Towards Throughput Aware and Energy Aware Traffic Load Balancing in Heterogeneous Networks with Hybrid Power Supplies

Towards Throughput Aware and Energy Aware Traffic Load Balancing in Heterogeneous Networks with Hybrid Power Supplies

Qiang Fan, *Student Member, IEEE*, and Nirwan Ansari, *Fellow, IEEE*

***Abstract*—Green communications has attracted much research attention for its environmental and economic benefits. For a cellular network, base stations (BSs) incur more than 50% of the energy consumption of the whole network. Therefore, BSs can be powered by green energy to reduce its on-grid power consumption. Meanwhile, the effective data rate (EDR) of a user's flow, which depends on both the user's channel condition and its BS's workload status, is an important metric for user performance. From the perspectives of a mobile provider, both the aggregated EDR (sum of EDRs of all users within the coverage area of a macro BS) and on-grid energy consumption should be taken into account in the traffic load balancing process. Therefore, we propose a Throughput aware and Energy Aware (TEA) traffic load balancing scheme for heterogeneous cellular networks to optimize the above two metrics. Since the EDR and energy consumption mutually affect each other, saving on-grid power is at the cost of sacrificing a certain loss of EDR. Thus, we employ an energy-throughput coefficient α to adjust the tradeoff between the two metrics based on the mobile provider's practical requirements. Simulation results demonstrate that TEA improves the aggregated EDR and significantly saves on-grid power.**

***Index Terms*—Throughput, green energy, user association, heterogeneous cellular network.**

I. INTRODUCTION

OWING to the direct impact of greenhouse gases on the environment and the climate change, curbing the energy consumption of mobile networks has attracted much attention. Driven by the proliferation of data-hungry devices and applications, mobile data traffic is expected to increase exponentially in the future [1]. In this situation, the increasing traffic not only calls for expansion of network capacity, but also intensifies the energy consumption [2]. Therefore, greening mobile networks is important to mitigate the environmental problems and reduce the operating cost of mobile operators [3], [4]. With the development of green energy technologies, green energy such as solar energy, wind energy and sustainable biofuels is being utilized to power base stations (BSs). However, owing to the unstable generation of green energy, hybrid energy supplies, consisting of both green energy and on-grid power, are a more practical option to power BSs [5]. Thus, green energy can be utilized to reduce the on-grid power consumption and therefore decrease the CO_2 emission, with the on-grid power as a backup power source [6]. Heterogeneous cellular networks (HCNs), in which the macro cells are overlaid with small cells, are promising to increase the total capacity of cellular networks [7]. Considering the dynamic workload distribution, small cell base stations (SCBSs) are placed in areas with high user density to facilitate more users to connect to a much closer BS, thus improving the channel conditions of users. Meanwhile, as the coverage of each SCBS is very small, the transmission power required by each SCBS is significantly smaller than those of traditional BSs [8], [9]. Therefore, the low power of SCBSs can potentially improve the spectral efficiency and energy efficiency of heterogeneous cellular networks [10].

Q. Fan and N. Ansari are with Advanced Networking Lab., Department of Electrical and Computer Engineering, New Jersey Institute of Technology, Newark, NJ, 07102 USA (Email: {qf4, nirwan.ansari}@njit.edu). This work was supported in part by the National Science Foundation (NSF) under grant no. CNS-1320468.

In a HCN with hybrid power supplies, the effective data rate (EDR) of a user's flow is based on both the channel condition of the user towards its BS and the BS's workload status [11]. As the user distribution is dynamic, if a user tends to associate with BSs only based on the channel condition or received power, it may connect to a congested BS, which degrades its EDR. Consequently, some BSs may be congested by the heavy traffic loads while other BSs are lightly loaded. The unbalanced workload distribution among BSs has a negative impact on user Quality-of-Service (QoS) in terms of the EDR. On the other hand, the main operating cost of mobile providers arises from the on-grid energy consumption. Owing to the dynamic traffic workload distribution among BSs, the energy demands of BSs may not match their available green energy, thus incurring the increment of on-grid energy consumption. In other words, while some BSs still have excessive green energy, others have drained their green energy and started to consume on-grid energy. To reduce the operating cost, traffic load balancing can be employed to reduce the gap between the energy demands of BSs and their green energy. Moreover, as mobile providers need to consider the gain of the aggregated EDR (sum of EDRs of all users within the coverage area of a macro BS) and the operating cost in terms of on-grid energy consumption simultaneously, the optimal traffic load balancing strategy should take into consideration of the above two factors. However, in the load balancing process, saving on-grid power is always at the cost of sacrificing an amount of EDR, i.e., the EDR and on-grid energy consumption exhibit a trade-off relationship. How to balance the traffic loads among BSs to optimize the aggregated EDR of the network and on-grid energy consumption still remains to be a critical problem.

In this paper, to solve the above problem, we propose a Throughput aware and Energy Aware (TEA) traffic load balancing scheme for heterogeneous networks to satisfy mobile providers' requirements by balancing traffic loads. The scheme not only optimizes the utilization of green energy in order to reduce the on-grid power consumption, but also optimizes the aggregated EDR of the network. Since the power consumption of a macro BS (MBS) is significantly larger than that of SCBS, associating users with SCBSs may reduce the on-grid power consumption. However, too many users associating with SCBSs may incur traffic congestion in SCBSs and thus degrades the EDRs of their users. The proposed TEA algorithm makes a tradeoff between the aggregated EDR of the network and on-grid energy consumption by assigning users to the suitable BSs. Below are the major contributions of this paper.

- We formulate the problem of making a tradeoff between the aggregated EDR and on-grid energy consumption by balancing traffic workloads among heterogeneous BSs. The mobile providers desire to improve the aggregated EDR while reducing on-grid energy consumption of the network. Since the user association aiming to increase the effective data may increase on-grid energy consumption, we need to balance these two factors in the scheme. Thus, we define an energy-throughput coefficient α to make a tradeoff between the aggregated EDR and on-grid energy cost, which can be predefined by each mobile provider based on its practical requirement.
- The workload status of a BS has a critical impact on the EDRs of its associated users. To guarantee the user QoS, we assume that the workload of each BS should be smaller than the BS's maximum workload threshold allowed by mobile providers.
- To solve the user association problem (i.e., load balancing) in each time slot, we propose a heuristic algorithm which iteratively moves users to suitable BSs. Then, we analyze the computational complexity of the algorithm. We also analyze some critical issues of the proposed algorithm in order to facilitate its practical implementation.

The rest of this paper is organized as follows. In Section II, we briefly review related works. In Section III, we define the system model. In Section IV, we formulate and analyze the user association problem, and propose the TEA scheme. Section V presents the heuristic TEA algorithm and its practical implementation. The viability of TEA is substantiated by simulation results in Section VI, and the concluding remarks are presented in Section VII.

II. RELATED WORKS

Some studies have considered powering mobile cellular networks with hybrid energy supplies [12]–[15]. To maximize the utilization of green energy, Fan *et al.* [8] proposed to offload the traffic loads from BSs depleting of green energy to other BSs with excessive green energy while satisfying the QoS requirements of users in terms of their data rates. Considering a network with multiple energy supplies, Han and Ansari [16] also proposed to optimize the utilization of green energy, and reduce the on-grid energy consumption in cellular networks by the cell size optimization.

In addition, there have been many research efforts on user association in cellular networks. Han *et al.* [17] proposed a heuristic user association algorithm to assign users to different green relays to maximize the minimum EDRs of users while taking into account of the green load capacities of the relay nodes. Ma *et al.* [18] studied the joint user association and resource allocation in the heterogeneous network with backhaul constraint in order to maximize the α-fairness network utility. Wang *et al.* [19] proposed user association algorithms to optimize the green energy utilization. In particular, they decomposed the problem into two components: first, users are allocated to different BSs based on the available green energy; then, the optimal bandwidth allocation enables the on-grid energy consumption of the network to be further reduced. Kong *et al.* [20] proposed a biased user association scheme with which users choose to associate with BSs in order to optimize the delay for users. Han and Ansari [21] proposed a distributed user association scheme named GALA for heterogeneous cellular networks that optimizes the trade-off between the on-grid power consumption and the average traffic delivery latency. Corroy *et al.* [22] developed a theoretical framework and proposed a dynamic cell association heuristic algorithm to maximize the sum rate of all users. Jo *et al.* [23] added an offset/bias for small cells to attract more users, referred to as the range expansion/extension association.

To the best of our knowledge, the existing published works on user association in heterogeneous networks have not considered to jointly optimize the EDRs of users and on-grid energy consumption in heterogeneous cellular networks. From the point view of mobile providers, both the aggregated EDR and green energy utilization of the network are critical components of their profits. Simply optimizing only one of them cannot satisfy the requirement of the mobile providers. Thus, we propose TEA to associate users to heterogeneous BSs by taking into account of user channel condition, BS workload status and available green energy. Since saving green energy is at the cost of sacrificing a certain amount of the EDR, the balance between the aggregated EDR and green energy can be determined by an energy-throughput coefficient α, an operating parameter determined by each mobile provider.

III. SYSTEM MODEL

In this paper, we consider an area consisting of one MBS and several SCBSs, as shown in Fig. 1. All these BSs are powered by both green energy (solar panel) and on-grid power. Here, the downlink user-BS association scenario is considered. Denote J and I as the set of BSs and the set of users, respectively, where j_0 is the MBS. In addition, we assume that orthogonal channels (OFDMA) are available. Both the MBS and SCBSs reuse the licensed spectrum to enhance the frequency efficiency.

The MBS is overlaid with multiple SCBSs in the two-tier heterogeneous network. All SCBSs are distributed around the MBS according to a homogeneous Spatial Poisson Point Process (PPP) with density λ_0, where λ_0 is the average number of SCBSs per unit area [8], [24].

A. Communications Model

Denote P_j as the transmit power of BS j and g_{ij} as the channel gain from BS j to user i. σ^2 is the noise power.

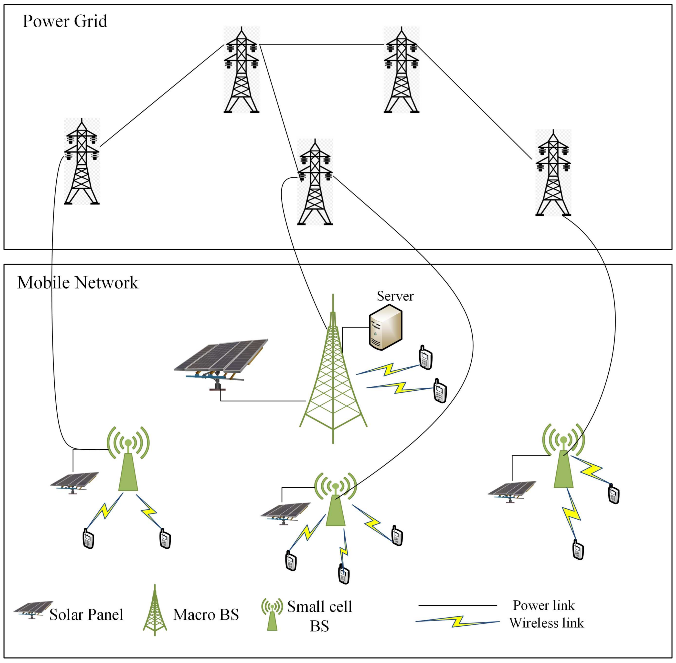


Fig. 1: Network architecture.

Thus, the signal to interference plus noise ratio (*SINR*) of BS j towards user i can be derived as

$$\gamma_{ij} = \frac{P_j g_{ij}}{\sigma^2 + \sum_{k \in J, k \neq j} P_k g_{i,k}}. \tag{1}$$

Then, the data rate C_{ij} of the ith user at the jth BS can be generally expressed as a logarithmic function of the perceived γ_{ij}, according to the Shannon Hartley theorem,

$$C_{ij} = W_j \ log(1 + \gamma_{ij}), \tag{2}$$

where W_j is the total bandwidth of the jth BS.

B. Traffic Model

Assume that the traffic for each user arrives at BS j according to a Poisson Process with the arrival rate λ_i and the lengths of flows have a general distribution with the average value of l_i. Then, the average traffic load density of user i in BS j is

$$\varrho_{ij} = \frac{\lambda_i l_i \eta_{ij}}{C_{ij}}. \tag{3}$$

Here, η_{ij} is a binary variable indicating whether the ith user is associated with the jth BS (1 if so; 0, otherwise).

By aggregating traffic load densities of users in a BS, we get the average traffic load density ρ_j in the BS. The value of ρ_j indicates the fraction of time during which BS j is busy.

$$\rho_j = \sum_{i \in I} \varrho_{ij}. \tag{4}$$

A BS is assumed to serve each user in the round-robin fashion. While the traffic arrival rate of each user follows the Poisson Process, the service time for the user's traffic satisfies the general distribution [21]. Then, a BS realizes a M/G/1-processor sharing (*PS*) queue. In the queue, a user in BS j is assumed to have the traffic load l_i and the communications data rate C_{ij}. In addition, to keep the queue stable, we need to guarantee ρ_j to be smaller than 1.

In an M/G/1-processor sharing queue of a BS, the average delivery time of user j's flows can be expressed as:

$$T_{ij} = \frac{l_i}{C_{ij}(1 - \rho_j)}. \tag{5}$$

Consequently, according to [17], [25], the EDR for each flow of user j (i.e., flow throughput) is derived as

$$C_{ij}^{eff} = C_{ij}(1 - \rho_j). \tag{6}$$

In the paper, we define the EDR of a user as the EDR for each flow of the user while the aggregated EDR of the network as the sum of EDRs of all users within a MBS's coverage area.

C. Energy Model

The power consumption of a BS consists of two parts: the static power consumption and the dynamic power consumption. The static power consumption p_j^s is used to keep a BS active even without traffic, while the dynamic power consumption is directly attributed to the traffic load in the BS. Here, the static power consumption of a BS is assumed constant. As a result, we focus on reducing the dynamic power which is related to the user association. The dynamic power consumption of a BS depends on its traffic load density. Denoting β_j as the linear coefficient which reflects the relationship between the traffic load and the dynamic power, then, the total power p_j of BS j is

$$p_j = \beta_j \rho_j + p_j^s. \tag{7}$$

While both the MBS and SCBSs are equipped with solar panels to generate green energy, all BSs can utilize both the on-grid power and green energy. If green energy of a BS is not sufficient, it may start to consume on-grid power. We need to maximize the utilization of available green energy of each BS in each time slot by assigning optimal user associations, in order to minimize on-grid energy consumption of BSs. Furthermore, according to [17], [21], the on-grid power consumption of the jth BS is

$$p_j^{on-grid} = \max(p_j - p_j^{green}, 0), \tag{8}$$

where p_j is the power consumption of BS j and p_j^{green} is the green energy generation rate of BS j.

IV. Problem Formulation

We have two goals in this paper. One is to maximize the aggregated EDR of the heterogeneous network. The other is to reduce the on-grid power consumption. For a BS, to reduce its on-grid power consumption, it has to shrink its coverage area. Therefore, the traffic loads are offloaded to its neighboring BSs that may result in traffic congestion in the neighboring BSs. Traffic congestion decreases the EDRs of users in those neighboring BSs. Considering both the aggregated EDR and on-grid power consumption, we can define the network utility of HCNs as follows:

$$u = \sum_{j \in J} \sum_{i \in I} C_{ij}^{eff} - \alpha \sum_{j \in J} p_j^{on-grid}(\eta), \tag{9}$$

where the first term is the aggregated EDR of the network as the gain of the mobile provider while the second term is the on-grid power consumption of the network as the cost. Moreover, α is an energy-throughput coefficient, which is given by mobile providers. α is an important system parameter. Increasing the value of α would increase the ratio of the on-grid power consumption to the aggregated EDR, and encourage users to associate with BSs with excessive green energy without considering the congestion problem. Thus, this parameter adjusts the tradeoff between the aggregated EDR and on-grid power consumption, and can be chosen via experiments based on mobile providers' detailed requirements in different time slots.

In order to minimize the network utility that takes into account of both the aggregated EDR and on-grid power consumption in the network, we formulate the problem as follows:

$$P1 : \underset{\eta}{Max} \sum_{j \in J} \sum_{i \in I} C_{ij} (1 - \sum_{i \in I} \frac{\lambda_i l_i \eta_{ij}}{C_{ij}}) \eta_{ij} - \alpha \sum_{j \in J} p_j^{on-grid}(\eta) \tag{10}$$

$$s.t. \sum_{j \in J} \eta_{ij} = 1, \forall i \in I \tag{11}$$

$$\rho_j \leq \rho_{max}, \forall j \in J. \tag{12}$$

Here, Constraint (11) imposes each user to be associated with only one BS; Constraint (12) imposes the workload of each BS to be lower than the maximum workload threshold of the BS, where ρ_{max} is less than 1 and predefined by mobile providers. The above optimization problem is a mixed integer quadratic constraint problem, which is challenging to achieve the optimal solution. In order to get the optimal user association assignment, a brute-force search leads to $O(M^N)$ iterations, where M and N represent the number of BSs and users, respectively. Obviously, the computational complexity of the brute-force search increases exponentially with respect to the total number of users. Therefore, it is not practical for real time applications especially for large scale networks. We have further proved the problem to be NP-hard.

Lemma 1. *The problem P1 is NP-hard.*

Proof:

Suppose there are two BSs, which are located at the same point, and thus the traffic load of a user towards either BS is the same. Then, to prove that P1 is a NP-hard problem, we need to demonstrate that its corresponding decision problem is NP-complete. The decision problem of P1 can be expressed as: given a positive value of b, is it possible to find a feasible solution $\eta = \{\eta_{ij} | i \in \mathcal{I}, j \in \mathcal{J}\}$ such that $\sum_{j \in J} \sum_{i \in I} C_{ij} (1 - \sum_{i \in I} \frac{\lambda_i l_i \eta_{ij}}{C_{ij}}) \eta_{ij} - \alpha \sum_{j \in J} p_j^{on-grid}(\eta) \leq b$, and Constraints (11), (12) are satisfied?

In order to prove that the above decision problem is NP-complete, only two BSs are considered and the traffic load threshold of either BS is set to be the same, i.e., $\rho_{max} = \frac{1}{2} \sum_{i \in \mathcal{I}} \varrho_{ij} + \epsilon$, where ϵ is a very small positive value, i.e., $\epsilon \ll \frac{1}{2} \min\{\varrho_{ij} | i \in \mathcal{I}\}$. Moreover, assume that $b \to +\infty$. Thus, $\sum_{j \in J} \sum_{i \in I} C_{ij} (1 - \sum_{i \in I} \frac{\lambda_i l_i \eta_{ij}}{C_{ij}}) \eta_{ij} - \alpha \sum_{j \in J} p_j^{on-grid}(\eta) \leq b$ is always satisfied for all solutions of η, and can be relaxed. To meet Constraint (12) (i.e., $\rho_j \leq \rho_{max} - \varepsilon, \forall j \in J$), we need to guarantee that $\sum_{i \in \mathcal{I}} \eta_{i1} \varrho_{i1} = \sum_{i \in \mathcal{I}} \eta_{i2} \varrho_{i2} = \frac{1}{2} \sum_{i \in \mathcal{I}} \varrho_{ij}$. Consequently, the decision problem can be transformed into a partition problem, i.e., whether the users can be partitioned into two sets to equalize the traffic loads of the two sets. Hence, the partition problem is reducible to the decision problem of P1. As the partition problem is a well-known NP-complete problem, the decision problem of P1 is also NP-complete, and thus P1 is NP-hard. ■

V. THE TEA ALGORITHM

A. Heuristic Algorithm

In this section, we propose a heuristic algorithm, TEA, which approaches the optimal solution with low computational complexity. The basic idea of the TEA algorithm is to iteratively select a suitable user and reallocate it to an alternative BS in order to improve the utility of the network, until all users cannot find better BSs. In the algorithm, both the green energy utilization and network throughput are considered.

We denote I' and J' as the set of unmarked users and the set of unmarked BSs, respectively. The network utility of the ith user in the jth BS can be expressed as:

$$u_{ij} = C_{ij}^{eff} - \alpha \frac{\varrho_{ij}}{\rho_j} p_j^{on-grid}. \tag{13}$$

The TEA algorithm, as shown in Algorithm 1 below, starts with an initial assignment, in which each user is associated with the BS providing the best SINR. At this time, all users are unmarked, i.e., $I' = \{i | i \in I\}$. Then, in each iteration, the TEA algorithm finds a user with the smallest utility among all unmarked users, and searches for a new BS by Algorithm 2 to improve the user's utility u_{ij}, while increasing the total utility of the whole network. If a new BS is found, TEA proceeds to the next iteration. Otherwise, the algorithm marks the user and continues to the next iteration. If all users are marked, the TEA algorithm terminates.

In Algorithm 2, we try to find a new BS for user i. The neighboring BSs, which can improve the utility of user i, are found and sorted in the decreasing order of the corresponding utility of user i. Then, each of these BSs is checked sequentially. If the traffic load of BS j that includes user i's traffic load is still within the limit of the BS's traffic load threshold and the current utility of the whole network has been improved as compared to the utility prior to this iteration, then BS j is qualified to be a new BS for user i. Otherwise, we will find the set of unmarked users in BS j and sort them in the increasing order of the user utility. Then, we check each user in this set to find out if the user can be moved to an alternative BS so as to reduce the traffic load of BS j. If so, the traffic load of BS j can be reduced and user i has a chance to be associated with BS j.

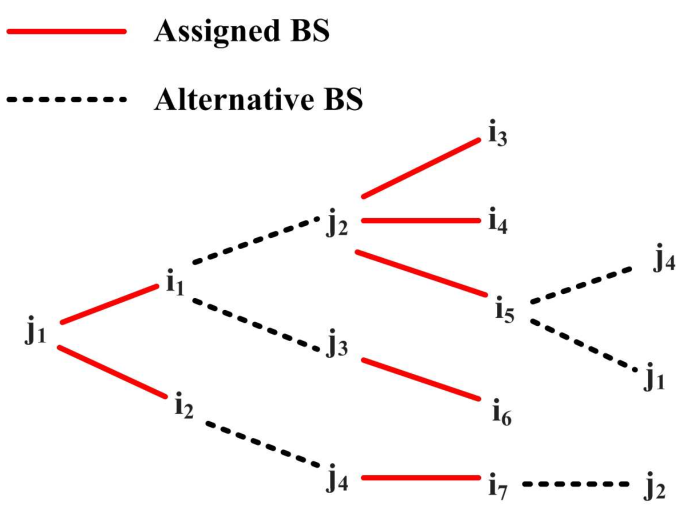


Fig. 2: An example of the algorithm.

Note that the main component of the proposed algorithm is Algorithm 2, which aims to find an alternative BS for a user. As Algorithm 2 is a recursive algorithm, $\hat{\eta}$ is denoted as the intermediate user association during the recursions. To ease understanding, we take an example in Fig. 2 to illustrate the recursive algorithm. As shown in Fig. 2, at the beginning, we suppose that user $i_1 \in I^{'}$ is the user with the minimum utility, and set all BSs to be unmarked. Then, we need to find a suitable alternative BS for user i_1 to increase its utility. The dotted lines indicate the potential alternative BSs for user i_1, i.e., BS j_2 and j_3 are alternative BSs for user i_1. If BS j_2 is not marked, the algorithm checks whether user i_1 can be associated to it successfully and marks it accordingly. In this case, if the total utility of the network is increased and the updated workload of BS j_2 is lower than the traffic load threshold ρ_{max}, the new suitable BS (i.e., j_2) is assigned to user i_1, and the user association matrix $\hat{\eta}$ is updated; otherwise, the algorithm checks whether users i_3, i_4 or i_5 of BS j_2 can be moved to other BSs to enable user i_1 to associate with BS j_2. Since user i_3 and i_4 can only be assigned to BS j_2, they cannot be reassigned to other BSs. For user i_5, to seek its new BS, Algorithm 2 will be called but with user i_5 as the input to check if it can be assigned to BS j_4 or j_1 to reduce the workload of BS j_2. Suppose it is not successful, then the algorithm will continue to check if user i_1 can be successfully assigned to BS j_3 by the same procedures above. If user i_1 cannot find a better BS in the end, the algorithm marks this user. Then, Algorithm 2 ends, and the TEA algorithm proceeds to the next iteration, i.e., seeking for an alternative BS for user $i_2 \in I^{'}$ that is unmarked and has the minimum utility.

We now analyze the computational complexity of the proposed TEA algorithm. In each iteration, due to the BS marking mechanism, which reduces the computational complexity, the number of BSs checked by the TEA algorithm can be M in the worst case. Therefore, the complexity of each iteration is $O(M)$. Since each user has a choice of up to M BSs and a user assigned to a BS can have N different data rates, the number of improvements for an individual user is limited by NM. Thus, in the worst case, the total number of iterations

Algorithm 1 The TEA Algorithm

Initialize the user association matrix η;
Set all users as unmarked, i.e., $I^{'} = \{i | i \in I\}$;
while $I^{'} \neq \emptyset$ **do**
  Set $flag = 0$;
  Find $i^* = \arg\min_{i \in I^{'}}(u_i)$ among unmarked users; let $u_{min} = u_{i^*}$
  Set all BSs as unmarked, i.e., $J^{'} = \{j | j \in J\}$;
  $(\hat{\eta}, flag) = find\_another\_bs(i^*, \eta, u_{total})$;
  if $flag == 1$ **then**
    Let $\eta = \hat{\eta}$;
    Update the total utility u_{total};
  else
    Mark user i^*, i.e., $I^{'} = I^{'}/i^*$;
  end if
end while
return η;

Algorithm 2 $(\hat{\eta}, flag) = find\_another\_bs(i^*, \eta, u_{total})$

Find $J_{i^*} = \{j | u_{i^*j} > u_{\min}, j \in J^{'}\}$ for user i^*; sort J_{i^*} in the decreasing order of u_{i^*j} ;
Set $flag = 0$;
while $J_{i^*} \neq \emptyset$ **do**
  Find the suitable BS by $j = \arg\min_{j \in J_{i^*}}(u_{i^*j})$;
  Mark BS j, i.e., $J^{'} = J^{'}/j$;
  Let η_{i^*j}=1;
  Calculate ρ_j and the current total utility u_{new} based on η;
  if $\rho_j < \rho_{max}$ **then**
    if $u_{new} > u_{total}$ **then**
      $\hat{\eta} = \eta$;
      $flag = 1$;
    end if
  else
    Find the set of unmarked users in BS j by $I_j = \{i | \eta_{ij} = 1 \& i \in I^{'}\}$, and sort it by the increasing utility u_{ij};
    for $k = 1 : |I_j|$ **do**
      if $\rho_j - \varrho_{k,j} < \rho_{max}$ **then**
        $(\bar{\eta}, flag) = find\_another\_bs(i_k, \eta, u_{total})$
        if $flag == 1$ **then**
          $\hat{\eta} = \bar{\eta}$, and break;
        end if
      end if
      k=k+1;
    end for
  end if
  if $flag == 1$ **then**
    break;
  end if
end while

that the algorithm can reach is $O(N^2M)$. As a result, the computational complexity of the heuristic TEA algorithm is $O(N^2M^2)$. When we fix the number of users deployed in the network, the computational complexity of the algorithm is polynomial with respect to the number of users.

B. Algorithm Implementation

In this section, we will present how to put the above algorithm into practice, and then discuss how to determine the energy-throughput coefficient in real applications from the perspectives of mobile providers. The proposed algorithm is operated in a controller located at each MBS to assign users among heterogeneous BSs (i.e., the MBS and its SCBSs). In a real HCN, in order to effectively balance the traffic workloads to avoid the situation that some BSs are congested while others are under-utilized, the controller should collect the workload distribution, user channel condition and green energy information of different SCBSs in advance. Then, based on the collected network information, the proposed scheme implemented in the controller optimizes the user association, and thus achieves the optimal traffic loads for different BSs (the MBS and its SCBSs). The optimization can be triggered either periodically or by some predefined events, e.g., a BS's traffic load exceeds a threshold or a BS's green energy utilization is lower than a threshold. What are the best strategies for triggering the algorithm can be determined by mobile providers. In this paper, we just assume that the traffic load balancing is executed in each time slot.

In the algorithm, users are just responsible for reporting their measurements, e.g., channel gain, rather than deciding the BS selection by themselves. From users' point of view, they may seek to maximize their QoS and violate the rule of the proposed algorithm. However, the EDR of each user not only depends on its capacity C_{ij}, but also on the traffic load of the selected BS ρ_j. Thus, the users, in fact, have no clue about which BS can improve their QoS, because the users do not have information about the traffic loads of BSs. In this case, simply selecting a BS with the best channel condition may direct users to a highly congested BS and thus degrades their EDRs. Thus, the centralized controller is employed to decide the user-BS association to improve the QoS of users and green energy utilization in the network by suitably balancing the traffic load among BSs.

From the point of view of mobile providers, the user association is optimized with the consideration of both the aggregated EDR and on-grid energy consumption. To balance the weight of these two metrics of the network utility, the proposed algorithm provides an energy-throughput coefficient α, which reflects each mobile provider's user association strategy. A mobile provider can determine the value of α for both MBS and SCBSs in a cell based on the comparison between the gain of the aggregated EDR and the cost of on-grid energy in different time slots. For a mobile provider, the gain of increasing one unit of EDR may be temporal and dynamic, i.e., it is remarkably higher during peak hours than idle hours. Therefore, in the peak hours, the EDR is the dominant factor of the network utility, and has a higher priority in the user association than the on-grid energy consumption. On the other hand, the price of on-grid energy always shows temporal and spatial dynamics. When the price of on-grid energy is very high in a certain slot, it is favorable to reduce the on-grid energy consumption in the user association as compared to improving the aggregated EDR. Thus, the mobile provider can flexibly provide the values of α to BSs in different areas at different time slots of one day. When a BS chooses a small α, the BS is EDR sensitive; in contrast, when α increases, the cost of on-grid energy becomes an important factor of the network utility, and thus the user association tends to be on-grid energy sensitive.

C. Overhead

As compared to the traditional user association mechanisms where users select their BSs based on metrics such as the received power or SINR, the proposed scheme needs to gather information from users and SCBSs in advance, and then run the algorithm in the MBS's controller. Specifically, SCBSs have to send their green energy states to the MBS while each user sends the information consisting of its channel condition and average traffic arrival rate to the MBS via a control channel. Hence, the collected network information enables the MBS's controller to execute the proposed algorithm in each time slot. However, the overhead arising from the proposed scheme remains low owing to following reasons: (1) the time slot is usually long (e.g., 1 min), and so the state information is not exchanged frequently; (2) all the information can be expressed in a few bits, thus incurring a short control message; (3) as each MBS is only in charge of assigning a few users (in its cell) among heterogeneous BSs, the scale of the network is limited, and hence the algorithm can be executed very fast.

VI. SIMULATION RESULTS

Simulations are set up to evaluate the performance of the proposed TEA algorithm in a heterogeneous network. In the simulation, one MBS is placed at the center of an area of 1000×1000 m^2, while four SCBSs are deployed in the MBS's coverage according to the PPP process. Meanwhile, users are uniformly distributed in the area. The network topology is shown in Fig. 3. The MBS's transmission power is 20 W, and each SCBS's transmission power is 5 W. We employ COST 231 Walfisch-Ikegami [26] as the propagation model with 9 dB Rayleigh fading and 5 dB shadowing fading. The carrier frequency is 2110 MHz, the bandwidth is 10 MHz, the antenna feeder loss is 3 dB, the transmitter gain is 1 dB, the noise power level is -104 dBm, and the receiver sensitivity is -97 dBm. The solar cell power efficiency is 17.4 percent. We assume that the weather condition is the standard condition which specifies a temperature of 25 °C, an air mass of 1.5, and an irradiance of 1000 W/m^2 [27]. Thus, the green power generation rate is 174 W/m^2. Meanwhile, the solar panel size of MBS is set as 4.6 m^2, while that of each SCBS is 0.7 m^2. In the simulation, the maximum workload threshold of each BS is set as $\rho_{max} = 0.95$. As the file (flow) length follows a general distribution, we set the average file length as 0.2 Mbits. Meanwhile, as the user file arrival rate follows the

Poisson process, we randomly choose the average file arrival rate between 0 and 2 (files/second).

In the simulation, we compare our algorithm with the heuristic green relay assignment (GRA) algorithm [17] and the Best SINR algorithm. In the Best SINR algorithm, users are associated to the BS with the best perceived SINRs. Thus, users are more likely to connect to the closest BS. Meanwhile, the GRA algorithm aims to iteratively maximize the minimum EDR among users.

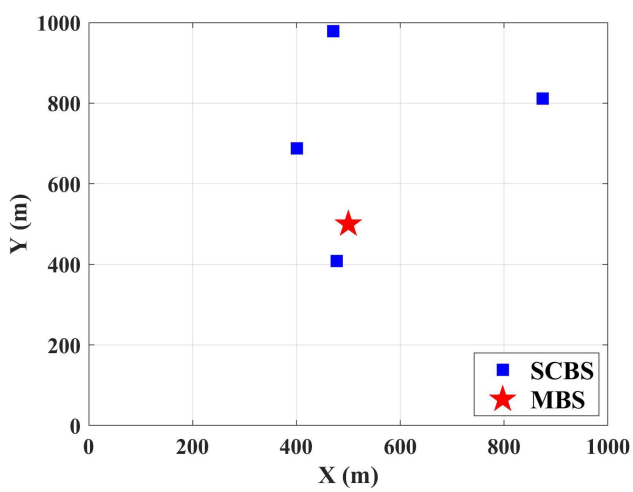


Fig. 3: Network topology.

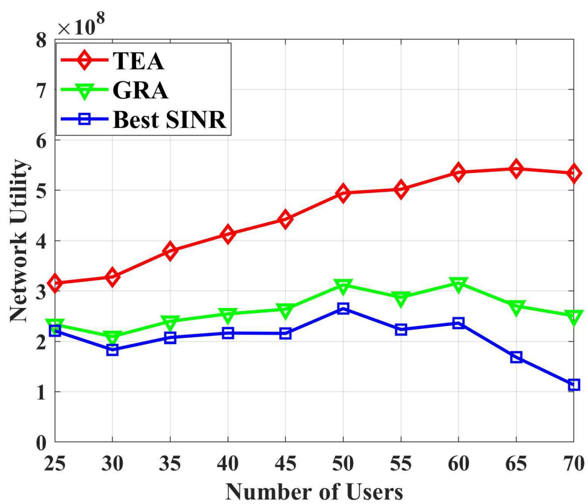


Fig. 4: The network utility comparison ($\alpha = 10^6$).

As shown in Fig. 4, the network utility of TEA is significantly higher than those of other algorithms as the number of users increases. The network utility is impacted by both the aggregated EDR and on-grid energy consumption. For GRA and Best SINR, as the number of users increases, the traffic loads of some BSs become overloaded quickly, and thus the aggregated EDR grows slowly. Meanwhile, the on-grid energy consumption keeps increasing with the traffic load. Thus, when the number of users is very large, the network utilities of GRA and Best SINR deteriorate. In contrast, since TEA considers

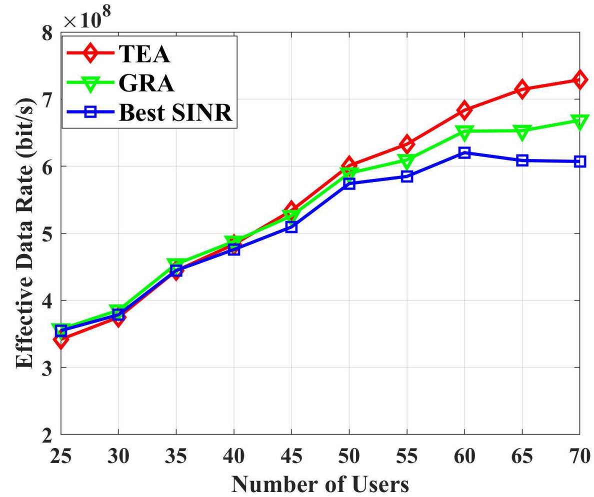


Fig. 5: The aggregated EDR comparison ($\alpha = 10^6$).

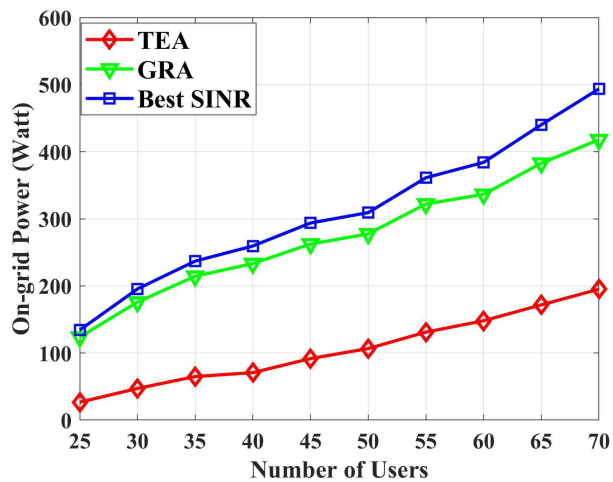


Fig. 6: The on-grid power comparison ($\alpha = 10^6$).

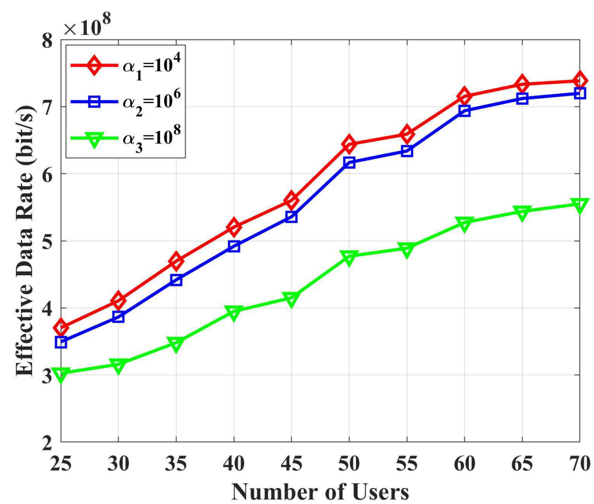


Fig. 7: The aggregated EDR comparison with different α.

both the aggregated EDR and on-grid energy consumption, when the aggregated EDR tends to be stable, the on-grid

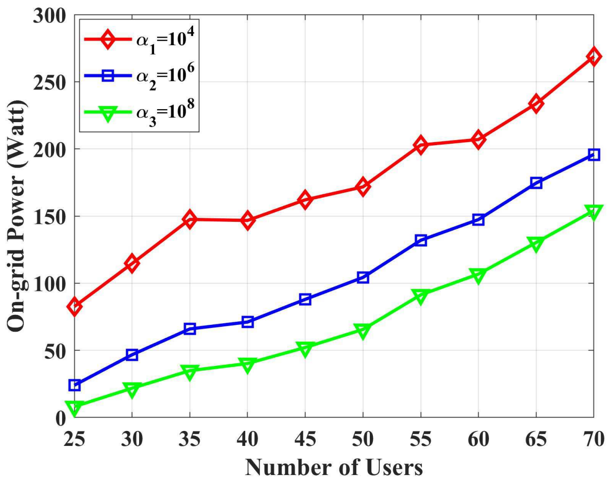


Fig. 8: The on-grid power comparison with different α.

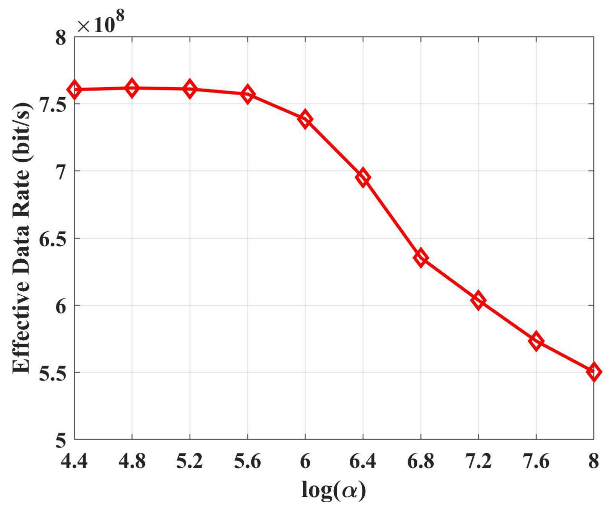


Fig. 9: The aggregated EDR v.s. α (user number=70).

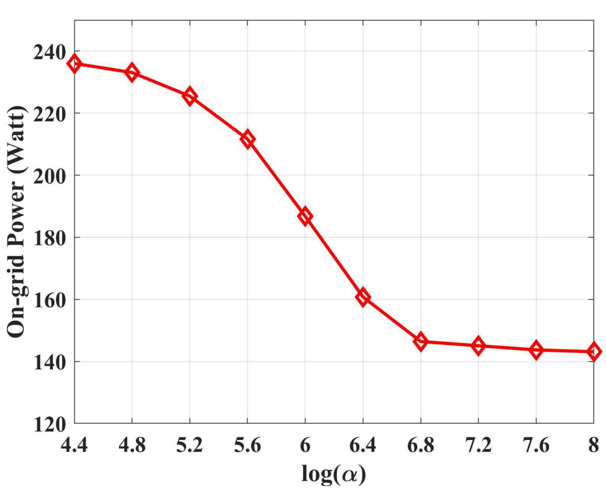


Fig. 10: The on-grid power v.s. α (user number=70).

energy consumption is still very low, thus achieving better network utilities as compared to GRA and SINR.

Fig. 5 and Fig. 6 compare the aggregated EDR and on-grid power consumption of the three algorithms with different numbers of users, respectively. From the figures, we can see that when the number of users is very small, the aggregated EDR of the proposed algorithm is a little lower than the other two algorithms. This phenomenon is attributed to the fact that when the traffic load is low (i.e., the traffic congestion of BSs is not an issue), the other two algorithms can associate users to BSs with the best channel conditions. However, the proposed scheme needs to consider the on-grid energy consumption, and thus assigns some users to the BSs with excess green energy instead of BSs providing the best channel conditions. Afterwards, as the number of users increases, the aggregated EDR of TEA is higher than those of the GRA and Best SINR algorithm, while the on-grid power consumption is remarkably lower than those of the other two algorithms. As we know, when the number of users increases, the traffic loads of BSs become an important factor for user EDR. Considering the tradeoff of the aggregated EDR and on-grid power consumption, TEA focuses on offloading the traffic from overloaded BSs to BSs with more resources, in order to optimize the performance of both the aggregated EDR and the on-grid power. In contrast, Best SINR just decides the user association based on user channel conditions, which may incur traffic congestions in BSs and degrade the aggregated EDR.

As shown in Fig. 7 and Fig. 8, we study the aggregated EDR and the on-grid power consumption of TEA under three different energy-throughput coefficients α when the number of users is increasing. A larger energy-throughput coefficient indicates that the BSs are more energy sensitive. Hence, as the number of users changes, TEA with a larger α maintains lower on-grid power consumption as compared to a smaller α. Correspondingly, TEA with a larger α will incur a lower aggregated EDR of the network.

Fig. 9 and Fig. 10 show the aggregated EDR and the on-grid power consumption of TEA when α changes continuously. In this case, the total number of users is fixed as 70. We can see that both the aggregated EDR and the on-grid power consumption decrease gradually when α increases. A larger energy-throughput coefficient indicates that the BSs are more energy sensitive. As a result, TEA achieves less on-grid power consumption. Meanwhile, the aggregated EDR is sacrificed in order to save the on-grid power. As shown in Fig. 9 and Fig. 10, the loss of the EDR is not significant as compared to the corresponding on-grid power savings. Then, the mobile provider can choose a desired α to balance the aggregated EDR and on-grid energy consumption in different time slots based on their requirements.

VII. Conclusion and Future Work

In this paper, we have proposed a throughput aware and energy aware traffic load balancing scheme, referred to as TEA, to optimize the trade-off between the aggregated EDR and on-grid energy consumption. The TEA algorithm not only considers the network throughput, but also considers the available green energy in each BS. By assigning user association, the TEA algorithm reduces the on-grid power consumption while avoiding the related BS congestion, which negatively influences the aggregated EDR of the network. As a result, the algorithm optimizes the aggregated EDR as well as the on-grid energy consumption of HCNs.

This work mainly focuses on the downlink communications of mobile networks. In the future, we will further explore to improve the EDRs of users in the uplink communications. In particular, as mobile edge computing has become a potential technology for various mobile applications [28], [29], the optimized EDR of the uplink communications will enable the workloads of users being readily offloaded to the edge computing resources (e.g., cloudlet or fog node).

REFERENCES

[1] "Cisco Visual Networking Index: Forecast and Methodology, 2014-2019 White Paper." [Online]. Available: http://www.cisco.com/c/en/us/solutions/collateral/service-provider/ip-ngn-ip-next-generation-network/white_paper_c11-481360.html

[2] M. Yousefvand, T. Han, N. Ansari, and A. Khreishah, "Distributed energy-spectrum trading in green cognitive radio cellular networks," *IEEE Transactions on Green Communications and Networking*, vol. 1, no. 3, pp. 253–263, Sept 2017.

[3] N. Ansari and T. Han, *Green Mobile Networks: A Networking Perspective*. Wiley-IEEE Press, ISBN: 978-1-119-12510-5, 2017.

[4] J. Xu, L. Duan, and R. Zhang, "Cost-aware green cellular networks with energy and communication cooperation," *IEEE Communications Magazine*, vol. 53, no. 5, pp. 257–263, 2015.

[5] N. B. Rached, H. Ghazzai, A. Kadri, and M. S. Alouini, "Energy management optimization for cellular networks under renewable energy generation uncertainty," *IEEE Transactions on Green Communications and Networking*, vol. 1, no. 2, pp. 158–166, June 2017.

[6] X. Sun, N. Ansari, and Q. Fan, "Green energy aware avatar migration strategy in green cloudlet networks," in *Proceedings - IEEE 7th International Conference on Cloud Computing Technology and Science, (CloudCom' 2015)*, Vancouver, Canada, Nov. 2015.

[7] B. Yang, G. Mao, X. Ge, M. Ding, and X. Yang, "On the energy-efficient deployment for ultra-dense heterogeneous networks with NLoS and LoS transmissions," *IEEE Transactions on Green Communications and Networking*, 2018, early access.

[8] Q. Fan and N. Ansari, "Green energy aware user association in heterogeneous networks," in *Proc. of IEEE Wireless Communications and Networking Conference (WCNC'2016)*, Doha, Qatar, Apr. 2016.

[9] H. S. Dhillon, R. K. Ganti, and J. G. Andrews, "Load-aware modeling and analysis of heterogeneous cellular networks," *IEEE Transactions on Wireless Communications*, vol. 12, no. 4, pp. 1666–1677, April 2013.

[10] C. Liu, B. Natarajan, and H. Xia, "Small cell base station sleep strategies for energy efficiency," *IEEE Transactions on Vehicular Technology*, vol. 65, no. 3, pp. 1652–1661, March 2016.

[11] Q. Fan and N. Ansari, "Throughput aware and green energy aware user association in heterogeneous networks," in *2017 IEEE International Conference on Communications (ICC)*, Paris, France, May 2017.

[12] J. Gong, J. S. Thompson, S. Zhou, and Z. Niu, "Base station sleeping and resource allocation in renewable energy powered cellular networks," *IEEE Trans. on Comm.*, vol. 62, no. 11, pp. 3801–3813, Nov. 2014.

[13] Q. Fan, N. Ansari, and X. Sun, "Energy driven avatar migration in green cloudlet networks," *IEEE Communications Letters*, vol. 21, no. 7, pp. 1601–1604, July 2017.

[14] M. J. Farooq, H. Ghazzai, A. Kadri, H. ElSawy, and M. S. Alouini, "A hybrid energy sharing framework for green cellular networks," *IEEE Transactions on Communications*, vol. 65, no. 2, pp. 918–934, Feb 2017.

[15] H. J. Hung, T. Y. Ho, S. Y. Lee, C. Y. Yang, and D. N. Yang, "Relay selection for heterogeneous cellular networks with renewable green energy sources," *IEEE Transactions on Mobile Computing*, vol. 17, no. 3, pp. 661–674, March 2018.

[16] T. Han and N. Ansari, "On optimizing green energy utilization for cellular networks with hybrid energy supplies," *IEEE Trans. on Wireless Communications*, vol. 12, no. 8, pp. 3872–3882, Aug. 2013.

[17] ——, "Heuristic relay assignments for green relay assisted device to device communications," in *Proceedings of IEEE Global Telecommunications Conference (GLOBECOM'13)*, Atlanta, GA, 2013.

[18] H. Ma, H. Zhang, X. Wang, and J. Cheng, "Backhaul-aware user association and resource allocation for massive mimo-enabled hetnets," *IEEE Comm. Letters*, vol. 21, no. 12, pp. 2710–2713, Dec 2017.

[19] B. Wang, Q. Kong, W. Liu, and L. T. Yang, "On efficient utilization of green energy in heterogeneous cellular networks," *IEEE Systems Journal*, vol. 11, no. 2, pp. 846–857, June 2017.

[20] F. Kong, X. Sun, V. C. M. Leung, and H. Zhu, "Delay-optimal biased user association in heterogeneous networks," *IEEE Transactions on Vehicular Technology*, vol. 66, no. 8, pp. 7360–7371, Aug 2017.

[21] T. Han and N. Ansari, "A traffic load balancing framework for software-defined radio access networks powered by hybrid energy sources," *IEEE/ACM Trans. on Networking*, vol. 24, no. 2, pp. 1038–1051, Apr. 2016.

[22] S. Corroy, L. Falconetti, and R. Mathar, "Dynamic cell association for downlink sum rate maximization in multi-cell heterogeneous networks," in *Proceedings of IEEE International Conference on Communications (ICC'12)*, Ottawa, Canada, Jun. 2012.

[23] H. S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous cellular networks with flexible cell association: A comprehensive downlink SINR analysis," *IEEE Transactions on Wireless Communications*, vol. 11, no. 10, pp. 3484–3494, 2012.

[24] Y. Kim, S. Lee, and D. Hong, "Performance analysis of two-tier femtocell networks with outage constraints," *IEEE Transactions on Wireless Communications*, vol. 9, no. 9, pp. 2695–2700, 2010.

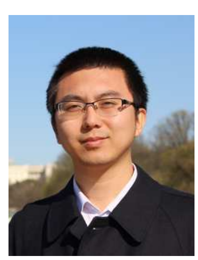
[25] L. Kleinrock, *Queueing Systems: Computer Applications*. Wiley-Interscience, 1976.

[26] "Evolution of land mobile radio (including personal) ccommunications: Cost 231." [Online]. Available: http://www.awe-communications.com/Propagation/Urban/COST/

[27] C. Riordan and R. Hulstron, "What is an air mass 1.5 spectrum? [solar cell performance calculations]," in *IEEE Conference on Photovoltaic Specialists*, vol. 2, May 1990, pp. 1085–1088.

[28] Q. Fan and N. Ansari, "Application aware workload allocation for edge computing based iot," *IEEE Internet of Things Journal*, 2018, DOI:10.1109/JIOT.2018.2826006, early access.

[29] ——, "Workload allocation in hierarchical cloudlet networks," *IEEE Comm. Letters*, vol. 22, no. 4, pp. 820–823, Apr. 2018.



Qiang Fan (S'15) received his M.S. degree in Electrical Engineering from Yunnan University of Nationalities, China, in 2013. He is currently a research assistant and a Ph.D. candidate in the Department of Electrical and Computer Engineering, New Jersey Institute of Technology (NJIT), Newark, New Jersey, USA. His research interests include mobile cellular networks, mobile cloud computing, Internet of things, and free space optical communications.

Nirwan Ansari (S'78-M'83-SM'94-F'09) received his B.S.E.E. degree from NJIT, M.S.E.E. degree from the University of Michigan, Ann Arbor, and Ph.D. degree from Purdue University, West Lafayette. He is Distinguished Professor of Electrical and Computer Engineering at NJIT. He has also been a visiting (chair) professor at several universities. He recently authored Green Mobile Networks: A Networking Perspective (Wiley-IEEE, 2017) with T. Han, and co-authored two other books. He has also (co-)authored more than 500 technical publications, over 250 in widely cited journals/magazines. He has guest-edited a number of special issues covering various emerging topics in communications and networking. He has served on the editorial/advisory board of over ten journals. His current research focuses on green communications and networking, cloud computing, and various aspects of broadband networks. Some of his recognitions include several Excellence in Teaching Awards, a few best paper awards, the NCE Excellence in Research Award, the ComSoc AHSN TC Technical Recognition Award, the IEEE TCGCC Distinguished Technical Achievement Recognition Award, the NJ Inventors Hall of Fame Inventor of the Year Award, the Thomas Alva Edison Patent Award, Purdue University Outstanding Electrical and Computer Engineer Award, and designation as a COMSOC Distinguished Lecturer. He has also been granted 36 U.S. patents.