Abstract—Cellular networks are now rapidly evolving to the renewable heterogeneous network (HetNet) architectures, where a mixture of macro base stations (BSs) and small cell BSs coexist. Particularly, these BSs are partly-powered or self-powered by renewable energy sources (RES). In this context, joint user association and resource allocation is challenging. In addition, wireless backhaul is rapidly becoming a reality for small cells, which generally has significantly less bandwidth and larger delay. Thus, the limited backhaul becomes a crucial constraint. In this paper, with the goal of maximizing the network utility, the backhaul-aware user association and resource allocation is jointly considered for energy-constrained HetNets. Based on the complementary slackness property, BSs’ efficient and meaningful resource allocation is achieved, and then a distributed user association algorithm is proposed. Simulation results demonstrate the improvement on network utility and load balancing with the proposed algorithm.

Index Terms—HetNets, renewable energy, backhaul-aware, user association, complementary slackness

I. INTRODUCTION

Nowadays, we have quantitative evidence that the data traffic demand in cellular networks is increasing at an exponential rate, and the wireless data explosion is real and will continue [1]. Moreover, due to the rapidly increasing power consumption of network equipments, which is for handling a much larger amount of traffic, the increasing trend of energy price, and the direct impact of greenhouse gases on the earth environment, it is important to manage cellular networks in an energy-efficient manner and reduce the amount of CO₂ emission levels simultaneously [2, 3].

In this context, cellular networks are now rapidly evolving to the renewable heterogeneous network (HetNet) architectures [4], where a mixture of macro base stations (BSs) and small cell BSs such as micro BSs, pico BSs, and femto BSs coexist, and each with different computational capabilities, power transmission methods, and coverage areas. Moreover, these BSs are partly-powered or self-powered by renewable energy sources (RES). HetNets have been widely demonstrated as an cost-effective approach to increase the network capacity [5]. Since RES could fully sustain the needs of low-power BSs, the HetNets can benefit from its adoption. In return, the advent of HetNets also marks a turning point for its viability of the use in cellular systems [4]. In addition, high-speed wireless backhaul is rapidly becoming a reality for small cells, which eliminates the need for other wired connections and is able to provide or extend coverage in a cost-effective manner [8]. Thus, the “drop and play” small cells are even more attractive.

However, HetNets, with their increased diversity in types and numbers of BSs, reopen the conventional challenges in cellular wireless networks. Among these challenges are cell association, resource allocation, and intercell interference coordination (ICIC) [6]. To address these, the appropriate association of users to different BSs and the rational resource allocation of these BSs are quite important. On one hand, for HetNets, even with a fairly uniform user distribution, the conventional cellular networks’ well applied association schemes can lead to a major load imbalance and some load balancing techniques may become exceedingly improper [9]. On the other hand, due to the intermittency and randomness of available energy, the integration of BSs with RES also adds complexity to the joint user association and resource allocation. Furthermore, while traditional BSs are connected to the core network through a large-capacity/low-delay backhaul link, the wireless backhaul for small cells generally has significantly less bandwidth and larger delay [4]. Thus, the limited backhaul becomes a crucial constraint for HetNets.

So far, load-aware user association for HetNets has been widely studied. In [10], the primary technical approaches to HetNet load balancing have been surveyed and compared. Mostly, these works only focus on user association, while there are works dealing with joint user association and resource allocation. For example, in [11], joint user association and power control problem is considered, and in [12], [13], joint user association and bandwidth (or resource blocks, RBs) allocation algorithms are proposed. It should be noted that, all the above works aim at unique user association, which means each user can be associated to at most one BS. This is the standard model for LTE without the advanced carrier aggregation (CA) feature. Also, in CA scenarios where users can be served by multiple BSs simultaneously, the works in [14–16] consider joint user association and resource allocation. Then, when it comes to HetNets with RES, some studies on user association and/or resource allocation are also presented. Specifically, with hybrid energy supplies, a distributed scheme...
to enable Green-energy Aware and Latency Aware (GALA) user-BS associations is proposed in [17] and a green energy optimization (GEO) problem is formulated and solved in [18]. In addition, for HetNets where BSs are solely powered by RES, a user association strategy to achieve load balancing is given in [19]. In summary, none of these works takes backhaul constraints into consideration. While for backhaul-aware joint user association and resource allocation in [14]–[16], they are for scenarios of CA and HetNets without energy constraints. Recently, a network utility aware (NUA) traffic load balancing scheme in backhaul-constrained small cell networks powered by hybrid energy sources is developed in [20]. However, the resource allocation is neglected.

In this paper, we deal with the backhaul-aware joint user association and resource allocation in hybrid energy powered HetNets, where the energy supplies are constrained. Specifically, unique user association is considered, and a network utility maximization problem in terms of a global network utility function reflecting proportion fairness (PF) is formulated. Based on the complementary slackness property, we obtain BSs’ efficient and meaningful resource allocation, and then propose a distributed user association algorithm. Furthermore, to reduce communication overhead over the air interface and avoid information leaking, we adopt the network function virtualization (NFV) and software-defined networking (SDN) architecture, and present a virtual user association and resource allocation (vUARA) scheme. Finally, simulation results demonstrate the convergence and effectiveness of the proposed algorithm.

The rest of the paper is organized as follows. In Section II, we describe the energy-constrained HetNet under consideration. In Section III, the backhaul-aware joint user association and resource allocation problem is formulated. Section IV presents solution to the formulated problem. The implementation of vUARA scheme is given in Section V. The performance evaluation is shown in Section VI, and concluding remarks are presented in Section VII.

II. SYSTEM MODEL

A. An Energy-constrained HetNet

Considering a HetNet composed of one macro BS and several small cell BSs. Particularly, the HetNet is powered by hybrid energy, i.e., solely traditional electric grid, solely RES, or hybrid electric grid and RES. For the BS with RES, RES collects energy from the environment (by solar panels and/or windmills) and charges the corresponding battery (each BS has its own RES and battery). This is especially important in rural areas, where accessing to the electric grid may be impossible or too expensive. A system model for such HetNet with backhaul layout is given in Fig. 1.

B. Transmission Model

The time duration of one association period depends on the dynamics of renewable energy generation and traffic intensity [18]. Here, we omit the specific definition of association periods, and focus on the $l$th association period, where $l \in \mathcal{L} = \{1, \cdots, L\}$. At time instant $t$ of the $l$th association period, we denote the channel power gain between user $k \in \mathcal{K} = \{1, \cdots, K\}$ and BS $n \in \mathcal{N} = \{1, \cdots, N\}$ by $H_{nk}(t)$, which embodies the effects of path loss, log normal shadowing, and antenna gains. Then, we have

$$H_{nk}(t) = G_{nk}(t) F_{nk}(t), \forall (n,k) \in \mathcal{N} \times \mathcal{K},$$

(1)

where “$\times$” denotes the Cartesian product. $G_{nk}(t)$ is assumed to be constant during one association period, while $F_{nk}(t)$ fluctuates fast enough such that user $k$ can average it out in its channel measurements.

Then, the instantaneous signal-to-interference-plus-noise ratio (SINR) from BS $n$ to user $k$ on an RB is

$$\text{SINR}_{nk}(t) = \frac{P_n(t) H_{nk}(t)}{\sum_{j \neq n} P_j(t) H_{jk}(t) + \sigma^2},$$

(2)

where $P_n(t)$ and $P_j(t)$ are transmission power of BS $n$ and $j$ on an RB, respectively, and $\sigma^2$ denotes the thermal noise spectral power.

Accordingly, the instantaneous achievable rate at user $k$, if it is served by BS $n$, can be written as

$$R_{nk}(t) = B_0 \log_2 (1 + \text{SINR}_{nk}(t)),$$

(3)

where $B_0$ denotes the bandwidth over which an RB is realized.

In general, since the mobility of the users between BSs takes place at longer time scales and the channel may vary during the whole association period, the association decision should be taken considering the average rate that a given user will obtain if it is associated with a certain BS. For an association period, the average rate is

$$R_{nk}(t) = B_0 \log_2 \left(1 + \frac{P_n(t) G_{nk}(t)}{\sum_{j \in \mathcal{N} \setminus n} P_j(t) G_{jk}(t) + \sigma^2}\right).$$

(4)
In addition, the number of users associated with a BS is usually more than one, and the users of the same BS need to share its resource. Thus, the long-term rate achieved by a user depends on the load of the BS and will only be a fraction of the rate $R_{n,k}^{(l)}$. Here, we assume unique association and define the user association indicators $x_{nk}^{(l)}$ as

$$x_{nk}^{(l)} = \begin{cases} 
1, & \text{if user } k \text{ is associated with BS } n, \\
0, & \text{otherwise}.
\end{cases} \quad (5)$$

Denoting the fraction of RBs over which user $k$ is served by BS $n$ as $y_{nk}^{(l)}$, we give the long-term rate $r_{nk}^{(l)}$ as

$$r_{nk}^{(l)} = y_{nk}^{(l)} R_{n,k}^{(l)} W_n, \quad \text{where } W_n \text{ is BS } n's \text{ overall RBs.} \quad (6)$$

### C. Models for Power Consumption and Energy Constraints

Firstly, adopting the model in EARTH project \[21\], the power consumption of BS $n$ in association period $l$ can be expressed as

$$C_n^{(l)} = s^{(l)} \left[ P_{n0} + \delta_n \left( \sum_{k \in K} x_{nk}^{(l)} y_{nk}^{(l)} W_n \right) P_{n1} \right], \quad (6)$$

where $s_{nk}^{(l)}$ is the time duration of the $l$th association period, $P_{n0}$ is the fixed part of BS $n's$ power consumption, $\delta_n$ is the variable power consumption slope of BS $n$, and $\delta_n \left( \sum_{k \in K} x_{nk}^{(l)} y_{nk}^{(l)} W_n \right) P_{n1}$ is the variable part related to BS $n's$ transmission power and consumed resource.

Then, due to RES adopted by some BSs, we analyze their battery dynamics. Specifically, let $B_{n}^{(l)}$ be the renewable energy stored at the battery of BS $n$ at the beginning of association period $l$. Then, at the beginning of association period $l + 1$, the battery level is updated as

$$B_{n}^{(l+1)} = f \left( B_{n}^{(l)}, C_{n,E}^{(l)}, E_{n}^{(l+1)} \right), \quad (7)$$

where $f(\cdot)$ depends upon the battery dynamics, such as storage efficiency and memory effects \[20\]. $C_{n,E}^{(l)}$ is BS $n's$ power consumption from RES in association period $l$, and $E_{n}^{(l+1)}$ is the amount of energy harvested at the beginning of association period $l + 1$. A common practice is to consider the battery update as

$$B_{n}^{(l+1)} = \max \left\{ 0, \min \left( B_{n}^{(l)}, C_{n,E}^{(l)} + E_{n}^{(l+1)} \right) \right\}, \quad (8)$$

where $B_{n,max}$ denotes BS $n's$ battery capacity, the inner minimization accounts for possible battery overflows, and the outer maximization assures the non-negativity of the battery levels. In general, $C_{n,E}^{(l)}$ will be limited by a function of the current battery level, i.e.,

$$C_{n,E}^{(l)} = g_n(B_{n}^{(l)}), \quad (9)$$

where $g_n(\cdot)$ limits the renewable energy that can be used in association period $l$ in order to spend the energy in a more conservative way. Simply, we can consider that only a given fraction of the battery is allowed to be used.

At the same time, we give BS $n's$ limited average electric power, i.e., the electric grid energy can be used freely on the condition that the average power over all the association periods is no longer than the threshold $G_{n,ave}$.

$$\frac{1}{L} \sum_{l \in L} C_{n,G}^{(l)} = \frac{1}{L} \sum_{l \in L} \left[ C_{n}^{(l)} - C_{n,E}^{(l)} \right] \leq G_{n,ave}. \quad (10)$$

Remark 1: In the above description, we didn’t give any notation for BS's specific energy supplies. This is due to the fact that, for BSs without RES, we can set $g_n(B_{n}) = 0$; while for BSs without energy supplies from electric grid, we have $G_{n,ave} = 0$.

### III. Problem Formulation

In practice, only causal information, i.e., information of the past and current channel states and energy harvesting, is available, thus, an online approach is quite desirable. Based on the models of power consumption and energy constraints, the network utility maximization problem can be decomposed into each association period. For simplicity, we omit the superscript $l$ in the following.

Firstly, we assume user $k$ obtains utility $U_k(R_k)$ when its overall rate is $R_k = \sum_{n \in N} x_{nk} y_{nk} R_{n,k} W_n$. To achieve a desired balance between network-wide performance and user fairness, we shall choose a continuously differentiable, monotonically increasing, and strictly concave utility function. Here, we consider the well-known PF, which is imposed by choosing the utility function as

$$U_k(R_k) = \log (R_k). \quad (11)$$

Then, the network utility maximization problem is given as:

$$\text{P1 : } \max_{x,y} \sum_{k \in K} \log \left( \sum_{n \in N} x_{nk} y_{nk} R_{n,k} W_n \right)$$

s.t. \quad C1 : \sum_{n \in N} x_{nk} = 1, \forall k \in K$

C2 : \sum_{k \in K} x_{nk} y_{nk} \leq 1, \forall n \in N$

C3 : \sum_{k \in K} x_{nk} y_{nk} R_{n,k} W_n \leq Z_{n,bh}, \forall n \in N$

C4 : C_{n} \leq g_n(B_{n}) + G_{n,ave}, \forall n \in N$

C5 : x_{nk} \in \{0,1\}, \forall n \in N, \forall k \in K$

C6 : y_{nk} \in [0,1], \forall n \in N, \forall k \in K \quad (12)$$

where C1 denotes that each user can only be served by one BS at a time; C2 specifies each BS's resource constraint; C3 gives backhaul constraints for all BSs, where $Z_{n,bh}$ is BS $n's$ backhaul capacity; C4 implies each BS's energy constraint based on (9) and (10); with each BS’s average electric power transformed into each association period ($C_{n,G} \leq G_{n,ave}$), C5 and C6 present the ranges of $x_{nk}$ and $y_{nk}$, respectively.

Moreover, since $x_{nks}$ take binary values and $\sum_{n \in N} x_{nk} = 1$, we have

$$\sum_{k \in K} \log \left( \sum_{n \in N} x_{nk} y_{nk} R_{n,k} W_n \right) = \sum_{k \in K} \sum_{n \in N} x_{nk} \log (y_{nk} R_{n,k} W_n).$$

Then, P1 can be
re-formulated as

\[
P_2 : \max_{x,y} \sum_{k \in K} \sum_{n \in N} x_{nk} \log (y_{nk} R_{nk} W_n)
\]

\[
s.t. \quad C1\text{-}C6 \text{ in } P1 \tag{13}
\]

Unfortunately, due to binary variables \(x_{nk}\)s and real variables \(y_{nk}\), \(P_2\) is a mixed integer nonlinear programming problem (MINLP), which is generally NP-hard. In the following section, we focus on solving it distributively.

IV. Solution

Obviously, the choices of \(y_{nk}\)s rely on the values of \(x_{nk}\)s. Given these coupled variables, we can apply the Primal Decomposition method to decompose \(P_2\) to the following problems in two levels. Firstly, by fixing variables \(x_{nk}\)s, we have the lower-level problem:

\[
P_3 : \max_{y} \sum_{n \in N} \sum_{k \in K_n} y_{nk} \log (y_{nk} R_{nk} W_n)
\]

\[
s.t. \quad C1 : \sum_{k \in K_n} y_{nk} \leq Q_{n,p}, \forall n \in N
\]

\[
C2 : y_{nk} R_{nk} W_n \leq Z_{n,bh}, \forall n \in N
\]

\[
C3 : 0 \leq y_{nk} \leq 1, \forall k \in K_n, \forall n \in N \tag{14}
\]

where \(K_n\) denotes the set of users that are associated with BS \(n\), and \(|K_n| = \sum_{k \in K} x_{nk}\); \(Q_{n,p} = \min \{1, \frac{1}{x_{nk} W_n R_n \log (x_{nk} W_n R_n)}\}\) denotes the normalized resource that BS \(n\) can accommodate with the constraint from available energy.

Then, when \(y_{nk}\)s are fixed, the higher-level problem (or the master problem) is given by

\[
P_4 : \max_{x} \sum_{k \in K} \sum_{n \in N} x_{nk} \log (y_{nk} R_{nk} W_n)
\]

\[
s.t. \quad C1 : \sum_{n \in N} x_{nk} = 1, \forall k \in K
\]

\[
C2 : x_{nk} \in \{0,1\}, \forall n \in N, \forall k \in K \tag{15}
\]

Since there are no couplings among the subproblems, \(P_3\) can be further decomposed into \(N\) subproblems:

\[
P_5 : \max_{y} \sum_{k \in K_n} \log (y_{nk} R_{nk} W_n)
\]

\[
s.t. \quad C1\text{-}C3 \text{ in } P3 \tag{16}
\]

Defining Lagrange multipliers \(\mu_n\) and \(\nu_n\), the Lagrangian function of \(P_5\) is given as

\[
L = \sum_{k \in K_n} \log (y_{nk} R_{nk} W_n) - \mu_n \left(\sum_{k \in K_n} y_{nk} - Q_{n,p}\right)
\]

\[
- \nu_n \left(\sum_{k \in K_n} y_{nk} R_{nk} W_n - Z_{n,bh}\right).
\]

Applying Karush-Kuhn-Tucker (KKT) conditions, the optimal solution can be obtained as

\[
y_{nk}^* = \frac{1}{\mu_{nk}^* + \nu_{nk}^* R_{nk} W_n}, \forall k \in K_n. \tag{18}
\]

It is easily observed that, \(y_{nk}\) is completely dependent on the dual variables \(\mu_n\) and \(\nu_n\). Intuitively, \(\mu_n\) and \(\nu_n\) can be interpreted as the prices of BS \(n\) determined by the load situation and the backhaul state, respectively. When BS \(n\) is over-loaded (\(\sum_{k \in K_n} y_{nk} \geq Q_{n,p}\)) or its backhaul is over-flowed (\(\sum_{k \in K_n} y_{nk} R_{nk} W_n \geq Z_{n,bh}\)), the corresponding price \(\mu_n\) or \(\nu_n\) goes up, then a user \(k\)'s resource fraction \(y_{nk}\) from associating with it decreases. Otherwise, the prices go down, \(y_{nk}\) increases, and BS \(n\) becomes more attractive.

Generally, this KKT system can be solved by finding the appropriate dual variables \((\mu_n, \nu_n)\). To achieve this, a two-dimensional search or classic constrained optimization techniques such as subgradient method or augmented Lagrangian can be applied. However, both of them are computationally intensive and might not be practical for large size problem. Therefore, we aim to solve this KKT system efficiently.

On one hand, we note that both the constraints \(C1\) and \(C3\) in \(P_3\) are monotonic functions of \(y_{nk}\), and \(y_{nk}\) is also a monotonic function of \(\mu_n\) when \(\nu_n\) is fixed and vice versa. On the other hand, the complementary slackness in constrained optimization states that, for inequality constraints \(f_i(x) \leq 0\) that are tight with equality, the associated dual variables are non-zero. Using this result, at a local optimum, each BS can be either energy-constrained or backhaul-constrained. Thus, we can perform our search on two single dual variables instead of a two-dimensional search.

Specifically, if BS \(n\) is energy-constrained, i.e.,

\[
\begin{align*}
\sum_{k \in K_n} y_{nk} &= Q_{n,p}, \\
\sum_{k \in K_n} y_{nk}^* R_{nk} W_n &\leq Z_{n,bh}, \tag{19}
\end{align*}
\]

then, there are

\[
\begin{align*}
\mu_{nk}^* &\neq 0, \\
\nu_{nk}^* &= 0. \tag{20}
\end{align*}
\]

Substituting (18) and (20) into the first equation of (19), we obtain

\[
y_{nk}^* = \frac{1}{\mu_{nk}^*} = \frac{Q_{n,p}}{|K_n|}, \forall k \in K_n, \tag{21}
\]

which means BS \(n\)'s optimal resource allocation is equal allocation among the associated users.

Else if BS \(n\) is backhaul-constrained, similarly, there is

\[
y_{nk}^* = \frac{1}{\nu_{nk}^* R_{nk} W_n} = \frac{Z_{n,bh}}{|K_n| R_{nk} W_n}, \forall k \in K_n, \tag{22}
\]

which means BS \(n\)'s optimal resource allocation can be obtained by averaging its backhaul capacity.

Substituting (21) and (22) into \(P_4\) respectively, we can obtain the corresponding solutions. Here, we take the resource allocation in (21) for example, and the corresponding optimization problem is:

\[
P_6 : \max_{x} \sum_{k \in K} \sum_{n \in N} x_{nk} \log \left(\frac{Q_{n,p} R_{nk} W_n}{\sum_{k \in K} x_{nk}}\right)
\]

\[
s.t. \quad C1\text{-}C2 \text{ in } P4 \tag{23}
\]

\(P_6\) is still combinatorial due to the binary variables \(x_{nk}\)s. Towards solving it, firstly, a set of auxiliary variables \(\phi_n = (\mu_n, \nu_n)\)
Moreover, since P5 is a convex problem, the local optimum resource allocation as well as the distributed user association can be written as:

$$\phi_n = \arg \max_{\phi_n \in K} \{ \log(Q_{n,p} R_n W_n) - v_n \}$$

where $n_k = \arg \max_n \{ \log(Q_{n,p} R_n W_n) - v_n \}$.

For $I(v)$, the optimum load $\phi_n$ can be calculated for each BS by applying KKT condition, we have

$$\phi_n = \min \{ \exp(v_n - 1), K \}.$$  

This is indeed an algorithm at BS's side.

Note that the dual function $G(v)$ is not differentiable, as $H(v)$ is a piecewise linear function and not differentiable. Therefore, we cannot use the usual gradient methods, instead, we will solve the dual problem using subgradient method.

It is easy to verify that $u_n(v_n) = \phi_n - \sum_{k \in K} x_{nk}$ is a subgradient of dual function $G(v)$ at point $v_n$. Thus, by subgradient method, we obtain the following algorithm for Lagrange multiplier $v_n$:

$$v_n(t+1) = \left[ v_n(t) - \eta_n(t) \left( \phi_n(t) - \sum_{k \in K} x_{nk}(t) \right) \right]^+,$$

where $\eta_n(\tau)$ is a positive scalar stepsize, and $+$ denotes the projection onto the set $\mathbb{R}^+$ of non-negative real numbers.

Similarly, the optimal user association for resource allocation in (22) can be written as:

$$x_{nk} = \begin{cases} 1 & \text{if } n = n_k \\ 0 & \text{otherwise,} \end{cases}$$

where $n_k = \arg \max_n \{ \log(Z_{n,bb}) - v_n \}$, $v_n$ is updated as in (29), and $\phi_n$ has the same expression as in (28).

In summary, by applying the complementary slackness property, we can easily obtain the efficient and meaningful resource allocation as well as the distributed user association. Moreover, since P5 is a convex problem, the local optimum is also the unique global optimum. Furthermore, we have the following proposition.

**Proposition 1:** Based on the overall rates of the associated users, BS $n$ can decide its resource partition:

$$y_{nk}^* \begin{cases} Q_n \frac{|R_n|}{|K|} & \text{if } k \in K, \\ Z_n \frac{R_n}{W_n} & \text{if } k \in K, \end{cases}$$

where $\frac{|K|}{|R_n|}$ is a positive scalar stepsize, and $y_{nk}^*$ is a positive scalar stepsize, and $G(v)$ is not differentiable, as $H(v)$ is a piecewise linear function and not differentiable.

**Proof:** If BS $n$ is energy-constrained, according to (21), we have

$$\sum_{k \in K_n} \frac{Q_n}{|K|} R_n W_n \leq Z_n,$$

and the second inequality in (21), we have

$$\sum_{k \in K_n} R_n W_n \leq Z_n.$$  

Therefore, if $\sum_{k \in K_n} R_n W_n \leq \frac{|K|}{|R|} Z_n$, BS $n$ divides its available resource equally among the associated users so as to maximize the network utility. Otherwise, BS $n$ averages its backhaul capacity.

Lastly, we give the distributed algorithm as in Algorithm 1.

**Algorithm 1** The Distributed Algorithm for Backhaul-aware Joint User Association and Resource Allocation in Energy-constrained HetNets

**Input:** $R_n(0)$, $Q_{n,p}$, and $Z_{n,bb}$, $\forall n \in N$, $\forall k \in K$.  

**repeat**  

Each user $k$ chooses its BS according to:

$$n_k^* = \begin{cases} \arg \max_n \{ \log(R_n Q_{n,p} W_n) - v_n(t) \} & \text{if } (33) \\ \arg \max_n \{ \log(Z_{n,bb}) - v_n(t) \} & \text{otherwise} \end{cases}$$

Each BS $n$ determines its resource partition according to (31), calculates $\phi_n$ and updates $v_n$: $\phi_n(t) = \min \{ \exp(v_n(t) - 1), K \}$, $v_n(t+1) = [v_n(t) - \eta_n(t) (\phi_n(t) - \sum_{k \in K} x_{nk}(t))]^+$, where $\eta_n(t)$ is BS $n$'s stepsize at iteration $t$.

**until** Convergence of Lagrange multiplier $v$

**Output:** User association and resource allocation $(x^*, y^*)$.

**Proposition 2:** Let $v^*$ denote an optimal value of the dual variable. With constant stepsize, Algorithm II is proved to converge statistically to $v^*$, i.e., if for any $\epsilon > 0$ there exists a stepsize $\eta$ such that $\lim_{t \to \infty} \sup_{t \to \infty} \{ G(v(t)) - G(v^*) \} \leq \epsilon$, where $v(t) = \frac{1}{t} \sum_{i=1}^t v(t)$; while for diminishing stepsize, Algorithm II is guaranteed to converge to the optimal value.

**Proof:** Based on results on the convergence of the subgradient method, we can show that, for constant stepsize, the above algorithm is guaranteed to converge to within a neighborhood of the optimal value; while for diminishing stepsize, the algorithm is guaranteed to converge to the optimal value. Here, for convergence analysis, we focus on constant stepsize; while for performance evaluation in Section VI, the convergence with diminishing stepsize is shown.
Given constant stepsize $\eta_n(t) = \eta$, according to the updating of $\upsilon$, we have
\[
\| \upsilon(t+1) - \upsilon^* \|^2_2 = \| \upsilon(t) - \eta_n(t) \upsilon(t) + \upsilon^* \|^2_2 \\
\leq \| \upsilon(t) - \eta_n(t) \upsilon(t) - \upsilon^* \|^2_2 \\
= \| \upsilon(t) - \upsilon^* \|^2_2 - 2\eta_n(t) \upsilon(t) \cdot (\upsilon(t) - \upsilon^*) + \eta_n^2 \| \upsilon(t) \|^2_2 \\
\leq \| \upsilon(t) - \upsilon^* \|^2_2 - 2\eta_n (G(\upsilon(t)) - G(\upsilon^*)) + \eta_n^2 \| \upsilon(t) \|^2_2 ,
\]
(34)
where the last inequality follows from the definition of sub-gradient. Applying the inequalities recursively, we obtain
\[
2 \sum_{t=1}^{T} \eta_n (G(\upsilon(t)) - G(\upsilon^*)) \\
= - \| \upsilon(t+1) - \upsilon^* \|^2_2 + \| \upsilon(1) - \upsilon^* \|^2_2 + \sum_{t=1}^{T} \eta_n^2 \| \upsilon(t) \|^2_2 \\
\leq \| \upsilon(1) - \upsilon^* \|^2_2 + \sum_{t=1}^{T} \eta_n^2 \| \upsilon(t) \|^2_2 .
\]
(35)

From this inequality, we obtain
\[
\frac{1}{t} \sum_{t=1}^{T} (G(\upsilon(t)) - G(\upsilon^*)) \leq \frac{\| \upsilon(1) - \upsilon^* \|^2_2}{2t} + \frac{\sum_{t=1}^{T} \eta_n \| \upsilon(t) \|^2_2}{2t} .
\]
(36)

Since $G(\upsilon)$ is a convex function, by Jensens inequality, we have
\[
G\left(\frac{\upsilon(t)}{t}\right) - G(\upsilon^*) \leq \frac{\| \upsilon(1) - \upsilon^* \|^2_2}{2t} + \frac{\sum_{t=1}^{T} \eta_n \| \upsilon(t) \|^2_2}{2t} .
\]
(37)

When both $\phi_n(l)$ and $\sum_{K=1}^{K} x_{nk}(l)$ are bounded, $\| \upsilon(t) \|^2_2$ is bounded too, i.e., $\sup_l \| \upsilon(t) \|^2_2 \leq c$ , where $c$ is a scalar. Then,
\[
G\left(\frac{\upsilon(t)}{t}\right) - G(\upsilon^*) \leq \frac{\| \upsilon(1) - \upsilon^* \|^2_2}{2t\eta} + \frac{\eta c}{2} .
\]
(38)

Therefore, $\lim \sup_{t \to \infty} G\left(\frac{\upsilon(t)}{t}\right) - G(\upsilon^*) \leq \epsilon$, where $\epsilon = \frac{\eta c}{2}$, i.e., given constant stepsize, the algorithm converges statistically to within $\eta c / 2$ of the optimal value. Besides, if stepsize $\eta$ is small enough, the algorithm converges to the optimal value.

**Remark 2:** In Algorithm 1, on one hand, each user chooses one BS to maximize the utility $\max \log (y_{nk} R_{nk} W_n)$, where $y_{nk}$ is decided by BS $n$’s resource partition based on Proposition 1. On the other hand, the multiplier $\upsilon$ can be regarded as a message between users and BSs, and interpreted as the service cost of BSs decided by the load distribution. When BS $n$ is over-loaded, i.e., $\sum_{k \in K} x_{nk} \geq \phi_n$, its price $\upsilon_n$ increases, otherwise, $\upsilon_n$ goes down to attract more users.

## VI. Performance Evaluation

Nowadays, the most relevant event at network level is the movement of data to the cloud so that it can be accessed from anywhere and via a variety of platforms. As such, NFV and SDN will become paramount. Briefly, NFV enables network functions that are traditionally tied to hardware appliances to run on cloud computing infrastructure in a data center, and in SDN architecture, the control and data planes are decoupled. Based on these concepts, we develop a vUARA scheme.

The implementation of vUARA is shown in Fig. 2. In the first phase of the vUARA scheme, the users and BSs measure and report their downlink SINRs, available energy, and backhaul capacity to the radio access networks controller (RANC). In the second phase, the RANC generates virtual users and virtual BSs (vBSs) by leveraging cloud computing and virtualization, and the virtual users and vBSs conduct the iterative user association and resource allocation adjustments. When the iterations converge, the optimal solution is obtained. Finally, in the third phase, the RANC informs BSs the decision. This procedure is shown in Fig. 2.

![Fig. 2. The implementation of vUARA](image)

In summary, firstly, it is the RANC that jointly optimizes the user association and resource allocation based on the measurements. Secondly, instead of exchanging information over the air interface, the virtual users and vBSs can iteratively update their BS selections and measurements via a wired link, e.g., a message bus. Lastly, with vUARA, the communication overhead over the air interface is significantly reduced, and the information leaking is avoided.

### V. IMPLEMENTATION

In this section, we evaluate the performance of the proposed algorithm. Without loss of generality, we consider a 3-tier HetNet with macro BS, micro BSs, and femto BSs. The location of macro BS is modeled to be fixed and it forms a conventional cellular structure, while other BSs and two kinds of users are randomly located in an orthohexagonal area with side length 500m. The simulation topology is given in Fig. 3, and the basic parameters are presented in Table III.
The convergence of the proposed algorithm with diminishing stepsize is shown in Fig. 4. With different number of random users, Fig. 5 presents the network utility and the percentage of users associated with macro BS. For max-SINR association, we adopt equal resource allocation and maximal achievable rate first (MARF) scheduling [9] that selects users in descending order of achievable rates with backhaul constraints. It is shown that, the proposed algorithm significantly improves the network utility and the load balancing.

![Fig. 3. The simulation topology](image)

**TABLE I**

| Parameter       | macro BS | micro BS | femto BS |
|-----------------|----------|----------|----------|
| Number          | 1        | 4        | 10       |
| Hotspot users   | 25       | 10       | 5        |
| $P_{t,0} (W)$   | 130      | 56       | 4.8      |
| $\delta_n$     | 4.7      | 2.6      | 8.0      |
| Output power(dBm)| 46       | 35       | 20       |
| Path loss(dB)   | $34 + 40 \log_{10}(d)$ | $37 + 30 \log_{10}(d)$ |
| $W_n$           | 500      | 100      | 50       |
| Available Energy(J/s) | 300 | 65       | 5.6      |
| $Z_{n,\text{ln}}$(Mbps) | 1000 | 100      | 10       |
| $\sigma^2$(dBm) | -111.45  |          |          |

The performance. Compared with max-SINR association, the proposed algorithm shows improvement on users’ overall data rates and fairness. Then, with either reduced backhaul capacity or available energy of BSs, users’ data rates decrease. It should be noted that, for max-SINR association with equal resource allocation and MARF scheduling, some users may be dropped because their BSs are running out of backhaul.

![Fig. 4. Lagrange multiplier $\nu$ and network utility versus iterations ($K_{\text{rand}} = 100$)](image)

![Fig. 5. Effects of different number of random users](image)

In addition, with the cumulative distribution of data rates over all users given in Fig. 6, we illustrate the effects of decreasing available energy or backhaul capacity on system performance. Compared with max-SINR association, the proposed algorithm shows improvement on users’ overall data rates and fairness. Then, with either reduced backhaul capacity or available energy of BSs, users’ data rates decrease. It should be noted that, for max-SINR association with equal resource allocation and MARF scheduling, some users may be dropped because their BSs are running out of backhaul.

![Fig. 6. Cumulative distribution of data rates over all users ($K_{\text{rand}} = 100$)](image)

**VII. CONCLUSIONS**

We focus on backhaul-aware joint user association and resource allocation for energy-constrained HetNets. Unlike most of the previous works on backhaul-aware, we consider unique user association together with resource allocation with energy constraints. Based on the complementary slackness property, BSs’ two kinds of resource partition are efficiently obtained, and a distributed user association algorithm is then proposed. Moreover, we develop a vUARA scheme, which significantly reduces the communication overhead on air interface and avoids the information leaking. Simulation results show the proposed algorithm’s convergence and effectiveness on improving the network utility and the load balancing.

**REFERENCES**

[1] J. G. Andrews, S. Buzzi, W. Choi, S. Hanly, A. Lozano, A. C. Soong, and J. C. Zhang, “What will 5G be?” IEEE J. Sel. Areas Commun., vol. 32, no. 6, pp. 1065–1082, June 2014.

[2] Z. Hasan, H. Boostanimehr, and V. K. Bhargava, “Green cellular networks: A survey, some research issues and challenges,” IEEE Commun. Surveys and Tutorials, vol. 13, no. 4, pp. 524–540, Fourth Quarter 2011.

[3] Y. S. Soh, T. Q. S. Quek, M. Kountouris, and H. Shin, “Energy efficient heterogeneous cellular networks,” IEEE J. Sel. Areas Commun., vol. 31, no. 5, pp. 840–850, May 2013.

[4] G. Piro, M. Miozzo, G. Forte, N. Baldo, L. Grieco, G. Boggia, and P. Dini, “HetNets powered by renewable energy sources: Sustainable next-generation cellular networks,” IEEE Trans. Internet Computing, vol. 17, no. 1, pp. 32–39, Jan. 2013.

[5] J. G. Andrews, “Seven ways that HetNets are a cellular paradigm shift,” IEEE Commun. Mag., vol. 51, no. 3, pp. 136–144, Mar. 2013.

[6] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, “A survey on 3GPP heterogeneous networks,” IEEE Trans. Wireless Commun., vol. 18, no. 3, pp. 10–21, June 2011.

[7] H. S. Dhillon, Y. Li, P. Nuggehalli, Z. Pi, and J. G. Andrews, “Fundamentals of heterogeneous cellular networks with energy harvesting,” IEEE Trans. Wireless Commun., vol. 13, no. 5, pp. 2782–2797, May 2014.

[8] S. Hur, T. Kim, D. J. Love, J. V. Krogmeier, T. A. Thomas, and A. Ghosh, “Millimeter wave beamforming for wireless backhaul and access in small cell networks,” IEEE Trans. on Commun., vol. 61, no. 10, pp. 4391–4403, Oct. 2013.

[9] T. Zhou, Y. Huang, and L. Yang, “QoS-aware user association for load balancing in heterogeneous cellular networks,” in Proc. IEEE VTC Fall, Sep. 2014, pp. 1–5.
[10] J. Andrews, S. Singh, Q. Ye, X. Lin, and H. Dhillon, “An overview of
load balancing in HetNets: old myths and open problems,” IEEE Trans. Wireless Commun., vol. 21, no. 2, pp. 18–25, Apr. 2014.
[11] K. Shen and W. Yu, “Distributed pricing-based user association for
downlink heterogeneous cellular networks,” IEEE J. Sel. Areas Commun., vol. 32, no. 6, pp. 1100–1113, June 2014.
[12] H. Boostanimehr and V. K. Bhargava, “Unified and distributed QoS-
driven cell association algorithms in heterogeneous networks,” IEEE Trans. Wireless Commun., vol. 14, no. 3, pp. 1650–1662, Jan. 2015.
[13] N. Wang, E. Hossain, and V. K. Bhargava, “Joint downlink cell associ-
ation and bandwidth allocation for wireless backhauling in two-tier HetNets with large-scale antenna arrays,” [arXiv:1501.00078v1].
[14] C. Kim, R. Ford, Y. Qi, and S. Rangan, “Joint interference and user association optimization in cellular wireless networks,” CoRR abs/1304.3977, 2013.
[15] C. Zhe and R. Adve, “Joint user association and resource allocation
in small cell networks with backhaul constraints,” in Proc. CISS, Mar. 2014, pp. 1–6.
[16] W.-C. Liao, M. Hong, H. Farmanba, X. Li, Z.-Q. Luo, and H. Zhang,
“Min flow rate maximization for software defined radio access net-
works,” IEEE J. Sel. Areas in Commun., vol. 32, no. 6, pp. 1282–1294, June 2014.
[17] T. Han and N. Ansari, “Green-energy aware and latency aware user associations in heterogeneous cellular networks,” in Proc. IEEE GLOBE-
COM, Dec. 2013, pp. 4946–4951.
[18] T. Han and N. Ansari, “On optimizing green energy utilization for
cellular networks with hybrid energy supplies,” IEEE Trans. Wireless Commun., vol. 12, no. 8, pp. 3872–3882, Aug. 2013.
[19] J. Rubio, A. Pascual Iserte, J. Del Olmo, and J. Vidal, “User association for load balancing in heterogeneous networks powered with energy harvesting sources,” in Proc. IEEE GLOBECOM, Dec. 2014, pp. 1227–1232.
[20] T. Han and N. Ansari, “Network utility aware traffic loading balancing in backhaul-constrained cache-enabled small cell networks with hybrid power supplies,” CoRR abs/1409.8267, 2014.
[21] G. Auer, V. Giannini, C. Desset, I. Godor, P. Skillermark, M. Olsson,
M. A. Imran, D. Sabelle, M. J. Gonzalez, O. Blume, and A. Fehske,
“How much energy is needed to run a wireless network?” IEEE Trans. Wireless Commun., vol. 18, no. 5, pp. 40–49, Oct. 2011.