

TRAFFIC-AWARE ASSOCIATION IN HETEROGENEOUS NETWORKS

Xiliang Luo

School of Information Science and Technology, ShanghaiTech University
319 Yueyang Rd., Shanghai, 200031, China

ABSTRACT

To meet the ever-growing mobile data traffic, network spatial densification with various low-power nodes in addition to the conventional high-power macro base stations (BS), a.k.a. heterogeneous network (HetNet), is regarded as one key enabling solution. Due to the unplanned nature, HetNets are very irregular and severe interference can happen without judicious designs of the user association rules. Conventional maximum downlink (DL) signal-to-interference-plus-noise ratio (SINR) association rule, on the other hand, can not fully release the traffic offloading capabilities of HetNets. Most prior work assumed best-effort (BE) traffic and sought the optimal association to maximize metrics such as the sum of log-rate. In this paper, we model the DL quality-of-service (QoS) flow to each mobile station (MS) explicitly. Aiming to minimize the network-wide packet delay, we investigate the optimal user association scheme and the corresponding resource allocation algorithm. Our proposed traffic-aware user association strategy enjoys low complexity and fast convergence, which is also corroborated with simulations.

Index Terms— Traffic, association, heterogeneous network

1. INTRODUCTION

A Heterogeneous network (HetNet) consists of a mixture of base stations¹ (BS) of different power classes, e.g. macro-BS at 40dBm, pico-BS at 30dBm, and femto-BS at 20dBm. It has great potential to meet the fast-growing mobile data traffic [1,2]. Unlike the conventional macro-based cellular networks, low power nodes are overlaid within the macro networks in HetNets to incur network spatial densification with lower capital expenditure cost [3,4].

In the conventional cellular networks, mobile stations (MS) are associated to the BSs offering the highest downlink (DL) signal-to-interference-plus-noise ratio (SINR) (a.k.a. Max-DL-SINR Association). In a HetNet, due to the huge disparity in the transmission power, the amount of MSs that become associated with the low-power nodes is not enough to balance the load with

the conventional Max-DL-SINR association. To address this issue, one simple approach named “cell range expansion” (CRE) was proposed in 3GPP, where a positive bias is added to the DL SINR from the low-power nodes. CRE thus allows the low-power nodes to serve more users even when the serving SINR is not the highest [4, 5]. But CRE fails to take into account the traffic load condition at each BS. In this paper, we will study the optimal association rule in HetNets when the traffic load information can be exploited.

A lot of works have been done regarding the user association in HetNets under different scenarios. In [6], a joint optimization of user association, transmission power, and channel selection was considered to minimize the sum of the inverse of the per-user throughput. In [7], the authors looked for the optimal association rule to maximize the number of admitted users while minimizing the total usage of the radio resources in a HetNet with relay nodes. In [8,9], joint resource allocation and user association was studied to maximize the proportional fair (PF) metric, i.e. the sum of log-rate. Optimal time resource partitioning and user association maximizing a weighted proportional fair metric was considered in [10] within the framework of 3GPP.

Unlike the aforementioned work, in this paper, we model the DL traffic to each MS explicitly and investigate the association scheme to optimize the network-wide packet delay performance. This metric is more relevant for a HetNet with quality-of-service (QoS) flows than the typically assumed PF metric [11]. Furthermore, we propose one low-complexity user association algorithm that observes the traffic load and enjoys fast convergence.

The rest of the paper is organized as follows: Section 2 describes the system model and the association problem; Section 3 addresses the optimal resource allocation and the Traffic-Aware Association Algorithm (TAAA) is proposed in Section 4; numerical simulations are provided in Section 5; and Section 6 concludes the paper.

2. SYSTEM MODEL AND PROBLEM STATEMENT

We are considering a HetNet of N BSs. These BSs are of different power class, i.e. some are macro-cells, some are pico-cells, and some are femto-cells. There are totally K MSs in coverage and served by this HetNet. In this paper, we consider the case when each MS: $k \in [1, K]$ has its own traffic arriving at the

This work was supported through the start-up funding from ShanghaiTech University.

¹In this paper, terms: “base station”, “cell”, and “node” are used interchangeably.

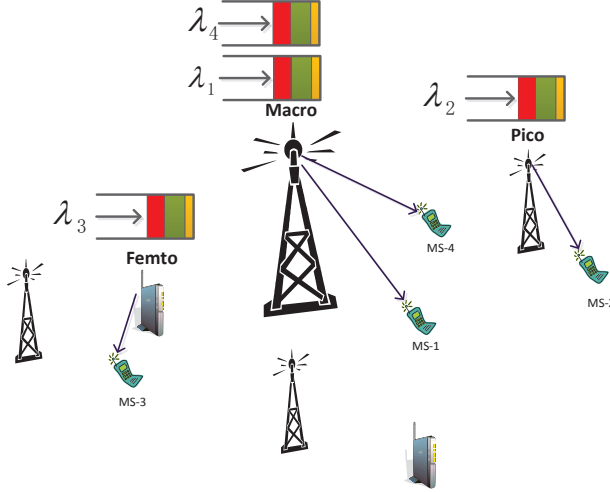


Fig. 1. HetNet with traffic.

network, which needs to be delivered to the MS in the DL. Fig. 1 provides an illustration of one such network.

Before proceeding to the problem description, we define the following quantities first:

- B : system bandwidth in Hz;
- $P_n, n = 1, \dots, N$: transmission power of BS- n ;
- $\lambda_k, L_k, k = 1, \dots, K$: packet arriving rate and average packet length in bits for MS- k ;
- $h_{k,n}, k = 1, \dots, K, n = 1, \dots, N$: channel gain between MS- k and BS- n ;
- $x_{k,n}, k = 1, \dots, K, n = 1, \dots, N$: indicator for the association of MS- k to BS- n , i.e. $x_{k,n} = 1(0)$ means MS- k is associated with BS- n (otherwise);
- $y_{k,n}, k = 1, \dots, K, n = 1, \dots, N$: resource allocation fraction when MS- k is associated with BS- n .

Meanwhile, we make the following assumptions:

- AS1: the inter-arrival times between the packets for MS- k are independent and exponentially distributed with mean $1/\lambda_k$ seconds;
- AS2: the lengths of the packets for MS- k are independent and exponentially distributed with mean L_k bits.
- AS3: each BS is always on even when there are no packets in its queue. This is the case when each BS has some best-effort traffic to serve;
- AS4: full frequency reuse is adopted across the whole network, which is typical in LTE networks, e.g. the co-channel eICIC [2, 4].

Under the above assumptions, when MS- k is associated with BS- n and occupying all the resources, the average achievable rate at the MS is given by

$$R_{k,n} = B \log \left(1 + \frac{P_n |h_{k,n}|^2}{\sum_{l=1, l \neq n}^N P_l |h_{k,l}|^2 + \sigma^2} \right), \quad (1)$$

where σ^2 denotes the thermal noise power at the MS. Taking into account the resource allocation fraction $y_{k,n}$, the service time for the traffic to MS- k is then exponentially distributed with mean $t_{k,n} = L_k / (y_{k,n} R_{k,n}) = 1 / (y_{k,n} r_{k,n})$ seconds, where $r_{k,n}$ denotes the normalized achievable rate with respect to the average packet length. Clearly, the queue for the traffic to MS- k is an M/M/1 queue [12]. Accordingly, the average delay for a packet in the queue to MS- k can be computed as

$$\tau_{k,n} = \frac{1}{1/t_{k,n} - \lambda_k} = \frac{1}{y_{k,n} r_{k,n} - \lambda_k}. \quad (2)$$

The problem we are interested is to find the optimal cell association such that the average packet delay across the whole HetNet is minimized. Mathematically, we want to find the optimal association scheme $\{x_{k,n}\}$ and the corresponding resource allocation scheme $\{y_{k,n}\}$ in the following optimization problem:

$$\text{minimize}_{x_{k,n}, y_{k,n}} \quad \frac{1}{\sum_{k=1}^K \lambda_k} \sum_{n=1}^N \sum_{k=1}^K \frac{x_{k,n} \lambda_k}{y_{k,n} r_{k,n} - x_{k,n} \lambda_k} \quad (3)$$

$$\text{subject to} \quad \sum_{n=1}^N x_{k,n} = 1.0, \forall k = 1, \dots, K \quad (4)$$

$$\sum_{k=1}^K y_{k,n} \leq 1.0, \forall n = 1, \dots, N \quad (5)$$

$$x_{k,n} \in \{0, 1\}, \forall k = 1, \dots, K, n = 1, \dots, N \quad (6)$$

$$y_{k,n} > \frac{x_{k,n} \lambda_k}{r_{k,n}}, \forall k = 1, \dots, K, n = 1, \dots, N. \quad (7)$$

In the above problem, the equality constraint in (4) and the constraint in (6) dictates that each MS will be associated to a single serving BS. Constraint in (5) represents the limit to the total amount of available resources. This problem turns out to be the classic knapsack problem and is NP-hard [13]. In our following developments, we will try to find low-complexity approximated solutions to this NP-hard problem.

3. OPTIMAL RESOURCE ALLOCATION AND LINEAR APPROXIMATION

In order to find the optimal solution to the problem in (3), we first obtain the optimal resource allocation $\{y_{k,n}\}$ assuming a fixed association scheme $\{x_{k,n}\}$. We can establish the following result:

Proposition 1: At BS- n , given user association $\{x_{k,n}\}$, when the following feasibility condition is satisfied:

$$\sum_{k=1}^K \frac{x_{k,n} \lambda_k}{r_{k,n}} < 1.0, \quad (8)$$

the optimal resource allocation minimizing the average packet delay is as follows:

$$y_{k,n} = \frac{x_{k,n} \lambda_k}{r_{k,n}} + \frac{1 - \sum_{u=1}^K \frac{x_{u,n} \lambda_u}{r_{u,n}}}{\sum_{u=1}^K \sqrt{\frac{x_{u,n} \lambda_u}{r_{u,n}}}} \sqrt{\frac{x_{k,n} \lambda_k}{r_{k,n}}}. \quad (9)$$

Here is the proof outline for the above proposition:

Proof. The Lagrangian for minimizing the average packet delay at BS- n is

$$L(\{y_{k,n}\}, \alpha) = \sum_{k=1}^K \frac{x_{k,n} \lambda_k}{y_{k,n} r_{k,n} - x_{k,n} \lambda_k} + \alpha \left(\sum_{k=1}^K y_{k,n} - 1.0 \right),$$

where α is the Lagrange multiplier for the total resource constraint in (5). From the duality theory [13], we know the optimal resource allocation should minimize the Lagrangian and we have:

$$\frac{\partial L}{\partial y_{k,n}} = \frac{-x_{k,n} \lambda_k r_{k,n}}{(y_{k,n} r_{k,n} - x_{k,n} \lambda_k)^2} + \alpha = 0.$$

Thus we know $\alpha > 0$ and we can get

$$y_{k,n} = \frac{x_{k,n} \lambda_k}{r_{k,n}} + \frac{1}{\sqrt{\alpha}} \sqrt{\frac{x_{k,n} \lambda_k}{r_{k,n}}}. \quad (10)$$

Since the total resource constraint is active, i.e. $\sum_k y_{k,n} = 1$, the result in Proposition 1 can be obtained naturally. \square

The optimal resource allocation contains the non-linear term: $\sqrt{x_{k,n} \lambda_k / r_{k,n}}$. Noticing that $|\sqrt{x} - x| \leq 1/4$ when $x \in [0, 1]$, we can approximate $\sqrt{x_{k,n} \lambda_k / r_{k,n}}$ as $\{x_{k,n} \lambda_k / r_{k,n}\}$. Then the resource allocation fraction becomes

$$y_{k,n} = \frac{1}{\sum_{u=1}^K \frac{x_{u,n} \lambda_u}{r_{u,n}}} \frac{x_{k,n} \lambda_k}{r_{k,n}}. \quad (11)$$

Substituting the resource allocation in (11) to (3), the problem in (3) becomes

$$\underset{x_{k,n}}{\text{minimize}} \quad \sum_{k=1}^K \sum_{n=1}^N \frac{x_{k,n} \lambda_k}{r_{k,n}} \frac{1}{1 - \sum_{u=1}^K \frac{x_{u,n} \lambda_u}{r_{u,n}}} \quad (12)$$

$$\text{subject to} \quad \sum_{n=1}^N x_{k,n} = 1, \forall k = 1, \dots, K \quad (13)$$

$$x_{k,n} \in \{0, 1\}, \forall k = 1, \dots, K, n = 1, \dots, N \quad (14)$$

$$\sum_{k=1}^K \frac{x_{k,n} \lambda_k}{r_{k,n}} < 1, \forall n = 1, \dots, N \quad (15)$$

where we have neglected the constant term in the objective function and the feasibility constraint in (15) dictates that each BS should be able to handle the traffic of all the MSs associated with it.

4. TRAFFIC-AWARE ASSOCIATION

The optimization problem in (12) is still NP-hard to solve. Hence, we first try to relax the constraint in (14), i.e. we allow one MS to be associated with multiple BSs so that $x_{k,n}$ can take fractional values. As a result, the problem in (12) becomes convex as follows:

$$\underset{x_{k,n}}{\text{minimize}} \quad f(\{x_{u,n}\}) = \sum_{k=1}^K f_k(\{x_{u,n}\}) \quad (16)$$

$$\text{subject to} \quad \sum_{n=1}^N x_{k,n} = 1, \forall k = 1, \dots, K \quad (17)$$

$$x_{k,n} \in [0, 1], \forall k = 1, \dots, K, n = 1, \dots, N \quad (18)$$

$$\sum_{k=1}^K \frac{x_{k,n} \lambda_k}{r_{k,n}} < 1, \forall n = 1, \dots, N \quad (19)$$

where the function $f_k(\{x_{u,n}\})$ is defined as:

$$\begin{aligned} f_k(\{x_{u,n}\}) &= \sum_{n=1}^N \frac{x_{k,n} \lambda_k}{r_{k,n}} \frac{1}{1 - \sum_{u=1}^K \frac{x_{u,n} \lambda_u}{r_{u,n}}} \\ &= \sum_{n=1}^N \frac{1}{1 - x_{k,n} \frac{\lambda_k}{r_{k,n} \delta_{k,n}}} - N, \end{aligned}$$

and $\delta_{k,n} := 1 - \sum_{u=1, u \neq k}^K \frac{x_{u,n} \lambda_u}{r_{u,n}}$ denotes the available load in BS- n for MS- k . Now we can establish the following main result regarding the optimal association for one MS:

Proposition 2: For MS- k , given other users' association pattern: $\{x_{u,n}\}, \forall u \neq k$, the optimal association pattern of MS- k minimizing the objective function in (16) is as follows:

$$x_{k,n} = \max \left\{ 0, \frac{r_{k,n} \delta_{k,n}}{\lambda_k} - \alpha \sqrt{\frac{r_{k,n}}{\lambda_k}} \right\}, \quad (20)$$

where α is chosen such that $\sum_n x_{k,n} = 1$.

Brief proof for the above proposition is as follows:

Proof. The Lagrangian for the convex optimization problem in (16) is

$$\begin{aligned} L(\{x_{u,n}\}, \{\nu_u\}, \{\mu_{u,n}\}) \\ &= \sum_k f_k(\{x_{u,n}\}) + \sum_k \nu_k \left(1 - \sum_n x_{k,n} \right) \\ &\quad + \sum_k \sum_n (-\mu_{k,n} x_{k,n}). \end{aligned}$$

According to the duality theory, the gradient of Lagrangian should vanish at the optimal primal and dual points, we have:

$$\frac{\partial L}{\partial x_{k,n}} = \frac{\frac{\lambda_k}{r_{k,n} \delta_{k,n}^2}}{\left(1 - \frac{x_{k,n} \lambda_k}{r_{k,n} \delta_{k,n}} \right)^2} - \nu_k - \mu_{k,n} = 0. \quad (21)$$

Then, we can find the optimal association scheme as

$$x_{k,n} = \frac{r_{k,n} \delta_{k,n}}{\lambda_k} - \frac{1}{\sqrt{\nu_k + \mu_{k,n}}} \sqrt{\frac{r_{k,n}}{\lambda_k}}.$$

Due to the complementary slackness, i.e. $\mu_{k,n} x_{k,n} = 0$, we can obtain the desired result in (20). \square

Now it is clear from Proposition 2 that the right cell association should observe the traffic: $\{\lambda_k\}$, the channel quality: $\{r_{k,n}\}$, and the load in each BS: $\{\delta_{k,n}\}$. Accordingly, we propose one iterative "Traffic-Aware Association Algorithm" (TAA) as follows:

1. **Feasible Start:** Find one feasible association $\{x_{k,n}\}$ by solving the following linear programming (LP) problem:

$$\underset{x_{k,n}, s}{\text{minimize}} \quad s \quad (22)$$

$$\text{subject to} \quad \sum_{n=1}^N x_{k,n} = 1, \forall k \quad (23)$$

$$x_{k,n} \in [0, 1], \forall k, n \quad (24)$$

$$\sum_{k=1}^K x_{k,n} \frac{\lambda_k}{r_{k,n}} - 1 \leq s, \forall n. \quad (25)$$

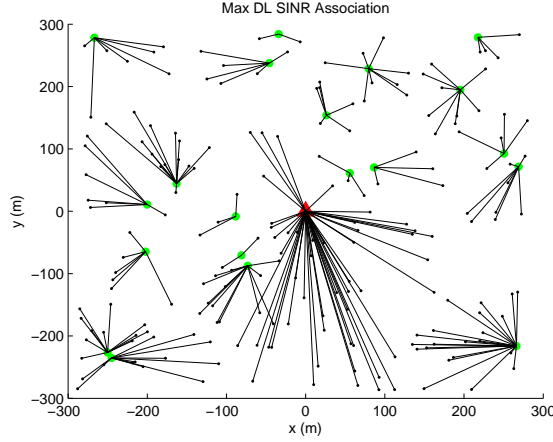


Fig. 2. Maximum DL SINR association (Red Triangle: Macro-BS; Green Circle: Pico-BS; Black Dot: MS).

If the minimum value of s is strictly less than 0, we have one feasible association to proceed. Otherwise, we can conclude that the network can not support the given traffic;

2. **Iterations:** During each iteration, we will update the association pattern of each MS one by one according to the rule specified in Proposition 2;
3. **Finalizing:** In a practical network, each MS is not allowed to be associated with multiple BSs. We use the following rule to generate one permitted association pattern from the fractional association pattern $\{x_{k,n}\}$ resulting from the “Iterations” step:

$$x_{k,n}^{(f)} = 1\{x_{k,n} > x_{k,l}, \forall l \neq n\}, \quad (26)$$

where $1\{\cdot\}$ denotes the indicator function, which is equal to 1 when the specified condition is met and 0 otherwise.

5. SIMULATIONS

We simulate the performance of our proposed TAAA using one HetNet with the following parameters: 1) the network consists of 1 macro-BS (40dBm transmission power (TxPwr)), 20 pico-BSs (30dBm TxPwr), and 200 MSs, which are uniformly dropped in a square specified by $[-288, 288]\text{m} \times [-288, 288]\text{m}$ (see Fig. 2); 2) the distance-dependent path-loss is modelled as: $PL_{dB} = 128.1 + 37.6 \log_{10}(d_{km})$, where d_{km} denotes the distance in kilometer; 3) each MS has one incoming traffic at a rate of 20 packets per second and each packet contains 20k bits on average; 4) the system BW is 10MHz. Conventional Max-DL-SINR association is shown in Fig. 2. The resulting association pattern of our proposed TAAA is plotted in Fig. 3, where we can clearly see that users are offloaded towards the pico-BSs when observing the high load in the macro-BS. In Fig. 4, we show the proposed TAAA can improve the average

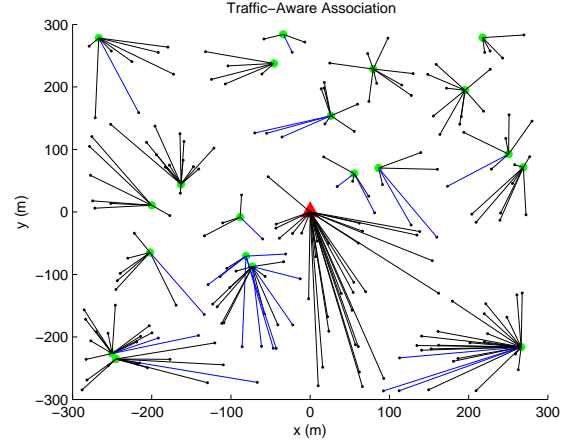


Fig. 3. Traffic-aware association (Blue links indicate the association changes w.r.t. the Max-DL-SINR rule).

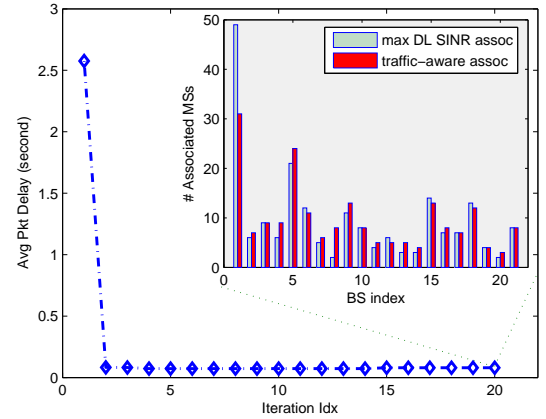


Fig. 4. Average packet delay and load of each BS.

delay performance of the network significantly, i.e. from 2.6s (corresponding to the Max-DL-SINR rule) to 0.07s, and converges quickly within several iterations. Meanwhile, we also depict the traffic-aware offloading statistics in Fig. 4.

6. CONCLUSIONS

In this paper, we have studied the problem of optimal user association in a HetNet with QoS flows. In this case, we are caring about the average packet delay performance other than the typically assumed proportional fair metric. To minimize the average packet delay across the whole network and avoid the need to solve one NP-hard knapsack problem, we proposed the traffic-aware association algorithm (TAAA) which enjoys low complexity and fast convergence. Network simulations demonstrated that our proposed association rule can significantly lower down the packet delay compared to the conventional Max-DL-SINR association. A more rigorous convergence analysis is to be carried out in the follow-up work.

7. REFERENCES

- [1] A. Ghosh, N. Mangalvedhe, R. Ratasuk, B. Mondal, M. Cudak, E. Visotsky, T. Thomas, J. Andrews, P. Xia, H. Jo, H. Dhillon, and T. Novlan, "Heterogeneous cellular networks: From theory to practice," *IEEE Commun. Mag.*, vol. 50, no. 6, pp. 54-64, Jun. 2012.
- [2] 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 Commun.*, vol. 18, no. 3, pp. 10-21, Jun. 2011.
- [3] N. Bhushan, J. Li, D. Malladi, R. Gilmore, D. Brenner, A. Damnjanovic, R. Sukhavasi, C. Patel, and S. Geirhofer, "Network densification: The dominant theme for wireless evolution into 5G," *IEEE Commun. Mag.*, vol. 52, no. 2, pp. 82-89, Feb. 2014.
- [4] S. Sesia, I. Toufik, and M. Baker, *LTE - The UMTS Long Term Evolution: From Theory to Practice*. John Wiley & Sons Ltd., 2nd Ed., 2011.
- [5] J. Andrews, S. Singh, Q. Ye, X. Lin, and H. Dhillon, "An overview of load balancing in HetNets: Old myths and open problems," *IEEE Wireless Commun.*, vol. 21, no. 2, pp. 18-25, Apr. 2014.
- [6] C. Chen, F. Baccelli, and L. Roullet, "Joint optimization of radio resources in small and macro cell networks," in *Proc. IEEE VTC Spring*, 2011.
- [7] Q. Li, R. Hu, G. Wu, and Y. Qian, "On the optimal mobile association in heterogeneous wireless relay networks" in *Proc. IEEE INFOCOM*, 2012.
- [8] D. Fooladivanda and C. Rosenberg, "Joint resource allocation and user association for heterogeneous wireless cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 1, pp. 248-257, Jan. 2013.
- [9] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. Andrews, "User association for load balancing in heterogeneous cellular networks," *IEEE Trans. Wireless Commun.*, vol. 12, no. 6, pp. 2706-2716, Jun. 2013.
- [10] S. Deb, P. Monogioudis, J. Miernik, and J. Seymour, "Algorithms for enhanced inter-cell interference coordination (eICIC) in LTE HetNets," *IEEE/ACM Trans. Netw.*, vol. 22, no. 1, pp. 137-150, Feb. 2014.
- [11] B. Zhuang, D. Guo, and M. Honig, "Traffic-driven spectrum allocation in heterogeneous networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2027-2038, Oct. 2015.
- [12] S. Ross, *Introduction to Probability Models*. Academic Press, 11th Ed., 2014.
- [13] D. Bertsekas, *Nonlinear Programming*. Athena Scientific, 2nd Ed., 1999.