

QoS Constrained Optimization of Cell Association and Resource Allocation for Load Balancing in Downlink Heterogeneous Cellular Networks

Gongchao Su^{1,2}, Bin Chen², Xiaohui Lin², Hui Wang² and Lemin Li¹

¹ School of Communication and Information Engineering, University of Electronic Science and Technology of China, Chengdu 611731 -China
[e-mail: gcsu@szu.edu.cn]

² Shenzhen Key Lab of Advanced Communications and Information Processing, Shenzhen University
Shenzhen 518060 -China

*Corresponding author: Gongchao Su

Received November 18, 2014; revised February 12, 2015; revised March 20, 2015; accepted April 21, 2015; published May 31, 2015

Abstract

This paper considers the optimal cell association and resource allocation for load balancing in a heterogeneous cellular network subject to user's quality-of-service (QoS) constraints. We adopt the proportional fairness (PF) utility maximization formulation which also accommodates the QoS constraints in terms of minimum rate requirements. With equal resource allocation this joint optimization problem is either infeasible or requires relaxation that yields a solution which is difficult to implement. Nevertheless, we show that this joint optimization problem can be effectively solved without any priori assumption on resource allocation and yields a cell association scheme which enforces single BS association for each user. We re-formulated the joint optimization problem as a network-wide resource allocation problem with cardinality constraints. A reweighted heuristic l_1 -norm regularization method is used to obtain a sparse solution to the re-formulated problem. The cell association scheme is then derived from the sparsity pattern of the solution, which guarantees a single BS association for each user. Compared with the previously proposed method based on equal resource allocation, the proposed framework results in a feasible cell association scheme and yields a robust solution on resource allocation that satisfies the QoS constraints. Our simulations illustrate the impact of user's minimum rate requirements on cell association and demonstrate that the proposed approach achieves load balancing and enforces single BS association for users.

Keywords: load balancing, heterogeneous cellular networks, optimization, regularization

This research is 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), and grants from the Foundation of Shenzhen City (JCYJ20140418095735590).

1. Introduction

With the proliferation of mobile devices and the continuous emergence of new services, the wireless traffic and data rate demands are growing at an unprecedented rate. To meet such demands, cellular networks are trending towards increasing heterogeneity. By overlaying small cells with macro cells, the heterogeneous cellular networks (HetNets) bring networks closer to users and have been shown to boost network throughput [1].

The paradigm shift towards heterogeneity brings many new challenges to cellular network design. One of the main challenges is to devise a cell association strategy that can actively offload users to less congested small cells, which provide more resources to their users [2]. Due to the low power nature of small cells, the fraction of user population that is offloaded to small cells is often limited under the traditional cell association strategy that allows a user to associate itself with the strongest base station (BS) in terms of received power. This results in large load disparity across macro and small cells and leads to suboptimal rate distribution across network users. On the other hand, the offloaded users usually experience degraded Signal-to-Interference-plus-Noise Ratio (SINR) since the strongest BS now contribute to interference. To prevent further degradation in user experience, resource allocation schemes need to be considered to combat that loss, since user rates are dictated by both the allocated resources and the SINR. These strategies are strongly coupled and require solutions from a network-wide perspective [3].

Fulfilling the user QoS requirements is a challenging task for radio resource management (RRM) in HetNets [4]. The load-aware cell association problem is also closely related with Quality of Service (QoS) provisioning in HetNets. The user QoS requirements, i.e. minimum rate requirements, dictate that users to be offloaded to a small cell with sufficient available radio resources. The resource allocation strategy adopted by the BS should also meet the distinct QoS requirements of its served users. Therefore, fulfilling the QoS requirements of every user requires more sophisticated solutions on cell association and resource allocation. In this paper we address this problem and show that the solutions can be readily found.

1.1 Motivation and Related Work

The cell association problem in HetNets has received significant recent attention. A game theoretical approach has been used to study the network selection problem in [5] which gives a solution without requiring coordination among different access networks. In [6], the network selection problem is formulated as a non-cooperative game and is shown to converge to the Nash Equilibrium. A popular suboptimal approach, referred to as cell range expansion (CRE), is used to offload users to small cells using an association bias and is part of 3GPP standardization efforts [7]. The coverage areas of small cells are artificially expanded with positive biasing factors. The gains from CRE can be increased if suitable interference avoidance strategies can be adopted to reduce the interference experienced by offloaded users. One such strategy is resource partitioning [8], wherein the transmission of macro tier BSs are muted on certain fraction of radio resources and the offloaded users are scheduled in those resources.

One way to find the more theoretically grounded solutions is to solve a system-level optimization problem that take into account both the cell association strategy and the resource allocation strategy. In [9], a proportional fairness utility maximization problem that takes into

account resource allocation and cell association is proposed. With log concave utility the optimal resource allocation is shown to be equal allocation. Therefore resource allocation is decoupled and the optimization problem is reformulated as a discrete optimization problem over cell association and can be tuned into a continuous one by relaxing the constraints. The problem is then solved in the dual domain using a subgradient method and the solution is rounded to ensure single BS association for users. In [10] a coordinate descent method is used to solve the dual problem. The utility maximization formulation is adopted in [11] to analyze load balancing and cell dormancy and in [12] to study the optimal cell load levels and the optimality conditions.

A few research address QoS-aware load balancing in cellular networks. In [13] QoS-aware load balancing is investigated in a macro-only multi-cell networks. A joint optimization problem of user association and resource allocation with QoS constraints is proposed in [14] to maximize network energy efficiency in an uplink HetNet. However, most of the previous work on the load balancing problem in a downlink HetNet does not consider user QoS requirements and employs a simple resource allocation strategy such as equal resource allocation at base stations. Equal resource allocation is not enough to cope with varying user QoS requirements. As mobile networks evolves towards increasing heterogeneity, it is important that the heterogeneous mobile networks can provide assurance of quality of experience to mobile users in order to satisfy the demand of emerging mobile services. Robust radio resource management schemes must be developed for the increasingly heterogeneous cellular networks, and for future mobile networks with convergence of different types of devices, protocols and services[15]. Traffic balancing and resource allocation are indispensable part of radio resource management schemes and it would be suboptimal to address them separately as they are strongly coupled. Moreover, these techniques directly influence the rate of users and are subject to user QoS constraints. Therefore it would be useful to explore the sensitivity of cell association and resource allocation to user QoS requirements.

1.2 Approach and Organization

This paper addresses the load balancing problem and analyze joint cell association and resource allocation with QoS constraints in downlink HetNets. We adopt the same proportional fairness utility function as in [9, 10] to achieve load balancing across macro and small cells. A set of minimum rate requirements are used as QoS constraints. We show that decoupling the joint optimization problem into a discrete optimization over cell association, which assumes equal resource allocation, does not yield feasible solutions under QoS constraints. The relaxed optimization cannot be solved in the dual domain using Lagrangian dual decomposition. We note that the equal resource allocation strategy is not robust enough to cope with user QoS requirements and the fractional user association employed in previous works is considered more difficult to implement. Our main contribution is that we propose a novel framework to solve the joint optimization problem without any priori assumption on resource allocation and the obtained solution on cell association guarantees single BS association for each user. Our proposed method differs from existing works which assume a resource allocation strategy first and then solve the relaxed cell association subproblem. QoS constraints in relation with resource allocation can be incorporated into our proposed framework, which is helpful for providing QoS support when offloading users to small cells in a heterogeneous cellular network. To do so we re-formulate the joint optimization problem as a global resource allocation problem with cardinality constraints, and derive the cell association scheme from the sparsity pattern of the resource allocation solution. We

re-formulate the constraints of the cell association as the cardinality inequality constraints of the resource allocation. Thus the resource allocation and cell association are decoupled, while the single-BS association is still enforced. Moreover, in order to obtain a sparse solution on resource allocation, we resort to the reweighted heuristic l_1 -norm regularization method [16,17] to solve this convex-cardinality problem, by adding a reweighted regularization term to the objective function. We note that a recent work [18] also uses the reweighted heuristic l_1 -norm technique to obtain sparse beamforming vectors when assigning users to a BS cluster in a cloud radio access network. The association strategy is directly derived from the sparsity pattern of the solutions. The obtained results provide insights for optimal cell association and resource allocation to achieve load balancing in HetNets with QoS constraints.

The rest of the paper is organized as follows. The HetNets system model is described in Section 2. Then we analyze the joint optimization problem with QoS constraints and solve our re-formulated convex-cardinality problem using reweighted heuristic l_1 -norm regularization in Section 3. In Section 4 numerical results are presented to validate our results. We give our conclusion in Section 5.

2. System Model

Consider a two-tier downlink heterogeneous cellular network with co-channelly deployed macro cell BSs and small cell BSs. The heterogeneous network covers a finite geographic area with a total of N users and M cell BSs, which are randomly and independently placed across the area. A frequency reuse factor of 1 is assumed across the network. We assume BSs of k^{th} ($k = 1,2$) tier transmit with the same transmit power P_k over the entire bandwidth. We assume frequency-flat channels and Rayleigh fading is used to model the random channel gain.

Let SINR_{ij} be the SINR of user i if it is associated BS j . Then the value of SINR_{ij} is

$$\text{SINR}_{ij} = \frac{P_j h_{ij} Z_{ij}^{-\alpha_j}}{\sum_{k \neq j} P_k h_{ik} Z_{ik}^{-\alpha_k} + \sigma^2} \quad (1)$$

where h_{ij} is the channel power gain, α_j is the path loss factor, Z_{ij} denotes the distance between user i and BS j , and σ^2 denotes the background noise power level. With frequency-flat channels the SINR remains constant during the cell association phase. We assume Shannon capacity is achieved for wireless links. We denote the achievable rate (in terms of bps/Hz) by c_{ij} , which is then

$$c_{ij} = \log_2(1 + \text{SINR}_{ij}) \quad (2)$$

Each user must be assigned to one BS. Since each BS generally serves more than one user, a user only gets a fraction of radio resources from its serving BS. Let y_{ij} be the portion of resources allocated to user i by BS j . Hence the long term rate of user i associated to BS j , denoted as R_{ij} , is defined as:

$$R_{ij} = y_{ij} c_{ij} \quad (3)$$

The overall rate of user i , denoted as R_i , is given by

$$R_i = \sum_j R_{ij} \quad (4)$$

We use a binary cell association indicator x_{ij} (1 or 0) to denote whether or not user i is associated with BS j . To make sure each user is assigned to exactly one BS, x_{ij} must satisfy

$$\sum_j x_{ij} = 1 \quad (5)$$

The requirement on unique association for user is also reflected on the resource allocation. If user i is associated with BS j , y_{ij} is strictly positive between 0 and 1. Otherwise y_{ij} equals to zero. We note that in order to enforce unique association for each user, the resource allocation indicator set $\{y_{ij}\}$ forms a sparse matrix, with exactly one nonzero element per row.

We use a set of minimum long term rate requirements as the user QoS requirements. The long term rate of user k in the network must satisfy

$$R_k = \sum_j R_{kj} = \sum_j y_{kj} c_{kj} \geq \theta_k \quad (6)$$

3. Problem Formulation

To achieve load balancing, this paper adopts a proportional fair network utility optimization framework of maximizing the sum of utilities across all network users. In [9, 10, 11, 12] a proportionally fair objective function (log-utility) is chosen as the utility function and the problem is formulated as the maximization of the sum of each user's utility. In this paper we also adopt the log utility function.

We first consider the optimization problem without QoS constraints. We follow the problem formulation in [9, 10]. The cell association and resource allocation problems can be interpreted as finding the optimal set $\{x_{ij}\}$ and $\{y_{ij}\}$ for the following optimization problem:

$$\begin{aligned} & \text{maximize}_{x,y} \quad \sum_i \log(\sum_j y_{ij} c_{ij}) & (7) \\ & \text{subject to} \quad \sum_j x_{ij} = 1 \\ & \quad \quad \quad \sum_i y_{ij} \leq 1 \\ & \quad \quad \quad 0 \leq y_{ij} \leq x_{ij}, x_{ij} \in \{0,1\} \end{aligned}$$

This problem is a combinatorial optimization problem with coupled variable x and y . Brutal force search method would require solving M^N convex problems with all possible x . The computation is essentially impossible for even a modest-sized cellular network. In [9], it is shown that for a given set of users associated with a BS, the optimal resource allocation is equal allocation. Let K_j denote the load of BS j , the optimization problem is then formulated as follows:

$$\begin{aligned} & \text{maximize}_x \quad \sum_i \sum_j x_{ij} \log(c_{ij}) - \sum_j K_j \log(K_j) & (8) \\ & \text{subject to} \quad \sum_j x_{ij} = 1 \\ & \quad \quad \quad \sum_i x_{ij} = K_j \\ & \quad \quad \quad \sum_j K_j = M \\ & \quad \quad \quad x_{ij} \in \{0,1\} \end{aligned}$$

Therefore the problem is formulated as a discrete optimization problem over cell association. The discrete optimization can be further relaxed into a continuous one by allowing fractional user association (FUA). As shown in [9, 10], the relaxed optimization problem is formulated as

$$\begin{aligned} & \text{maximize}_x \quad \sum_i \sum_j x_{ij} \log(c_{ij}) - \sum_j K_j \log(K_j) & (9) \\ & \text{subject to} \quad \sum_j x_{ij} = 1 \\ & \quad \quad \quad \sum_i x_{ij} = K_j \\ & \quad \quad \quad \sum_j K_j \leq M \\ & \quad \quad \quad 0 \leq x_{ij} \leq 1 \end{aligned}$$

Its dual problem can be decomposed into two convex problems, which are solved separately by each user and BS. The obtained solution is then rounded to ensure unique association for each user.

3.1 Cell Association with QoS Constraints

We now add QoS constraints to the optimization problem (8). With equal resource allocation, the minimum rate requirement (6) of user k is then

$$R_k = \sum_j R_{kj} = \sum_j \frac{x_{kj} c_{kj}}{\sum_i x_{ij}} \geq \theta_k \quad (10)$$

The optimization problem is formulated as

$$\begin{aligned} & \text{maximize}_x && \sum_i \sum_j x_{ij} \log(c_{ij}) - \sum_j K_j \log(K_j) \\ & \text{subject to} && \sum_j x_{ij} = 1 \\ & && \sum_i x_{ij} = K_j \\ & && \sum_j K_j = M \\ & && x_{ij} \in \{0,1\} \\ & && \sum_j \frac{x_{ij} c_{ij}}{K_j} \geq \theta_i \end{aligned} \quad (11)$$

The inequality constraint function of (11) is a linear combination of linear-fractional functions and is quasiconvex. Its domain $\{x | x_{ij} \in \{0,1\}\}$ is not a convex set. The resulting optimization problem is not convex. The feasibility of the problem is very limited.

Theorem 1: The optimization problem (11) is strictly infeasible if there exist more than M cell edge users whose SINR satisfy $\max_j C_{ij} < 2\theta_i$.

Proof: To satisfy the constraints in (11) every user should be associated with exactly one BS. If there are more than M cell edge users in the networks, at least 2 of them will be associated with the same BS. Suppose there are two cell edge users associated with BS ℓ and their SINR satisfy $\max_j C_{ij} < 2\theta_i$, following (10), the load K_ℓ of the BS satisfies

$$K_\ell \leq \frac{c_{i\ell}}{\theta_i} \leq \frac{\max_j C_{ij}}{\theta_i} < 2 \quad (12)$$

This contradicts the assumption. Therefore single BS association requirement cannot be satisfied and the problem is infeasible.

We comment that for a HetNet with N users and M BSs, generally there are quite a few cell edge users with low SINR, since typically $N \gg M$. Then in many cases the optimization problem (11) is infeasible. This means that with equal resource allocation we cannot obtain a practical cell association scheme which enforces unique association for users by solving the optimization problem (11).

If FUA is allowed as proposed in [9, 10], the relaxed optimization (9) with QoS constraints can be tuned into a convex optimization problem since the quasiconvex inequality constraints (10) can be replaced by a convex inequality constraint [16]. The feasibility of the relaxed optimization

$$\begin{aligned} & \text{find} && x, K \\ & \text{subject to} && \sum_j x_{ij} = 1 \\ & && \sum_i x_{ij} = K_j \\ & && \sum_j K_j \leq M \\ & && 0 \leq x_{ij} \leq 1 \\ & && \sum_j \frac{x_{ij} c_{ij}}{K_j} \geq \theta_i \end{aligned} \quad (13)$$

is also a convex feasibility problem. The feasibility of the problem is dependent on the rate c_{ij} and the minimum rate requirement θ_i . However, even if the problem is feasible, FUA requires a user to be simultaneously associated with multiple BSs and is difficult to implement. With QoS constraints FUA is more frequent because if the equally allocated portion of resources from a serving BS is not enough to satisfy the QoS constraints, the user has to ask for more resources from other BSs. The physical relaxation of constraints also results in an upper bound on the cell association problem with equal resource allocation.

We also note that the dual problem of the relaxed optimization (9) with QoS constraints (10) cannot be decomposed due to the coupling QoS constraint. To see that we can write the Lagrangian function as

$$\mathcal{L}(X, K, \lambda, \mu, v) = \sum_i \sum_j x_{ij} \log(c_{ij}) - \sum_j K_j \log(K_j) - \sum_i \lambda_i \left(\sum_j \frac{x_{ij} c_{ij}}{K_j} - \theta_i \right) - \sum_j \mu_j (\sum_i x_{ij} - K_j) - v(\sum_j K_j - M) \quad (14)$$

which is not separable. This means that under QoS constraints we cannot follow the fractional-rounding approach in [9, 10, 11] to obtain a unique BS association scheme by rounding the solutions of the dual problem.

In general we show that with the assumption of equal resource allocation it is difficult to find a practical solution on the cell association problem under QoS constraints. To enforce single BS association and satisfy user QoS requirements, a robust resource allocation strategy which is sensitive to user SINR is needed.

3.2 Problem Reformulation as a Convex-Cardinality Problem

Without the assumption of equal resource allocation we now revisit the optimization problem (7) under QoS constraints (6). It can be noted from the utility maximization problem that the objective function of (7) does not include the cell association indicator x . Furthermore, the two constraints in relation to cell association indicator x

$$\sum_j x_{ij} = 1 \quad (15)$$

$$0 \leq y_{ij} \leq x_{ij}, x_{ij} \in \{0,1\} \quad (16)$$

can be interpreted as requirements for single BS association for users. Under single BS association, the resource allocation indicator y under single BS association is sparse, with exactly one nonzero element for each row of y . The coupled inequality constraint function (16) indicates that the cell association indicator x is exactly the sparsity pattern of the resource allocation indicator y , and along with the equality constraint (15), single BS association is enforced. This motivate us to re-formulate the single BS association requirement as the constraints on the resource allocation indicator y . The following theorem inherits the single BS association merit:

Theorem 2: Let $\text{card}(y)$ denote the cardinality of the resource allocation indicator y . Single BS association is enforced if y satisfies:

$$0 \leq y_{ij} \leq 1 \quad (17)$$

$$\sum_j y_{ij} > 0 \quad (18)$$

$$\text{card}(y) \leq N \quad (19)$$

Proof: Consider $y \in \mathbb{R}_+^{N \times M}$. The constraint (17) and (18) indicate that there must be at least one positive element for each row of y , therefore $\text{card}(y_i) \geq 1$. If $\text{card}(y_i) > 1$, then $\text{card}(y) = \sum_i \text{card}(y_i) > N$. This violates (19). Now we have $\text{card}(y_i) = 1$ and obtain a solution of y with exactly one nonzero element in each row, which enforces single BS association for each user.

The constraints of the cell association indicator x can now be replaced by the corresponding constraints of resource allocation indicator y . Hence cell association is decoupled from the optimization problem. Note that the inequality constraint (18) can be readily replaced by the QoS constraint (6) since c_{ij} and θ_i are all positive and y_{ij} is nonnegative.

We now re-formulate the utility maximization problem with QoS constraints as

$$\begin{aligned} & \text{maximize}_y && \sum_i \log(\sum_j y_{ij} c_{ij}) && (20) \\ & \text{subject to} && \sum_i y_{ij} \leq 1 \\ & && 0 \leq y_{ij} \leq 1 \\ & && \sum_j y_{ij} c_{ij} \geq \theta_i \\ & && \text{card}(y) \leq N \end{aligned}$$

Note that the optimization problem (11) with equal resource allocation can be considered as a special case of our proposed optimization problem (20) by substituting y_{ij} with $\frac{x_{ij}}{K_j}$. The resulting problem is a utility maximization problem over resource allocation only and can be interpreted as a network-wide resource allocation problem. The cell association strategy can be directly derived from the solution of (20), since the cell association indicator is the sparsity pattern of the resource allocation indicator. Theoretically the problem (20) is equivalent to the original coupled maximization problem (7). However it is a continuous optimization problem and is a convex problem with a cardinality constraint. Convex-cardinality problems are frequently addressed in modern optimization theories and this motivates us to turn to the reweighted heuristic l_1 -norm method that works well on these problems.

3.3 The Reweighted Heuristic l_1 -norm Regularization

We use $\|\cdot\|_1$ to denote the l_1 -norm function. Under the basic l_1 -norm regularization method[16], the cardinality constraint (19) is replaced by extracting regularization term $\beta\|y\|_1$ from the objective function. The basic l_1 -norm regularization can be further refined as the reweighted l_1 -norm regularization[17] by replacing $\beta\|y\|_1$ with a weighted regularization term $\sum \beta_i |y_i|$. The regularized problem is formulized as

$$\begin{aligned} & \text{maximize}_y && \sum_i \log(\sum_j y_{ij} c_{ij}) - \sum_i \sum_j \beta_{ij} |y_{ij}| && (21) \\ & \text{subject to} && \sum_i y_{ij} \leq 1 \\ & && 0 \leq y_{ij} \leq 1 \\ & && \sum_j y_{ij} c_{ij} \geq \theta_i \end{aligned}$$

where β_{ij} is the constant weight associated with the resource allocation variable y_{ij} . With large values of the weight β_{ij} , some elements of y with are heavily penalized. This results in a lower cardinality of the solution.

The regularized optimization problem (21) is strictly convex. Its feasibility problem can be written as

$$\begin{aligned} & \text{find} && y && (22) \\ & \text{subject to} && \sum_i y_{ij} \leq 1 \\ & && 0 \leq y_{ij} \leq 1 \\ & && \sum_j y_{ij} c_{ij} \geq \theta_i \end{aligned}$$

We now compare the feasibility of the regularized optimization with the feasibility of the relaxed optimization with FUA. Let $\frac{x_{ij}}{K_j} = y_{ij}$ and substitute all x_{ij} in (13) and we have the feasibility problem of relaxed optimization as:

$$\begin{aligned} & \text{find} && y && (23) \\ & \text{subject to} && \sum_j K_j y_{ij} = 1 \\ & && \sum_i y_{ij} = 1 \\ & && \sum_j K_j \leq M \\ & && 0 \leq K_j y_{ij} \leq 1 \\ & && \sum_j y_{ij} c_{ij} \geq \theta_i \end{aligned}$$

The feasible set on y of (23) is strictly a subset of the feasible set of (22). This means we are more likely to find a feasible solution by solving the regularized optimization (20) than solving the relaxed optimization.

We can now analyze the optimality condition of the convex optimization problem (21). For a given value of β , the regularized optimization problem of (21) is a constrained convex optimization problem. Let $f(y) = -\sum_i \log(\sum_j y_{ij} c_{ij}) + \sum_i \sum_j \beta_{ij} |y_{ij}|$. We can now vectorize the optimization problem. Let y_i denote the i th element of the solution, $i \in \mathcal{L} = \{1, 2, \dots, N \times M\}$. Let $A = \{n | (n \bmod M) = (i \bmod M), n \in \mathcal{L}, n \neq i\}$ and $B = \{m | k \times M < i < (k+1) \times M, k \times M < m < (k+1) \times M, m \in \mathcal{L}, m \neq i\}$. Suppose y^* is a local optimal point, then y^* satisfies the Karush-Kuhn-Tucker (KKT) condition:

$$\begin{aligned} & 0 \leq y_i^* \leq 1 \\ & \sum_{n \in A} y_n^* + y_i^* \leq 1, \\ & \sum_{m \in B} y_m^* c_m + y_i^* c_i \geq \theta_i \\ & \lambda_i, u_i, \phi_i, v_i \geq 0 \\ & \lambda_i y_i^* = 0 \\ & u_i (y_i^* - 1) = 0 \\ & \phi_i (\sum_{n \in A} y_n^* + y_i^* - 1) = 0 \\ & v_i (\theta_i - \sum_{m \in B} y_m^* c_m - y_i^* c_i) = 0 \\ & \nabla_i f(y^*) + \sum_i (-\lambda_i + u_i + \phi_i - c_i v_i) = 0 \end{aligned} \quad (24)$$

Although the objective function $f(y)$ contains a l_1 norm function which is usually non smooth, we note that any feasible y is strictly nonnegative. Then the regularization term can be written as

$$\sum_i \beta_i |y_i| = \sum_i \beta_i y_i \quad (25)$$

Now we have a continuously differentiable objective function $f(y)$ and its first order partial derivative $\nabla_i f(y)$ can be formulated as

$$\nabla_i f(y) = \beta_i - \frac{c_i}{\sum_{m \in B} y_m^* c_m + y_i^* c_i} \quad (26)$$

Substitute (26) into (24) and we now have the optimality condition with relation to β as:

$$\beta_i - \frac{c_i}{\sum_{m \in B} y_m^* c_m + y_i^* c_i} + \sum_i (-\lambda_i + u_i + \phi_i - c_i v_i) = 0 \quad (27)$$

It can be observed from (27) that larger values of β lead to smaller values of y_i and result in a higher sparsity of the solution. In order to choose appropriate values of β and achieve the desired sparsity of the solution, we adopt the reweighted heuristic l_1 norm approach proposed in [17,18]. In the heuristic approach the regularization parameter β can be found heuristically by iteratively updating β and solving the corresponding problem (21). The iteration can be stopped when a solution of (21) with the desired sparsity pattern is found.

In each iteration, the weight β_{ij} is updated according to

$$\beta_{ij} = \frac{1}{|y_{ij}| + \tau} \quad (28)$$

where $\tau > 0$ is a constant small factor and y_{ij} is the solution obtained from the previous iteration. It can be seen from (28) that β_{ij} is inversely proportional to y_{ij} and heavy penalties are given to small y_{ij} , which is further reduced in the iteration.

As shown in [17], the reweighting iteration process is robust to the choice of the constant factor τ . In addition, in order to determine whether or not the obtained solution satisfy the cardinality constraint, we define an arbitrarily small cardinality threshold ε and if $y_{ij} \leq \varepsilon$ then $y_{ij} = 0$. Thus we have $\text{card}(\mathbf{y}) = \sum_i \sum_j 1(y_{ij} > \varepsilon)$. The cell association strategy can be directly obtained by rounding the solution. The reweighted heuristic algorithm is described in **Table 1**.

Table 1. Reweighted Heuristic l_1 -norm Regularization Method

<p>Step 1: Initialization: Set $\beta_{ij} = 1, \forall i, j$. Choose $\tau > 0$ and $\varepsilon > 0$</p> <p>Step 2: Determine the feasibility of QoS requirements by solving the convex feasibility problem (22).</p> <p>Step 3: If problem (21) is feasible Repeat solve the convex problem (21) and find the solution \mathbf{y} update β_{ij} according to (28) until $\text{card}(\mathbf{y}) = \sum_i \sum_j 1(y_{ij} > \varepsilon) \leq N$ End if</p> <p>Step 4: Set the cell association indicator x_{ij} according to :</p> $x_{ij} = \begin{cases} 1, & \text{if } j = j^{(i)} \\ 0, & \text{if } j \neq j^{(i)} \end{cases}, \text{ where } j^{(i)} = \arg \max_j y_{ij} \quad (29)$

Since β_{ij} is initialized to 1, the first iteration is the basic l_1 norm regularization. If the obtained solution does not satisfy the cardinality constraint, our algorithm iteratively updates the weight β_{ij} and increases the relative weight on small y_{ij} until the cardinality constraint is satisfied.

From the heuristic algorithm we obtain a solution of resource allocation with desired sparsity, which leads to load balancing across the networks. The corresponding cell association is guaranteed to be a single BS association for every user.

Due to the regularization item the obtained solution is suboptimal. It can be interpreted as a tradeoff between the feasibility in terms of the required sparsity and the optimality of the solution. From our numerical results we show that the obtained cell association achieves close to optimal performance in load balancing.

The computational complexity largely depends on the number of reweighting iterations needed until the solution satisfies the cardinality constraint, as in each iteration a standard convex optimization problem is solved. Since the weight β_{ij} is updated iteratively and inversely proportional to the obtained y_{ij} , in a few iterations the small y_{ij} is reduced below the cardinality threshold ε . Thus it only takes a few reweighting iterations until the algorithm reaches convergence.

3.4 Discussion

The main point of this paper is that a viable single BS association strategy can be found by formulating a global resource allocation problem and solving the regularized form of the problem. There is no priori restriction on resource allocation. Hence the cell association problem based on equal resource allocation in previous work [9, 10] can be considered as a special case of our formulated problem. With QoS constraints our formulated problem is more likely to yield a feasible solution that achieves load balancing. The regularized optimization can be interpreted as a bi-criterion problem. The log-concave proportional fairness function encourages cell edge users to select small BSs as their candidate serving BSs, since there are more resources available at these BSs. On the other hand, the regularization item encourages a user to ask for more resources from one of its candidate serving BS by inducing a sparse solution on resource allocation. For any user, the resources allocated from all BSs form a sparse vector with exactly one large element and the rest smaller than the threshold ϵ . Then the user selects the BS from which it receives the largest portion of resources. Hence a single BS association strategy is obtained.

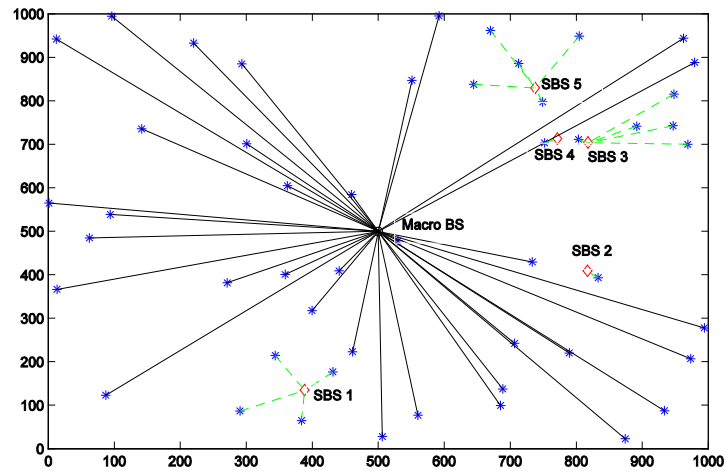
Moreover, our method can be extended to cases with more general proportional fairness utility functions. In [19, 20] it has been shown that proportional fairness can be achieved if the utility function is concave, monotonically increasing and continuously differentiable. With more general utility functions the optimal resource allocation would not be an equal resource allocation. Hence there is no predetermined restriction on resource allocation. In our method the convex-cardinality reformulation of the utility maximization problem is independent of the utility function and yields a robust resource allocation solution which guarantees single BS association. It can be used to derive cell association and resource allocations schemes for the general utility maximization problem.

4. Numerical Results

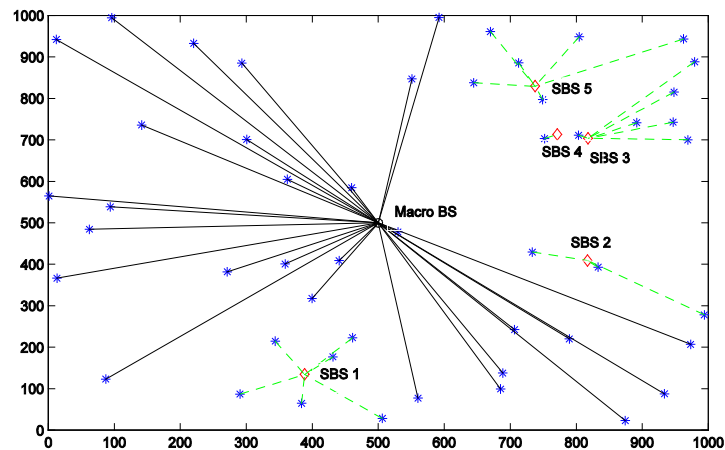
We validate our results in a two tier cellular heterogeneous networks with a finite cell coverage area of the size $1000\text{m} \times 1000\text{m}$. The numbers of macro BSs, small cell BSs and UEs are fixed to $\{1,5,50\}$. We assume the location of the macro BS is fixed, and the locations of the small cell BSs as well as users are randomly and independently distributed across the area. This corresponds to the operator deployed macro cell and customer deployed small cells. The transmit power of macro and small cell BSs are fixed to $\{50,30\}$ dBm respectively. The random channel gains are assumed to be Rayleigh distributed with $E[H] = 1$. Path loss factor is assumed to be 4 for all the wireless links. Without loss of generality, thermal noise is neglected. The cardinality threshold ϵ is set to 10^{-5} . A minimum long term rate requirements of $\theta \geq 0.08$ bps/Hz is chosen as the QoS constraint.

Fig. 1 depicts the cell association schemes for various cell association schemes. In **Fig. 1 (a)** the max SINR scheme is depicted. Due to the large power difference between macro cell and small cells, many users are associated with the macro BS, while most of the small cells BSs (SBS) remain lightly loaded. In **Fig. 1 (b)** the association scheme obtained via fractional-rounding method is depicted. Since there is no minimum rate requirement for users, fractional-rounding method can be used to obtain a single BS association strategy with equal resource allocation and more users are offloaded to small cell BSs, resulting in more balanced loads. **Fig. 1 (c)** depicts the fractional user association scheme under minimum rate constraints. With minimum rate constraints fractional-rounding method is no longer viable. The relaxation of fractional user association and the assumption of equal resource allocation result in many

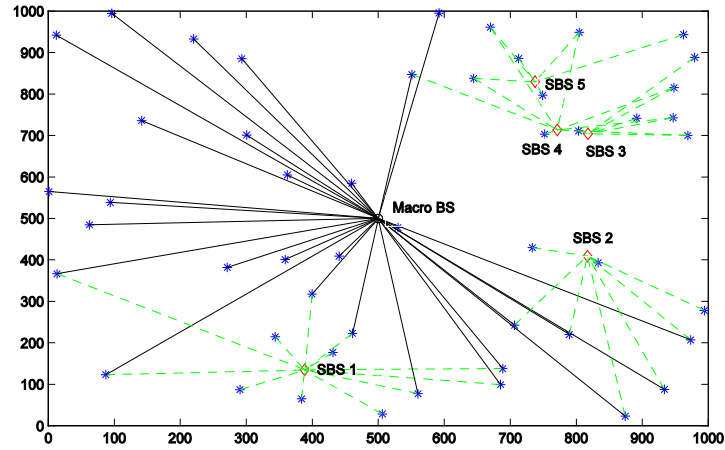
users simultaneously asking for resources from multiple cells. **Fig. 1 (d)** depicts the cell association scheme obtained from the reweighted heuristic l_1 -norm regularized optimization with QoS constraints. The result provides a single BS association for users and achieves a similar performance in load balancing compared with the fractional rounding approach. Note that the cell association in **Fig. 1 (d)** is different from that in **Fig. 1 (b)**, although both guarantee single BS association. That shows the impact of minimum rate constraints on cell association. The QoS constraints dictate that a user must be assigned to the BS with sufficient resources to meet the rate requirements.



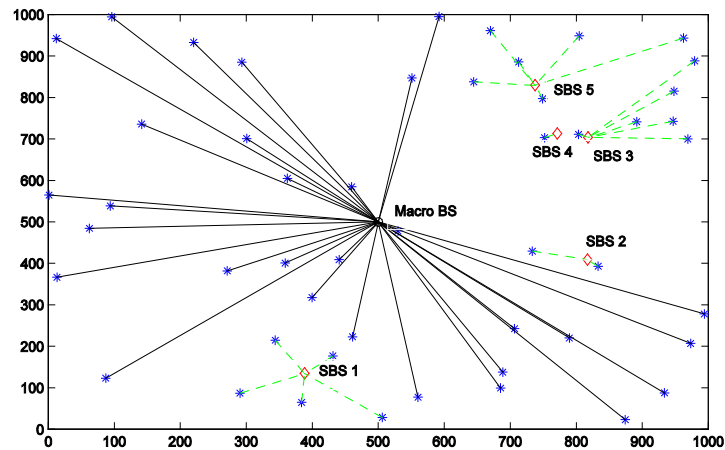
(a) max-SINR association



(b) fractional-rounding without QoS constraints



(c) fractional user association with QoS constraints



(d) regularized association with QoS constraints

Fig. 1. Different cell association strategies in a two-tier HetNets. The solid lines represent the association between users and the macro cell, while the dashed lines represent the association between users and the small cells.

Fig. 2 shows the load situations among different association schemes. In max SINR association scheme the loads among the macro cell and small cells are very unbalanced. In the other three association schemes the load on the macro cell is reduced, which suggests the proportional fairness objective function leads to more balanced loads across the network. Compared with fractional-rounding scheme without minimum rate requirements, the regularized association with minimum rate constraints yields a very similar performance in load balancing. However, in fractional user association (FUA), the number of users served by the small cells is much larger. This shows that under the minimum rate requirement and equal resource allocation some users are simultaneously associated with multiple BSs, including small cell BSs and the macro cell BS. To implement this association coordination among small cells or between the macro cell and small cells is needed and that also introduces more operational overhead.

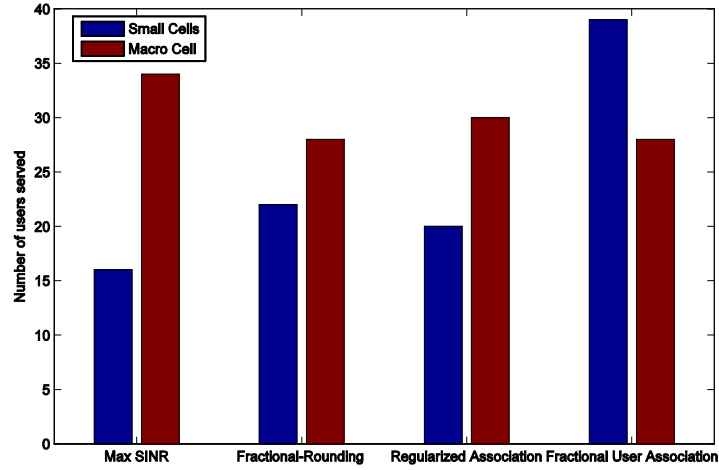


Fig. 2. Number of users served per tier in a two-tier HetNets. In regularized association and fractional user association a minimum long term rate $\theta \geq 0.08$ bps/Hz is required.

Fig. 3 shows the cumulative distribution function (CDF) of the resource allocation variable y_{ij} obtained from our regularized optimization with QoS constraints. The majorities (more than 80%) of the values are kept below 10^{-8} and the rest are kept within the range of 10^{-2} and 1. Note that in the HetNets consisting of 6 cells, single BS association for users would require that about 83% (or $\frac{5}{6}$) of the values of the resource allocation variables are very small. Our solution obtained from the regularized optimization is a sparse solution and satisfy the requirement.

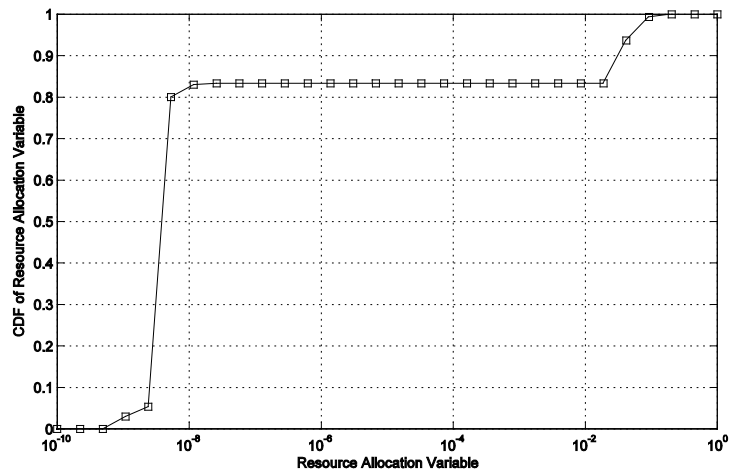


Fig. 3. The CDFs of resource allocation variable obtained from our regularized optimization in a two-tier HetNets consisting of 50 users, 1 macrocell and 5 small cells.

Fig. 4 shows the CDFs of long term user rates with different association schemes. The simulation is done across more than 2000 different realizations of random two-tier HetNets. In max SINR association the distribution of user rates are quite uneven, resulting in a large rate gap between cell edge users and high rate users. The fractional-rounding method with equal

resource allocation offers substantial rate improvement to low rate users and guarantees single BS association for users. However, it is not enough to satisfy the minimum rate requirement as the bottom 15% of the user rates are lower than that. With the minimum rate requirement, our regularized approach guarantees single BS association for users, achieves similar performance in user fairness and increases rates of cell edge users to satisfy the rate requirement. Note that there is a rate decrease for high rate users since more resources are allocated to cell edge users to satisfy the rate requirement. The fractional user association method, which allows multiple BS association for users, satisfies the rate requirement and achieves a higher overall rate compared with fractional-rounding method and our regularized optimization approach, due to the relaxation of single BS association constraints. That shows relaxing the single BS association constraints results in a higher bound on the network performance.

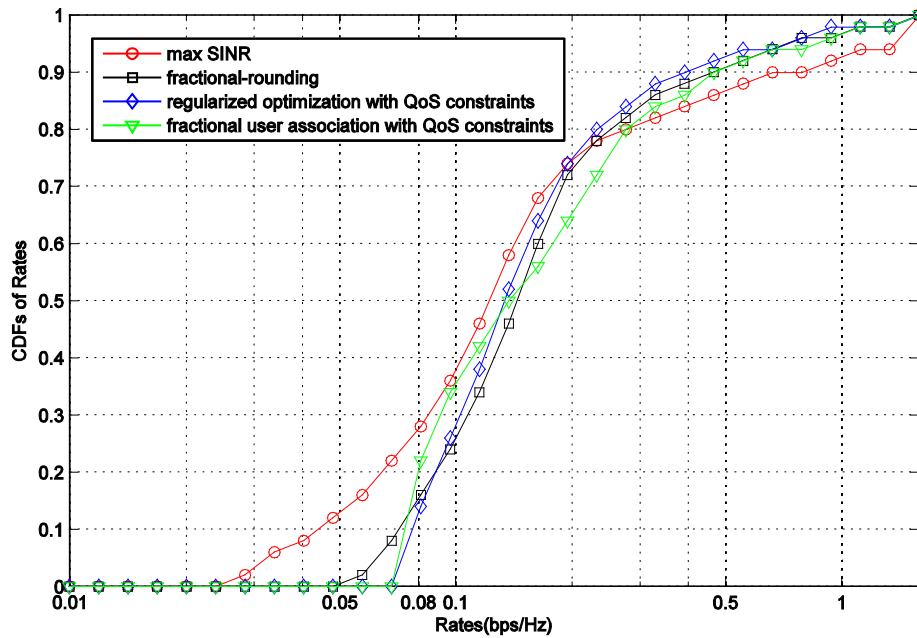


Fig. 4. The CDFs of long term user rates in a two-tier HetNet consisting of 50 users, 1 macrocell and 5 small cells. The minimum rate requirement is 0.08bps/Hz.

Fig. 5 shows the average number of reweighting iterations until convergence versus different choice of τ . We run 200 trials for each choice of τ . As we can see, the impact of different choices of τ on the number of reweighting iterations is very limited. Since the first iteration is the basic l_1 norm regularization with β initialized to 1, our algorithm usually converges after few iterations of reweighting process. That shows our algorithm is reasonably robust to the choice of τ and the computational complexity is low.

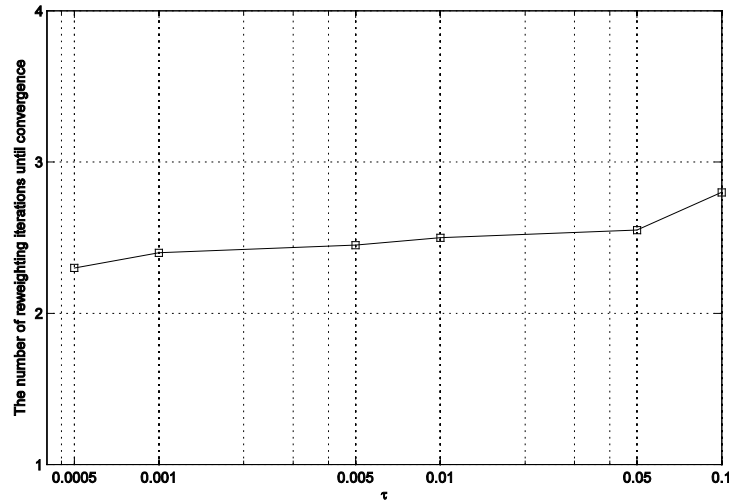


Fig. 5. The average number of reweighting iterations until convergence versus different choices of τ . The cardinality threshold is 10^{-5} .

5. Conclusion

In this paper, we investigate the load balancing problem with minimum rate requirements in a heterogeneous cellular network. We incorporate the minimum rate constraints with a utility maximization problem that considers cell association and resource allocation jointly. Then we show that with the assumption of equal resource allocation and the requirement of single BS association for users the maximization problem is often infeasible. The relaxation of constraints, which allows multiple BS association for users, yields a solution that is difficult to implement. To address this we propose a novel framework that eliminates any priori assumption on resource allocation and guarantees single BS association for users. We reformulate the joint maximization problem into a global resource allocation problem with cardinality constraints. The reweighted heuristic l_1 -norm regularization method is then used to obtain a solution with desired sparsity. A cell association scheme is then derived from the sparsity pattern of the solution, which guarantees a single BS association for users.

The main benefit of our approach is that we do not take any priori assumption on resource allocation and our approach can work with a variety of objective functions. Our approach can be extended to load balancing problems with other QoS constraints that can be presented in a convex form. Our results are helpful to design a cell association and resource allocation scheme that can accommodate realistic user QoS requirements.

References

- [1] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocell: Past, present, and future," *IEEE Journal on Selected Areas in Communications*, vol. 30, pp. 497-508, 2012. [Article \(CrossRef Link\)](#)
- [2] J. G. Andrews, S. Singh, Q. Ye, X. Lin, and H. Dhillon, "An overview of load balancing in HetNets: Old myths and open problems," *IEEE Wireless Communications Magazine*, vol. 21, no. 2, pp. 18-25, 2014. [Article \(CrossRef Link\)](#)
- [3] S. Singh, H. S. Dhillon, and J. G. Andrews, "Offloading in heterogeneous networks: Modeling, analysis, and design insights," *IEEE Transaction on Wireless Communications*, vol. 12, no. 5, pp. 2482-2497, 2013. [Article \(CrossRef Link\)](#)

- [4] Y. L. Lee, T. C. Chuah, J. Loo and A. Vinel, "Recent advances in radio resource management for heterogeneous LTE/LTE-A Networks," *IEEE Communications Surveys&Tutorials*, no. 99, pp. 1-39, 2014. [Article \(CrossRef Link\)](#)
- [5] S. Elayoubi, E. Altman, M. Haddad and Z. Altman, "A hybrid decision approach for the association problem in heterogeneous networks," in *Proc. IEEE Int. Conf. on Computer Communications*, pp. 1-5, 2010. [Article \(CrossRef Link\)](#)
- [6] E. Aryafar, A. K. Haddad, M. Wang and M. Chiang, "Rat selection games in HetNets," in *Proc. IEEE Int. Conf. on Computer Communications*, pp. 998-1006, 2013. [Article \(CrossRef Link\)](#)
- [7] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Communications Magazine*, vol. 18, pp. 10-21, 2011. [Article \(CrossRef Link\)](#)
- [8] S. Singh, J. G. Andrews, "Joint resource partitioning and offloading in heterogeneous cellular networks," *IEEE Transaction on Wireless Communications*, vol. 13, no. 2, pp. 888-901, 2014. [Article \(CrossRef Link\)](#)
- [9] 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 Transaction on Wireless Communications*, vol. 12, no. 6, pp. 2706-2716, 2013. [Article \(CrossRef Link\)](#)
- [10] Kaiming Shen and Wei Yu, "Downlink cell association optimization for heterogeneous networks via dual coordinate descent," in *Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing*, pp. 4779-4783, 2013. [Article \(CrossRef Link\)](#)
- [11] N. Prasad, M. Arslan, and S. Rangarajan, "Exploiting cell dormancy and load balancing in LTE HetNets: Optimizing the proportional fairness utility," *IEEE Transaction on Communications*, vol. 62, no. 10, pp. 3706-3722, 2014. [Article \(CrossRef Link\)](#)
- [12] I. Siomina and D. Yuan, "On optimal load setting of load-coupled cells in heterogeneous LTE networks," in *Proc. IEEE Int. Conf. on Computer Communications*, pp. 1254-1259, 2014. [Article \(CrossRef Link\)](#)
- [13] H. Wang, L. Ding, P. Wu, Z. Pan, N. Liu and X. You, "Qos-aware load balancing in 3GPP Long Term Evolution multi-cell networks," in *Proc. IEEE Int. Conf. on Computer Communications*, pp. 1-5, 2011. [Article \(CrossRef Link\)](#)
- [14] H. Pervaiz, L. Musavian and Q. Ni, "Joint user association and energy-efficient resource allocation with minimum-rate constraints in a two-tier HetNets," in *Proc. IEEE Int. Symp. On Personal, Indoor and Mobile Radio Communications*, pp. 1634-1639, 2013. [Article \(CrossRef Link\)](#)
- [15] Minh Jo, T. Maksymyuk, R. L. Batista, T. F. Maciel, A. L. F. de Almeida and M. Klymash, "A Survey of Converging Solutions for Heterogeneous Mobile Networks," *IEEE Wireless Communications*, vol. 21, no.8, pp.54-62, 2014. [Article \(CrossRef Link\)](#)
- [16] S. Boyd and L. Vandenberghe, *Convex Optimization*, Cambridge University Press, 2004.
- [17] E. Candes, M. Wakin and S. Boyd, "Enhancing sparsity by reweighted l_1 minimization," *Journal of Fourier Analysis and Applications*, vol. 14, no. 5, pp. 877-905, 2008. [Article \(CrossRef Link\)](#)
- [18] 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. [Article \(CrossRef Link\)](#)
- [19] T. Bu, L. Li and R. Ramjee, "Generalized proportional fair scheduling in third generation wireless data networks," in *Proc. IEEE Int. Conf. on Computer Communications*, pp. 1-12, 2006. [Article \(CrossRef Link\)](#)
- [20] H. Kim, G. de Veciana, X. Yang and M. Venkatachalam, "Distributed α -Optimal user association and cell load balancing in wireless networks," *IEEE Transaction on Networking*, vol. 20, no. 1, pp. 177-190, 2012. [Article \(CrossRef Link\)](#)



Gongchao Su received his B.S and M.S degrees from the University of Electronic Science and Technology of China (UESTC), in 2000 and 2003, respectively, all in communications engineering. He is now a Ph.D. candidate in School of Communication and Information Engineering, UESTC. He has been a lecturer in the Faculty of Information Engineering, Shenzhen University since 2003. His research interests include modeling, design and performance analysis of wireless heterogeneous networks.



Bin Chen received his B.E. and M.E. degrees in electrical engineering from Lanzhou University, P.R. China, in 1997 and 2002, respectively. In 2007, He received his Ph.D. degree from the School of Electrical and Electronic Engineering, Nanyang Technology University, Singapore. Currently, he is an associate professor in the Faculty of Information Engineering, Shenzhen University, China. His research interests include IP over WDM networks, wireless networks and network traffic engineering.



Xiaohui Lin received his B.S. and M.S. degrees in Electronics and Information Science from the Lanzhou University, in 1997 and 2000, respectively. He got his Ph.D. degree in Electrical and Electronic Engineering from the University of Hong Kong in 2003. He is now an professor in the Faculty of Information Engineering, Shenzhen University in Guangdong, China. His research interests include mobile computing, ad hoc wireless network, and multimedia communication. In these fields, he has published more than 40 papers in international leading journals and refereed conferences.



Hui Wang received his B.S., M.S. and Ph.D. degrees from Xi'an Jiaotong University, in 1990, 1993, and 1996, respectively, all in telecommunication. He is now a professor and dean of the Faculty of Information Engineering, Shenzhen University. His research interests include wireless communication, signal processing, and distributed computing systems, in which, he is author or co-author of more than 50 international leading journals, conferences and book chapters. He is a member of IEEE.



Leming Li received the BS degree in electrical engineering from Shanghai Jiaotong University, China in 1952. From 1952 to 1956, he was with the Department of Electrical Communications at Shanghai Jiaotong University. Since 1956 he has been with the University of Electronic Science and Technology of China (UESTC). From August 1980 to August 1982, he was a Visiting Scholar in the Department of Electrical Engineering and Computer Science at the University of California at San Diego, USA, doing research on digital and spread spectrum communications. He is currently a Professor of the School of Communication and Information Engineering, UESTC. He is a member of the Chinese Academy of Engineering. His present research interests are in the areas of communication networks including broadband networks and wireless networks.