DOWNLINK CELL ASSOCIATION OPTIMIZATION FOR HETEROGENEOUS NETWORKS VIA DUAL COORDINATE DESCENT

Kaiming Shen and Wei Yu

Electrical and Computer Engineering Dept.
University of Toronto, Toronto, Ontario, Canada M5S 3G4
Email: {kshen,weiyu}@comm.utoronto.ca

ABSTRACT

This paper considers the optimal association of remote user terminals to different cells in a heterogeneous network for load balancing. Assuming fixed transmit powers at the base-stations, we adopt a network utility maximization formulation with a proportional fairness objective and show that the downlink user association problem can be solved efficiently using a pricing approach where the prices are updated in the dual domain via coordinate descent. As compared to the previously proposed subgradient method, the proposed coordinate descent algorithm does not require the base-stations to synchronize in their price updates, while still guaranteeing convergence, which makes it particularly suitable for distributed implementation. Simulations show that the proposed method has fast convergence while achieving near-optimal solution.

1. INTRODUCTION

Heterogeneous network is an emerging wireless cellular architecture in which femto or pico base-stations (BSs) are deployed throughout the geographical area to off-load traffic from the macro BSs. By splitting the conventional cellular structure into small cells, the heterogeneous network architecture allows more aggressive reuse of frequencies as well as improved coverage and higher overall throughput for the entire network.

One of the main challenges in the design and implementation of heterogenous network architecture is the appropriate setting of the transmit power levels at the femto/pico BSs and the definition of each cell's coverage area. In most practical deployments, the power levels at the femto/pico BSs are fixed *a priori* (typically at 10-30dB below the macro BS power). The downlink coverage areas of the small cells are then defined according to the received signal-to-interference-and-noise ratio (SINR). From the the user terminals' perspective, each user is simply associated with the BS with the highest SINR. This downlink cell association rule is often referred to as the max-SINR rule.

A key problem with the max-SINR cell association is that it does not account for the varied data traffic pattern in the net-

work, hence it does not address the load balancing problem effectively. Load balancing is essential for a small-cell environment, because femto/pico BSs are often deployed to alleviate traffic "hot-spots" with higher-than-average user density. In these situations, a heuristic that adds a bias term to the reference SINR often needs to be used in practice to address the load imbalance.

This paper addresses the downlink coverage and user-cell association problem from a proportionally fair utility maximization perspective. We assume that the BS powers are fixed, and use a pricing approach for BS association. Our main contribution is a distributed dual price update method based on coordinate descent which is guaranteed to converge and is shown to achieve near-optimal performance in practice.

The BS association problem has been considered extensively in the literature. For example, the network utility maximization approach has been considered in [1], where heuristic pricing strategies are considered for the joint power control and cell-site selection problem. The joint power control and cell association problem is also considered in [2], which gives an optimal solution to the problem but only under certain restricted conditions and only for the case where the number of users and the number of BSs are the same. In [3], a different special case of only a single pair of macro and pico BSs is investigated. In [4], an approximate solution to the joint power control, beamforming, and BS association problem is given. In [5], a game theoretical approach is used, and the Nash Equilibrium of the problem is found.

A common method for solving the BS association problem is the greedy method [6, 7, 8]. However, the performance of the greedy algorithm is not easy to control. The convergence is slow, if only a single user improves its BS association at each step. But, the algorithm may exhibit oscillatory behavior if too many users switch BSs at the same time [7].

In this paper, we use the duality theory in optimization to tackle the BS association problem. This approach is first taken in [9], where the BS association problem under the proportionally fair utility is solved in the dual domain using a subgradient method. The dual variables have a pricing interpretation. The users determine their BS association based on

their achievable utilities minus the BS price, and the BSs update their prices based on a subgradient.

This paper uses the same problem formulation as in [9], but instead of using subgradient update, we propose a price update based on coordinate descent in the dual domain. As compared to the subgradient method, the advantages of the proposed method are two fold. First, the performance of the subgradient method is often sensitive to parameter setting, while the proposed method is not. Second and more importantly, the implementation of subgradient update requires synchronization among the BSs in the sense that all BSs must update their prices using the same step size at the same time in order to ensure convergence. In contrast, the proposed method requires only minimal coordination among the BSs and is therefore more suited for distributed implementation.

2. PROBLEM FORMULATION

Consider the downlink of a heterogeneous cellular network with L BSs with fixed (but possibly different) transmission power spectrum density (PSD) levels, and a total of K user terminals across a geographic area (possibly nonuniformly). A maximum frequency reuse factor of 1 is assumed; the total available bandwidth is W. Let SINR_{ij} be the SINR of user i if it is associated with BS j. To simplify the problem, we assume frequency-flat channels and a flat transmit PSD level so that the SINR is constant across the frequencies. Further, we assume that if a total of K_j users are associated with the BS j, then each user simply gets $1/K_j$ of the total bandwidth[1]. In this case, the data rate of the user i associated with BS j can be calculated as

$$R_{ij} = \frac{W}{K_j} \log(1 + \text{SINR}_{ij}). \tag{1}$$

This paper adopts a proportionally fair (or log-utility) objective function, and seeks user-cell associations that maximize the sum of each user's utility, $\log(R_{ij})$.

We follow the problem formulation of [9]. Let x_{ij} be binary variables (1 or 0) denoting whether or not user i is associated with BS j, and let K_j be the number of users assigned to BS j. The user association problem can be written as:

$$\underset{x_{ij}, K_j}{\text{maximize}} \quad \sum_{ij} x_{ij} \log \left(\frac{W}{K_j} \log(1 + \text{SINR}_{ij}) \right) \tag{2}$$

$$\text{subject to} \quad \sum_j x_{ij} = 1, \ i = 1, \dots, K \tag{3}$$

$$\sum_i x_{ij} = K_j, \ j = 1, \dots, L \tag{4}$$

$$\sum_j K_j = K \tag{5}$$

$$x_{ij} \in \{0, 1\} \tag{6}$$

[1]This is justified in [9] as Round-robin scheduling can be shown to maximize the log-utility.

where (3) states that each user can only associate with one BS, and (5) states that all users in the network must be served. To simplify the notation, we introduce a_{ij} as

$$a_{ij} = \log \left(W \log(1 + \text{SINR}_{ij}) \right). \tag{7}$$

The objective function (2) can now be written as

$$f(x_{ij}, K_j) = \sum_{ij} a_{ij} x_{ij} - \sum_j K_j \log(K_j) \tag{8}$$

3. ALGORITHM

A. Lagrangian Dual Analysis

The BS association problem (2) is a discrete optimization problem, which is typically difficult to solve. However, it turns out that its dual problem can be explicitly derived. This forms the basis of the pricing update algorithm of this paper.

Introduce dual variables $\mu = (\mu_1, \dots, \mu_L)$ for the constraint (4) and ν for (5). The Lagrangian function of the optimization problem with respect to these two constraints is

$$L(x_{ij}, K_j, \mu_j, \nu) = \sum_{ij} a_{ij} x_{ij} - \sum_j K_j \log(K_j)$$

$$- \sum_j \mu_j \left(\sum_i x_{ij} - K_j \right) - \nu \left(\sum_j K_j - K \right). \tag{9}$$

The dual function $g(\cdot)$ can now be stated as

$$g(\mu_j, \nu) = \begin{cases} \max & L(x_{ij}, K_j, \mu_j, \nu) \\ \text{s.t.} & \sum_j x_{ij} = 1; \ x_{ij} \in \{0, 1\} \end{cases} \tag{10}$$

The above optimization problem has the following explicit analytic solution:

$$x_{ij} = \begin{cases} 1, & \text{if } j = j^{(i)} \\ 0, & \text{if } j \neq j^{(i)} \end{cases} \quad \text{where } j^{(i)} = \arg\max_j \left(a_{ij} - \mu_j \right) \tag{11}$$

and

$$K_j = e^{\mu_j - \nu - 1}. \tag{12}$$

Substituting (11) and (12) back into (10), we obtain the dual objective

$$g(\mu_j, \nu) = \sum_i \max_j \left(a_{ij} - \mu_j \right) + \sum_j \left(e^{\mu_j - \nu - 1} \right) + \nu K. \tag{13}$$

The dual problem is now just the minimization of the above $g(\mu_j, \nu)$ over μ_j and ν. Note that if $j^{(i)}$ in (11) is not unique, x_{ij} can be assigned value 1 for any of the BSs with maximum $(a_{ij} - \mu_j)$ without affecting the value of $g(\mu_j, \nu)$ in (13).

The main point of this paper is that the optimization of $g(\mu_j, \nu)$ can be done via coordinate descent. This approach is inspired by the development of auction algorithm by Bertsekas [10]. The BS association problem considered in this paper can be thought of as a generalization of the 1-to-1 assignment problem solved by the auction algorithm [10] to the K-to-1 case.

B. Subgradient Method

We first review the subgradient method for optimizing $g(\mu_j, \nu)$ as proposed in [9]. Observe first that if μ_j's are fixed, $g(\cdot)$ is a differentiable convex function of ν, and the optimal ν can be found as

$$\nu^{(t+1)} = \log \frac{\sum_j e^{\mu_j^{(t)}-1}}{K}. \tag{14}$$

However, $g(\cdot)$ is not a differentiable function of μ_j, so instead of taking the gradient with respect to μ_j, the subgradient method updates μ_j in each step according to

$$\mu_j^{(t+1)} = \mu_j^{(t)} - \alpha_t \left(e^{\mu_j^{(t)}-\nu^{(t)}-1} - K_j^{(t)} \right), \ j = 1, \dots, L \tag{15}$$

where α_t is the step size and $K_j^{(t)} = \sum_i x_{ij}$, where x_{ij} is as defined in (11). Note that again x_{ij} may not be unique because $j^{(i)}$ in (11) may not be unique, but any valid set of x_{ij}'s gives a valid subgradient.

One problem with the subgradient method is that its convergence speed depends heavily on the choice of step sizes α_t. Possible choices of α_t include constant step size (but the constant is difficult to choose) or diminishing step sizes (which guarantee convergence but can be quite slow in practice). As a baseline for comparison, this paper adopts the self-adaptive scheme of [11] as suggested in [9]. We refer the detailed algorithm description to [11], and only mention that the scheme involves quite a few parameters, namely γ_t, $\rho \geq 1$, $\beta < 1$, as well as δ_1 and δ.

Note that because all the μ_j's need to be updated at the same time using the same step size (in order to ensure convergence), the distributed implementation of the subgradient method requires synchronized price updates across the BSs. The dual coordinate descent algorithm proposed in the next section removes such a requirement.

C. Dual Coordinate Descent

The basic idea of the proposed approach is to recognize that the dual function (13) is in a closed form, and it can be optimized in a coordinate descent fashion in closed form. Fixing all the μ_j's, we see that ν can be updated according to (14). Fixing ν and all μ_j's except one of them, we see that $g(\mu_j, \nu)$ is in fact the sum of a continuous piece-wise linear function and an exponential function. So we can take its left and right derivatives and choose μ_j to be such that the left-derivative at μ_j is less than or equal to zero, and the right derivative is greater than or equal to zero. Mathematically, define functions

$$f_1(\mu_j) = |\mathcal{C}_j|, \text{where } \mathcal{C}_j = \left\{ i \left| a_{ij} - \mu_j = \max_s \{a_{is} - \mu_s\} \right. \right\} \tag{16}$$

and

$$f_2(\mu_j) = e^{\mu_j - \nu - 1}. \tag{17}$$

(a) f_1 and f_2 have intersection   (b) no intersection
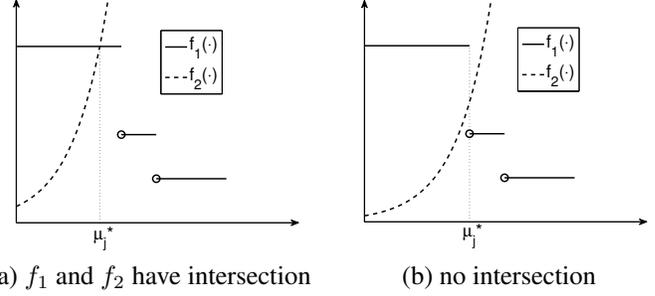
Fig. 1. Two cases of updating μ_j^* in dual coordinate descent

It is easy to see that the left partial derivative of $g(\mu_j, \nu)$ with respect to μ_j is just $f_2(\mu_j) - f_1(\mu_j)$. So, fixing all other dual variables, the μ_j that minimizes $g(\cdot)$ is just

$$\mu_j^{(t+1)} = \sup \{\mu_j | f_2(\mu_j) - f_1(\mu_j) \leq 0\} \tag{18}$$

The dual coordinate descent algorithm is described below:

Algorithm 1 Dual Coordinate Descent

Initialization: Set $\mu_j = 0, \forall j$. Set $\nu = \log \frac{\sum_j e^{\mu_j - 1}}{K}$;
repeat
  1) For $j = 1, \dots, L$, update μ_j according to (18);
  2) Update ν according to (14);
until Convergence;
Set user-BS association according to (11). Resolve ties if necessary.

The dual coordinate descent algorithm is quite intuitive. The dual variable μ_j is the price of the BS j, while a_{ij} is the utility of the user i if it is associated with BS j. Each user maximizes its utility minus the price among all possible BSs, while the BSs choose their prices in an iterative fashion to balance their loads. Fig. 1 illustrates the price update condition, which seeks μ_j^* to match $f_1(\mu_j^*)$ and $f_2(\mu_j^*)$. Here, $f_1(\cdot)$ is a step function. The functions $f_1(\cdot)$ and $f_2(\cdot)$ may not intersect, but the optimal μ_j^* can always be determined uniquely.

As mentioned earlier, the main advantage of the dual coordinate descent algorithm as compared to the subgradient method is that the BSs do not need to synchronize in their price updates. In fact, the order of price updates in Algorithm 1 can be arbitrary. Since each dual update step always decreases the dual objective, the iterative algorithm is always guaranteed to converge.

To completely specify the algorithm, we also need to describe how to recover the primal variables x_{ij} from a set of dual solutions. This is done using (11), but a user may have several BSs with equal $(a_{ij} - \mu_j)$. To resolve such ties, one may use greedy search or relaxation methods. In our simulations, only a very small number of users are typically involved in ties, so tie-breaking by exhaustive search is also feasible.
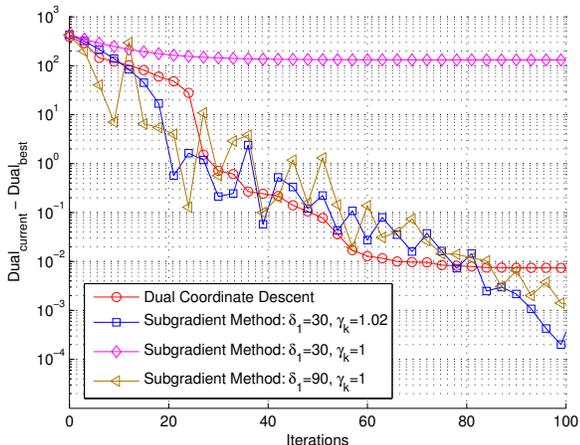
Fig. 2. Convergence: Dual coordinate descent vs Subgradient

Fig. 3. Data rate CDF after 56 iterations

D. Duality Gap

Because of the fact that $g(\mu_j, \nu)$ is not differentiable everywhere, the coordinate descent algorithm is not guaranteed to converge to the optimal dual variables. Further, because of the integer constraints, there may be a non-zero optimal duality gap between the primal and the dual problems. For the BS association problem considered in this paper, the duality gap can actually be characterized in closed form.

Let (μ_j, ν) be a dual feasible solution, and (x_{ij}, K_j) be the primal solution recovered by (11) and tie-breaking. It is possible to show that

$$f(x_{ij}, K_j) = g(\mu_j, \nu) - \sum_j K_j \log\left(\frac{K_j}{e^{\mu_j - \nu - 1}}\right) \quad (19)$$

By weak duality, $g(\cdot)$ is greater than or equal to the primal optimum $f^*(\cdot)$. Thus, the difference between the current primal objective $f(x_{ij}, K_j)$ and the true optimum is bounded by $\sum_j K_j \log\left(\frac{K_j}{e^{\mu_j - \nu - 1}}\right)$. Note that whenever $K_j = e^{\mu_j - \nu - 1}$ for a BS j, as in Fig. 1(a), it does not contribute to the duality gap. When a BS is involved in ties, the duality gap is minimized when K_j is made as close to $e^{\mu_j - \nu - 1}$ as possible.

4. NUMERICAL RESULTS

We validate the proposed algorithm in a 7-cell wrap-around topology with 7 macro BSs and 3 femto/pico BSs per cell, for a total of 28 BSs. The macro-BSs are 1.4km apart. The femto BSs are located 0.5km from the macro BSs. A total of 840 users are distributed uniformly across the network. The transmit power spectrum densities (PSDs) of macro-BSs and femto-BSs are set to be -27dBm/Hz and -47dBm/Hz, respectively, over a 10MHz bandwidth. PSD of background noise is assumed to be -174dBm/Hz. The channels are modeled with a distance-dependent path-loss given by $128.1 +$
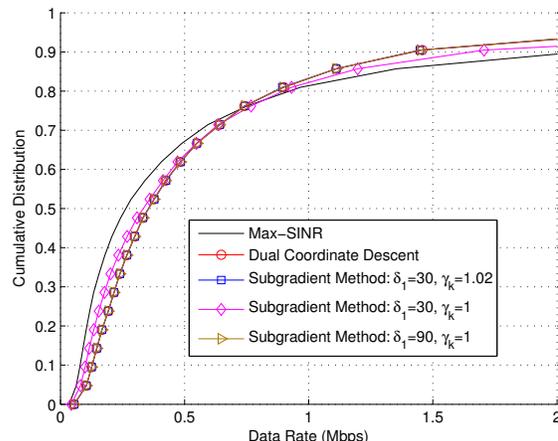
$37.6 \log_{10}(d)$dB (where d is in km) and a log-normal shadowing with standard deviation of 8dB. An antenna gain of 15 dBi is assumed.

Fig. 2 compares the convergence behavior of the dual coordinate descent algorithm with that of the adaptive subgradient method. Each iteration here refers to either a single update of μ_j or a subgradient update of all μ_j's. We see that the dual coordinate descent converges to within 10^{-2} of the optimum with only three rounds of iterations per BS, while the convergence of the subgradient method is very sensitive to parameters. Here, we set $\rho = 1.2$, $\beta = 0.9$, and $\delta = 0.002$ in adaptive subgradient [11] and see that different settings of δ_1 and γ_k can result in quite different convergence behaviors.

We notice in Fig. 2 that the dual coordinate descent algorithm does not converge to the optimum. This is due to the fact that it is possible for coordinate descent to get stuck in a suboptimal point. This gap is quite small, however.

Fig. 3 shows the cumulative distribution of data rates after 56 iterations for the various algorithms. We see that both the dual coordinate descent and the subgradient algorithms, when converged, offer substantial rate improvement as compared to the max-SINR BS association rule. The 50th-percentile rate is increased by about 30%. This is a consequence of off-loading traffic from the macro-BSs to the femto-BSs.

5. CONCLUSIONS

This paper proposes a new cell association algorithm for the heterogenous network based on the use of the pricing approach for BS association and dual coordinate descent for price updates. Comparing to previous approaches, the proposed algorithm has faster convergence and leads to a solution structure more suitable for distributed implementation. It also leads to more balanced traffic loads and proportionally fairer allocation of resources than the SINR-based cell association.

6. REFERENCES

[1] J.-W. Lee, R. R. Mazumdar, and N. B. Shroff, "Joint resource allocation and base-station assignment for the downlink in CDMA networks," *IEEE/ACM Trans. Netw.*, vol. 14, no. 1, pp. 1–14, Jan. 2006.

[2] R. Sun, M. Hong, and Z.-Q. Luo, "Optimal joint base station assignment and power allocation in a cellular network," in *IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC)*, June 2012, pp. 234–238.

[3] S. Corroy, L. Falconetti, and R. Mathar, "Cell association in small heterogeneous networks: Downlink sum rate and min rate maximization," in *IEEE Wireless Commun. Networking Conf. (WCNC)*, 2012, pp. 888–892.

[4] M. Sanjabi, M. Razaviyayn, and Z.-Q. Luo, "Optimal joint base station assignment and downlink beamforming for heterogeneous networks," in *IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, Mar. 2012, pp. 2821–2824.

[5] L. Jiang, S. Parekh, and J. Walrand, "Base station association game in multi-cell wireless networks," in *IEEE Wireless Commun. Networking Conf. (WCNC)*, 2008, pp. 1616–1621.

[6] R. Madan, J. Borran, A. Sampath, N. Bhushan, A. Khandekar, and T. Ji, "Cell association and interference coordination in heterogeneous LTE-A cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 28, no. 12, pp. 1479–1489, Dec. 2010.

[7] T. Bu, L. Li, and R. Ramjee, "Generalized proportional fair scheduling in third generation wireless data networks," in *INFOCOM*, Apr. 2006, pp. 1–12.

[8] K. Son, S. Chong, and G. de Veciana, "Dynamic association for load balancing and interference avoidance in multi-cell networks," *IEEE Trans. Wireless Commun.*, vol. 8, no. 7, pp. 3566–3576, July 2009.

[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," submitted for publication [Online]. Available: http://arxiv.org/abs/1205.2833, May 2012.

[10] D. P. Bertsekas, "The auction algorithm: A distributed relaxation method for the assignment problem," *Annals of Operations Research*, vol. 14, pp. 105–123, Dec. 1988.

[11] D. P. Bertsekas, *Convex Optimization Theory*, Athena Scientific, 2009.