IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received May 1, 2019, accepted May 26, 2019, date of publication May 31, 2019, date of current version August 15, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2920149

Load-Aware Energy Efficiency With Unequal User Priority in Downlink Heterogeneous Network System

RUI LI, NING CAO, MINGHE MAO, AND YIFAN HU
School of Computer and Information, Hohai University, Nanjing 210098, China

Corresponding author: Ning Cao (caoning@vip.163.com)

This work was supported in part by the National Natural Science Youth Foundation of China under Grant 61701167, in part by the Fundamental Research Funds for the Central Universities under Grant 2019B00814, and in part by the National Natural Science Foundation of China under Grant 41830110.

ABSTRACT This paper considers a load-aware energy efficiency maximization problem for downlink heterogeneous cellular networks with attention to unequal user priorities, load balance, quality of service (QoS) requirements, and power control. The load-aware energy efficiency of the networks is defined as the ratio of the effective rate and power consumption and mathematically formulated as a mixed-integer optimization problem in fractional forms with multi-constraints. A two-layer iterative algorithm with low computational complexity is proposed to address this optimization problem. In the outer layer, the objective function is converted into an equivalent problem in subtractive form by using the Dinkelbach method. The joint resource allocation involves cell-user association and power control that are decoupled over the independent portions of the system. In the inner layer, we search the association indices and transmit power parameters via solving a class of convex optimization problems governed by constraints. Specifically, by introducing some new auxiliary variables and using the Lagrangian dual method, the closed-form optimal association strategy and power control solution are obtained. The numerical simulations show that the proposed algorithm is guaranteed to converge and gives rise to higher energy efficiency than the existing one.

INDEX TERMS Heterogeneous networks, association rules, load management, energy efficiency.

I. INTRODUCTION

The global number of wireless subscribers is expected to exceed 6.5 billion by 2030 - an average of over a million new subscribers per day [1]. All the benefits that come from the rapid development of global wireless communication technology. Considering that thousands and millions of devices are connected, wireless cellular networks are required to effectively utilize the precious radio spectrum resources and existing infrastructures so as to support reliable and powerful communication service. As such, cellular networks are evolved from the static one-layer deployment into the dynamic multi-layer deployment [2]. A typical one is known as Heterogenous Networks (HetNets), which is composed of macrocells, picocells, or/and femtocells to provide the seamless wireless connection over the whole areas from an opening environment to office buildings, small rooms, even underground space. Low-power flexible base stations (BS) both complement and extend the tower-mounted macro cellular networks' coverage. Such a utilization of small cell coupled with conventional macro BSs supports a plethora of different types of mobile services and potentially enhances the overall capacity of cellular networks. Compared to the traditional single-tier cellular networks, the advanced network architecture, HetNets, is likely to become one of the most promising technologies in the field of wireless communication.

Furthermore, dense small cells are deployed, which substantially shortens the distance between the BS and terminal and thus emigrates the path-loss effect. On the other hand, in such an architecture, low-power cells move the burden of transmitting messages from the receiver to the macro cell base station as well as further decrease the total energy expenditure. Distinct from the conventional homogeneous

The associate editor coordinating the review of this manuscript and approving it for publication was Xiaofei Wang.

networks, it is possible for users to be covered by multiple cells, therefore, effective association strategies for load balance and power control schemes for performance enhancement are still significant and meaningful topics in HetNets [3].

In wireless communication systems, a user is usually allocated to BSs according to the max-SINR, min-pass loss or min-interference criterions. However, these traditional methods are unsuitable for HetNets since the transmit power variance of macro and low-power may naturally lead to imbalanced traffic distribution and insufficient resource utilization. To improve the rate of usage of BS, [4]–[6] investigated a user-cell allocation method with load-aware, formulated as the aggregate logarithmic utility maximization problem for downlink HetNets and proved that the user association rule had a large impact on performance and the experience of users. In addition, some other utility functions are studied in [7]–[9]. These studies showed that load-aware user association scheme achieves a better tradeoff between the load balance, resource utilization and performance. The mentioned researches have typically been viewed as maximizing the sum effective data rate of users without power control, where the transmit power of BSs is at the maximum level. From a practical perspective, meeting the required rate of mobile devices is sometimes more important than maximizing the system sum-rate. The concept of optimal power strategy has thus emerged and already has become a main approach for capacity maximization, energy saving and resource management in green communications [10]. Besides, power allocation is always associated with the energy consumption, inter-cell and intra-cell interference control and terminals' requirements of the networks.

User association and power control are two dominant themes of resource allocation in wireless communication systems. Joint analysis and optimization of user association and power allocation provide a more comprehensive view of the system performance improvement. [11]–[13] studied the interplay of user association and interference control which incorporated the total power consumption of the system in the throughput optimization. In [14]–[17], energy management policies were devised in downlink HetNets system where it adopted the ratio of effective rate and aggregate power consumption as the objective function. To further handle the uneven distribution of traffic load, a novel energy-efficient user association scheme was investigated in [18] to reduce the amount of system transmission power while satisfying the traffic demand of access points. Reference [19] proposed a duality algorithm for the load-aware energy efficiency problem with quality of service (QoS) constraints by relaxing the logarithmic utility function as a concave one and achieved a lower bound of the problem via an approximate simplification. Reference [20] considered an achieved utility-based energy efficiency problem, which was formulated as the joint optimization of association and power control. The non-convex optimization problem was solved by an iterative algorithm with a near-optimal performance.

In most realistic scenarios, power control is always subjected to at least one power constraint, e.g. the maximum power constraints, the total power constraints, cross-tier interference constraints. In addition, the user priority and QoS constraints among terminal represent its unique characteristics which could be further accord with the practical situation. Priority-based optimization has been widely discussed in many scenarios of wireless communications [21], [22]. In [23], it proposed a distributed belief propagation (BP) algorithm to maximize the unequal priority user association problem in HetNets. However, load-aware energy efficiency for downlink HetNets, jointly considering load balance, power control, user priority, and QoS requirements, has not been studied in previous work.

In this paper, we put forward a load-aware energy efficiency maximization framework for downlink HetNets system. The objective function is formulated as the ratio of logarithmic utility and overall power consumption based on the considerations of the load balance, unequal user priority and QoS requirements. The proposed model achieves a tradeoff between throughput and power consumption. Mathematically, it is a mixed-integer programming problem with polynomial complexity and is very difficult to directly obtain its optimal solutions. We solve these problems by a two-layer iteration algorithm which comprised of Dinkelbach's algorithm, Lagrangian dual decomposition and gradient descent method. We first transform the original function into a parametric non-fractional form and then search the energy efficiency parameter via Dinkelbach's method. The solution involves user association and power allocation that are decoupled over the independent portions of the networks. Then, we alternatively optimize the problems, e.g. the weighted logarithm utility maximization and power consummation minimization, under the fixed energy efficiency. The two problems can be relaxed as the convex optimization by introducing the auxiliary variables and solved via Lagrangian dual decomposition.

The rest of this paper is organized as follows. Section II describes the system model and formulates the average utility-based energy efficiency (UEE) function. Section III details a two-layer iteration algorithm for the maximization of the UEE under multi-constrain condition. The complexity analysis is validated at the end of section. Simulations are presented in Section IV. Finally, Section V concludes the paper.

II. SYSTEM MODEL AND PROBLEM FORMULATION

In this paper, we focus on a two-layer downlink HetNet where macro BSs (MBSs) and pico BSs (PBSs) serve all the users within the cover scope of a cell. Let $\mathcal{I} = \mathcal{I}_m \cup \mathcal{I}_p$ be the set of all BSs, where $\mathcal{I}_m$ is the set of MBSs and $\mathcal{I}_p$ is the set of PBSs, respectively. $\mathcal{J}$ is the set of users. Given that the inter-cell and inner-cell interference result from frequency reuse, we consider the average signal-to-interference-plus-noise ratio (SINR) from BS i to user j over long-term as

$$\mathrm{SINR}_{ij} = \frac{p_i g_{ij}}{\sum_{k\in\mathcal{I}\setminus\{i\}} p_k g_{kj} + \sigma^2}, \tag{1}$$

where p_i is the transmit power of BS i and σ^2 is the power of additive white Gaussian noise (AWGN). g_{ij} represents the channel gain from BS i to user j. Then, the spectral efficiency on terminal j is

$$\gamma_{ij} = log(1 + \mathrm{SINR}_{ij}). \tag{2}$$

Furthermore, user priority is taken into consideration. In practice, it is common that users are deployed independently and have different characteristics. BS devotes some portion of its resource to the associated users according to proportional fairness. We denote φ_{ij} as the optimal transmission time assigned to user j by BS i. Mathematically, it is given by

$$\varphi_{ij} = \frac{w_j x_{ij}}{\sum_{k\in\mathcal{J}} w_k x_{ik}}, \tag{3}$$

where w_j is the weight factor of the user j and the binary variable x_{ij} is the user association indicator. Further details see [24, Th. 3]. In this paper, we assume that each user has an exclusive association with a given BS. When user j associates with the BS i, $x_{ij} = 1$, otherwise, it equals 0. For specific, the optimal resource allocation is equal when users have the same priority [4]. Thus, the rate that user j achieves can be expressed as

$$R_j = \sum_{i\in\mathcal{I}} B x_{ij}\varphi_{ij}\gamma_{ij}, \tag{4}$$

where B is the bandwidth of BS i.

III. OPTIMIZATION FRAMEWORK

In this section, we devise an energy-efficient resource allocation scheme to augment the performance of downlink HetNets system. Combined with the balanced load, we define the overall energy efficiency as the ratio of the sum utility rate to total power consumption, which is formulated as

$$\max_{\boldsymbol{x},\boldsymbol{p}} \; F(\boldsymbol{x},\boldsymbol{p}) = \frac{\sum_{j\in\mathcal{J}} U(R_j)}{\sum_{i\in\mathcal{I}} \rho_i p_i + Pc} \tag{5a}$$

$$s.t. \; \sum_{i\in\mathcal{I}} x_{ij} \le 1, \quad \forall j \in \mathcal{J} \tag{5b}$$

$$x_{ij} \in \{0, 1\}, \quad \forall i \in \mathcal{I}, j \in \mathcal{J} \tag{5c}$$

$$\sum_{i\in\mathcal{I}} x_{ij}\mathrm{SINR}_{ij} \ge \tau_j, \quad \forall j \in \mathcal{J} \tag{5d}$$

$$0 \le p_i \le p_i^{\max}, \quad \forall i \in \mathcal{I} \tag{5e}$$

where ρ_i and $p_i^{\max}$ are the transmit power gain and maximum transmit power of BS i, respectively. Pc is the static operational power consumed by electronic circuits and $U(\cdot)$ is the utility function which is used to retain some degree of fairness. τ_j is the minimal SINR that user j should meet. The above constraints are set up as follows: (5b) and (5c) ensure that no two or more BSs are allocated to same user in a time slot. Constraints (5d) and (5e) are the minimum QoS requirement of each user and the allowable power constraint of each BS, separately. Different QoS constraints and power constraints effect the performance in load balance, the amount of consumed resource and energy efficiency.

In this paper, we utilize the logarithmic utility function $U(x) = log(x)$ for the user fairness which is concave and monotonic decreasing [25], [26]. By introducing the utility function, we have the following simplification

$$\begin{aligned}\sum_{j\in\mathcal{J}} U(R_j) &= \sum_{j\in\mathcal{J}} w_j \log(\sum_{i\in\mathcal{I}} B x_{ij}\varphi_{ij}\gamma_{ij}) \\ &= \sum_{j\in\mathcal{J}} w_j \log(\sum_{i\in\mathcal{I}} B x_{ij}\gamma_{ij} \frac{w_j x_{ij}}{\sum_{k\in\mathcal{J}} w_k x_{ik}}) \\ &= \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} w_j x_{ij} \log(B\gamma_{ij} \frac{w_j}{\sum_{k\in\mathcal{J}} w_k x_{ik}}).\end{aligned} \tag{6}$$

Applied (6) to (5a), a compromised model in improving the system throughput and load balance is formulated, and the transmit power consumption can be controlled as the sufficiently small in the solution. However, this model is a typical fractional programming problem which results in primal function (5a) is hard to be tackled. We first transform the fractional problem into a linear programming via non-linear variable transformation [27]. With an auxiliary variable, problem (5a) is formed as the subtractive one

$$\begin{aligned}F(\zeta) = \max_{\boldsymbol{x},\boldsymbol{p}} &\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} w_j x_{ij} \log(B\gamma_{ij} \frac{w_j}{\sum_{k\in\mathcal{J}} w_k x_{ik}}) \\ &- \zeta(\sum_{i\in\mathcal{I}} \rho_i p_i + Pc) \\ &s.t. \; (5b), (5c), (5d), (5e).\end{aligned} \tag{7}$$

The problem (7) is equivalent to (5a), if and only if when $F(\zeta^*) = 0$. In this way, the solution $\{\zeta^*, x^*, p^*\}$ obtained from (7) is also optimal for problem (5a), where ζ^* is the optimal energy efficiency. The efficient Dinkelbach method is imposed to find the solution of $F(\zeta) = 0$ and the convergence has been proved in theory [27].

Since the optimization variables of (7) are still in a coupling form, the joint optimization can be broken down into sequential problems and preserve the optimality. Thus, we first handle the user-cell association problem with fixed $\boldsymbol{p}$, and then solve the power allocation problem with fixed $\boldsymbol{x}$. The optimal value would be achieved by means of the alternative optimization. The detailed process is presented in **Algorithm 1** where T_1 is the maximal number of iteration.

Algorithm 1 Two-layer iteration algorithm for UEE

Require:
The initialization of energy efficiency, ζ,
The set of maximal iteration, T_1.
Ensure:
The optimal energy efficiency ζ^*.
repeat
Updata $\boldsymbol{x}$ according to **Algorithm** 2;
Updata $\boldsymbol{p}$ according to **Algorithm** 3;
$t_1 = t_1 + 1$.
until Convergence $F(\zeta)$ or $t_1 = T_1$.

A. LOAD-AWARE USER ASSOCAITION FOR BS

Given any fixed $\boldsymbol{p}$ and ζ, the original problem is transformed into a logarithmic utility maximization problem, as

$$\max_{\boldsymbol{x}} \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} w_j x_{ij}\log(Bw_j\gamma_{ij}\varphi_{ij}) \tag{8a}$$

$$s.t. \sum_{i\in\mathcal{I}} x_{ij} \leq 1, \quad \forall j\in\mathcal{J} \tag{8b}$$

$$\sum_{i\in\mathcal{I}} x_{ij}\mathrm{SINR}_{ij} \geq \tau_j, \quad \forall j\in\mathcal{J} \tag{8c}$$

$$x_{ij}\in\{0,1\}, \quad \forall i\in\mathcal{I}, j\in\mathcal{J} \tag{8d}$$

The problem in (8) is a NP-hard combinatorial optimization problem due to the binary constraint (8d). For tractability, we first relax the binary variable as a continuous one and simplify the complicated problem (8a) by introducing an auxiliary variable $\boldsymbol{y}$. It can be then developed as follows

$$\max_{\boldsymbol{x},\boldsymbol{y}} \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} w_j x_{ij}\log(Bw_j\gamma_{ij}) - \sum_{i\in\mathcal{I}} y_i\log y_i \tag{9a}$$

$$s.t. \sum_{i\in\mathcal{I}} x_{ij} \leq 1, \quad \forall j\in\mathcal{J} \tag{9b}$$

$$\sum_{i\in\mathcal{I}} x_{ij}\mathrm{SINR}_{ij} \geq \tau_j, \quad \forall j\in\mathcal{J} \tag{9c}$$

$$\sum_{k\in\mathcal{J}} w_k x_{ik} = y_i, \quad \forall i\in\mathcal{I} \tag{9d}$$

$$0\leq x_{ij}\leq 1, \quad \forall i\in\mathcal{I}, j\in\mathcal{J} \tag{9e}$$

It is hard to directly solve the problem due to the coupling constraint (9d). We first decouple it by utilizing the Lagrangian dual decomposition method and the corresponding Lagrange function is given by

$$\begin{aligned}\mathcal{L}(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda},\boldsymbol{\mu}) = &\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} w_j x_{ij}\log(Bw_j\gamma_{ij}) - \sum_{i\in\mathcal{I}} y_i\log y_i \\ &+\sum_{i\in\mathcal{I}}\mu_i(y_i - \sum_{j\in\mathcal{J}} w_j x_{ij}) \\ &+\sum_{j\in\mathcal{J}}\lambda_j(\sum_{i\in\mathcal{I}} x_{ij}\mathrm{SINR}_{ij} - \tau_j),\end{aligned} \tag{10}$$

where $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ are the Lagrangian multipliers correspond to constraint (9c) and (9d). More specifically, the dual form of the primal problem (9a) can be written as

$$\min_{\boldsymbol{\lambda},\boldsymbol{\mu}}\max_{\boldsymbol{x},\boldsymbol{p}} L(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\lambda},\boldsymbol{\mu}) = \min_{\boldsymbol{\lambda},\boldsymbol{\mu}}(G_1(\boldsymbol{\lambda},\boldsymbol{\mu}) + G_2(\boldsymbol{\lambda},\boldsymbol{\mu})), \tag{11}$$

where

$$G_1(\boldsymbol{\lambda},\boldsymbol{\mu}) = \max_{\boldsymbol{x}} \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} x_{ij}u_{ij}, \tag{12}$$

and

$$G_2(\boldsymbol{\lambda},\boldsymbol{\mu}) = \max_{\boldsymbol{y}} \sum_{i\in\mathcal{I}} y_i(\mu_i - \log y_i). \tag{13}$$

For simplification, we denote $u_{ij} = w_j\log(Bw_j\gamma_{ij}) - \mu_i w_j + \lambda_j\mathrm{SINR}_{ij}$. Then, we can obtain the optimal solution of $\{\boldsymbol{x},\boldsymbol{y}\}$ via taking the derivative of convex function (12) and (13), separately. Since the user association index is a binary value, the optimal user selection is obtained by

$$x_{ij} = \begin{cases} 1 & i = i^* \\ 0 & \text{otherwise} \end{cases}$$

where

$$i^* = \arg\max_{i\in\mathcal{I}} u_{ij}, \forall j\in\mathcal{J} \tag{14}$$

and we have the optimum resource occupancy in t-th iteration

$$y_i^t = \exp(\mu_i^t - 1). \tag{15}$$

Further, the Lagrange multiplier $\boldsymbol{\lambda}$ and $\boldsymbol{\mu}$ can be updated by the following formula:

$$\lambda_j = \max\{0, \lambda_j^t - \delta^t(\sum_{i\in\mathcal{I}}\mathrm{SINR}_{ij} - \tau_j)\}, \tag{16}$$

and

$$\mu_i = \mu_i^t - \delta^t(y_i^t - \sum_{j\in\mathcal{J}} w_j x_{ij}), \tag{17}$$

where δ^t is the t-th step length for the possible solution. The objective value increases with optimal solution found in each iteration and converges to a unique value [19]. The proposed algorithm is detailed in **Algorithm** 2 and T_2 represents the maximal number of iterations.

Algorithm 2 Load-aware based association algorithm for UEE

Require:
The initial transmit power of BSs, $\boldsymbol{p}$,
The set of maximal iteration, T_2,
The initialization of auxiliary dual, $\boldsymbol{\lambda}$.
Ensure:
The optimal allocation index $\boldsymbol{x}^*$.
repeat
Allocate some BS to each user by applying (14);
Calculate y_i according to (15);
Update λ_j and μ_i by applying (16) and (17);
$t_2 = t_2 + 1$.
until Convergence or $t_2 = T_2$.

B. TRANSMIT POWER EFFICIENCY FOR BS

Given a fixed user association $\boldsymbol{x}$, we further consider the optimum transmit power of each BS to meet the QoS requirement and suppress inter-cell interference. The objective function towards $\boldsymbol{p}$ can be expressed as

$$\max_{\boldsymbol{p}} \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} x_{ij}w_j \log(\log(1+\frac{p_i g_{ij}}{\sum\limits_{k\in\mathcal{I}\backslash\{i\}} p_k g_{kj}+\sigma^2})) - \zeta\sum_{i\in\mathcal{I}} p_i\rho_i$$

$$s.t.\ 0\le p_i\le p_i^{\max},\quad \forall i\in\mathcal{I} \tag{18a}$$

$$\mathrm{SINR}_j\ge\tau_j,\quad \forall j\in\mathcal{J} \tag{18b}$$

where $\mathrm{SINR}_j = \{\sum_i x_{ij}\mathrm{SINR}_{ij}, \forall j\in\mathcal{J}\}$. The power allocation problem (18a) has high computational complexity because of the non-convexity of the objective in terms of $\boldsymbol{p}$ [28]. We first introduce an auxiliary variable ν, and thus we have

$$\max_{\boldsymbol{p},\boldsymbol{\nu}} \sum_{j\in\mathcal{J}} w_j\log(\log(1+\nu_j)) - \zeta\sum_{i\in\mathcal{I}} p_i\rho_i \tag{19a}$$

$$s.t.\ 0\le p_i\le p_i^{\max},\quad \forall i\in\mathcal{I} \tag{19b}$$

$$\mathrm{SINR}_j\ge\nu_j,\quad \forall j\in\mathcal{J} \tag{19c}$$

$$\mathrm{SINR}_j\ge\tau_j,\quad \forall j\in\mathcal{J} \tag{19d}$$

Up to now, the objective function is still concave because of the nonlinear constraint (19c) and (19d). Then, we deal with the problem by transforming it into an equivalent convex one via the following exponential variable transformation. Let $\nu = e^{\theta}$ and $p = e^{q}$, and thus (19a) can be further simplified as

$$\max_{\boldsymbol{q},\boldsymbol{\theta}} \sum_{j\in\mathcal{J}} w_j\log(\log(1+e^{\theta_j})) - \zeta\sum_{i\in\mathcal{I}} e^{q_i}\rho_i \tag{20a}$$

$$s.t.\ 0\le q_i\le \log(p_i^{\max}),\quad \forall i\in\mathcal{I} \tag{20b}$$

$$\log(\frac{e^{q_i}g_{ij}}{\sum\limits_{k\neq i} e^{q_k}g_{kj}+\sigma^2})\ge\theta_j,\quad \forall i\in\mathcal{I}, j\in U_i \tag{20c}$$

$$\log(\frac{e^{q_i}g_{ij}}{\sum\limits_{k\neq i} e^{q_k}g_{kj}+\sigma^2})\ge\log(\tau_j),\quad \forall i\in\mathcal{I}, j\in U_i \tag{20d}$$

where $U_i = \{x_{ij}=1 \mid j\in\mathcal{J}\}$ denotes the set of users associated with BS i. Instead of using the interior point method, we adopt the dual method to realize an optimal solution. The Lagrange dual function of (20a) can be expressed as

$$\begin{aligned}\mathcal{L}(\boldsymbol{q},\boldsymbol{\theta},\boldsymbol{\alpha},\boldsymbol{\beta}) &= \sum_{j\in\mathcal{J}} w_j\log(\log(1+e^{\theta_j})) - \zeta\sum_{i\in\mathcal{I}} e^{q_i}\rho_i \\ &\quad +\sum_{i\in\mathcal{I}}\sum_{j\in U_i}\alpha_j(\log(\frac{e^{q_i}g_{ij}}{\sum\limits_{k\neq i} e^{q_k}g_{kj}+\sigma^2}) - \log(\tau_j)) \\ &\quad +\sum_{i\in\mathcal{I}}\sum_{j\in U_i}\beta_j(\log(\frac{e^{q_i}g_{ij}}{\sum\limits_{k\neq i} e^{q_k}g_{kj}+\sigma^2}) - \theta_j)\end{aligned} \tag{21}$$

where $\boldsymbol{\alpha} = \{\alpha_j, \forall j\in\mathcal{J}\}$ and $\boldsymbol{\beta} = \{\beta_j, \forall j\in\mathcal{J}\}$ are the nonnegative Lagrange multipliers corresponding to the constraint (20c) and (20d). The optimal $\boldsymbol{q}$ should meet the Karush-Kuhn-Tucker (KKT) condition as

$$\frac{\partial\mathcal{L}}{\partial q_i} = (\zeta + \sum_{j\in U_i}\frac{(\alpha_j+\mu_j)g_{ij}}{\sum_{k\neq i} e^{q_k}g_{kj}+\sigma^2})e^{q_j} - \sum_{j\in U_i}\alpha_j = 0. \tag{22}$$

From (22), the optimum transmit power is thus obtained as

$$p_i^{t+1} = \frac{\sum_{j\in U_i}(\alpha_j+\mu_j)}{\rho_i\zeta + \sum_{j\in U_i}\frac{(\alpha_j+\mu_j)g_{ij}}{\sum_{k\neq i} e^{q_k}g_{kj}+\sigma^2}}. \tag{23}$$

The iterative expression of (23) is a two-sided scalable function [29], which converges to an optimal point given any initial value. The value of α_j is solved by the gradient method, therefore, we have:

$$\alpha_j^{t+1} = \alpha_j^t - \delta^t(\log(\mathrm{SINR}_j) - \log\tau_j). \tag{24}$$

When the dual optimal solution $\boldsymbol{\alpha}^*$ is given, the problem (21) can be easily simplified into

$$\mathcal{L}_1(\boldsymbol{\theta},\boldsymbol{\beta}) = \sum_{j\in\mathcal{J}} w_j\log(\log(1+e^{\theta_j})) - \sum_{j\in\mathcal{J}}\beta_j(\log(\mathrm{SINR}_j)-\theta_j). \tag{25}$$

An efficient numerical approach, Newton gradient method, is employed to update θ and β, which has the two-order convergence [30] [31]. The dual variables at each iteration are updated as

$$\theta_j^{t+1} = \theta_j^t - \delta^t\frac{\phi(\theta_j^t)}{\phi'(\theta_j^t)}, \tag{26}$$

and

$$\beta_j^{t+1} = \beta_j^t - \delta^t\frac{\psi(\beta_j^t)}{\psi'(\beta_j^t)}, \tag{27}$$

where ϕ' and ψ' are the first-order partial derivative of ϕ and ψ in terms of θ and β, and

$$\psi(\beta_j) = \log\frac{e^{q_i}g_{ij}}{\sum_{k\neq i} e^{q_k}g_{kj}+\sigma^2} - \theta_j,\quad \forall i,j \tag{28}$$

and

$$\phi(\theta_j) = \frac{w_je^{\theta_j}}{\log(1+e^{\theta_j})(1+e^{\theta_j})} - \beta_j.\quad \forall i,j \tag{29}$$

This is derived from the partial KKT conditions of (25) as follows

$$\frac{\partial\mathcal{L}_1}{\partial\theta_j} = \frac{w_je^{\overline{\theta}_j}}{\log(1+e^{\overline{\theta}_j})(1+e^{\overline{\theta}_j})} - \overline{\mu}_j = 0, \tag{30}$$

and

$$\overline{\beta}_j\frac{\partial\mathcal{L}_1}{\partial\beta_j} = \log\frac{e^{\overline{q}_i}g_{ij}}{\sum_{k\neq i} e^{\overline{q}_k}g_{kj}+\sigma^2} - \overline{\theta}_j = 0. \tag{31}$$

where $\overline{\theta}_j$ and $\overline{\beta}_j$ is the optimal solution of (25). Combine (30) and (31) and we have

$$\overline{\theta}_j = \log \frac{e^{\overline{q}_i} g_{ij}}{\sum_{k \neq i} e^{\overline{q}_k} g_{kj} + \sigma^2}, \quad \forall i \in \mathcal{I}, j \in \mathcal{U}_i \tag{32}$$

and

$$\overline{\beta}_j = \frac{w_j e^{\overline{\theta}_j}}{\log(1 + e^{\overline{\theta}_j})(1 + e^{\overline{\theta}_j})}, \quad \forall i \in \mathcal{I}, j \in \mathcal{U}_i \tag{33}$$

Algorithm 3 outlines the steps in search of the optimal power allocation under fixed BS association, where T_3 and T_4 are the numbers of maximum iterations of inner and outer loops respectively.

Algorithm 3 Transmit power allocation algorithm for UEE

Require:
The initial transmit power of BSs, $\boldsymbol{p}$,
The set of maximal iteration, T_3, T_4,
The initialization of dual variable, $\boldsymbol{\theta}$, $\boldsymbol{\alpha}$, $\boldsymbol{\beta}$.
Ensure:
The optimal power allocation $\boldsymbol{p}^*$.
repeat
Obtain the optimal transmit power p_i according to (23);
Upada α_j according to (24), $t_3 = t_3 + 1$.
repeat
Updata $\theta_j^{t+1} = \theta_j^t - \delta^t \frac{\phi(\theta_j^t)}{\phi'(\theta_j^t)}$, $\forall j$;
Updata $\beta_j^{t+1} = \beta_j^t - \delta^t \frac{\psi(\alpha_j^t)}{\psi'(\beta_j^t)}$, $\forall j$;
$\beta_j^{t+1} = \beta_j^{t+1} / \sum_{j \in \mathcal{J}} \beta_j^{t+1}$, $t_4 = t_4 + 1$.
until Convergence or $t_3 = T_3$.
until Convergence or $t_4 = T_4$.

C. COMPLEXITY ANALYSIS

In this section, we analyze the complexity of the proposed two-layer iterative algorithm. The major complexity lies in solving two subproblems: user association problem and power control problem. The user association problem is solved by **Algorithm** 2. In this procedure, each user selects a specific BS for utility maximization, thus the total complexity is $\mathcal{O}(MN)$. Each BS updates its transmit power by using the rule (23) and results in a complexity of $\mathcal{O}(MN)$. Then, **Algorithm** 3 has a complexity of $\mathcal{O}(M^2N)$. Hence, the total computational complexity of proposed algorithm is $\max\{\mathcal{O}(T_1T_2MN, T_1T_3T_4M^2N)\}$. Since the value of T_1, T_2, T_3 and T_4 can be relatively small, the total complexity of the proposed algorithm is formatted as $O(M^2N)$. The proposed scheme has a lower complexity, compared with the conventional interior-point method [32] and is appropriate for the practical application scenarios.

IV. SIMULATION RESULTS AND ANALYSIS

In this section, we present the convergence of our proposed algorithm and analyze the simulation results via numerical simulation. A two-layer downlink HetNets is modeled where one MBS is located in the central area and PBSs are spatially distributed within the coverage of MBS. We assume that 50 users are distributed in a random order. The maximum transmit power of MBS is set as 46 dBm. The circuit power is {10 W, 0.1 W} for MBS and PBS respectively. Besides, the power amplification ratio of BS is assumed as {4,2}. For propagation, the large-scale pass loss of the MBS and PBS are $L = 128.1 + 37.6 \log 10(d)$ and $L = 140.7 + 36.7 \log 10(d)$ separately, where d is the distance between the BS and user. Besides, the shadow effect follows logarithmic normal distribution, where the standard deviation is 8 dB. The noise power of AWGN is -104 dBm and the bandwidth of the system is 10MHz. We carry out the analytical results via Monte Carlo simulations.

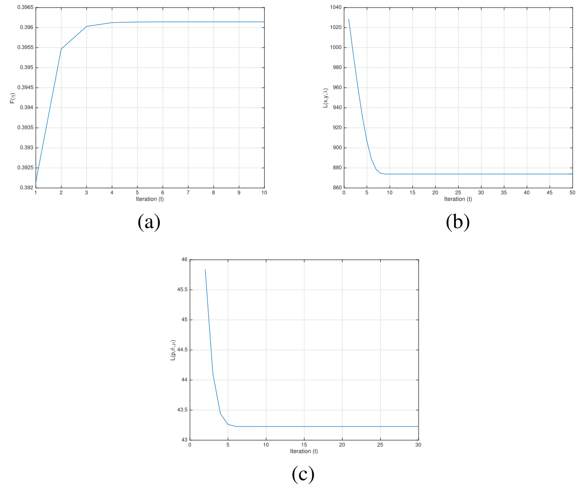


FIGURE 1. The convergence of proposed algorithm. (a) The convergence of two-layer iteration algorithm. (b) The convergence of user association algorithm. (c) The convergence of power allocation algorithm.

A. STIMULATION FOR CONVERGENCE

Fig.1 illustrate the convergence performance of the proposed UEE algorithm. Fig.1 (a) shows the outer algorithm converges to the optimum value of energy efficiency rapidly. Fig.1 (b) and (c) show the convergence of the load-aware user association algorithm and the power allocation algorithm respectively. Since the user association problem and power allocation problem are both convergent, the inner loop of the proposed algorithm is bound to converge. The numerical results show that the optimization method has good convergence and performance.

B. ANALYSIS AND COMPARISON OF PERFORMANCE

Fig.2 shows that the user association of the proposed UEE algorithm compared to the maximum throughput with power control (MTWPT) association scheme. We assume that a number of users and 6 PBSs are uniform distributed in a macrocell radius of 500m and the minimum QoS requirement of all users are set to 0db. It exemplifies that more users are served by PBSs of UEE algorithm than that of MTWPT. For MTWPT association, the users are pushed to the MBSs for a

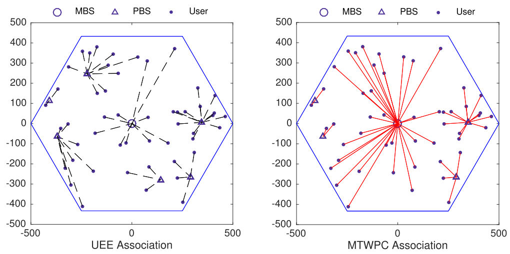


FIGURE 2. The comparison of two association scheme.

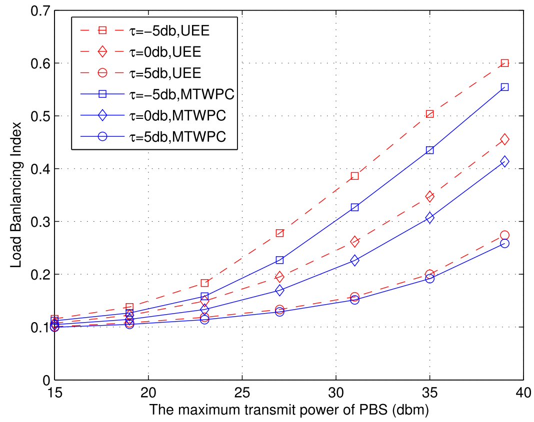


FIGURE 3. The load balancing index of HetNets system.

better performance while increasing the burden of message transmission. Obviously, the UEE association scheme can realize rate fairness and has a level of resources utilization than MTWPT.

Fig.3 compares the load balancing index of the UEE algorithm with MTWPT association scheme in terms of maximum transmit power of PBS under different QoS constraints. In this paper, we measure the load balancing level by using the Jain's fairness index [33]. A larger load balancing index indicates a more fairness user association distribution. Since MTWPT association is mainly considered the throughputs, it have a lower load balancing index than UEE algorithm. That is, the MBSs in MTWPT algorithm are overloaded with the most users, while the advantage of PBS cannot be fully embodied. The proposed UEE algorithm provides networks with the ability to improve load balancing index by moving workloads from servers that are heavily utilized to more lightly utilized workloads. With the increase of the maximal transmit power of PBSs, the gap between MBSs and PBSs are gradually reduced, thus, users associated with the overload MBS are assigned to the less congested PBSs. Furthermore, in the domain of high QoS requirement, some users are shifted to MBSs for meeting the minimal rate requirement. Overall, the proposed UEE algorithm could dynamically improve the data traffic distribution to maintain a balance between overall demand and supply.

Fig.4 compares the average energy efficiency of the proposed UEE algorithm with MTWPT algorithm with respect to the change of the available transmit power of PBS under different QoS constraints. Since our proposed algorithm performs well in user fairness, the average energy efficiency of our proposed algorithm also obtains a higher value; in other words, the load-index is becoming hugely important to the overall energy efficiency. For the two schemes, the energy efficiency increases when the available range of transmit power of PBSs enlarge. This is due to that, with increasing maximal transmit power of PBSs, more users have been allocated to the PBSs and the co-channel interference could be refrained effectively. And the higher QoS requirement results in the increasing of transmit power consumption, and the performances of the system will be decreased. Above all, the proposed algorithm performs better than the MTWPT in terms of the different transmit power and QoS requirements.

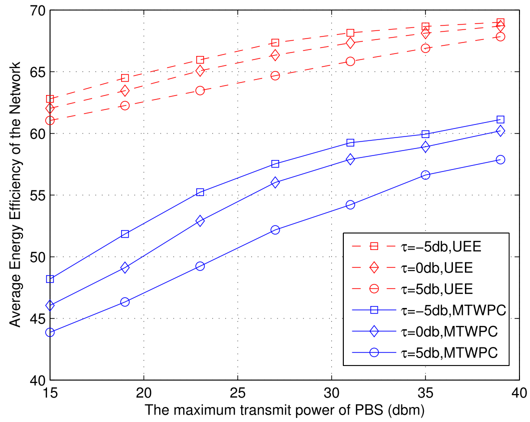


FIGURE 4. The energy efficiency of HetNets system.

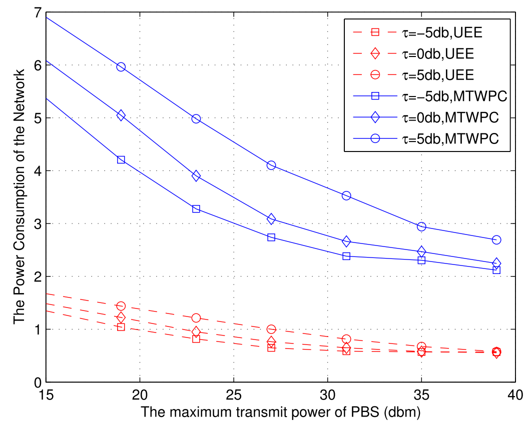


FIGURE 5. The total power consumption of HetNets system.

Fig.5 shows the impact of different maximum transmit power of PBS on power consumption in different algorithms under different QoS constraints. There is a slight reduction in power consumption for UEE algorithm when compared with the MTWPT algorithm. This improvement is obtained from the deployment of PBSs and the insufficient resource can be utilized rationally. For MTWPT, achieving for higher throughput always results in the added cost of power consumption, meanwhile, leading to less energy efficiency. Further increment in values of transmit power of PBSs leads to a lower power consumption, because of more low-power BSs are involved. The gap between the proposed algorithm and

MTWPT algorithm is increased in terms of power consumption, because the difference of utilization of PBSs between two algorithms has become wider. With high requirement for effective user rate, the overall power consumption of the networks increases for maintaining the performance of the users. Comparison with MTWPT algorithm demonstrates that the proposed optimization framework reduces energy consumption and enhance the throughput of the system.

V. CONCLUSION

In this paper, we have investigated the optimal energy efficiency problem for load-coupled HetNets with unequal user priorities and QoS requirement. By jointly considering the user association and power allocation, we formulated the original problem as a multi-variable iterative optimization problem and transformed it into an equivalent convex problem. The reformulated problem can be solved by using the Lagrangian dual method separately. Numerical results demonstrate that the proposed algorithm achieves better performance in load-balance and power consumption than conventional algorithms.

REFERENCES

[1] (2015). *eSustainability Initiative (GeSI) T G. SMARTer2030: ICT Solutions for 21s Tcentury Challenges [EB/OL]*. [Online]. Available: https://gesi.org/report/detail/smarter-2030

[2] H. S. Dhillon, R. K. Ganti, F. Baccelli, and J. G. Andrews, "Modeling and analysis of K-tier downlink heterogeneous cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 550–560, Apr. 2012.

[3] D. Liu, L. Wang, Y. Chen, M. Elkashlan, K. K. Wong, R. Schober, and L. Hanzo, "User association in 5G networks: A survey and an outlook," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1018–1044, 2nd Quart., 2016.

[4] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. G. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2706–2716, Jun. 2013.

[5] T. Zhou, Y. Huang, W. Huang, S. Li, Y. Sun, and L. Yang, "QoS-aware user association for load balancing in heterogeneous cellular networks," in *Proc. IEEE 80th Veh. Technol. Conf. (VTC)*, Vancouver, BC, Canada, Sep. 2014, pp. 1–5.

[6] 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, Jun. 2014.

[7] H. Cheng, F. Chen, N. Liu, Z. Pan, and X. You, "Utility-based energy-efficient resource allocation in downlink heterogeneous OFDMA networks with mixed traffic," in *Proc. 8th Int. Conf. Wireless Commun. Signal Process. (WCSP)*, Yangzhou, China, Oct. 2016, pp. 1–6.

[8] H. Kim, G. de Veciana, X. Yang, and M. Venkatachalam, "Distributed α-optimal user association and cell load balancing in wireless networks," *IEEE/ACM Trans. Netw.*, vol. 20, no. 1, pp. 177–190, Feb. 2012.

[9] K. Son, S. Lee, Y. Yi, and S. Chong, "Practical dynamic interference management in multi-carrier multi-cell wireless networks: A reference user based approach," in *Proc. 8th Int. Symp. Modeling Optim. Mobile, Ad Hoc, Wireless Netw.*, Avignon, France, May/Jun. 2010, pp. 186–195.

[10] M. J. Neely, E. Modiano, and C.-P. Li, "Fairness and optimal stochastic control for heterogeneous networks," *IEEE/ACM Trans. Netw.*, vol. 6, no. 2, pp. 396–409, Apr. 2008.

[11] K. Chitti, Q. Kuang, and J. Speidel, "Joint base station association and power allocation for uplink sum-rate maximization," in *Proc. IEEE SPAWC*, Jun. 2013, pp. 6–10.

[12] V. N. Ha and L. B. Le, "Distributed base station association and power control for heterogeneous cellular networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 1, pp. 282–296, Jan. 2014.

[13] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, and T. Ji, "Cell association and interference coordination in heterogeneous LTE-A cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 9, pp. 1479–1489, Dec. 2010.

[14] W. Wang and G. Shen, "Energy efficiency of heterogeneous cellular network," in *Proc. IEEE VTC-Fall*, Sep. 2010, pp. 1–5.

[15] M. Usman, A. Vastberg, and T. Edler, "Energy efficient high capacity HETNET by offloading high QoS users through femto," in *Proc. IEEE Int. Conf. Netw. (ICON)*, Dec. 2011, pp. 19–24.

[16] Y. Cui, W. Xu, H. Shen, H. Zhang, and X. You, "Optimal energy efficient association for small cell networks with QoS requirements," in *Proc. IEEE Wireless Commun. Netw. Conf. (WCNC)*, Doha, Qatar, Apr. 2016, pp. 1–6.

[17] J. Zhang, L. Xiang, D. W. K. Ng, M. Jo, and M. Chen, "Energy efficiency evaluation of multi-tier cellular uplink transmission under maximum power constraint," *IEEE Trans. Wireless Commun.*, vol. 16, no. 11, pp. 7092–7107, Nov. 2017.

[18] H. Zhu, S. Wang, and D. Chen, "Energy-efficient user association for heterogenous cloud cellular networks," in *Proc. IEEE Globecom Workshops (GC Wkshps)*, Dec. 2012, pp. 273–278.

[19] T. Zhou, Y. Huang, and L. Yang, "Energy-efficient user association in downlink heterogeneous cellular networks," *IET Commun.*, vol. 10, no. 13, pp. 1553–1561, Dec. 2016.

[20] H. Huang, W. Xu, H. Shen, H. Zhang, and X. You, "Utility-energy efficiency oriented user association with power control in heterogeneous networks," *IEEE Wireless Commun. Lett.*, vol. 7, no. 4, pp. 526–529, Aug. 2018.

[21] Z. Yang, W. Xu, J. Shi, H. Xu, and M. Chen, "Association and load optimization with user priorities in load-coupled heterogeneous networks," *IEEE Trans. Wireless Commun.*, vol. 17, no. 1, pp. 324–338, Jan. 2018.

[22] N. Roseveare and B. Natarajan, "An alternative perspective on utility maximization in energy-harvesting wireless sensor networks," *IEEE Trans. Veh. Technol.*, vol. 63, no. 1, pp. 344–356, Jan. 2014.

[23] Y. Chen, J. Li, Z. Lin, G. Mao, and B. Vucetic, "User association with unequal user priorities in heterogeneous cellular networks," *IEEE Trans. Veh. Technol.*, vol. 65, no. 9, pp. 7374–7388, Sep. 2016.

[24] W. Li, S. Wang, Y. Cui, X. Cheng, R. Xin, M. A. Al-Rodhaan, and A. Al-Dhelaan, "AP association for proportional fairness in multirate WLANs," *IEEE/ACM Trans. Netw.*, vol. 22, no. 1, pp. 191–202, Feb. 2014.

[25] H. J. Kushner and P. A. Whiting, "Asymptotic properties of proportional-fair sharing algorithms," in *Proc. 40th Annu. Allerton Conf.*, 2002, pp. 1051–1059.

[26] A. Jalali, R. Padovani, and R. Pankaj, "Data throughput of CDMA-HDR a high efficiency-high data rate personal communication wireless system," in *Proc. IEEE VTC*, May 2000, pp. 1854–1858.

[27] W. Dinkelbach, "On nonlinear fractional programming," *Manage. Sci.*, vol. 13, no. 7, pp. 492–498, Mar. 1967. [Online]. Available: http://www.jstor.org/stable/2627691

[28] S. Boyd and L. Vandenberghe, *Convex Optimization*. Cambridge, U.K.: Cambridge Univ. Press, 2004.

[29] C. W. Sung and K.-K. Leung, "A generalized framework for distributed power control in wireless networks," *IEEE Trans. Inf. Theory*, vol. 51, no. 7, pp. 2625–2635, Jul. 2005.

[30] T. Zhou, Z. Liu, J. Zhao, C. Li, and L. Yang, "Joint user association and power control for load balancing in downlink heterogeneous cellular networks," *IEEE Trans. Veh. Technol.*, vol. 67, no. 3, pp. 2582–2593, Mar. 2018.

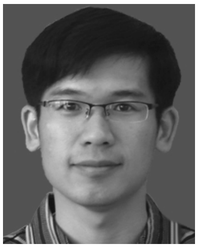
[31] Y. Jong. (2012). *An Efficient Global Optimization Algorithm for Nonlinear Sum-of-Ratios Problem*. [Online]. Available: www.optimizationonline.org

[32] *Ipopt: Interior Point OPTimizer*. Accessed: Oct. 20, 2011. [Online]. Available: https://projects.coin-or.org/IpoptS

[33] A. B. Sediq, R. H. Gohary, R. Schoenen, and H. Yanikomeroglu, "Optimal tradeoff between sum-rate efficiency and Jain's fairness index in resource allocation," *IEEE Trans. Wireless Commun.*, vol. 12, no. 7, pp. 3496–3509, Jul. 2013.

RUI LI received the B.S. degrees from the School of Computer and Information, Hohai University, Nanjing, China, in 2016, where she is currently pursuing the Ph.D. degree in information and communication engineering. Her current research interests include heterogeneous networks, simultaneous wireless information and power transfer (SWIPT), and wireless communications.

NING CAO received the B.S. and M.S. degrees in electrical engineering from Southeast University, Nanjing, China, in 1984 and 1990, respectively. He is currently a Professor with the College of Computer and Information, Hohai University, Nanjing. His current research interests include statistical signal processing, sensor array processing, and wireless communications.



MINGHE MAO received the B.S. and Ph.D. degrees from the School of Computer and Information, Hohai University, Nanjing, China, in 2009 and 2016, respectively, where he is currently a Postdoctoral Fellow with the Department of Electronic Information Engineering, School of Computer and Information. His research interests include cognitive radios, multi-hop relaying, heterogeneous networks for wireless communications, as well as biomimetic robotics control and communication. He has been a recipient of the National Natural Science Foundation of China, in 2017.

YIFAN HU received the B.S. and M.S. degrees from the School of Computer and Information, Hohai University, Nanjing, China, in 2011 and 2013, respectively, where he is currently pursuing the Ph.D. degree with the School of Computer and Information. His research interests include the relaying networks, cognitive radios, and energy harvesting in wireless communications.