Flexible Multiple Base Station Association and Activation for Downlink Heterogeneous Networks

Kaiming Shen[1], Ya-Feng Liu[2], David Y. Ding[1], and Wei Yu[1]

[1]Electrical and Computer Engineering Department
University of Toronto

[2]Institute of Computational Mathematics and Scientific/Enginnering Computing
Chinese Academy of Sciences

July 2017

Motivation

- Consider a downlink HetNet where macro and pico BSs coexist.
- Macro has much higher Tx power & on-power than pico.
- Central questions in the network design:
  - Load balancing between macro and pico
  - Fair resource allocation within the cell
  - Cross-cell interference control
  - Power saving by BS turnoff
- This paper: A system-level optimization approach for the entire network.

Flexible BS Association

- Most current works [Ye et al.,'13] [Shen-Yu,'14] focus on *single-BS* association, i.e., the association of each user is fixed at one BS.
- We propose a more flexible policy for BS association:
    - User-BS association can be different across the frequency bands.
    - A user can be associated with multiple BSs across different bands.
    - We optimize $x_{k\ell}^n$: the fraction of time BS ℓ serves user k in band n.
- Our work further introduces *power saving* in the system design.

Problem Formulation

- Per-link rate (between user k and BS ℓ in band n):

$$r_{k\ell}^n = \frac{W}{N} \log \left(1 + \frac{g_{k\ell}^n p_\ell^n}{\sigma^2 + \sum_{\ell' \neq \ell} g_{k\ell'}^n p_{\ell'}^n} \right).$$

- Per-user rate:

$$R_k = \sum_{\ell=1}^{L} \sum_{n=1}^{N} x_{k\ell}^n r_{k\ell}^n, \;\; k = 1, 2, \ldots, K.$$

- Each BS ℓ has a power consumption (Tx power plus on-power):

$$Q_\ell(\mathbf{p}_\ell) = \mathbf{e}_N^T \mathbf{p}_\ell + \psi_\ell \left\| \mathbf{p}_\ell \right\|_0$$

where ψ_ℓ is on-power, $\mathbf{p}_\ell = (p_\ell^1, \ldots, p_\ell^N)$ the Tx power across bands.

Problem Formulation (cont.)

- Consider maximizing log-utility and minimizing power jointly (given $\lambda \geq 0$):

$$\begin{aligned}
\max_{\mathbf{X},\mathbf{P}} \quad & \sum_{k=1}^{K} \log\left(R_k\right) - \lambda \sum_{\ell=1}^{L} Q_\ell\left(\mathbf{p}_\ell\right) && \text{(1a)}\\
\text{s.t.} \quad & \mathbf{0} \leq \mathbf{p}_\ell \leq \bar{\mathbf{p}}_\ell,\ \ \ell = 1, 2, \ldots, L && \text{(1b)}\\
& \mathbf{X}^n \mathbf{e}_L \leq \mathbf{e}_K,\ \ n = 1, 2, \ldots, N && \text{(1c)}\\
& \left(\mathbf{X}^n\right)^T \mathbf{e}_K \leq \mathbf{e}_L,\ \ n = 1, 2, \ldots, N && \text{(1d)}\\
& \mathbf{X}^n \geq \mathbf{0},\ \ n = 1, 2, \ldots, N && \text{(1e)}
\end{aligned}$$

- Constraints:
  - (1b): Tx power constraint
  - (1c): Total fraction a user is served from all BSs in one band ≤ 1.
  - (1d): Total fraction each BS allocates to all its users in one band ≤ 1.

Iterative Reweighting

- It is difficult to optimize the ℓ_0-norm term in Q_ℓ directly.

- We propose to approximate ℓ_0-norm as ℓ_2-norm via iterative reweighting, i.e., $\|\mathbf{p}_\ell\|_0 \approx w_\ell \|\mathbf{p}_\ell\|_2$, where $w_\ell(t+1) = \frac{1}{\|\mathbf{p}_\ell(t)\|_2 + \tau(t)}$.

  Intuition: If some $\mathbf{p}_\ell$ is small, we raise its w_ℓ to encourage setting it to $\mathbf{0}$.

- This leads us to a new problem objective: $\max\ f(\mathbf{X}, \mathbf{P}) - h(\mathbf{P})$, which is composed of the smooth part

$$f(\mathbf{X}, \mathbf{P}) = \sum_{k=1}^{K} \log(R_k) - \lambda \mathbf{e}_L^T \mathbf{P} \mathbf{e}_N$$

and the non-smooth part

$$h(\mathbf{P}) = \lambda \sum_{\ell=1}^{L} \psi_\ell w_\ell ||\mathbf{p}_\ell||_2.$$

Gradient Projection for Updating $\mathbf{X}$

- We first consider optimizing $\mathbf{X}$ alone with $\mathbf{P}$ held fixed, that is

$$\max_{\mathbf{X}} \quad \sum_k \log\left(\sum_{\ell,n} x_{k\ell}^n r_{k\ell}^n\right) \tag{2a}$$

$$\text{s.t.} \quad \mathbf{X}^n \mathbf{e}_L \leq \mathbf{e}_K,\ n = 1, 2, \ldots, N \tag{2b}$$

$$\left(\mathbf{X}^n\right)^T \mathbf{e}_K \leq \mathbf{e}_L,\ n = 1, 2, \ldots, N \tag{2c}$$

$$\mathbf{X}^n \geq \mathbf{0},\ n = 1, 2, \ldots, N. \tag{2d}$$

This problem is convex!

- To attain the optimum, we iteratively update $\mathbf{X}$ to $\tilde{\mathbf{X}}$ in the gradient direction, then project $\tilde{\mathbf{X}}$ to the constraint: i.e., $\min_{\mathbf{X}^n} \frac{1}{2}\|\mathbf{X}^n - \tilde{\mathbf{X}}\|_2^2$ s.t. (2b)-(2d).
- The projection can be done more easily in the dual Lagrangian domain.

Projection in the Lagrangian Dual Domain

- The Lagrangian function is (where $\mathbf{y}^n$ and $\mathbf{z}^n$ are dual variables)

$$\begin{aligned}\mathcal{L}(\mathbf{X}^n, \mathbf{y}^n, \mathbf{z}^n) &= \frac{1}{2}\|\mathbf{X}^n - \tilde{\mathbf{X}}^n\|_2^2 + (\mathbf{X}^n\mathbf{e}_L - \mathbf{e}_K)^T\mathbf{y}^n + ((\mathbf{X}^n)^T\mathbf{e}_K - \mathbf{e}_L)^T\mathbf{z}^n \\ &= \frac{1}{2}\|\mathbf{X}^n\|_2^2 - \mathrm{tr}(\mathbf{X}^n(\tilde{\mathbf{X}}^n)^T) + \mathbf{e}_L^T(\mathbf{X}^n)^T\mathbf{y}^n + \mathbf{e}_K^T\mathbf{X}^n\mathbf{z}^n \\ &\quad - \mathbf{e}_K^T\mathbf{y}^n - \mathbf{e}_L^T\mathbf{z}^n + \mathrm{const}(\tilde{\mathbf{X}}^n)\end{aligned}$$

- Find $(\mathbf{X}^n)^* = \tilde{\mathbf{X}}^n - \mathbf{y}^n\mathbf{e}_L^T - \mathbf{e}_K(\mathbf{z}^n)^T$; substitute $(\mathbf{X}^n)^*$ back in $\mathcal{L}$ yields

$$\min_{\mathbf{X}^n} \mathcal{L} = -\frac{1}{2}\|\tilde{\mathbf{X}}^n - \mathbf{y}^n\mathbf{e}_L^T - \mathbf{e}_K(\mathbf{z}^n)^T\|_2^2 - \mathbf{e}_K^T\mathbf{y}^n - \mathbf{e}_L^T\mathbf{z}^n + \mathrm{const}(\tilde{\mathbf{X}}^n)$$

- Thus, the dual problem $\max_{\mathbf{y}^n\geq\mathbf{0},\mathbf{z}^n\geq\mathbf{0}} \min_{\mathbf{X}^n} \mathcal{L}$ is equivalent to

$$\min_{\mathbf{y}^n\geq\mathbf{0},\mathbf{z}^n\geq\mathbf{0}} \frac{1}{2}\|\tilde{\mathbf{X}}^n - \mathbf{y}^n\mathbf{e}_L^T - \mathbf{e}_K(\mathbf{z}^n)^T\|_2^2 + \mathbf{e}_K^T\mathbf{y}^n + \mathbf{e}_L^T\mathbf{z}^n$$

Solving this problem is more computationally efficient than the direct projection because its constraints are much easier!

Proximal Gradient for Updating $\mathbf{P}$

- For fixed $\mathbf{X}$, the optimization of $\mathbf{P}$ is

$$\max_{\mathbf{P}} \quad f(\mathbf{P}) - h(\mathbf{P}) \tag{3a}$$
$$\text{s.t.} \quad \mathbf{0} \le \mathbf{p}_\ell \le \bar{\mathbf{p}}_\ell,\ \ell = 1, 2.\ldots, L. \tag{3b}$$

Recall that function h is non-smooth, so gradient does not work here.

- We follow a proximal gradient approach [Parikh-Boyd,'14] by first updating $\mathbf{P}$ by gradient step with respect to the smooth part f only:

$$\tilde{\mathbf{p}}_\ell = \mathbf{p}_\ell + \beta_\ell \nabla_{\mathbf{p}_\ell} f,\ \ell = 1, \ldots, L.$$

We then further update $\mathbf{P}$ to the proximal projection of $\tilde{\mathbf{P}}$:

$$\mathbf{p}_\ell^{\text{new}} = \arg\min_{\mathbf{u}_\ell} t_\ell h(\mathbf{u}_\ell) + \frac{1}{2}\|\mathbf{u}_\ell - \tilde{\mathbf{p}}_\ell\|_2^2.$$

- Intuition: The gradient of the smooth part $\tilde{\mathbf{p}}_\ell$ does not account for $h(\mathbf{P})$. The proximal gradient step adds a $\frac{1}{2}\|\mathbf{u}_\ell - \tilde{\mathbf{p}}_\ell\|_2^2$ term to min of $h(\mathbf{P})$.

Proximal Gradient for Updating $\mathbf{P}$ (cont.)

- A nice feature of our problem is that it has closed-form proximal projection!
- We solve $\min_{\mathbf{u}} t\|\mathbf{u}\|_2 + \frac{1}{2}\|\mathbf{u} - \tilde{\mathbf{p}}\|_2^2$. By setting its subgradient to $\mathbf{0}$, we get: $(t/\|\mathbf{u}\|_2 + 1)\mathbf{u} = \tilde{\mathbf{p}}$, so $\mathbf{u}^*$ and $\tilde{\mathbf{p}}$ are of the same direction, i.e., $\mathbf{u}^* = \lambda\tilde{\mathbf{p}}$, $\lambda \geq 0$.
- We then consider $\min_\lambda t\lambda\|\tilde{\mathbf{p}}\|_2 + \frac{(\lambda-1)^2}{2}\|\tilde{\mathbf{p}}\|_2^2$. Scalar variable λ can be optimally determined as $\lambda^* = [1 - t/\|\tilde{\mathbf{p}}\|_2]^+$. More specifically,

$$\mathbf{p}_\ell^{\text{new}} = \max\left\{1 - \frac{t_\ell}{\|\tilde{\mathbf{p}}_\ell\|_2}, 0\right\}\tilde{\mathbf{p}}_\ell.$$

- In summary, we iteratively update $\mathbf{X}$ by gradient and update $\mathbf{P}$ by proximal. Convergence is guaranteed by proper step size (e.g., by backtracking).

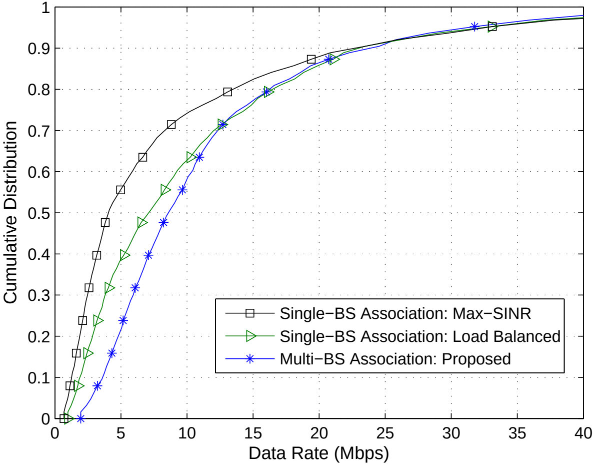
Benchmark

- We introduce a common greedy heuristic for comparison purpose.
- Greedily turn off one BS at a time; fix the on-off pattern (so the non-smooth part h can be ignored) and then optimize $\mathbf{X}$ and $\mathbf{P}$ by gradient.
- Drawbacks:
  - Greedy can turn off (at most) one BS at each iteration.
  - How BSs are ordered in the sequence of greedy test is critical to the performance; but it is hard to decide the sequence in practice.

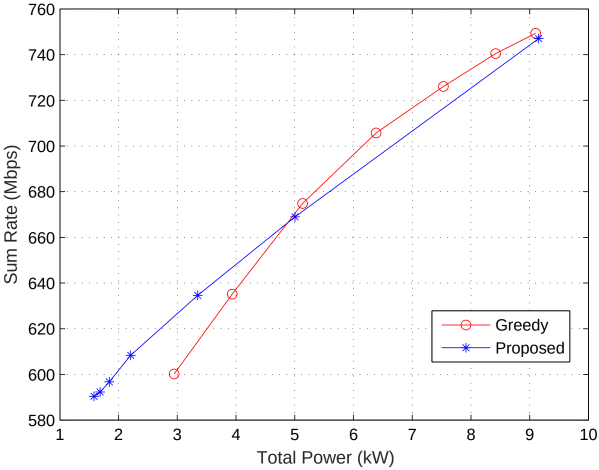
Simulation Model

- 7 wrapped-around cells
- 1 macro BS and 3 pico BSs per cell
- 9 users uniformly distributed per cell
- Total bandwidth of 10MHz dived into 4 equal bands
- Max Tx power spectrum density: macro is -27dBm/Hz, pico is -47dBm/Hz
- ON-power: macro is 1450W, pico is 21.32W

Rate CDF: Single-BS vs. Multiple-BS Association



BS Turning-Off: Greedy vs. Proposed Method



Conclusion

- We propose a novel multiple-band multiple-BS association, which is more flexible than the conventional single-BS association.
- We formulate a novel utility-minus-power problem to take both network throughput and power consumption into account.
- We propose a gradient projection to optimize the time-sharing variable $\mathbf{X}$.
- We apply iterative reweighting and proximal gradient to optimize the power variable $\mathbf{P}$ in its sparse and non-differentiable objective function .
- Numerical results show that the proposed method enables effective balancing between network throughput and power consumption.