a user scheduling algorithm⁴ at time scale t to obtain the instantaneous resource allocation variables (powers and codes), see Fig. 1. The final output of the previous optimization problem is solely the variable $x_{ji}(\tau)$.

In the previous optimization problem, $\bar{P}_{\text{BS}}^i(C_i(\tau)) \triangleq (g_i(C_i(\tau))/T_e - ((P_{\text{CPICH}}^i + P_c^i))_0^\infty)$ is the maximum power that the BS can use for traffic and pilot channels taking into account that the overall radiated power is limited and also the current battery level. Note that, even though at a particular epoch just the current battery is present in the formulation, the past harvesting and battery spendings also affect the current performance of the epoch as they appear in $C_i(\tau)$ through (2) (even though in (2) we used the time dependence t). Note that constraint C6 assures

that if $x_{ji}^* = 0$, then $n_{ji}^* = 0$ and the rate $R_{ji}^* = 0$. In this case, the allocated power will also be zero, $p_{ji}^* = 0$, as it does not improve the objective function but wastes power if $p_{ji}^* > 0$. Note that, at the optimum, C4 is attained with equality. Otherwise, if C4 is not fulfilled with equality, we could re-scale all the powers with a common positive factor higher than 1 until C4 is fulfilled with equality. This would increase the objective function and all the other constraints would still be fulfilled. Then, $P_{\text{RAD}}^{i*}(\tau) = \bar{P}_{\text{BS}}^i(C_i(\tau)) + P_{\text{CPICH}}^i$ and we can eliminate constraint C5 in problem (12).

Note that problem (12) is a convex optimization problem that can be solved using standard numerical algorithms [30]. However, less complex algorithms can be derived by making use of the analytical structure of the problem. According to this, we propose in the following a more efficient algorithm based on the dual problem and the subgradient method.

4.1.1 Primal-dual solution

In this subsection, we develop the association algorithm based on the dual problem of (12). The optimal solution will be presented first as a function of the Lagrange multipliers (or dual variables). A gradient-type scheme will later be developed for computing the optimum multipliers. Let $\mathbf{v} = \{v_j, j \in \mathcal{U}_V\}$, $\boldsymbol{\beta} = \{\beta_j, j \in \mathcal{U}_T\}$, $\boldsymbol{\mu} = \{\mu_i, i \in \mathcal{B}\}$, $\boldsymbol{\lambda} = \{\lambda_i, i \in \mathcal{B}\}$, $\boldsymbol{\pi} = \{\pi_{ji}, j \in \mathcal{U}_D, i \in \mathcal{B}\}$, and $\boldsymbol{\xi} = \{\xi_{ji}, j \in \mathcal{U}_V, i \in \mathcal{B}\}$ denote the vectors of dual variables associated to constraints C1, ..., C4, C6, and C7 in problem (12). We collect all the Lagrange multipliers in $\boldsymbol{\Psi} = \{\mathbf{v}, \boldsymbol{\beta}, \boldsymbol{\mu}, \boldsymbol{\lambda}, \boldsymbol{\pi}, \boldsymbol{\xi}\}$.

For a given set of Lagrange multipliers, $\boldsymbol{\Psi}$, we need to minimize the Lagrangian of problem (12), denoted by $\mathcal{L}(\mathbf{p}, \check{\mathbf{p}}, \mathbf{n}, \mathbf{x}, \boldsymbol{\Psi})$, w.r.t. the primal variables [30]. As it will be shown next, the structure of $\mathcal{L}(\mathbf{p}, \check{\mathbf{p}}, \mathbf{n}, \mathbf{x}, \boldsymbol{\Psi})$ allows the minimization w.r.t. \mathbf{p} and \mathbf{n} to be found in closed-form. Because $\mathcal{L}(\mathbf{p}, \check{\mathbf{p}}, \mathbf{n}, \mathbf{x}, \boldsymbol{\Psi})$ is strictly convex and differentiable w.r.t. \mathbf{p} and \mathbf{n} , minimization w.r.t. these variables require equating the corresponding partial derivatives of $\mathcal{L}(\mathbf{p}, \check{\mathbf{p}}, \mathbf{n}, \mathbf{x}, \boldsymbol{\Psi})$ to zero. Differentiating the Lagrangian

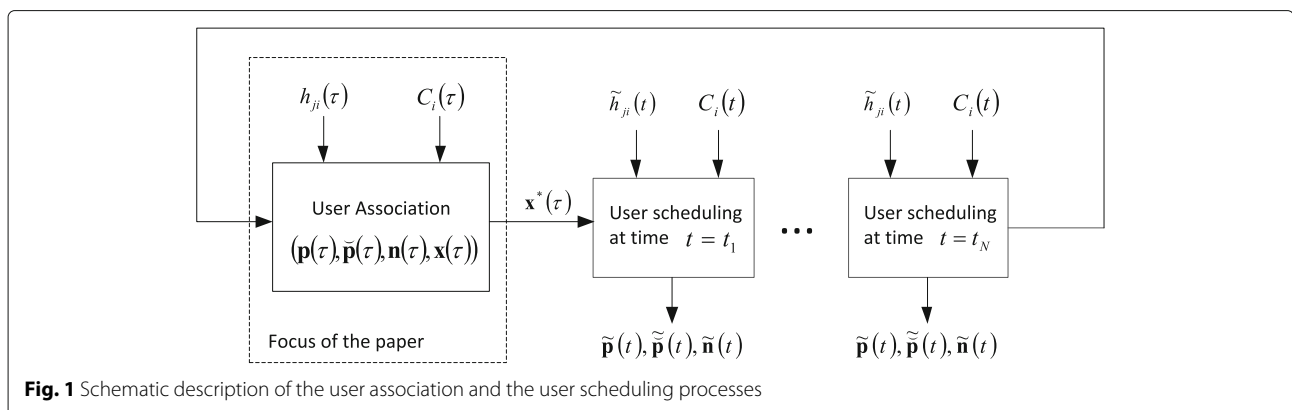


Fig. 1 Schematic description of the user association and the user scheduling processes

w.r.t. the data powers, equating the derivatives to zero and solving such expression for the data powers yields

$$p_{k\ell}^*(\mathbf{n}, \mathbf{p}, \Psi) = \left(\frac{G_{k\ell}(\mathbf{n}, \mathbf{p}) n_{k\ell}}{\ln(2)\lambda_\ell} - \frac{(\theta P_{\text{RAD}}^{\ell*} + A_{k\ell}) n_{k\ell}}{M_D} \right)_0^\infty, \quad (13)$$

where the projection guarantees that the allocated powers are nonnegative (constraint C10 in (12)) and $G_{k\ell}(\mathbf{n}, \mathbf{p}) \triangleq \dot{U}_k \left(s_{k\ell} + n_{k\ell} \log_2 \left(1 + \frac{M_D p_{k\ell}}{n_{k\ell} (\theta P_{\text{RAD}}^{\ell*} + A_{k\ell})} \right) \right)$, where $\dot{U}_k(\cdot)$ is the derivative of function $U_k(\cdot)$ and $s_{k\ell} = \sum_{i \in \mathcal{B}/\{\ell\}} n_{ki} \log_2 \left(1 + \frac{M_D p_{ki}}{n_{ki} (\theta P_{\text{RAD}}^{\ell*} + A_{ki})} \right)$ is a constant term for all users connected to BS ℓ . Notice that the bisection method can be used to calculate numerically the optimum power allocation (13). Proceeding similarly in the calculation of the optimum code allocation, we calculate the partial derivatives of the Lagrangian w.r.t. n_{ji} , equate them to zero and solve such equations for the codes, obtaining the expression

$$n_{k\ell}^*(\mathbf{n}, \mathbf{p}, \Psi) = \begin{cases} \left(\frac{G_{k\ell}(\mathbf{n}, \mathbf{p}) M_D p_{k\ell} ((\theta P_{\text{RAD}}^{\ell*} + A_{k\ell}) \ln(2))^{-1}}{G_{k\ell}(\mathbf{n}, \mathbf{p}) \log_2 \left(1 + \frac{M_D p_{k\ell}}{n_{k\ell} (\theta P_{\text{RAD}}^{\ell*} + A_{k\ell})} \right) - \mu_\ell - \pi_{k\ell}} - \frac{M_D p_{k\ell}}{\theta P_{\text{RAD}}^{\ell*} + A_{k\ell}} \right)_0^\infty & \text{if } \ell \in S_k, \\ 0 & \text{if } \ell \notin S_k, \end{cases} \quad (14)$$

where the projection guarantees that the codes satisfy the nonnegative constraints C6 and the definition by parts of the function of the codes is due to constraint C8. Notice that the bisection method can also be used in this case to compute the optimum code allocation, $n_{k\ell}$.

Unfortunately, the minimization of the Lagrangian w.r.t., the association variables x_{ji} , and the voice powers \check{p}_{ji} cannot be obtained by differentiating the Lagrangian as they appear through linear terms. In order to obtain the optimum power allocation for voice users and the association variables, we employ an iterative projected gradient approach to minimize the Lagrangian. The gradient for the j -th voice user power connected to the i -th BS is given by

$$s_{ji} \triangleq \nabla_{\check{p}_{ji}} \mathcal{L}(\mathbf{p}, \check{\mathbf{p}}, \mathbf{n}, \mathbf{x}, \Psi) = \frac{v_j M_D}{\theta P_{\text{RAD}}^{\ell*} + A_{ji}} + \lambda_i + \xi_{ji}, \quad \forall j \in \mathcal{U}_V, \forall i \in \mathcal{B}, \quad (15)$$

and the update equation of the projected gradient method is given by

$$\check{p}_{ji}^{(k+1)}(\tau) = \left(\check{p}_{ji}^{(k)}(\tau) - \delta^{(k)} s_{ji}(\tau) \right)_0^\infty, \quad \forall j \in \mathcal{U}_V, \forall i \in \mathcal{B}, \quad (16)$$

where k is the iteration index, the projection guarantees the nonnegativity constraints in (12), and $\delta^{(k)} = \frac{K}{\sqrt{k} \|\nabla \mathcal{L}\|_2}$ is the step size chosen such that the diminishing conditions are fulfilled, i.e., $\lim_{k \rightarrow \infty} \delta^{(k)} = 0$, $\sum_{k=1}^\infty \delta^{(k)} = \infty$, being $\nabla \mathcal{L}$ the gradient of the Lagrangian (w.r.t. all variables) [31]. Note that we have introduced the time dependence in order to clearly show the different time scales. For the association variables, the gradient is given by

$$t_{ji} \triangleq \nabla_{x_{ji}} \mathcal{L}(\mathbf{p}, \check{\mathbf{p}}, \mathbf{n}, \mathbf{x}, \Psi) = \beta_j - \xi_{ji} \bar{P}_{\text{BS}}^i(C_i) - \pi_{ji} n_D^{(i)}, \quad \forall j \in \mathcal{U}_T, \forall i \in \mathcal{B}, \quad (17)$$

and the update equation of the projected method is

$$x_{ji}^{(k+1)}(\tau) = \left(x_{ji}^{(k)}(\tau) - \delta^{(k)} t_{ji}(\tau) \right)_0^\infty, \quad \forall j \in \mathcal{U}_T, \forall i \in \mathcal{B}. \quad (18)$$

At this point, we have the expressions of the optimum primal variables, either in closed-form in (13) and (14) or iteratively in (16) and (18), for given dual variables. If the dual variables were optimum, then the expressions for the primal variables would yield the optimum values. The optimum Lagrange multipliers can be obtained from the dual problem by maximizing the dual function in (not presented for brevity) [30]. As the dual function is concave and generally not differentiable, we can apply any subgradient-type algorithm to find the optimum solution. A valid supergradient for each particular dual variable is given by the constraint it is associated with [31]. The update equations of the dual variables are not presented due to space constraints.

Once we know the optimal dual variables Ψ^* , we can obtain the optimum associations $\mathbf{x}^*(\Psi^*)$. The last step requires the quantization or rounding of such variables since the solution of (12) will provide, in general, a value of $x_{ji}^* \in [0, 1]$, but, for the actual implementation of the user association, we require $x_{ji}^* \in \{0, 1\}$ with $\sum_{i \in \mathcal{B}} x_{ji}^* = 1$, $\forall j$.

The proposed user association algorithm is based on the primal-dual block coordinate descent method for the update of the primal variables p_{ji} and n_{ji} (see Algorithm 1). The optimal association policy for any user j can be written in closed form using the indicator function as

$$x_{ji}^*(\boldsymbol{\beta}^*, \boldsymbol{\xi}^*, \boldsymbol{\pi}^*) = \mathbb{1}_{\{i = \arg \max_{i'} \{x_{ji'}(\boldsymbol{\beta}^*, \boldsymbol{\xi}^*, \boldsymbol{\pi}^*)\}\}}. \quad (19)$$

If multiple maximums exist simultaneously in (19), we propose to select just one randomly.

Algorithm 1 Primal-dual general user association algorithm

```

1: input:  $C_i(\tau), \forall i \in \mathcal{B}$ 
2: repeat (index  $\tau$ )
3:   compute  $P_{\text{RAD}}^{i*}(\tau) = \bar{P}_{\text{BS}}^i(C_i(\tau)) + P_{\text{CPICH}}^i, \forall i \in \mathcal{B}$ 
4:   calculate  $\mathcal{S}_j(\tau), \forall j \in \mathcal{U}_T$ 
5:   primal-dual algorithm:
6:     initialize  $\mathbf{v}(\tau) \geq 0, \boldsymbol{\beta}(\tau), \boldsymbol{\mu}(\tau) \geq 0, \boldsymbol{\lambda}(\tau) \geq 0, \boldsymbol{\pi}(\tau) \geq 0, \boldsymbol{\xi}(\tau) \geq 0$ 
7:     repeat (index  $q$ )
8:       initialize  $\mathbf{n}(\tau) \geq 0$ 
9:       repeat (index  $k$ )
10:         $p_{ji}^{(q,k+1)}(\tau) = p_{ji}^*(\mathbf{n}^{(q,k)}, \mathbf{p}^{(q,k)}, \boldsymbol{\Psi}^{(q)}, \tau)$  using (13)  $\forall j, i$ 
11:         $n_{ji}^{(q,k+1)}(\tau) = n_{ji}^*(\mathbf{n}^{(q,k)}, \mathbf{p}^{(q,k+1)}, \boldsymbol{\Psi}^{(q)}, \tau)$  using (14)  $\forall j, i$ 
12:        until  $p_{ji}^{(q,k+1)}(\tau)$  and  $n_{ji}^{(q,k+1)}(\tau)$  converge
13:        repeat (index  $k$ )
14:           $\check{p}_{ji}^{(q,k+1)}(\tau) = \left( \check{p}_{ji}^{(q,k)}(\tau) - \delta^{(k)} s_{ji}^{(q)}(\tau) \right)_0^\infty \quad \forall j, i$ 
15:          until  $\check{p}_{ji}^{(q,k+1)}(\tau)$  converges
16:          repeat (index  $k$ )
17:             $x_{ji}^{(q,k+1)}(\tau) = \left( x_{ji}^{(q,k)}(\tau) - \delta^{(k)} t_{ji}^{(q)}(\tau) \right)_0^\infty \quad \forall j, i$ 
18:            until  $x_{ji}^{(q,k+1)}(\tau)$  converges
19:          update dual variables with subgradient method using  $p_{ji}^{(q)}(\tau), \check{p}_{ji}^{(q)}(\tau), n_{ji}^{(q)}(\tau), x_{ji}^{(q)}(\tau)$  [31]
20:          until  $\mathbf{v}^{(q+1)}(\tau), \boldsymbol{\beta}^{(q+1)}(\tau), \boldsymbol{\mu}^{(q+1)}(\tau), \boldsymbol{\lambda}^{(q+1)}(\tau), \boldsymbol{\pi}^{(q+1)}(\tau), \boldsymbol{\xi}^{(q+1)}(\tau)$  converge
21:        end of primal-dual algorithm:
22:        quantization of  $x_{ji}(\tau)$  :
23:         $\forall j \in \mathcal{U}_T \longrightarrow i^* = \arg \max_{i \in \mathcal{B}} x_{ji}(\tau)$ 
24:         $x_{ji^*}^*(\tau) = 1, x_{ji}^*(\tau) = 0, \forall i \neq i^*$ 
25:        update batteries with power allocated by the scheduler:
26:         $C_i(\tau + 1) = (C_i(\tau) - E_i(\tau) + H_i(\tau))_0^{C_{\text{max}}^i}, \forall i \in \mathcal{B}$ 
27:      for all epochs  $\tau \in \Upsilon$ 
28:    end algorithm

```

4.2 Suboptimum user association strategies: low complexity solutions

In this section, we develop some strategies with lower complexity than the solution presented in the previous section. The goal is to provide algorithms with a reduced computational burden that could be of interest in scenarios where there is high mobility and the association procedure has to be executed more frequently.

4.2.1 Association of voice and data users separately

The first approach to consider is to split the two sets of users, namely the sets of voice and data users, and solve the association procedure for each group separately, first the voice users and then the data users. This is a good approach if the users are more or less spread out throughout the network. The assumption would not be valid if most of the data users are concentrated around just one BS and the voice users are connected to that BS, even though

they could be connected to the neighboring BSs due to their overlapping coverage areas. In any case, if the number of voice users is relatively low compared to the number of data users, such approximation is reasonably fair.

Voice users association We first perform the association of voice users. As we saw earlier, each particular voice user has to fulfill a minimum quality constraint in terms of SINR to guarantee a fixed rate service:

$$\sum_{i \in \mathcal{B}} x_{ji}(\tau) \frac{M_V \check{p}_{ji}(\tau)}{\theta P_{\text{RAD}}^{i*}(\tau) + A_{ji}(\tau)} \geq \gamma_j, \quad \forall j \in \mathcal{U}_V. \quad (20)$$

In order not to waste power unnecessarily, the previous constraint should be tight at the optimum. Given that, in the following, we present the procedure for the association of the voice users: for all users $j \in \mathcal{U}_V$, find the BS i' such that $\theta P_{\text{RAD}}^{i'*}(\tau) + A_{ji'}(\tau)$ is minimum and $i' \in \mathcal{S}_j(\tau)$, i.e., constraint C8 from problem (12) is not violated.

According to this, the association of voice users can be written in closed form using the indicator function as

$$x_{ji}^*(\tau) = \mathbb{1} \left\{ i = \arg \min_{i'} \left\{ \theta P_{\text{RAD}}^{i'}(\tau) + A_{ji'}(\tau) \mid i' \in \mathcal{S}_j(\tau) \right\} \right\}, \quad \forall j, i. \quad (21)$$

The optimum power to be allocated to voice user j associated to BS i' is given by

$$\check{p}_{ji'}^*(\tau) = \begin{cases} \gamma_j \frac{\theta P_{\text{RAD}}^{i'}(\tau) + A_{ji'}(\tau)}{M_V}, & \text{if } x_{ji}^*(\tau) = 1, \\ 0, & \text{if } x_{ji}^*(\tau) = 0. \end{cases} \quad (22)$$

The association is feasible if $P_V^{(i)}(\tau) \triangleq \sum_{j \in \mathcal{U}_V} \check{p}_{ji}^*(\tau) \leq \bar{P}_{\text{BS}}^i(C_i(\tau))$, $\forall i \in \mathcal{B}$, otherwise there is not enough energy at the batteries to fulfill all the minimum SINR constraints. In such case, a few users should be dropped from the system or their target SINR, γ_j , should be reduced.

Data users association Assuming that the voice user association is feasible, we can focus on the association of the data users. We model the association procedure through the following convex problem:

$$\text{maximize}_{\mathbf{p}(\tau), \mathbf{n}(\tau), \mathbf{x}(\tau)} \sum_{j \in \mathcal{U}_D} U_j \left(\sum_{i \in \mathcal{B}} \bar{R}_{ji}(\tau) \right)$$

subject to

$$C1 : \sum_{j \in \mathcal{U}_D} p_{ji}(\tau) \leq \bar{P}_{\text{BS}}^i(C_i(\tau)) - P_V^{(i)}(\tau), \quad \forall i \in \mathcal{B} \\ \text{and } C2, C3, C6, C8, C9, C10 \text{ from problem (12)}. \quad (23)$$

Notice the similarities of the previous optimization problem w.r.t. (12). We have eliminated the constraints involving the voice users. Strong duality also holds for this problem and, therefore, it can be solved using the primal-dual approach presented in the previous section. The details of the procedure employed to solve this problem will not be presented in the paper. This association strategy will be referred as first-voice then-data (FVTD).

4.2.2 Fixed resources per data user

We can go one step further and reduce the complexity of the previous algorithm by considering that the number of codes or the power assigned to data users are fixed. In this case, similar to the approach from [12], only one variable (code or power) will be used as a load metric of the BS. There is a slight conceptual difference between these two approaches (fixed codes and fixed power) when compared to the previous approach or the general approach in (12). In the two previous cases, the association strategy did not force to assign resources to all users, i.e., some data users could be assigned to a particular BS but the average rate they are assigned could be zero. However, as in this case a

given user is already assigned a portion of the total power or a certain number of codes, the associated rate will be greater than zero, i.e., the BS will spend resources as long as such user is associated to it.

Fixed power allocation per data user In this case, we consider that all users get a given portion of the power available at the BS, i.e., $p_i(\tau) \triangleq \frac{\bar{P}_{\text{BS}}^i(C_i(\tau)) - P_V^{(i)}(\tau)}{K_i}$, $\forall i \in \mathcal{B}$, where K_i is the fixed maximum number of users that the BS i wants to accept. In the most general case, this portion is different for each BS, where such portion together with the total available power constrains the maximum number of users that are allowed to be connected to each particular BS. The user association strategy is formulated through the following convex optimization problem:

$$\text{maximize}_{\mathbf{n}(\tau), \mathbf{x}(\tau)} \sum_{j \in \mathcal{U}_D} U_j \left(\sum_{i \in \mathcal{B}} \bar{R}_{ji}(\tau) \right)$$

subject to

$$C1 : \sum_{j \in \mathcal{U}_D} x_{ji}(\tau) \leq \frac{\bar{P}_{\text{BS}}^i(C_i(\tau)) - P_V^{(i)}(\tau)}{p_i(\tau)}, \quad \forall i \in \mathcal{B}$$

and C2, C3, C6, C8, C9 from problem (12).

(24)

We will not present the details of the solution of this problem for the sake of space as we have already presented the methodology for the previous two association problems. We will just evaluate the performance of the strategy in the simulations section.

Fixed code allocation per data user Now we consider that a given number of codes are assigned to each particular data user, i.e., $n_i \triangleq \frac{n_D^{(i)}}{K_i}$, $\forall i \in \mathcal{B}$, where K_i is again the maximum number of users that BS i wants to accept. In this case, the power is the resource that is optimized. The association strategy is, thus, formulated as

$$\text{maximize}_{\mathbf{p}(\tau), \mathbf{x}(\tau)} \sum_{j \in \mathcal{U}_D} U_j \left(\sum_{i \in \mathcal{B}} \bar{R}_{ji}(\tau) \right)$$

subject to

$$C1 : \sum_{j \in \mathcal{U}_D} x_{ji}(\tau) \leq \frac{n_D^{(i)}}{n_i}, \quad \forall i \in \mathcal{B}$$

$$C2 : \sum_{j \in \mathcal{U}_D} p_{ji}(\tau) \leq \bar{P}_{\text{BS}}^i(C_i(\tau)) - P_V^{(i)}(\tau), \quad \forall i \in \mathcal{B}$$

$$C3 : 0 \leq p_{ji}(\tau) \leq x_{ji}(\tau) \left(\bar{P}_{\text{BS}}^i(C_i(\tau)) - P_V^{(i)}(\tau) \right), \quad \forall i, j$$

and C2, C8, C9 from problem (12).

(25)

Also in this case, we will not present the details of the solution of the problem as we have already presented the methodology for the previous two association problems. We will just evaluate the performance of the strategy in the numerical sections.

4.3 Distributed algorithm

In this section, we extend the concept presented in previous sections where the association strategy was executed by a centralized entity that required the knowledge of all the parameters involved in the association decision, such as the channels of all users to all BSs, the current battery levels of all BSs, etc. Usually, the distributed solutions may overcome some of the difficulties of the centralized approaches. Some of the advantages of the distributed algorithms are presented in the sequel:

- All link connections among the BSs and the central entity are required to operate in the centralized approach. If one link fails, it is not possible to solve the optimization problem. The distributed algorithms are usually more robust in this sense since they require only connections with neighbor BSs and if one link does not work the rest of the network can still compute the user association.
- Usually, the information sharing in the distributed approach is only required among neighbor nodes, which are physically close. Hence, there is no need to deploy long backhaul connections from all BSs to the central entity. However, in the particular solution that we will present, backhaul connections are not needed since BSs do not need to exchange any message among them.
- The computational load required at each individual element to solve the distributed approach is lower than its centralized counterpart since fewer variables are considered.
- For the aforementioned issues, the distributed solution is more scalable than the centralized one.

It is important to note that we decided to present first the centralized approach and then the distributed one. This is so since distributed solutions are derived from centralized optimization problems and, therefore, it is required to first mathematically model the centralized approach and then derive the distributed one. The distributed solution has some drawbacks as well. For example, it is known that distributed solutions require more iterations than the centralized ones to converge

and more frequent message exchange among BSs and users [32], which means that the centralized approach is preferable to the distributed one when the convergence speed is important, that is, when the dynamics of the users are high and the network should adapt quickly. Also, if a centralized entity exists collecting all the needed information, then there is no reason to resort to a distributed approach since this would imply increasing the computational effort of the individual entities (both BSs and users in the distributed approach proposed in this section) unnecessarily. In other words, both the centralized and the distributed approaches have pros and cons, and the final decision on what approach to apply depends on the different aspects mentioned before and the concrete requirements and characteristics of the scenario at hand.

In this section, we propose a distributed algorithm of the general problem presented in (12) via Lagrange dual decomposition [32]. We can apply a dual decomposition to the original problem in (12) and solve it separately by users and BSs. This implies that only local information is required to solve the subproblems, i.e., users only need to know the available resources of the BSs they are able to connect to and the individual propagation channels with them.

Having introduced the motivation behind the distributed solution, we proceed now to explain the steps of the procedure. As the objective function of the problem cannot be decomposed for each BS, the users will be the ones carrying out the computation of the association. The role of the BSs will be to provide certain *prices* to the users as a function of the association demands. Because each user selects the preferred BS, the criterion to be employed for the voice users is the one explained in Section 4.2.1. As a consequence, the voice users and data users must select the BSs separately, first the voice users and then the data users.

Once the voice users have selected the BSs they prefer and the problem is feasible, we focus on the data users by applying a dual decomposition method to problem (23). The coupling constraints are $C1$ and $C3$ as we have just one constraint for all users not allowing to split the optimization problem for each user. This motivates us to dualize both constraints and decouple the problem as explained in the sequel. Let $\mu(\tau) = \{\mu_i(\tau), i \in \mathcal{B}\}$ and $\lambda(\tau) = \{\lambda_i(\tau), i \in \mathcal{B}\}$ be the Lagrange multipliers associated to constraints $C1$ and $C3$ and let $\mathbf{p}_j(\tau) = \{p_{ji}(\tau), i \in \mathcal{S}_j(\tau)\}$, $\mathbf{n}_j(\tau) = \{n_{ji}(\tau), i \in \mathcal{S}_j(\tau)\}$, and $\mathbf{x}_j(\tau) = \{x_{ji}(\tau), i \in \mathcal{S}_j(\tau)\}$. Then, the j -th data user must solve the following optimization problem (where, for simplicity in the notation and the resulting algorithm, we have eliminated the terms that do not depend on the optimization variables, i.e., constant terms, in the objective function as they do not affect the optimization variables):

$$\begin{aligned}
 & \underset{\mathbf{p}(\tau), \mathbf{n}_j(\tau), \mathbf{x}_j(\tau)}{\text{maximize}} && U_j \left(\sum_{i \in \mathcal{S}_j(\tau)} \bar{R}_{ji}(\tau) \right) - \sum_{i \in \mathcal{S}_j(\tau)} \mu_i(\tau) n_{ji}(\tau) - \sum_{i \in \mathcal{S}_j(\tau)} \lambda_i(\tau) p_{ji}(\tau) \\
 & \text{subject to} && \\
 & C1 : && \sum_{i \in \mathcal{S}_j(\tau)} x_{ji}(\tau) = 1 \\
 & C2 : && 0 \leq n_{ji}(\tau) \leq x_{ji}(\tau) n_D^{(i)}, \quad \forall i \in \mathcal{S}_j(\tau) \\
 & C3 : && x_{ji}(\tau) \geq 0, \quad \forall i \in \mathcal{S}_j(\tau) \\
 & C4 : && p_{ji}(\tau) \geq 0, \quad \forall i \in \mathcal{S}_j(\tau).
 \end{aligned} \tag{26}$$

Notice that the knowledge of the Lagrange multipliers $\boldsymbol{\mu}(\tau)$ and $\boldsymbol{\lambda}(\tau)$ is required to solve the previous problem. Notice also that only local information is needed, which is represented by the set of available BSs for each particular user given by $\mathcal{S}_j(\tau)$ for the j -th user. The previous optimization problem is convex and can be solved using any standard procedure as the one explained in this document (primal-dual iteration) in Section 4.1.1.

4.3.1 Complexity, exchange messages, and convergence

It is true that the complexity of distributed approaches is usually larger than the complexity of centralized ones since more iterations are required to obtain the optimum solution. However, the exchange of these messages in each iteration between BSs and users only require local information. Besides, in this particular case, BSs do not need to exchange information among them and

hence there is no need to deploy backhaul connections, see Fig. 2.

In terms of exchange messages among users and BSs, when data users execute the user association strategy (26), they convey the results of the optimum variables, $(\mathbf{p}_j^*(\tau), \mathbf{n}_j^*(\tau), \text{ and } \mathbf{x}_j^*(\tau))$, to the BSs they can connect to. Then, the BSs are responsible for iterating to solve the dual problem to obtain asymptotically the optimum Lagrange multipliers $\boldsymbol{\mu}(\tau)$ and $\boldsymbol{\lambda}(\tau)$ using the expressions

$$\mu_i^{(q+1)}(\tau) = \left(\mu_i^{(q)}(\tau) + \delta^{(q)} \left(\sum_{j \in \mathcal{U}_D^{(i)}} n_{ji}^{(q)}(\tau) - n_D^{(i)} \right) \right)_0^\infty, \quad \forall i \in \mathcal{B} \tag{27}$$

$$\begin{aligned}
 & \lambda_i^{(q+1)}(\tau) \\
 & = \left(\lambda_i^{(q)}(\tau) + \delta^{(q)} \left(\sum_{j \in \mathcal{U}_D^{(i)}} p_{ji}^{(q)}(\tau) - \bar{P}_{\text{BS}}^i(C_i(\tau)) + P_V^{(i)}(\tau) \right) \right)_0^\infty, \quad \forall i \in \mathcal{B}.
 \end{aligned} \tag{28}$$

Then, the BSs broadcast the information of the updated values of the Lagrange multipliers to the users, and then users execute again the association procedure.

It is important to note that since the dual problem is convex, any subgradient-type algorithm yields asymptotically the optimum solution. Because Slater's condition is satisfied, strong duality holds and the original problem (23) can be equivalently solved in a distributed fashion

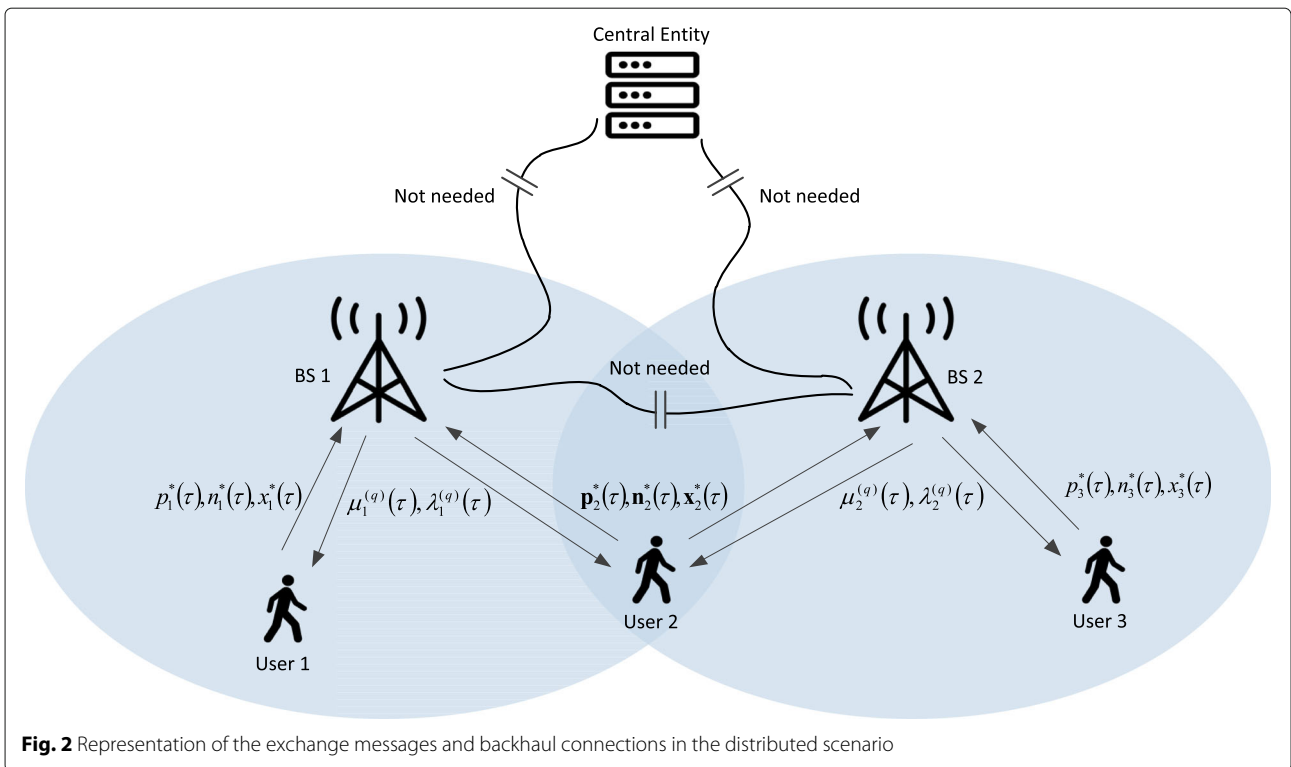


Fig. 2 Representation of the exchange messages and backhaul connections in the distributed scenario

without coordination among the users or the BSs and the convergence is guaranteed.

4.3.2 Detailed steps of the distributed algorithm

The final distributed algorithm is shown in Algorithm 2. In the algorithm, $\delta^{(q)}$ is the step size. The multipliers $\mu(\tau)$ and $\lambda(\tau)$ work as messages between data users and BSs. They represent the *price* of the resources of the particular BS determined by the load situation. If, for example, the demand of code resources for a given BS i , $\sum_j n_{ji}(\tau)$, is larger than the maximum number that the BS can offer, $\sum_j n_{ji}(\tau) > n_D^{(i)}$ then, the price μ_i increases, which means that the users will have to pay more to utilize its code resources making the users consider other BS options with lower prices that would improve its objective function

(and, thus, its throughput). The same reasoning can be applied to the power. If a given BS is underutilized, then its corresponding prices will be low making it appealing for the users achieving a well load-balanced network.

5 Asymptotic analysis of the battery evolution

In this section, we present the asymptotic behavior of the battery evolution of the system. This analysis will help us understand how the system works when the number of epochs grows up to infinity. Note that the performance of an energy-limited network, such as the one considered in this paper, depends on the sizes of the batteries. According to this, the usefulness of the analysis in this section relies on the fact that we can study which is the energy that is asymptotically stored in the batteries. This could be taken

Algorithm 2 User association strategy based on the distributed algorithm

```

1: input:  $C_i(\tau), \forall i \in \mathcal{B}$ 
2: repeat (index  $\tau$ )
3:   compute  $P_{\text{RAD}}^{i*}(\tau) = \bar{P}_{\text{BS}}^i(C_i(\tau)) + P_{\text{CPICH}}^i, \forall i \in \mathcal{B}$ 
4:   calculate  $S_j(\tau), \forall j \in \mathcal{U}_T$ 
5:   distributed algorithm:
6:   voice users:
7:      $x_{ji}^*(\tau) = \mathbb{1}_{\{i = \arg \min_{i'} \{\theta P_{\text{RAD}}^{i'*}(\tau) + A_{ji'}(\tau) \mid i' \in S_j(\tau)\}\}}, \forall j \in \mathcal{U}_V, \forall i \in \mathcal{B}$ 
8:      $\check{p}_{ji}^*(\tau) = x_{ji}^*(\tau) \frac{\gamma_i(\theta P_{\text{RAD}}^{i*}(\tau) + A_{ji}(\tau))}{M_V}, \forall j \in \mathcal{U}_V, \forall i \in \mathcal{B}$ 
9:      $P_V^{(i)}(\tau) \triangleq \sum_{j \in \mathcal{U}_V} \check{p}_{ji}^*(\tau), \forall i \in \mathcal{B}$ 
10:    if  $P_V^{(i)}(\tau) > \bar{P}_{\text{BS}}^i(C_i(\tau))$ 
11:      find any  $j'$  such that  $x_{j'i}^*(\tau) = 1$ 
12:      set  $x_{ji}^*(\tau) = 0, x_{j'k}^*(\tau) = \mathbb{1}_{\{k = \arg \min_{k'} \{\theta P_{\text{RAD}}^{k'*}(\tau) + A_{j'k'}(\tau) \mid k' \in S_j(\tau) \setminus \{i\}\}}$ 
13:      go to 8
14:    end if
15:  initialize  $\mu(\tau) \geq 0, \lambda(\tau) \geq 0$ 
16:  repeat (index  $q$ )
17:    data users:
18:    solve problem (26) and obtain  $\mathbf{p}_j^{(q)}(\lambda^{(q)}, \tau)$  and  $\mathbf{n}_j^{(q)}(\mu^{(q)}, \tau), \forall j \in \mathcal{U}_D$ 
19:    each user sends  $p_{ji}^{(q)}(\lambda^{(q)}, \tau)$  and  $n_{ji}^{(q)}(\mu^{(q)}, \tau)$  to the corresponding BSs
20:    BSs:
21:    update dual variables using  $\mathbf{p}_j^{(q)}(\tau)$  and  $\mathbf{n}_j^{(q)}(\tau)$ , and broadcast them:
22:     $\mu_i^{(q+1)}(\tau) = \left( \mu_i^{(q)}(\tau) + \delta^{(q)} \left( \sum_{j \in \mathcal{U}_D^{(i)}} n_{ji}^{(q)}(\tau) - n_D^{(i)} \right) \right)_0^\infty, \forall i \in \mathcal{B}$ 
23:     $\lambda_i^{(q+1)}(\tau) = \left( \lambda_i^{(q)}(\tau) + \delta^{(q)} \left( \sum_{j \in \mathcal{U}_D^{(i)}} p_{ji}^{(q)}(\tau) - \bar{P}_{\text{BS}}^i(C_i(\tau)) + P_V^{(i)}(\tau) \right) \right)_0^\infty$ 
24:  until  $\mu^{(q+1)}(\tau)$  and  $\lambda^{(q+1)}(\tau)$  converge
25:  quantization of  $x_{ji}(\tau)$  :
26:   $\forall j \in \mathcal{U}_D \longrightarrow i^* = \arg \max_{i \in \mathcal{B}} x_{ji}(\tau)$ 
27:   $x_{j i^*}^*(\tau) = 1, x_{ji}^*(\tau) = 0, \forall i \neq i^*$ 
28:  update batteries with power allocated by the scheduler:
29:   $C_i(\tau + 1) = (C_i(\tau) - E_i(\tau) + H_i(\tau))_0^{C_{\text{max}}^i}, \forall i \in \mathcal{B}$ 
30: for all epochs  $\tau \in \Upsilon$ 
31: end algorithm

```

as a basis for an initial calculation of the sizes of the batteries that will be needed to avoid overflowing, that is, as a basis study for the energy-dimensioning of the network so that all the energy that is harvested can be used effectively and not lost due to overflowing.

We will assume that function $g_i(\cdot)$ is linear (or affine) w.r.t. the battery, and that the BSs do not have physical limitations ($P_{BS_i}^{\max} = \infty$), i.e., $g_i(C_i(\tau)) = \alpha \cdot C_i(\tau)$, $\forall i$. This implies that the lower limit of the battery is never reached as long as $\mathbb{E}[H_i(\tau)] > 0$. Let us consider that the harvesting is stationary (otherwise no convergence is guaranteed unless the time window where the harvesting intensity does not change is sufficiently large). We will also assume for simplicity in the development that the battery never reaches its maximum limit, C_{\max}^i and, thus, the battery dynamics equation that governs the battery evolution is given by

$$C_i(\tau+1) = C_i(\tau) - E_i(\tau) + H_i(\tau) = (1-\alpha)C_i(\tau) + H_i(\tau), \quad (29)$$

where in the right hand side of (29) we have assumed that $E_i(\tau) = \alpha_i \cdot C_i(\tau)$ (this is true because constraint C4 in (12) is fulfilled with equality). Recall that $H_i(\tau)$ is a Bernoulli stochastic process such that its components $\{H_i(\tau)\}_\tau$ are identically distributed (i.i.d.) random variables and, thus, $C_i(\tau)$ is also a stochastic process. If $0 < \alpha_i < 1$, then the recursive relation in (29) is stable since the input, $H_i(\tau)$, is bounded and the solution to (29) is given by

$$C_i(\tau+1) = \sum_{j=0}^{\infty} (1-\alpha_i)^j H_i(\tau-j), \quad (30)$$

where the previous sum converges in quadratic mean since $|1-\alpha_i| < 1$ implies that $\sum_{j=0}^{\infty} |(1-\alpha_i)^j| = \frac{1}{1-\alpha_i} < \infty$. Given this, the expected value of the battery in convergence is, thus, given by

$$\lim_{\tau \rightarrow \infty} \mathbb{E}[C_i(\tau)] = \frac{\mathbb{E}[H_i(\tau)]}{\alpha_i} = \frac{p_i \cdot e_i}{\alpha_i}. \quad (31)$$

This implies that the expected energy allowed to be extracted from the battery is limited by the energy collected through harvesting, $\mathbb{E}[H_i(\tau)]$, as $\mathbb{E}[g(C_i(\tau))] = \mathbb{E}[H_i(\tau)]$. As a result, the expected performance in terms of throughput does not depend on the value of α_i that is configured or the initial battery level. In fact, this is true if the upper limit of the battery is never reached or $C_{\max}^i = \infty$. Notice that, if $\alpha \approx 0$, then overflows may occur, incurring a loss of energy and, thus, a loss in throughput performance. Therefore, in reality, there is an optimum value of $\alpha_i^* > 0$ that achieves the maximum throughput performance. Unfortunately, an analytical expression of the optimum value of α has not been found and numerical

simulations are needed (see [29]). The variance of the expected battery in convergence is expressed as

$$\text{var}(C_i(\tau)) = \text{var}\left(\sum_{j=0}^{\infty} (1-\alpha_i)^j H_i(\tau-j)\right) \quad (32)$$

$$= \sum_{j=0}^{\infty} (1-\alpha_i)^{2j} \text{var}(H_i(\tau-j)) \quad (33)$$

$$= \sum_{j=0}^{\infty} (1-\alpha_i)^{2j} p_i(1-p_i)e_i^2 \quad (34)$$

$$= \frac{p_i(1-p_i)e_i^2}{1-(1-\alpha_i)^2}. \quad (35)$$

6 Simulations results

In this section we evaluate the performance of the proposed strategy⁵. The scenario under consideration is composed of 15 BSs with four tiers. The deployment layout is shown in Fig. 3, where we also show the number of available BS at each point of the scenario (at a given epoch). BS 1 belongs to tier 1, BSs {2, ..., 9} belong to tier 2, BSs {11, ..., 13} belong to tier 3, and BSs {14, 15} belong to tier 4. The maximum radiated power, $P_{BS_i}^{\max} + P_{CPICH}^i$, is 46 dBm, 24 dBm, 20 dBm, and 13 dBm for each tier. The pilot power, P_{CPICH}^i , is 5% of $P_{BS_i}^{\max}$. The fixed power, P_c^i , is 33 dBm, 17 dBm, 13 dBm, and 6 dBm for each tier. The system contains 85 data users and 15 voice users. All the users in the system are mobile with a speed of 4 km/h. The instantaneous channel, \hat{h}_{ji} , incorporates Rayleigh fading with unitary power and a path loss based on Okumura-Hata for open areas. The orthogonality factor is $\theta = 0.35$ [33]. The code gain of data codes is $M_D = 16$ and the minimum SINR normalized by gain, $\frac{\gamma}{M_V} = -13.7$ dB. The minimum SINR for pilot signals is $\gamma_{CPICH} = -20$ dB. The noise power is $\sigma^2 = -102$ dBm. The battery

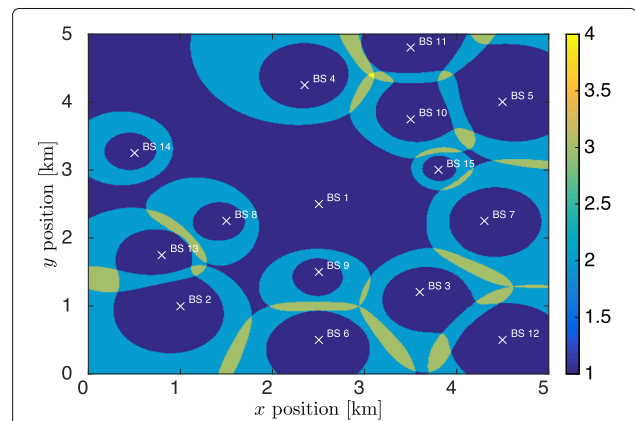


Fig. 3 Reference scenario and available BSs at a given epoch. The color bar represents the number of BSs available at each point

capacities, C_{\max}^i , are 1200 Joules, 30 Joules, 12 Joules, and 2.5 Joules, for each tier and the quantity of energy in an energy harvesting packet is, $e_i = C_{\max}^i/10$ Joules. The number of epochs considered is 300 and the time between epochs is $T_e = 10$ s. The utility function considered is $U_j(\cdot) = \log(\cdot)$ since the log(rates) tends to provide fairness among user rates [26]. In the simulations, for each particular user association, Proportionally Fair (PF) scheduling with instantaneous channels is run at each particular BS for assigning the instantaneous resources. Thus, in this section, instantaneous data rates will be evaluated and shown in the figures and the batteries will be updated with the actual (instantaneous) powers radiated by the BSs. For a more complete description of the deployment, see [5].

In the max-SINR strategy, the users are associated with the strongest BS in terms of received pilot power without considering the actual load of the BS nor its current battery level. Figure 4 presents the user association for the first epoch, considering that the batteries are full and $\alpha = 0.1$. As we can conclude from the figures, some of the users connected to the BS1 (macro BS) are transferred to smaller cells if a proper load balancing mechanism is considered (see for example BS8, BS11, BS14). If we compare the number of users associated to the macro BS (BS1) for both approaches, i.e., max-SINR and our strategy, we conclude from the figures that, with the max-SINR approach, the macro is serving 40 more users than with the proposed strategy. This makes the macro BS saturate and assign very little resources to the users while the small cells are operating with few users.

Figure 5a presents the cumulative distribution function (CDF) of the rates for the more complex strategy proposed in this paper in (12), denoted as ‘G’ (or ‘General’ in some subsequent figures) in the legend, and the maximum SINR association strategy, denoted as ‘MS’, for different values of α and harvesting p^6 . Notice that $\alpha = 1$ allows to use all the battery at each particular epoch. As we can see, the proposed scheme provides two to three times higher rates in the low-rate regime compared to the max-SINR

and outperforms the traditional approach up to rates in the order of 0.4 bits/s/Hz, which represents around 70% of the user rates. Note that the CDF in the figure shows that, by using the proposed approach, the probability of a user receiving a very low rate is much lower than with the max-SINR strategy, which means that more users will have a connection with a minimum quality. Another conclusion is that the proposed approach provides a fair load balancing network in terms of user distribution which is translated into an improvement of the overall network throughput. However, this comes with a small reduction in terms of peak rates (lower than 20%). An interesting insight is that for larger energy harvested values, the influence of the value of α is lower. However, if more energy is available, a better load-balanced network is created. This is due to the fact that small cells have more power and the coverage radius are larger offering more possibilities for the users to associate to BSs. Contrarily, for larger values of α we obtain larger peak rates, but a worse balanced network. Thus, from a load-balancing perspective, it is better to control how the energy is being used at different epochs.

Figure 5b presents the CDF of the rates of the data users with $\alpha = 0.1$ and $p = 0.6$ for the low-complexity approaches derived from (24) and (25), the distributed solution from (26), and the max-SINR. In this case, the distributed solution yields the same performance as the FVTD strategy (even though it is not shown in the figure). This is expected as both problems attempt to solve the same optimization problem. As we can see, even if we use the simple low-complexity solution fixing the power or the codes to be allocated, the performance obtained is quite close to the rest of the strategies proposed in this paper. All the strategies outperforms the max-SINR providing a better balanced network. The results obtained from (24) and (25) are similar to the results that would be obtained from the method in [12] as the measure of the load is only based on one variable, but now we also consider battery limitations and time evolution.

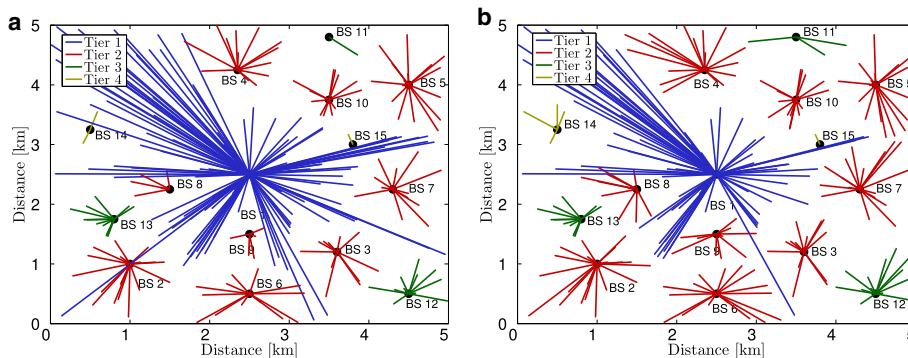


Fig. 4 Snapshot of the user association for different policies. **a** User association policy: max-SINR. **b** User association policy: problem (12)

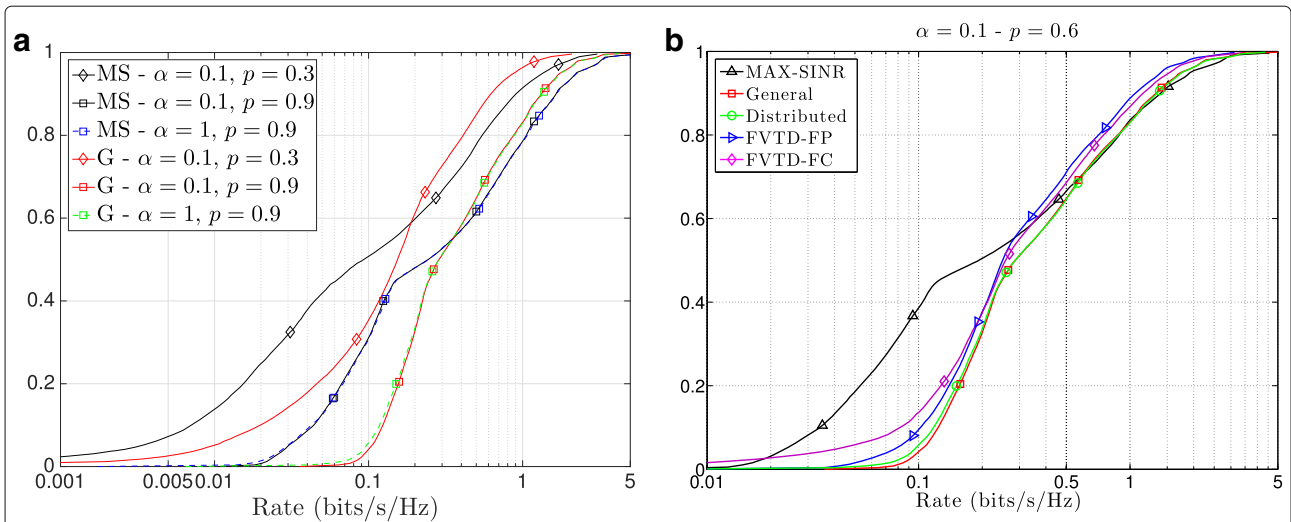


Fig. 5 CDF of the instantaneous user rates for different user association strategies. **a** CDF of the user rates for different energy harvesting intensities ($p = 0.3$; $p = 0.6$; and $p = 0.9$) and the value of α for the proposed strategy and the max-SINR strategy. **b** CDF of the user rates for the different proposed strategies and the max-SINR

Figures 6 and 7 depict the battery evolution of the BSs in tier 1, tier 2, tier 3, and tier 4 for different values of harvesting p and α considering the traditional max-SINR approach and the proposed strategy. As expected, if the harvesting increases, the expected battery also increases. Also the value of α impacts the battery evolution. For larger values of α , we obtain lower residual battery levels

at convergence (after the transient period). In fact, it can be shown that if the battery converges in average terms, i.e., $0 < \lim_{t \rightarrow \infty} \mathbb{E}[C_i(\tau)] < C_{\max}^i$, then, as we mentioned before, $\lim_{t \rightarrow \infty} \mathbb{E}[C_i(\tau)] = \frac{\mathbb{E}[H_i(\tau)]}{\alpha_i} = \frac{p_i \cdot e_i}{\alpha_i}$. The variance can also be checked to fulfill the expression in (35). Notice, however, that this is not always the behavior of tier 4. Recall that, through the development in Section 5,

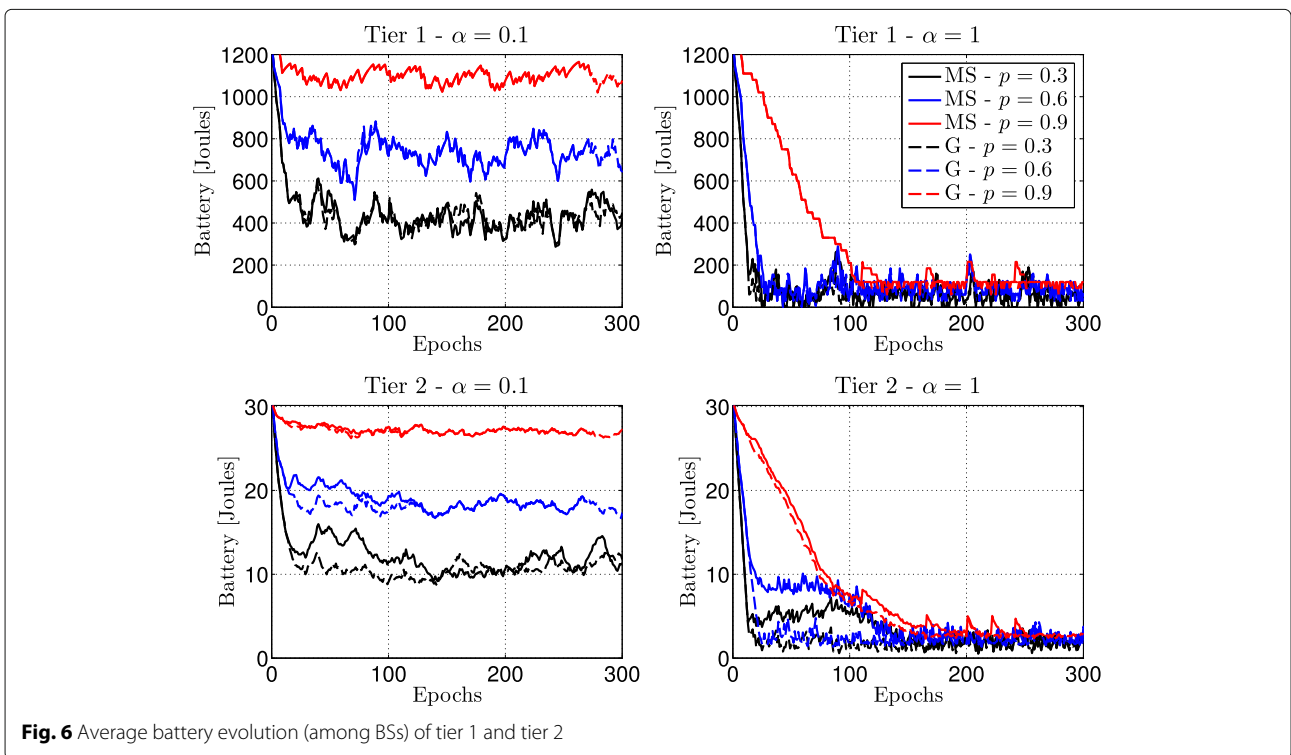


Fig. 6 Average battery evolution (among BSs) of tier 1 and tier 2

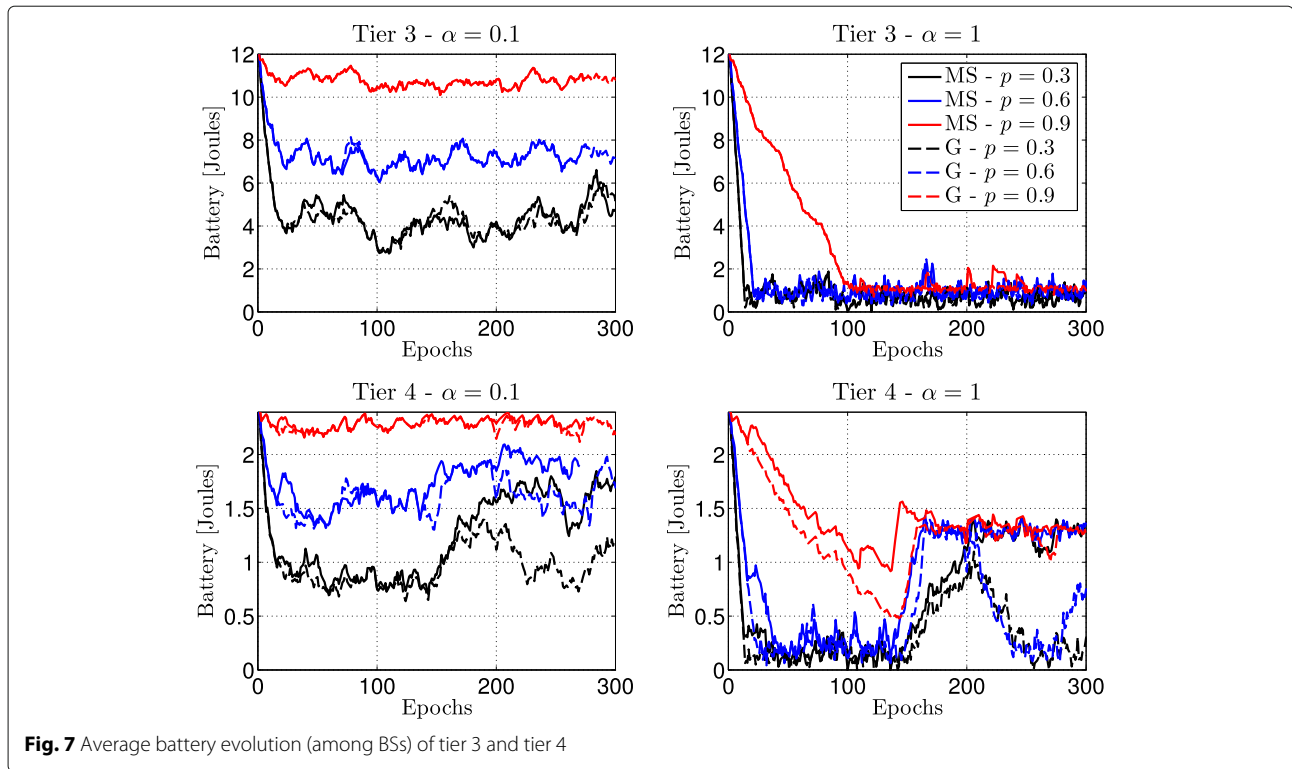


Fig. 7 Average battery evolution (among BSs) of tier 3 and tier 4

we assumed that constraint $C4$ was attained with equality. This will only happen if there is at least one data user connected to any BS of the tier. Because there are just two BSs in tier 4, it is very likely that no data users are within the coverage area of these two BSs at a given epoch. This phenomenon is what we can see in the figure. The batteries increase due to harvesting.

The evolution of the average number of users associated to a particular tier as time evolves (average among the BSs of the same tier) is shown in Figs. 8 and 9 for different values of α , different harvesting intensities, and different association strategies. Each vertical cut of these plots would yield the snapshot user association presented in Fig. 4 (epoch 1 was considered in those previous plots). The main insight that we obtain from these figures is that, for the max-SINR strategy, the tier 1 tends to have more users than the proposed strategies. As time evolves, the macro BS (tier 1) tends to reduce its battery level (as it is widely used) and the number of users that associate with it decreases a little. This effect is more notable with the proposed strategies. If we have a look at tier 2, for example, the number of users increases as time evolves. In general terms, users associated to the macro BS when the max-SINR strategy is considered are transferred to other tiers if a load balancing technique is implemented. Users associated to tier 4 are almost double in average terms when comparing max-SINR strategy with the proposed one. In terms of the macro BS, there are roughly 20 users less

in average, if we compare the proposed approach with the max-SINR strategy. The solutions of the user association for the general case derived from (12) and the FVTD strategy are quite similar. Another insight is that for larger values of α we see that the fluctuation of the curves increases. This is because the batteries of the BSs tend to run out of energy and users tend to associate to other BSs more frequently.

Finally, Figs. 10 and 11 show the computational complexity associated with the centralized approaches proposed in the paper. We show the computational complexity⁷ in relative terms with respect to the most computational demanding solution. We show two different scenarios to see the scalability of the algorithms. As it can be seen in the figures, the low complexity solutions presented in the paper require half the computational burden of the most general approach whereas the FVTD approach demands a computational burden between 20 to 30% lower than the most general solution. Hence, in situations with limited computational resources or where the algorithms must be run quite frequently (due to, for example, high mobility), the low complexity approaches represent a practical solution with good performance.

7 Conclusions

In this paper, we have proposed several user association strategies driven by DL metrics in a scenario composed

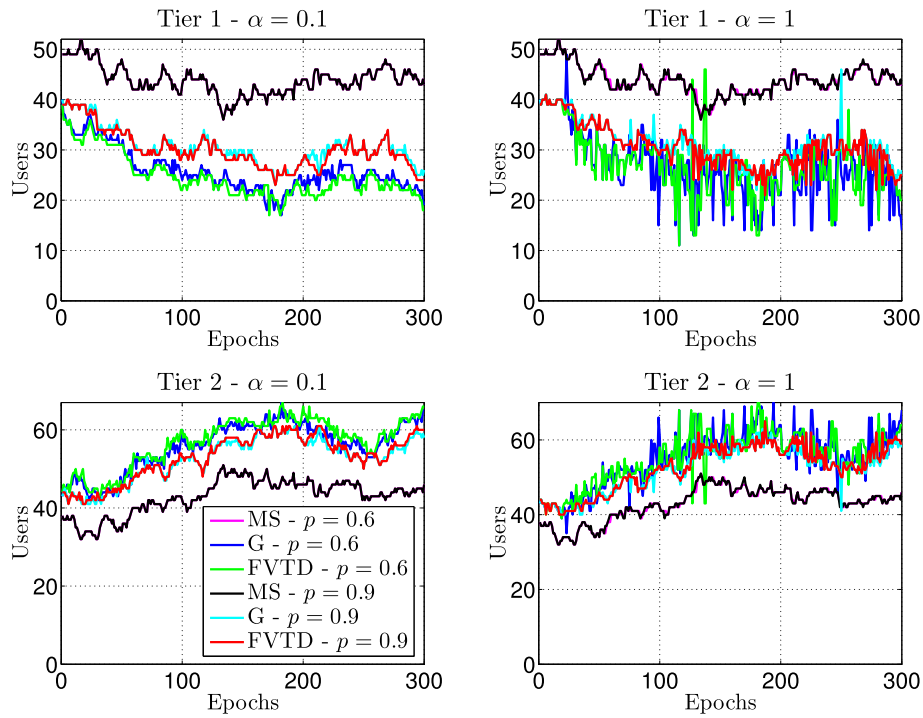


Fig. 8 Evolution of user association in tier 1 and tier 2

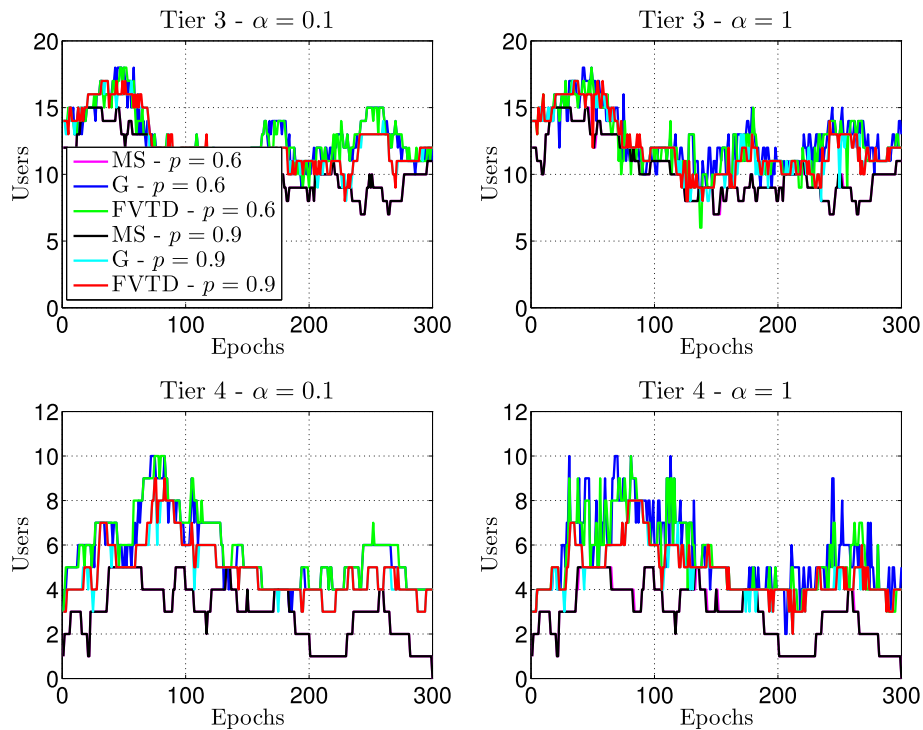
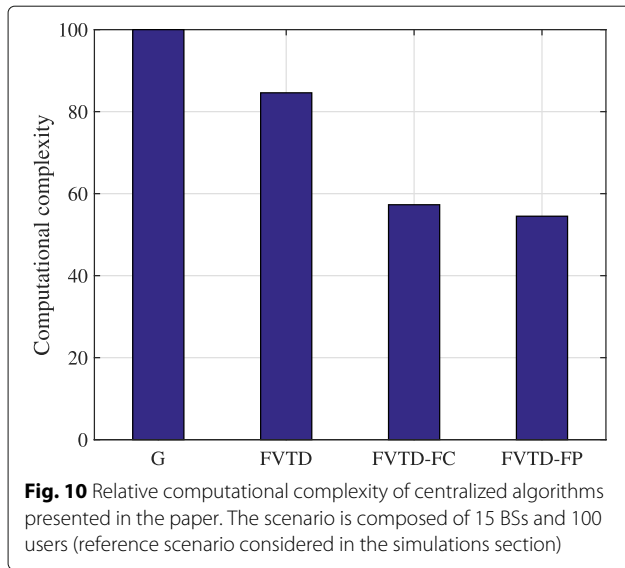
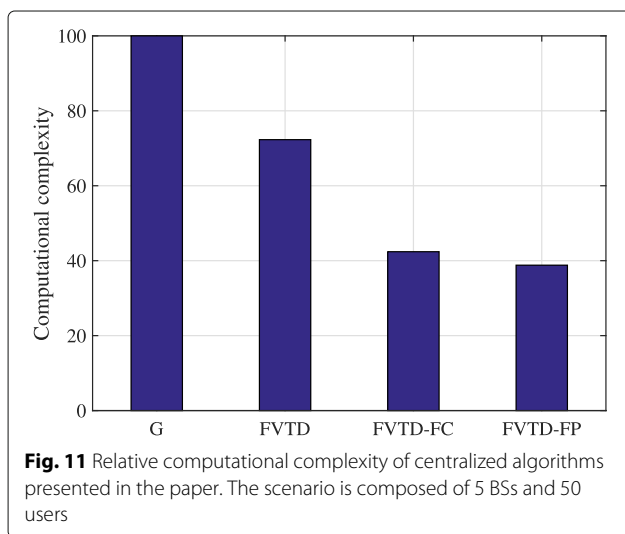


Fig. 9 Evolution of user association in tier 3 and tier 4



of several BSs to achieve load balancing in terms of balanced network throughput for heterogeneous networks. In this scenario, the BSs were solely powered with finite batteries and energy harvesting sources that allowed them to recharge their batteries. An iterative solution based on a primal-dual approach has been derived that considers explicitly the current and the past energy at the batteries. We have also proposed a distributed solution to be solved by each user and the BSs with just local information. Several low-complexity solutions aimed at high mobility scenarios have also been developed. We have compared the proposed strategies with the classical max-SINR approach and showed that improvement in terms of load-balancing is possible if a proper balancing technique is designed and the information of the battery status is considered in the user association procedure.



Endnotes

¹From now on, we will assume that the users that request a fixed service rate are voice users, but this set could indeed contain any type of service requiring fixed service rates.

²If the number of users is relatively high, then $P_{RAD}^i(\tau) \gg p_{ji}(\tau)$, and the approximation is fair. In any case, the approximation provides a lower bound of the actual SINR value.

³It is a worst-case solution as the BSs may not necessarily be transmitting at full power all the time.

⁴The modelling of the user scheduling is out of the scope of this paper. In the simulations section of this paper, we comment on the type of user scheduling used to run the simulations.

⁵The scenario considered has been taken similar to that taken as reference in our previous conference paper [1].

⁶The CDFs are presented instead of the average values of the rates since this allows to evaluate the rates of all the users in the scenario. For example, the CDFs allow to evaluate which is the probability of a user of receiving a very low rate for the different strategies, which could not be done by resorting only to the average values.

⁷The computational complexity shown in these figures has been measured in terms of the 'simulation time' of each algorithm in relative terms with respect to the general solution 'G'. These results have been obtained through simulations running each algorithm in the same server without any other process running simultaneously. The simulation time includes, thus, not only the complexity in terms of mathematical operations, but also the time required to manage the data, the memory, etc. Of course, if a different server or a different way of implementing the optimization procedures were used, the results could change slightly, although the main conclusions would remain the same. Note that an analytical study of the complexity is not possible, or at least extremely complicated, in this case since it should account for the number of iterations needed for convergence, the complexity of the numerical method used to solve the problems, the scenario, etc.

Abbreviations

BS: Base station; CDF: Cumulative distribution function; CDMA: Code division multiple access; CSI Channel state information; DL: Downlink; DP: Dynamic programming; FVTD: First-voice then-data; HetNet: Heterogeneous network; LTE: Long-term evolution; OFDMA: Orthogonal frequency-division multiple

access; PDF: Probability density function; PF: Proportionally fair; SINR: Signal to interference plus noise ratio; SISO: Single-input single-output; WCDMA: Wide-band code division multiple access

Funding

This work has been partially funded by the Spanish Ministry of Economy, Industry and Competitiveness through the projects 5G&B RUNNER-UPC (TEC2016-77148-C2-1-R (AEI/FEDER, UE)), GRE3N-LINK-MAC (TEC2011-29006-C03-02), and the FPI grant BES-2012-052850; the European Commission in the framework of the FP7 Network of Excellence in Wireless COMMunications NEWCOM# (Grant agreement no. 318306) and the project TUCAN3G (Grant agreement no. ICT-2011-601102); and the Catalan Government (AGAUR) through the grant 2014 SGR 60.

Authors' contributions

JR put forward the idea and wrote the manuscript. API, JdO, and JV took part in the discussion and they also guided, reviewed, and checked the writing. JR and JdO carried out experiments. All authors contributed to the interpretation of the results and read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 12 December 2016 Accepted: 13 October 2017

Published online: 02 December 2017

References

- J Rubio, A Pascual-Iserte, J Del Olmo, J Vidal, in *Proc. IEEE Global Communications Conference (GLOBECOM)*. User association for load balancing in heterogeneous networks powered with energy harvesting sources, (2014). doi:10.1109/GLOCOMW.2014.7063604
- A Damnjanovic, et al., A survey on 3GPP heterogeneous networks. *IEEE Wirel. Comm. Mag.* **18**(3), 10–21 (2011)
- X Wu, B Murherjee, D Ghosal, Hierarchical architectures in the third-generation cellular network. *IEEE Wirel. Commun.* **11**(3), 62–71 (2004)
- A Damnjanovic, et al., Multitier cell design. *IEEE Commun. Mag.* **35**(8), 60–64 (1997)
- ICT-601102 STP, in *Deliverable D42*. TUCAN3G: optimization and monitoring of HNB network, (2014). Available online: <http://www.ict-tucan3g.eu>
- K Son, S Chong, G Veciana, Dynamic association for load balancing and interference avoidance in multi-cell networks. *IEEE Trans. Wirel. Commun.* **8**(7), 3566–3576 (2009)
- A Sang, X Wang, M Madhian, RD Gitlin, Dynamic association for load balancing and interference avoidance in multi-cell networks. *Wirel. Netw.* **14**, 103–120 (2008)
- SK Das, SK Sen, R Jayaram, A dynamic load balancing strategy for channel assignment using selective borrowing in cellular mobile environment. *Wirel. Netw.* **3**(5), 333–347 (1997)
- E Yanmaz, OK Tonguz, Dynamic load balancing and sharing performance of integrated wireless networks. *IEEE J. Sel. Areas Comm.* **22**(5), 862–872 (2004)
- R Madan, et al., Cell association and interference coordination in heterogeneous LTE-A cellular networks. *IEEE J. Sel. Areas Comm.* **28**(9), 1479–1489 (2010)
- K Shen, W Yu, in *IEEE International Conference on Acoustics, Speech, and Signal Processing*. Downlink cell association optimization for heterogeneous networks via dual coordinate descents (IEEE, Vancouver, 2013)
- Q Ye, et al., User association for load balancing in heterogeneous cellular networks. *IEEE Trans. Wirel. Commun.* **12**(6), 2706–2716 (2013)
- J-W Lee, RR Mazumbar, NB Shroff, Joint resource allocation and base station assignment for the downlink in CDMA networks. *IEEE/ACM Trans. Netw.* **14**(1), 1–14 (2006)
- M Hong, A Garcia, Mechanisms design for base station association and resource allocation in downlink OFDMA network. *IEEE J. Sel. Areas Comm.* **30**(11), 2238–2250 (2012)
- X Hongzhang, Z Tiankui, Z Zhimin, L Dantong, in *IEEE Wireless Communications and Networking Conference Workshops (WCNCW)*. Distributed user association for delay-load tradeoffs in energy harvesting enabled HetNets (IEEE, New Orleans, 2015)
- L Dantong, et al., in *21st International Conference on Telecommunications (ICT)*. Adaptive user association in HetNets with renewable energy powered base stations (IEEE, Lisbon, 2014)
- L Dantong, et al., Two dimensional optimization on user association and green energy allocation for HetNets with hybrid energy sources. *IEEE Trans. Commun.* **63**, 4111–4124 (2015)
- T Zhang, et al., User association for energy-load tradeoffs in HetNets with renewable energy supply. *IEEE Commun. Lett.* **19**(12), 2214–2217 (2015)
- A Goldsmith, *Wireless communications*. (Cambridge University Press, Cambridge, 2005)
- J Laiho, A Wacker, T Novosad, *Radio Network Planning and Optimisation for UMTS*. (Wiley, Hoboken, 2006)
- C Ho, R Zhang, Optimal energy allocation for wireless communications with energy harvesting constraints. *IEEE Trans. Signal Process.* **60**(9), 4808–4818 (2012)
- J Yang, O Ozel, S Ulukus, Broadcasting with an energy harvesting rechargeable transmitter. *IEEE Trans. Wirel. Commun.* **11**(2), 571–583 (2012)
- M Gregori, A Pascual-Iserte, M Payaró, in *IEEE Wireless Comm. Networking Conference (WCNC)*. Mutual information maximization for a wireless energy harvesting node considering the circuitry power consumption (IEEE, Shanghai, 2013)
- O Ozel, K Tutuncuoglu, J Yang, S Ulukus, A Yener, Transmission with energy harvesting nodes in fading wireless channels: optimal policies. *IEEE J. Sel. Areas Comm.* **29**(8), 1732–1743 (2011)
- DP Bertsekas, *Dynamic programming and optimal control*, third edition. (Athena Scientific, Belmont, 2005)
- X Wang, GB Giannakis, AG Marques, A unified approach to QoS-guaranteed scheduling for channel-adaptive wireless networks. *Proc. IEEE.* **95**(12), 2410–2431 (2007)
- Q Ye, M Al-Shalash, C Caramanis, JG Andrews, in *IEEE Global Communications Conference (GLOBECOM)*. On/Off macrocells and load balancing in heterogeneous cellular networks (IEEE, Atlanta, 2013)
- L Wei, Y Xu, RQ Hu, Y Qian, in *IEEE Global Communications Conference (GLOBECOM)*. An algebraic framework for mobile association in wireless heterogeneous networks (IEEE, Atlanta, 2013)
- J Rubio, A Pascual-Iserte, Energy-aware broadcast MU-MIMO precoder design with imperfect channel and battery knowledge. *IEEE Trans. Wirel. Comm.* **13**(6), 3137–3152 (2014)
- S Boyd, L Vandenbergue, *Convex optimization*. (Cambridge University Press, Cambridge, 2004)
- DP Bertsekas, *Nonlinear programming*, second edition. (Athena Scientific, Belmont, 1999)
- D Palomar, M Chiang, A tutorial on decomposition methods for network utility maximization. *IEEE J. Sel. Areas Comm.* **24**(8), 1439–1451 (2006)
- O Awoniyi, NB Mehta, LJ Greenstein, Characterizing the orthogonality factor in WCDMA downlinks. *IEEE Trans. Wirel. Commun.* **4**(2), 621–625 (2003)