Hindawi Publishing Corporation
Mathematical Problems in Engineering
Volume 2016, Article ID 4567625, 11 pages
http://dx.doi.org/10.1155/2016/4567625

Research Article

A Submodular Optimization Framework for Outage-Aware Cell Association in Heterogeneous Cellular Networks

Gongchao Su,[1,2] Bin Chen,[2] Xiaohui Lin,[2] Hui Wang,[2] and Lemin Li[1]

[1] *School of Communication and Information Engineering, University of Electronic Science and Technology of China, Chengdu 611731, China*
[2] *Shenzhen Key Laboratory of Advanced Communications and Information Processing, Shenzhen University, Shenzhen 518060, China*

Correspondence should be addressed to Gongchao Su; gcsu@szu.edu.cn

Received 3 November 2015; Revised 3 February 2016; Accepted 11 February 2016

Academic Editor: Luciano Mescia

In cellular heterogeneous networks (HetNets), offloading users to small cell base stations (SBSs) leads to a degradation in signal to interference plus noise ratio (SINR) and results in high outage probabilities for offloaded users. In this paper, we propose a novel framework to solve the cell association problem with the intention of improving user outage performance while achieving load balancing across different tiers of BSs. We formulate a combinatorial utility maximization problem with weighted BS loads that achieves proportional fairness among users and also takes into account user outage performance. A formulation of the weighting parameters is proposed to discourage assigning users to BSs with high outage probabilities. In addition, we show that the combinatorial optimization problem can be reformulated as a monotone submodular maximization problem and it can be readily solved via a greedy algorithm with lazy evaluations. The obtained solution offers a constant performance guarantee to the cell association problem. Simulation results show that our proposed approach leads to over 30% reduction in outage probabilities for offloaded users and achieves load balancing across macrocell and small cell BSs.

1. Introduction

In recent years, wireless cellular networks are evolving towards increasing heterogeneity to cope with the exponential growth of mobile devices and data traffic. With the proliferation of low power base stations (BSs), such as picocells and femtocells, heterogeneous networks (HetNets) can significantly boost network capacity by providing more radio resources and allowing more aggressive frequency reuse. HetNets with dense deployments of transmission points promise large gains in area spectral efficiency and are envisioned as one of the key technologies to achieve 1000x increase in data rates in future 5G wireless communication systems [1].

The dense deployment of small cell BSs (SBSs) makes cell association quite challenging. Conventional schemes such as max signal to interference plus noise ratio (SINR) association, where users are simply associated with the strongest BS in terms of received signal strength, do not work well under the HetNets scenarios. Due to the massive difference in transmission power of SBSs and macrocell BSs (MBSs), this scheme leads to severe load imbalance and renders the small cells underutilized [2]. Hence novel cell association schemes that can proactively offload users to small cells need to be developed to achieve load balancing among MBSs and SBSs.

However, offloading users to small cells can bring undesired consequences. In a HetNet scenario where SBSs are cochannelly deployed with MBSs, offloaded users suffer from strong interference from the MBS which leads to a degradation in SINR. In fast fading channels this degradation in user SINR results in a higher probability of outage, which occurs when the instantaneous SINR falls below a given threshold. This will greatly degrades the user Quality of Service (QoS) experiences. Therefore, it will be desired that we take outage performance into consideration when devising cell association schemes.

Cell association in the context of HetNets has received much recent attention. In [3], several popular approaches are highlighted to cope with the load balancing problem in HetNets. A practical and simple method is Cell Range Expansion [4], which applies a predetermined bias to the received reference signal power from the small cells in order to extend their coverage area. Although this method is simple to implement, it is difficult to prescribe the optimal bias values. Another approach is to adopt a game theoretical framework, which allows decentralized solutions. For example, in [5] the cell association problem is modeled as a noncooperative game and a distributed learning algorithm is developed to find a suboptimal user association. Furthermore, the authors of [6] formulate the user association problem as a many to one matching game with externalities and an iterative deferred acceptance algorithm is developed to find a stable matching. However, these methods aim to find the solutions that achieve Nash equilibrium or stable outcomes but they do not necessarily yield optimal solutions. Another popular approach is to adopt a utility optimization framework and leverage optimization techniques. In such a framework cell association can be studied along with other resource management schemes such as frequency reuse, beamforming, power control, and resource allocation. There exist a wide variety of choices for the utility functions. A linear programming (LP) problem with respect to cell association is formulized in [7] to maximize average user throughput. In [8] a joint optimization problem over beam forming vectors and user scheduling is proposed to maximize overall network throughput. Another joint optimization problem over cell association and power control is formulized in [9] to maximize system throughput and reduce overall energy consumption. However, these problem formulations do not directly encourage load balancing. To achieve load balancing we can choose a problem formulation with objective functions that promote fairness among users. Such fairness would encourage offloading users to less congested small cells. For example, max-min fairness objective function is chosen in [10] to achieve fairness through joint cell association and power control. A more popular choice is a broad class of utility functions referred to as α-fairness utilities [11, 12], where α is a tunable fairness parameter to achieve different tradeoffs between network throughput and user fairness.

The optimization over cell association is in essence a combinatorial optimization problem due to the requirement that each user must be assigned to a single BS. A straightforward way is to circumvent this constraint and assume that users can be simultaneously associated with multiple BSs. The relaxation in constraints results in a convex optimization problem. In [13] the relaxed optimization problem with proportional fairness utility is solved via Lagrangian dual decomposition. A coordinate descend method is used to solve the dual problem in [14]. A dynamic cell association scheme and range extension algorithm are proposed in [15]. In [16] the relaxed optimization problem is extended to HetNets with massive MIMO BSs. However, the result is a fractional solution and multiple BSs association is difficult to implement. Moreover, the obtained solution is upper bounded. Instead of relaxing the constraints, a recent work [17] shows that the combinatorial cell association problem can be reformulated as a submodular maximization problem with matroid constraint and can be solved efficiently using greedy methods with low complexity. The obtained solution is an integer solution and provides a constant factor approximation for the problem. The result is extended to more general cases of α-fairness utility functions in [18].

To alleviate the QoS degradation for offloaded users, cell association can be studied alongside other techniques such as power control and intercell interference coordination (ICIC) using Almost Blank Subframes (ABS) [19–21]. For example, enhanced ICIC (eICIC) is usually used in combination with CRE to help offloading traffic to small cells and reduce interference for offload users. Macrocell BSs can be muted on ABS subframes and a high bias value can be applied to SBSs so as to attract more users, which are immune to the strong intercell interference from macrocells. Determining the optimal ABS density is crucial as macrocell BSs stop transmission in ABS subframes. A high ABS density benefits small cell users but causes performance degradations for macrocell users. A low ABS density may not be enough to mitigate interference for offloaded users. However, jointly determining the optimal cell-specific bias and ABS density is shown to be intractable [22]. Moreover, the eICIC schemes do not address intratier interference between neighboring small cells. As HetNets are expected to become increasingly dense, interference between small cells may become a dominating factor [23]. Another approach for interference management is to jointly optimize user association and BS transmission powers [24]. The resulting optimization problem is a joint optimization problem subject to QoS constraints such as minimum SINR or rate constraints. The joint optimization problem is usually solved in an alternating fashion. However, this approach has some limitations. The joint optimization problem has a high computational complexity. When dealing with the QoS constraints this complexity can rapidly increase and the feasibility of the solution is difficult to guarantee. Furthermore implementing power control or ICIC requires extensive cooperation among all BSs in HetNets.

In addition, few researches on cell association address the outage performance in fast fading channels. A recent work [25] considers the outage as the event of the instantaneous rate falling below a threshold and formulizes a joint optimization problem over cell association and resource allocation. The joint optimization problem is then decomposed into two subproblems: user association subproblem subject to long term QoS constraints and rate-based outage minimization subproblem over resource allocation. Since rate is a combination of allocated resources and spectral efficiency, the rate-based outage probability is minimized by allocating an optimal number of resource blocks to users. Both subproblems are combinatorial optimization problems and they are relaxed to convex optimization problems and then solved iteratively using an alternating optimization approach. Although their method reduces rate-based outage probabilities for offload users, the degradation in SINR is not addressed as the rate-based outage can always be alleviated by allocating more resources.

This paper addresses the cell association problem with the aim of achieving load balancing and alleviating QoS degradation for offloaded users. We intend to improve the outage performance for offloaded users with the absence of other techniques such as power control and ICIC. To do so we need to determine whether a user should be offloaded and then choose the proper candidate BS for offloading. Hence it is desirable to devise a load balancing scheme that takes into consideration the QoS performance such as outage, along with the loads of BSs. Towards this end, we summarize our contributions as follows:

(i) We propose a novel problem formulation for load balancing that incorporate both loads of BSs and outage performance of user-BS links. We adopt the log utility as the system wide utility and propose a weighted split term formulation by assigning a set of weights to the cost incurred by the loads of BSs. The weight is defined as a barrier function with respect to outage probability. Such a problem formulation inherits the load balancing capabilities of the utility maximization framework and discourages assigning users to a small cell with a high probability of outage.

(ii) For the formulated problem we first prove that the underlying utility function is a normalized nonmonotone submodular function. We then show that the nonmonotone submodular maximization problem can be reformulated into a monotone submodular maximization problem and approximated by a greedy algorithm. The greedy algorithm has a low computational complexity and is useful for densely deployed HetNets with a large number of users and BSs.

(iii) Finally, we compare our results with the log utility maximization scheme and the baseline max SINR scheme via extensive simulation over a two-tier HetNet topology and highlight the significant reduction in outage performance for the offloaded users.

In general we present a user QoS aware cell association framework that takes into account base station transmission power levels, base station loads, and user outage performance. We show that making user association decisions aware of QoS requirements benefits HetNet users. Our results provide a way to combat SINR degradation for offloaded users in a scenario where ICIC and power control are not employed.

The rest of this paper is organized as follows. The system model is presented in Section 2. We formulate our maximization problem in Section 3. In Section 4 we analyze the submodularity of the utility maximization problem in a set function form. The two-stage greedy algorithm is presented in Section 5. Numeral results are presented in Section 6 and Section 7 concludes the paper.

2. System Model

We consider a multitier downlink HetNet consisting of MBSs and a set of SBSs of various types. We consider the coverage area as a finite Euclidean plane with N users and M BSs. Let $U = \{1, \ldots, N\}$ denote the set of users and $B = \{1, \ldots, M\}$ denote the set of BSs. Each BS $j \in B$ has a fixed transmission power P_j. A frequency reuse factor of 1 is assumed. We also assume that SBSs are densely deployed across the coverage area so that the HetNet is interference limited.

For each user $i \in U$, the power received from BS j is given by $G_{ij}F_{ij}P_j$ where G_{ij} models the large scale fading components including path loss, shadowing, and antenna gains, and F_{ij} represents the small scale fading component, which models Rayleigh fading. Hence F_{ij} is an exponentially distributed random variable and we assume it has a unit variance. Note that G_{ij} remains constant during cell association phase. The received power is also an exponentially distributed random variable with mean value:

$$E\left[G_{ij}F_{ij}P_j\right] = G_{ij}P_j. \quad (1)$$

In an interference limited HetNet, SINR is reduced to signal to interference ratio (SIR). The instantaneous SIR seen from BS j is given by

$$\mathrm{SIR}_{ij} = \frac{G_{ij}F_{ij}P_j}{\sum_{k\in B/j} G_{ik}F_{ik}P_k}. \quad (2)$$

Cell association is assumed to be carried out over a large time scale compared to the channel fluctuations [13]. Thus fast fading is averaged out and the SIR remains as a constant during the entire association time. The constant SIR, denoted as $\overline{\mathrm{SIR}}$, is referred to as the long term SIR and given by [22]

$$\overline{\mathrm{SIR}_{ij}} = \frac{G_{ij}P_j}{\sum_{k\in B/j} G_{ik}P_k}. \quad (3)$$

Accordingly, the long term spectral efficiency for user i served by BS j, denoted as s_{ij}, is also a function of the long term SIR, which is written as [22]

$$s_{ij} = \log_2\left(1 + \overline{\mathrm{SIR}_{ij}}\right) = \log_2\left(1 + \frac{G_{ij}P_j}{\sum_{k\in B/j} G_{ik}P_k}\right). \quad (4)$$

The long term rate for user i served by BS j, denoted as r_{ij}, is given by

$$r_{ij} = y_{ij}s_{ij}, \quad (5)$$

where y_{ij} denotes the portion of resources allocated by BS j to user i. Specifically, with an equal resource allocation policy, we can rewrite r_{ij} as

$$r_{ij} = \frac{s_{ij}}{\sum_{i\in U} x_{ij}}, \quad (6)$$

where $x_{ij} \in \{0, 1\}$, $(i, j) \in U \times B$ is a binary indicator that denotes whether or not user i is associated with BS j. To enforce a single BS association for each user, we have

$$\sum_{j\in B} x_{ij} = 1, \quad \forall i \in U. \quad (7)$$

The utility maximization framework for the load balancing problem involves finding the appropriate set of $\{x_{ij}\}$ that maximize the aggregate utilities of user rate r_{ij}. Clearly such a framework does not involve the dynamics of channel fluctuation and the related outage performance. The outage event occurs when the instantaneous SIR at user i served by BS j falls below a given threshold γ^{th}. Let P_{ij}^{out} denote the outage probability for user i served by BS j. In an interference limited network with Rayleigh fading channels, the probability of outage is given in [26] and is written as

$$P_{ij}^{\text{out}} = \text{Prob}\left(\text{SIR}_{ij} \leq \gamma^{\text{th}}\right) = 1 - \prod_{k \in B/j} \frac{\lambda_{ik}}{\lambda_{ik} + \gamma^{\text{th}} \lambda_{ij}}, \quad (8)$$

where $\lambda_{ij} = 1/E[G_{ij}F_{ij}P_j] = 1/G_{ij}P_j$ is a constant during cell association phase. Note that the deterioration in average received power $G_{ij}P_j$ leads to an increase of the outage probability P_{ij}^{out} and vice versa. When users are offloaded to small cells, they suffer from severe degradations in SINR and higher probabilities of outage. Therefore, we should discourage offloading users to a small cell with a high probability of outage.

We note that the above system model corresponds to a worst case scenario where MBSs and SBSs are cochannelly deployed and interference mitigation schemes are not employed. In such cases cell association needs to be devised in order to balance traffic loads as well as alleviate SINR degradations for offloaded users without the use of ICIC and power control schemes. The system model corresponds to a HetNet scenario with Single-Input-Single-Output (SISO) configuration. With massive MIMO enabled small cells, the scenario is much different since massive MIMO leads to an increased spectral efficiency and interference seen by a user includes not only the intercell interference but also the intracell interference which is dependent on the channel coefficients of all the antenna of a BS. Interference may also behave differently with advanced beamforming [27]. We leave that topic for future research.

3. Problem Formulation

We adopt the log function as the utility for the load balancing problem. It has been shown in [13] that, with the log utility, the optimal resource allocation policy is equal allocation. Therefore, the aggregate utility function is given by

$$\sum_{i \in U} \sum_{j \in B} x_{ij} \log\left(s_{ij}\right) - \sum_{j \in B} \left(\sum_{i \in U} x_{ij}\right) \log\left(\sum_{i \in U} x_{ij}\right), \quad (9)$$

which is a split term formulation. The first term is an assignment cost with respect to user spectral efficiency, and the second term is a penalty term with respect to BS loads. The solution is obtained by solving the following optimization problem [13, 14, 17, 20–22]:

$$\begin{aligned} \max_{x} \quad & \sum_{i \in U} \sum_{j \in B} x_{ij} \log\left(s_{ij}\right) - \sum_{j \in B} \left(\sum_{i \in U} x_{ij}\right) \log\left(\sum_{i \in U} x_{ij}\right) \\ \text{s.t.} \quad & x_{ij} \in \{0, 1\} \\ & \sum_{j \in B} x_{ij} = 1. \end{aligned} \quad (10)$$

It can be seen from (10) that each user-BS pair (i, j) is associated with a penalty $\log(\sum_{i \in U} x_{ij})$. Assigning users to more congested BSs triggers a larger penalty. Hence the optimal solution always discourage selecting a highly congested BS for user i. As the MBS is usually much more congested than the SBSs, the optimal solution will then offload some users to less congested SBSs. However, the penalty is only dependent on BS loads and does not involve the outage performance. We can deduce that in some cases the optimal solution will assign users to an underutilized BS with low SIR and a high probability of outage. We show an example of such instances.

Consider a simple network consisting of two BSs $B = \{j, k\}$. Without loss of generality we assume the load of BS j, denoted as L_j, is much larger than the load of BS k, denoted as L_k. We also assume that for user i the received signal power from BS k is weak compared to the power of signals from BS j. Thus we have $L_j \gg L_k$ and $\text{SIR}_{ij} \gg \text{SIR}_{ik}$. Since the optimization problem (10) is a linear programming (LP) problem with respect to association indicators x_{ij}, the optimal solution will choose the user-BS pair with a larger weight which is denoted as $\log(s_{ij}/\sum_{i \in U} x_{ij})$. Therefore, the optimal solution will assign user i to BS k if

$$\log\left(\frac{s_{ik}}{L_k}\right) > \log\left(\frac{s_{ij}}{L_j}\right). \quad (11)$$

Substituting (4) into (11) we have

$$\text{SIR}_{ik} > \left(1 + \text{SIR}_{ij}\right)^{L_k/L_j} - 1. \quad (12)$$

We can see from (12) that if $L_j \gg L_k$ we have $L_k/L_j \to 0$ and $(1 + \text{SIR}_{ij})^{L_k/L_j} \to 1$. This indicates that (12) is satisfied once $\text{SIR}_{ik} > 0$. Thus the optimal solution may assign user i to BS k with a very low SIR, resulting in a high outage probability. If BS j is a heavily congested MBS and BS k is a heavily underutilized SBS, we end up with high outage probabilities for users offloaded to the SBS.

The above example shows that the utility maximization framework itself cannot provide satisfactory QoS guarantees for offloaded users. To address this problem the utility maximization problem can be reformulated into a joint optimization problem over cell association and other network resources management schemes such as power and interference control. However, the joint optimization problem is more difficult to solve and the solution is more difficult to implement.

In order to address the QoS degradation for offloaded users, we need to incorporate user outage performance with our problem formulation. When the outage constraint dictates that the outage probability for a user must be lower than a threshold, the cell association scheme needs to guarantee that a user must be assigned to a BS which satisfies the constraint. Without power control this combinatorial optimization with outage constraint is very difficult to solve and possibly infeasible due to limited solution spaces. We also note that the SINR based outage probability is given by a per link basis and we have to guarantee that a user must be served by a single BS. This is different from the rate-based outage event in [25] because when evaluating outage in a rate basis we can safely assume a user can be associated with multiple BSs and the rate for this user is the sum of rates from multiple links as it is in [25]. For SINR based outage this cannot be the case. Hence we do not relax the combinatorial optimization problem into a convex one. In order to find a feasible solution for this combinatorial problem with constraints we propose a split term formulation with weighted loads that incorporates user outage performance. The intention of the problem formulation is to discourage assigning users to BSs with high outage probabilities. Towards this end, we note that the penalty term in (10) is only dependent on BS loads. To incorporate user outage performance, we assign a set of outage-dependent weighting parameters β_{ij} to BS loads. We define β_{ij} as

$$\beta_{ij} = \frac{1}{1 - P_{ij}^{\text{out}}} \tag{13}$$

and the optimization problem is reformulated as

$$\begin{aligned} \max_{x} \quad & \sum_{i\in U}\sum_{j\in B} x_{ij} \log\left(s_{ij}\right) \\ & - \sum_{j\in B}\left(\sum_{i\in U} x_{ij}\right) \log\left(\beta_{ij}\sum_{i\in U} x_{ij}\right) \\ \text{s.t.} \quad & x_{ij} \in \{0,1\} \\ & \sum_{j\in B} x_{ij} = 1. \end{aligned} \tag{14}$$

We can see from the split term problem formulation (14) that each user-BS pair (i, j) is now associated with a weighted penalty that depends not only on the load of BS j, but also on the weighting parameter β_{ij}. It can be seen from definition (13) that β_{ij} is monotone increasing with P_{ij}^{out}. In fact, we can see that $\beta_{ij} \geq 1$ always holds which indicates that any selection of heavily congested BSs is always penalized. If $P_{ij}^{\text{out}} \to 0$ then we have $\beta_{ij} \to 1$. Then the formulated problem (14) is identical to the utility maximization problem (10). Since $P_{ij}^{\text{out}} \to 0$ indicates that no outage events occur, we preserve the load balancing formulation of (10). However, if $P_{ij}^{\text{out}} \to 1$ we have $\beta_{ij} \to \infty$. Then any choice of (i, j) pair with large outage probability is heavily penalized. Substituting (13) into (14) we have the triple term formulation of the optimization problem as

$$\begin{aligned} \max_{x} \quad & \sum_{i\in U}\sum_{j\in B} x_{ij} \log\left(s_{ij}\right) - \sum_{j\in B}\left(\sum_{i\in U} x_{ij}\right) \log\left(\sum_{i\in U} x_{ij}\right) \\ & - \sum_{i\in U}\sum_{j\in B} x_{ij} \log\left(\beta_{ij}\right) \\ \text{s.t.} \quad & x_{ij} \in \{0,1\} \\ & \sum_{j\in B} x_{ij} = 1, \end{aligned} \tag{15}$$

where the last term is analogous to the log barrier function [28] with respect to outage probability P_{ij}^{out}. Compared with the nonweighted optimization problem (10) we note that the only difference between (15) and (10) is that we have replaced $\log(s_{ij})$ with $\log(s_{ij}) - \log(\beta_{ij})$.

In practice in order to avoid computation of division by zero we define β_{ij} as

$$\beta_{ij} = \frac{1}{1 - P_{ij}^{\text{out}} + \varepsilon}, \tag{16}$$

where $0 < \varepsilon \ll 1$ is a constant with a small value. Then as $P_{ij}^{\text{out}} \to 1$ we have $\beta_{ij} \to 1/\varepsilon$.

The reformulated problem (15) is a combinatorial optimization problem. It is shown in [3, 18] that the sum log utility maximization problem (10) is NP hard. If there exists an optimal polynomial time algorithm for (15), we can solve (10) in polynomial time by replacing the constant $\log(s_{ij})$ with $\log(s_{ij}) - \log(\beta_{ij})$. Hence (15) is also NP hard. To solve the problem we adopt the submodular maximization approach proposed in [19], which is used to solve the log utility maximization problem (10) and extended to α-fairness utility maximization in [18]. The benefit of submodular optimization is that it yields a discrete solution on cell association, which is more practical to implement than the fractional solution obtained through relaxed convex optimization.

4. Submodular Function Maximization

We proceed to analyze the submodularity of the reformulated problem (15). Towards this end, the objective function in (15) is interpreted as a set function. Let $\mathscr{L} \subseteq 2^V$ denote a collection of subsets of V so that the (i, j) pairs in the subsets have mutually distinctive users. We define a ground set $V = \{(i, j) : i \in U, j \in B\}$ which consists of all possible user-BS associations. We use $|\cdot|$ to denote the cardinality function. Let $\mathscr{G}$ denote the selected (i, j) pairs for the cell association. Clearly $\mathscr{G} \in \mathscr{L}$. We use a ground set $V^{(j)} = \{(i, j) : i \in U\}$ to denote all the possible user-BS association for a given BS j. Similarly, we use another ground set $V^{(i)} = \{(i, j) : j \in B\}$ to denote all the possible user-BS association for a given user i.

We first give some basic definitions for submodular set functions [29].

Definition 1. A set function $f : 2^V \to \mathbb{R}$ is submodular if for every $A \subseteq B \subseteq V$ and $a \in V \setminus B$ it holds that

$$f(A \cup \{a\}) - f(A) \leq f(B \cup \{a\}) - f(B), \tag{17}$$

and it is modular if it holds that

$$f(A \cup \{a\}) - f(A) = f(B \cup \{a\}) - f(B). \tag{18}$$

Definition 2. A set function $f : 2^V \to \mathbb{R}$ is a normalized set function if $f(\varnothing) = 0$ where $\varnothing$ denotes the empty set. Further, it is monotone if, for every $A \subseteq B \subseteq V$, $f(A) \leq f(B)$.

We now rewrite the objective function of (15) into a set function form. Note that s_{ij} and β_{ij} are all constants in cell association. Let $v \in \mathscr{G}$ denote a selected user-BS pair and define its weight function $w_\beta(v) = \log(s_{ij}) - \log(\beta_{ij})$; we can rewrite (15) as

$$\begin{aligned} \max_{\mathcal{G}\subseteq\mathcal{L}} \quad & \sum_{v\in\mathcal{G}} w_\beta(v) - \sum_{j\in B}\left|\mathcal{G}\cap V^{(j)}\right|\log\left(\left|\mathcal{G}\cap V^{(j)}\right|\right) \\ \text{s.t.} \quad & \left|\mathcal{G}\cap V^{(i)}\right| = 1. \end{aligned} \tag{19}$$

Proposition 3. *The objective function in (19) is a submodular set function.*

Proof. It can be easily verified that $g(\mathcal{G}) = \sum_{v\in\mathcal{G}} w_\beta(v)$ is a modular function since for every $A \subseteq B \subseteq \mathcal{G}$ and $a \in \mathcal{G}\setminus B$ it always holds that

$$g(A\cup\{a\}) - g(A) = w_\beta(a) = g(B\cup\{a\}) - g(B). \tag{20}$$

Moreover, it has been proven in [17] that $-|\mathcal{G}|\log|\mathcal{G}|$ is submodular. Due to the fact that submodularity is preserved under restriction and the objective function in (19) is a mutual sum of submodular functions and modular functions, we can conclude that the objective function in (19) is a submodular set function. □

Corollary 4. *The objective function in (19) is a normalized nonmonotone submodular set function.*

Proof. We adopt the convention that $0\log 0 = 0$. Then we have $f(\varnothing) = 0$ for the objective function. Hence it is a normalized set function. Furthermore, as the weight $w_\beta(v) = \log(s_{ij}) - \log(\beta_{ij})$ is not nonnegative, we can conclude that the objective function in (19) is nonmonotone. □

Therefore, the optimization problem (19) is a nonmonotone submodular maximization problem with equality constraints on a partition matroid. We now reformulate (19) into a monotone submodular maximization problem with inequality constraints on a partition matroid.

We first revisit the optimization problem (15) and we have the following proposition.

Proposition 5. *The optimization problem (15) is equivalent to*

$$\begin{aligned} \max_{x} \quad & \sum_{i\in U}\sum_{j\in B} x_{ij}\left(K + \log\left(s_{ij}\right) - \log\left(\beta_{ij}\right)\right) \\ & - \sum_{j\in B}\left(\sum_{i\in U} x_{ij}\right)\log\left(\sum_{i\in U} x_{ij}\right) \\ s.t. \quad & x_{ij} \in \{0,1\} \\ & \sum_{j\in B} x_{ij} = 1, \end{aligned} \tag{21}$$

where K is a constant with any arbitrary value.

Proof. From the integer constraint $\sum_{j\in B} x_{ij} = 1$ we can see that it always holds that

$$\sum_{i\in U}\sum_{j\in B} x_{ij} = N. \tag{22}$$

Hence the newly added term $\sum_{i\in U}\sum_{j\in B} x_{ij}K = KN$ is a constant. We can conclude that (21) is equivalent to (19). □

We can now rewrite (21) into a set function form. Let $v \in \mathcal{G}$ and $w_{\beta,K}(v) = K + \log(s_{ij}) - \log(\beta_{ij})$; we have the following proposition.

Proposition 6. *The submodular maximization problem (19) is equivalent to*

$$\begin{aligned} \max_{\mathcal{G}\subseteq\mathcal{L}} \quad & \sum_{v\in\mathcal{G}} w_{\beta,K}(v) - \sum_{j\in B}\left|G\cap V^{(j)}\right|\log\left(\left|G\cap V^{(j)}\right|\right) \\ s.t. \quad & \left|G\cap V^{(i)}\right| \le 1, \end{aligned} \tag{23}$$

if K satisfies

$$\begin{aligned} K > {} & N\log N - (N-1)\log(N-1) \\ & - \min_{(i,j)}\left\{\log s_{ij}\right\} - \log\varepsilon. \end{aligned} \tag{24}$$

Proof. Since $\beta_{ij} \le 1/\varepsilon$, we can see that if (24) holds then it holds that

$$\begin{aligned} K > {} & N\log N - (N-1)\log(N-1) \\ & - \min_{(i,j)}\left\{\log s_{ij} - \log\beta_{ij}\right\}. \end{aligned} \tag{25}$$

We note that the only difference between (15) and (10) is that we have replaced $\log(s_{ij})$ with $\log(s_{ij}) - \log(\beta_{ij})$. For the nonweighted load optimization problem ($\beta_{ij} = 1$) it has been shown in [17] that it is monotone submodular if

$$\begin{aligned} K > {} & N\log N - (N-1)\log(N-1) \\ & - \min_{(i,j)}\left\{\log s_{ij}\right\}. \end{aligned} \tag{26}$$

Hence we can deduce that the objective function in (23) is also monotone submodular. The optimal solution to (23) will always assign a user to a BS $j \in B$, since assigning a user to a BS always brings positive gains. Hence it holds that $|\mathcal{G}\cap V^{(i)}| > 0$. Therefore, the optimal solution of (23) satisfies $|\mathcal{G}\cap V^{(i)}| = 1$. We can conclude that (23) is equivalent to (19). □

The reformulated problem (23) is a monotone submodular maximization problem subject to a partition matroid constraint. For monotone submodular maximization problems, we can leverage the greedy algorithm [29] which provides a near-optimal solution. The greedy algorithm yields an integer solution to cell association, which guarantees a single BS association for each user.

5. The Greedy Algorithm with Lazy Evaluations

In this section we consider approximating the submodular maximization problem (23) through a greedy algorithm. The greedy algorithm starts with an empty set $\mathcal{G}_0$, and in ith iteration it adds the element v that maximize the gain $\Delta(v|\mathcal{G}_{i-1}) = f(\{v\}\cup\mathcal{G}_{i-1}) - f(\mathcal{G}_{i-1})$ where $f(\mathcal{G})$ denote

(1) **Initialize:** Define ground set V, set $\mathcal{G} = \varnothing$, $k = N$. Set $\{\rho(v) : v \in V\}$ to ∞.
(2) **Repeat**
(3) Sort $\{\rho(v)\}$ in descending order and return the list V'.
(4) **for** $v \in V'$
    $\Delta_{\max} \leftarrow 0$.
    if $\rho(v) \geq \Delta_{\max}$
        Compute $\Delta(v \mid \mathcal{G}) = f(\{v\} \cup \mathcal{G}) - f(\mathcal{G})$.
            $\rho(v) \leftarrow \Delta(v \mid \mathcal{G})$.
            $\Delta_{\max} \leftarrow \max\{\Delta_{\max}, \Delta(v \mid \mathcal{G})\}$.
        else
            break
    End if
End for
(5) Determine $v \in \arg\max_{v \in V}\{\rho(v)\}$ and the corresponding user i. Find $\{v_1 : (i,k) \mid k \in B\}$
(6) Update $\mathcal{G} = \mathcal{G} \cup \{v\}$, $V = V \setminus \{v_1\}$, $k = k - 1$ and update $\{\rho(v) : v \in V\}$
(7) **Until** $k = 0$.
(8) **Output** $\mathcal{G}$.

Algorithm 1: The greedy algorithm with lazy evaluations.

the objective function in (23). Catering to the partition matroid constraint $|\mathcal{G} \cap V^{(i)}| \leq 1$, once an element $v \in V$ is selected we need to remove all (i,j) pairs that share the same user with v from the ground set V. Furthermore, due to the monotonicity of the submodular function, we can speed up the greedy process through lazy evaluations [23]. The submodularity of the objective function in (23) guarantees that the incremental gains from any element $v \in V$ are monotonically nonincreasing during the iterations of the algorithm; that is, $\Delta(v \mid \mathcal{G}_i) \geq \Delta(v \mid \mathcal{G}_j)$, $i < j$. The greedy algorithm maintains a list of upper bounds $\rho(v)$ on the incremental gains sorted in decreasing order. Initially $\rho(v)$ is set to ∞ and updated with $\Delta(v \mid \mathcal{G}_0)$ at first iteration. In ith iteration, we have the solution set $\mathcal{G}_{i-1}$ and the updated ground set V from the previous iteration. Instead of computing $\Delta(v \mid \mathcal{G}_{i-1})$ for every $v \in V$, the algorithm extracts the first element $\{v\}$ from the ordered list and then updates the corresponding bound $\rho(v)$ with the real gain $\Delta(v \mid \mathcal{G}_{i-1})$ for current iteration. It then proceeds to evaluate the next element v' from the ordered list and if $\rho(v') \leq \rho(v)$ then $\Delta(v \mid \mathcal{G}_{i-1}) \geq \rho(v') = \Delta(v' \mid \mathcal{G}_j) \geq \Delta(v' \mid \mathcal{G}_{i-1})$, $j < i-1$, $\forall v' \in V \setminus \{v\}$. The element with the maximal incremental gain in current iteration is identified and the algorithm can now proceed to the next iteration without computing $\Delta(v' \mid \mathcal{G}_{i-1})$ for the remaining $v' \in V$. The algorithm stops only if every user is associated with a BS. The algorithm is described in Algorithm 1.

It can be seem from Algorithm 1. that in each iteration a user is assigned to a serving BS. Hence the greedy algorithm terminates after N iterations. The computational complexity is only dependent on the number of users and the complexity is low. Therefore, it is useful in scenarios with dense HetNets deployment with a large number of users and BSs. It is well known that for the monotone submodular maximization problem the greedy algorithm provides a $1 - 1/e$ (or equivalently 0.631) approximation [29]. Since the submodular maximization problem (23) is equivalent to (19) and the objective function of (23) is the sum of the objective function of (19) and a constant, we have the following proposition.

Proposition 7. *Let $\mathcal{G}$ be the solution obtained via the greedy algorithm and $\mathcal{G}^*$ be the optimal set for the submodular optimization problem (21). Let $f(\cdot)$ be the objective submodular function of (21). Then we have*

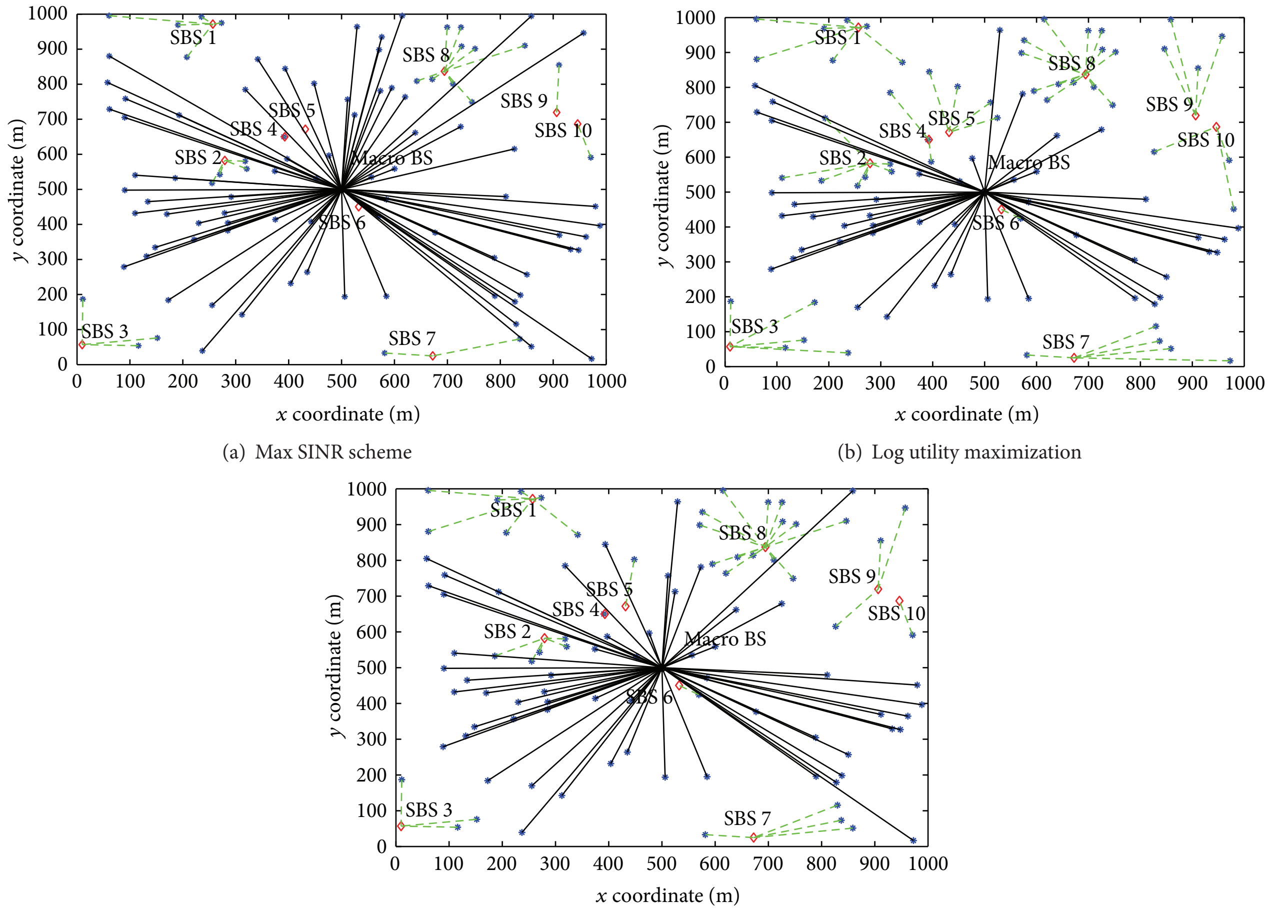
$$f(\mathcal{G}) \geq 0.631 \times f(\mathcal{G}^*). \tag{27}$$

Hence we can conclude that the greedy algorithm yields a near-optimal solution, which provides a performance guarantee for the user association problem. For dense HetNets with a large number of users and BSs, the greedy algorithm is useful for its low computational complexity and near-optimality.

6. Numerical Results

We conduct our simulations in MATLAB over a two-tier HetNet topology. The coverage area is limited to a Euclidean plane with the size 1000 m × 1000 m. The numbers of macrocell BSs, small cell BSs, and users are fixed to $\{1, 10, 100\}$, respectively. The transmission power of the macrocell BS and small cell BSs are fixed to $\{50, 30\}$ dbm, respectively. We assume the location of the macro-BS is fixed at the center of coverage area. All the small cell BSs and users are uniformly and independently distributed across the coverage area. Path loss with a fixed path loss factor of 4 and the Rayleigh fading with unit variance are used to model the channel power gains. Thermal noise is neglected.

Figure 1 depicts the user associations under various association schemes in one simulated two-tier HetNet. Figure 1(a) shows the user associations for the max SINR scheme. Since users are assigned to the strongest BS, we can see from Figure 1(a) that the majority of the SBSs are lightly loaded and some of the SBSs that are placed near the macro-BS do not serve any users at all. The loads across the macro-BS and SBSs are heavily imbalanced and the macro-BS is



(a) Max SINR scheme

(b) Log utility maximization

(c) Log utility maximization with weighted loads

Figure 1: A visual illustration of user associations under different association scheme. The simulated HetNet consists of 1 macro-BS, 10 SBSs, and 100 users. The outage threshold is 0 dB. The dark solid lines denote associations with the macro-BS and green dashed lines denote associations with SBSs.

heavily congested. The majority of user population is still associated with the macro-BS. Figure 1(b) shows the user associations obtained for the log utility maximization problem (11). Since the utility maximization framework penalizes associations with heavily loaded BSs, quite a few users previously associated with the macro-BS are now offloaded to SBSs, including some users adjacent to the macro-BS. However, those offloaded users adjacent to the macro-BS will suffer from strong cross-tier intercell interference, which leads to high outage probabilities. Figure 1(c) shows the users associations obtained via the greedy algorithm for the utility maximization problem with weighted loads proposed by this paper. Since the penalty is dependent on both BS loads and outage probabilities, we can see that some of the previously offloaded users are now reassigned to the macro-BS or other SBSs with higher received power in order to improve outage performance for those users. This shows the tradeoff between load balancing and outage performance.

Figure 2 shows the percentage of user population associated SBSs versus numbers of SBSs under different association schemes. The simulation is done over 200 different realizations of a two-tier HetNet. Both the log utility association and our weighted load association can offload more users to SBSs compared with max SINR association. In log utility association scheme more than 20% of user population is offloaded from the macro-BS to SBSs, compared with max SINR scheme. However, in weighted load association scheme the percentage of offloaded user population is reduced. This is because the penalty on outage now discourages assigning users to some SBSs with high outage probabilities. Hence some users previously offloaded to SBSs in log utility scheme are now reassigned to the macro-BS in order to improve outage performance. This also shows the tradeoff between outage performance and load balancing.

Figure 3 shows the cumulative distributive functions (CDFs) of user long term SIRs for the simulated HetNet. It can be seen from Figure 3 that in the log utility association scheme about 40% of the user population suffer from degradations in SIR compared with the max SINR scheme. The degradation in SIR is more severe for the offloaded cell edge users with low SIR. When these users are offloaded to SBSs they suffer from strong interference from the macro-BS resulting in further degradation in SIR. Subsequently these users suffer from high outage probabilities due to the SIR degradation.

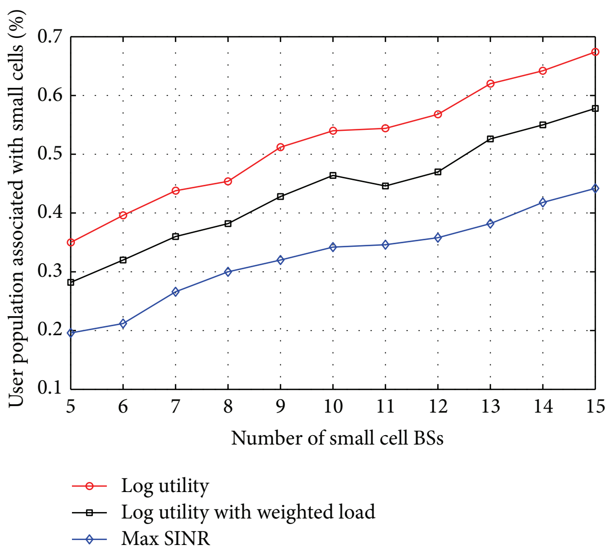


Figure 2: Percentage of user population associated with SBSs versus numbers of SBSs under different association schemes. The outage threshold is 0 dB.

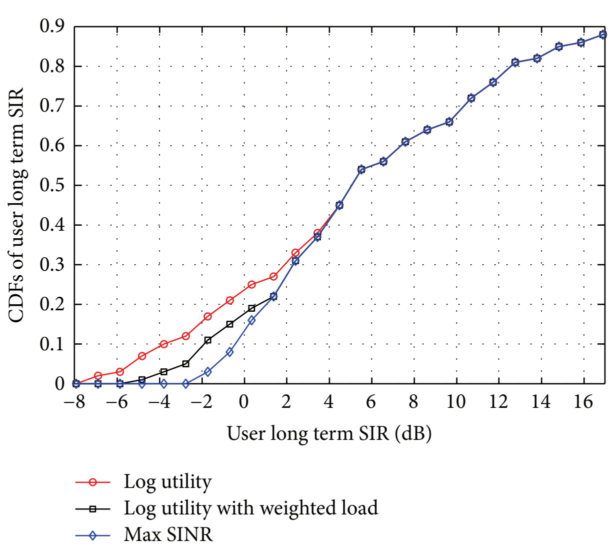


Figure 3: The CDFs of user long term SIRs in a two-tier HetNet consisting of 1 macro-BS, 10 SBSs, and 100 users. The outage threshold is 0 dB.

This shows the negative impact of offloading users to SBSs. However, under our weighted load association scheme, this SIR degradation has been largely alleviated and we can see that now about 20% of the user population suffers from degradations in SIR compared with the max SINR association scheme. Furthermore, for the offloaded users the magnitude of the SIR degradation is reduced by almost 50 percent compared with the log utility association scheme. This shows that our proposed approach alleviate the SIR degradations for offloaded users.

Figure 4 shows the system outage performance versus different outage thresholds under various association schemes. Similar to [21], we define the system outage probability

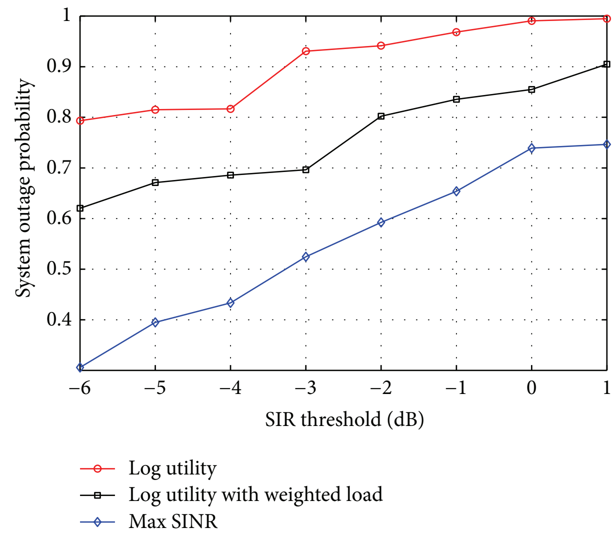


Figure 4: System outage probabilities versus SIR thresholds in a two-tier HetNet with 1 macro-BS, 10 SBSs, and 100 users.

P^{out} as the worst outage probability over all users, $P^{out} = \max_{i \in U} P_i^{out}$. It can be seen that under log utility association scheme, the system outage probability remains high over a wide range of SIR thresholds. The system outage probability under max SINR association scheme is halved compared with the log utility association scheme. With our weighted load association scheme, there is a 30 percent decrease in the system outage probabilities compared with the log utility association scheme. This shows the benefit of devising association schemes that can incorporate user outage performance.

7. Conclusions

In this paper, we consider the cell association problem in the context of downlink HetNets. In order to alleviate the SINR degradation and outage deterioration for offloaded users, we propose a split term problem formulation with weighted BS loads that incorporates user outage performance. We propose a formulation of the weighting parameters that penalizes the selections of user-BS pairs with high outage probabilities. Moreover, we show that the resulting combinatorial optimization problem can be reformulated into a monotone submodular maximization problem and can be approximated via a greedy algorithm with lazy evaluations. The greedy algorithm has a low computational complexity and is useful in densely deployed HetNets. The obtained cell association scheme guarantees single BS association for each user. Simulation results show that our proposed method preserves the offloading capabilities of the log utility association scheme while significantly reducing user outage probabilities. Our approach strikes a tradeoff between load balancing and outage performance and is useful for improving user QoS experiences in HetNets.

Conflict of Interests

The authors declare that there is no conflict of interests regarding the publication of this paper.

Acknowledgment

This research was partially supported by research grants from National Key Basic Research Program of China (973 Program) under Grant no. 2013CB329103, grants from the Natural Science Foundation of China (61301182, 61372078, and 61171071), grant from Guangdong Natural Science Foundation (2015A030313552), and grants from the Foundation of Shenzhen City (JCYJ20140418095735590, GJHS20120621143440025, and ZDSY20120612094614154).

References

[1] J. G. Andrews, S. Buzzi, W. Choi et al., "What will 5G be?" *IEEE Journal on Selected Areas in Communications*, vol. 32, no. 6, pp. 1065–1082, 2014.

[2] J. G. Andrews, "Seven ways that hetnets are a cellular paradigm shift," *IEEE Communications Magazine*, vol. 51, no. 3, pp. 136–144, 2013.

[3] J. G. Andrews, S. Singh, Q. Ye, X. Lin, and H. S. Dhillon, "An overview of load balancing in HetNets: old myths and open problems," *IEEE Wireless Communications*, vol. 21, no. 2, pp. 18–25, 2014.

[4] A. Damnjanovic, J. Montojo, Y. Wei et al., "A survey on 3GPP heterogeneous networks," *IEEE Wireless Communications*, vol. 18, no. 3, pp. 10–21, 2011.

[5] Z. Mlika, E. Driouch, W. Ajib, and H. Elbiaze, "A completely distributed algorithm for user association in HetNets," in *Proceedings of the IEEE International Conference on Computer Communication (INFOCOM '15)*, pp. 2172–2177, London, UK, April 2015.

[6] N. Namvar, W. Saad, B. Maham, and S. Valentin, "A context-aware matching game for user association in wireless small cell networks," in *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP '14)*, pp. 439–443, Florence, Italy, May 2014.

[7] W. Bao and B. Liang, "Structured spectrum allocation and user association in heterogeneous cellular networks," in *Proceedings of the 33rd IEEE Conference on Computer Communications (INFOCOM '14)*, pp. 1069–1077, Toronto, Canada, April 2014.

[8] B. Dai and W. Yu, "Sparse beamforming and user-centric clustering for downlink cloud radio access networks," *IEEE Access*, vol. 2, pp. 1326–1339, 2014.

[9] L. P. Qian, Y. J. A. Zhang, Y. Wu, and J. Chen, "Joint base station association and power control via Benders' decomposition," *IEEE Transactions on Wireless Communications*, vol. 12, no. 4, pp. 1651–1665, 2013.

[10] R. Sun, M. Hong, and Z. Luo, "Joint downlink base station association and power control for max-min fairness: computation and complexity," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 6, pp. 1040–1054, 2015.

[11] X. Lin and N. B. Shroff, "The impact of imperfect scheduling on cross-layer rate control in wireless networks," in *Proceedings of the IEEE International Conference on Computer Communications (IEEE INFOCOM '05)*, vol. 3, pp. 1804–1814, Miami, Fla, USA, March 2005.

[12] J. Wildman, Y. Osmanlioglu, S. Weber, and A. Shokoufandeh, "Delay minimizing user association in cellular networks via hierarchically well-separated trees," in *Proceedings of the IEEE International Conference on Communications (ICC '15)*, pp. 4005–4011, London, UK, June 2015.

[13] 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 Transactions on Wireless Communications*, vol. 12, no. 6, pp. 2706–2716, 2013.

[14] K. Shen and W. Yu, "Downlink cell association optimization for heterogeneous networks via dual coordinate descent," in *Proceedings of the 38th IEEE International Conference on Acoustics, Speech, and Signal Processing, (ICASSP '13)*, pp. 4779–4783, IEEE, Vancouver, Canada, May 2013.

[15] S. Corroy, L. Falconetti, and R. Mathar, "Dynamic cell association for downlink sum rate maximization in multi-cell heterogeneous networks," in *Proceedings of the IEEE International Conference on Communications (ICC '12)*, pp. 2457–2461, Ottawa, Canada, June 2012.

[16] D. Bethanabhotla, O. Y. Bursalioglu, H. C. Papadopoulos, and G. Caire, "User association and load balancing for cellular massive MIMO," in *Proceedings of the IEEE Information Theory and Applications Workshop (ITA '14)*, San Diego, Calif, USA, February 2014.

[17] N. Prasad, M. Arslan, and S. Rangarajan, "Exploiting cell dormancy and load balancing in LTE HetNets: optimizing the proportional fairness utility," *IEEE Transactions on Communications*, vol. 62, no. 10, pp. 3706–3722, 2014.

[18] V. Singh, N. Prasad, M. Y. Arslan, and S. Rangarajan, "Optimizing user association and activation fractions in heterogeneous wireless networks," in *Proceedings of the 13th International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt '15)*, pp. 283–290, IEEE, Mumbai, India, May 2015.

[19] 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 Journal on Selected Areas in Communications*, vol. 28, no. 9, pp. 1479–1489, 2010.

[20] Q. Ye, M. Al-Shalashy, C. Caramanis, and J. G. Andrews, "On/off macrocells and load balancing in heterogeneous cellular networks," in *Proceedings of the IEEE Global Communications Conference (GLOBECOM '13)*, pp. 3814–3819, IEEE, Atlanta, Ga, USA, December 2013.

[21] Y. Ikeda, S. Okasaka, M. Hoshino, and K. Higuchi, "Proportional fair-based joint optimization of cell association and inter-cell interference coordination for heterogeneous networks," in *Proceedings of the 80th IEEE Vehicular Technology Conference (VTC '14)*, pp. 1–5, Vancouver, Canada, September 2014.

[22] S. Deb, P. Monogioudis, J. Miernik, and J. P. Seymour, "Algorithms for enhanced inter-cell interference coordination (eICIC) in LTE HetNets," *IEEE/ACM Transactions on Networking*, vol. 22, no. 1, pp. 137–150, 2014.

[23] ZTE, "Mechanisms for interference avoidance and coordination in small cell scenarios," in *Proceedings of the 3GPP TSG RAN WG1 Meeting #72*, R1-130135, January 2013.

[24] E. Hossain, M. Rasti, H. Tabassum, and A. Abdelnasser, "Evolution toward 5G multi-tier cellular wireless networks: an interference management perspective," *IEEE Wireless Communications*, vol. 21, no. 3, pp. 118–127, 2014.

[25] H. Boostanimehr and V. K. Bhargava, "Unified and distributed Qos-driven cell association algorithms in heterogeneous networks," *IEEE Transactions on Wireless Communications*, vol. 14, no. 3, pp. 1650–1662, 2015.

[26] S. Kandukuri and S. Boyd, "Optimal power control in interference-limited fading wireless channels with outage-probability specifications," *IEEE Transactions on Wireless Communications*, vol. 1, no. 1, pp. 46–55, 2002.

[27] V. Jungnickel, K. Manolakis, W. Zirwas et al., "The role of small cells, coordinated multipoint, and massive MIMO in 5G," *IEEE Communications Magazine*, vol. 52, no. 5, pp. 44–51, 2014.

[28] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, Cambridge, UK, 2004.

[29] S. Fujishige, *Submodular Functions and Maximizations*, Elsevier Science, Philadelphia, Pa, USA, 2005.