A Proposed Network Balance Index for Heterogeneous Networks

Yifei Huang, Salman Durrani, Pawel Dmochowski, and Xiangyun Zhou

Abstract—Load balancing and fairness are used in the heterogeneous network literature to describe how the users or user rates are distributed across the network. While quantitative metrics of fairness exist, there is no formal metric for quantifying load balance. In this letter, we demonstrate that a "fair" network may not be a balanced one, since fairness is from a user perspective, while load balance is from a network perspective. We propose a new network balance index (NBI) metric to measure the load balance across the network, which accounts for transmit powers, bias values for cell range expansion, and multiplicatively weighted Voronoi coverage areas of heterogeneous base stations. We show an application of the proposed metric by implementing a user association refinement algorithm, which aims to improve the NBI metric. Our mathematical analysis and simulations establish that when users are heavily clustered, increasing the NBI metric using the proposed algorithm also increases the sum rate despite decreasing fairness. This highlights the usefulness of the proposed NBI metric in network planning.

Index Terms—Load balancing, heterogeneous networks, user association, fairness, multiplicatively weighted Voronoi cells.

I. Introduction

LOAD balancing, fairness and user association are key practical considerations to help unlock the potential of heterogeneous networks (HetNets) to meet the coverage and service demands of future wireless networks. It is well known that when users associate with the strongest base station in a HetNet, the huge disparity in the base station transmit powers leads to highly unequal base station loads [1]. Many excellent approaches have been proposed in the literature to address this issue, e.g., using cell biasing to increase the likelihood of users associating with small cells rather than the macro-cells [2] or using load-aware or quality-of-service aware user association policies [3], [4]. In future fifth-generation (5G) networks, clustering and heterogeneous user distributions are expected to lead to more non-uniform networks [5]. In order to avoid congested base stations serving a larger number of users, load balancing becomes even more critical in non-uniform networks [6].

To ensure quality of service for all users, fairness has also been an important consideration, e.g., when designing user association algorithms [3] which also impact the load balancing or encouraging load balancing via optimising proportional fair utility functions [7]. In this regard, the quantitative measures of fairness used in the literature include Jain's fairness index (JFI) and the more generalized α-fairness [7], [8]. In the case of JFI, the fairest network occurs when all users receive the same rate. However, a network where all users receive similar but very low rates (e.g., due to load imbalance and base station being congested) can be deemed 'fair' even though it is undesirable from a load balancing perspective. Thus, in general, fairness metrics cannot be used as a reliable indication of good load balance.

While there have been many efforts in the literature towards methods that facilitate load balancing in HetNets [2], the degree of balance achieved is not explicitly quantified. To the best of the authors' knowledge, there is no formal definition of a balanced network load in the literature, although it has been noted in some papers that *balancing* network load is not necessarily the same as *equalizing* network load [7]. In addition, there is no quantitative measure of network balance, and as such, no objective or comparative way of determining how balanced a network is (regardless of how the user association is achieved). Addressing these two issues is the main focus of this letter.

Letter Contributions: We consider a downlink HetNet where resources are allocated in round robin fashion and each user initially associates with the closest base station, subject to cell-biasing. The novel contributions of the letter are: (i) We propose a network balance index (NBI) metric that quantifies the deviation of the current load distribution to the expected (i.e., ideally balanced) load distribution, the latter being determined using multiplicatively weighted Voronoi cell areas. While multiplicatively weighted Voronoi cells have been used to analyse the coverage areas of HetNets in [5] and [9], their use to describe network balance has not been considered. (ii) We propose an algorithm that uses NBI to improve user association in HetNets. We show analytically and via simulations that when users are heavily clustered, increasing the NBI metric also increases the sum rate as underloaded base stations can better serve edge users.

II. System Model and Problem Formulation

Consider a region with M fixed location base stations and N users, with N_j denoting the number of users associated to base station j. Each base station has transmit power P_j with bias β_j (such that the effective power is $P_j\beta_j$), with the macro bias set to 0 dB. We assume that each base station transmits at its maximum power constantly, and that each user associates with only one base station at a time. We study the downlink scenario in this letter. Users are associated according to their closest weighted Voronoi cells, where the weights are the biased transmit powers, i.e., $w_j = P_j\beta_j$.

We consider the signal-to-noise-ratio, and assume that interference can be dealt with through interference

coordination and orthogonal resource allocation. We assume the proportional fair scheduling scheme, and that time and frequency resources are allocated in round robin fashion, such that if a user i is connected to base station j, that user's rate is inversely proportional to the number of users also connected to that base station [1], i.e.,

$$\frac{1}{N_j} \log_2 \left(1 + \frac{P_j \psi_{j,i} d_{j,i}^{-\alpha} |h_{j,i}|^2}{\sigma_n^2} \right) = \frac{r_{j,i}}{N_j}, \qquad (1)$$

where $|h_{j,i}|^2$ is the Rayleigh channel gain from base station j to user i, $d_{j,i}^{-\alpha}$ is the pathloss due to distance $d_{j,i}$ with pathloss exponent α, $\psi_{j,i}$ represents lognormal shadowing with mean 0 dB and variance σ_s^2, σ_n^2 is the Gaussian noise variance and $r_{j,i}$ is the rate without load considerations.

Our performance metric is sum rate, which is the total of all the users' rates in bits/s/Hz. We also impose a maximum user rate such that sum rate and fairness values are not skewed by users associated with lightly loaded base stations.

III. PROPOSED NETWORK BALANCE INDEX

As illustrated in Section I, fairness alone is not a reliable indication of network balance. The reason is twofold: (i) fairness metrics do not inherently consider network balance as they do not take into account user and base station density and geography, and (ii) fairness metrics do not capture the under or over utilization of base stations. *While fairness measures the relative spread or similarity of user rates (user centric), network balance should measure the even distribution of network resources according to network topology (network centric).* A balanced network is desired from a load balancing perspective. For instance, an overloaded base station may not be an option for new entering users, hence forcing those users to have to connect to some less desirable base station. In addition, if we reduce the load on a congested base station, it can better serve its remaining users.

In order to define the proposed network balance index, we need to first define load and expected (i.e., balanced) load. We define load as the percentage of users associated with a base station. In an ideal balanced network, the expected load of a base station should be (i) proportional to the biased transmit power of each base station, and (ii) inversely proportional to the density of surrounding cells, i.e., the more base stations around a particular cell, the less load that cell should have due to competing base stations. Note that expected load must be defined prior to any knowledge about user distribution or association. Otherwise, it is possible to construct a definition of network balance that will be high for any arbitrary user distribution.

To mathematically model these properties, we use the area of a multiplicatively weighted Voronoi cell [10] to represent coverage areas and to help define the expected load. If the area of the weighted Voronoi cell of base station i is x_i, its expected load portion expressed as a percentage is

$$e_i = 100 \times \frac{x_i}{\sum_{j=1}^{M} x_j}. \qquad (2)$$

Unfortunately, x_i has no known closed form expression [10], but can be estimated by geographical data in real applications. In our simulations we will use Monte Carlo methods to approximate x_i. This will be explained in detail in Section V.

Let $\mathbf{e}$ be a $1 \times M$ vector, with elements e_i from (2), denoting the expected load distributions a_i. Let $\mathbf{a}$ be a $1 \times M$ vector containing the actual base station load distribution. We propose a network balance index as follows.

Definition 1: The *network balance index* (NBI) is a measure of the deviation of the current load distribution from the expected load distribution, i.e.,

$$\text{NBI} = 1 - \frac{\|\mathbf{e} - \mathbf{a}\|_1}{200}, \qquad (3)$$

where $\|\cdot\|_1$ is the ℓ_1 norm. The vector $\|\mathbf{e} - \mathbf{a}\|_1$ is divided by 200 because the maximum possible deviation is 200%. Note that like JFI, NBI has values in the range $[0, 1]$.

To illustrate, consider a network with one macro and three picos, and that their expected load distributions are $[40, 20, 20, 20]$. If the actual distributions are $[80, 10, 10, 0]$, then the NBI would be

$$1 - \frac{|-40| + |10| + |10| + |20|}{200} = 1 - \frac{80}{200} = 0.6, \qquad (4)$$

indicating that the actual network is 60% balanced with respect to the expected load distribution.

Network balance can provide information about the network that fairness alone does not. For instance, even if sum rate has been optimized with a minimum fairness, improving this sum rate further cannot be done without knowing whether to increase or decrease the fairness constraint. With network balance however, we can identify certain scenarios where increasing NBI will also lead to increasing fairness.

IV. SUM RATE IMPROVEMENT ALGORITHM USING NBI

To show an application of the proposed metric in network planning, we propose an algorithm that uses NBI as an indicator to refine the initial user association. Note that a refinement algorithm cannot aim to increase sum rate, as sum rate is not normalized and therefore optimal values cannot be known beforehand. Thus, we aim to increase NBI, as it can identify which base stations are overloaded and which ones should receive offloaded users.

Our algorithm first associates users to the base station of the weighted Voronoi cell they are located in and then computes the NBI. Next, we denote the most overloaded base station and its users by $s_i \in O$, $i = 1, \ldots, |O|$, and the closest empty base stations followed by the most underloaded base station (ordered from most empty/underloaded to least if there are multiple), by $\mathcal{U}_j$. At each step, the user from O that is closest to $\mathcal{U}_j$ is reassociated, such that

$$\left\{ s_i \notin O, s_i \in \mathcal{U}_j | a_j \le e_j \right\}. \qquad (5)$$

Each base station $\mathcal{U}_j$ will gain a user from O until $\lfloor \theta N \rfloor$ users are reassociated, or until any $\mathcal{U}_j$ reaches its expected load e_j. $\theta \in [0, 1]$ is the maximum fraction of total users that can be offloaded, rounded down to the nearest integer (line 4 in Algorithm table).

Condition for Increasing NBI and Sum Rate: We can analytically show for a clustered user network that our algorithm increases sum rate by increasing NBI through offloading users. Let $r_{\{\text{over,under}\},i}$ be the rate received by the ith user associated with an overloaded or underloaded base station without load considerations. Suppose one user is being offloaded from an

---

Algorithm 1 Reassociating Users Based on NBI

1: Associate users using (biased) min-distance association.
2: Determine most overloaded (O) and most empty and underloaded base stations ($\mathcal{U}_j$).
3: **Initialize** reassociate $= 0$
4: **while** reassociate $< \lfloor \theta N \rfloor$ and $a_j < e_j$ **do**
5:      Determine user $s_i \in O$ closest to $\mathcal{U}_j$.
6:      Offload user from O to $\mathcal{U}_j$, i.e., $s_i \notin O$, $s_i \in \mathcal{U}_j$.
7:      reassociate = reassociate + 1.
8:      Move onto next $\mathcal{U}_j$
9: **end while**

---

overloaded base station with N_{over} initial users to an under-loaded base station with N_{under} initial users. Using (1), for sum rate to improve, the difference in sum rate before and after offloading must be greater than 0, i.e.,

$$
\underbrace{\left(\sum_{i=1}^{N_{\text{over}}-1} \frac{r_{\text{over},i}}{N_{\text{over}}-1} + \sum_{i=1}^{N_{\text{under}}+1} \frac{r_{\text{under},i}}{N_{\text{under}}+1} \right)}_{\text{After offloading}}
$$

$$
- \underbrace{\left(\sum_{i=1}^{N_{\text{over}}} \frac{r_{\text{over},i}}{N_{\text{over}}} + \sum_{i=1}^{N_{\text{under}}} \frac{r_{\text{under},i}}{N_{\text{under}}} \right)}_{\text{Before offloading}} > 0. \quad (6)
$$

It is clear that the rates of those $N_{\text{over}}-1$ users still associated with the overloaded base station will improve, since

$$
\sum_{i=1}^{N_{\text{over}}-1} \frac{r_{\text{over},i}}{N_{\text{over}}-1} > \sum_{i=1}^{N_{\text{over}}-1} \frac{r_{\text{over},i}}{N_{\text{over}}}. \quad (7)
$$

Therefore, for sum rate to improve, we require that

$$
\underbrace{\sum_{i=1}^{N_{\text{under}}+1} \frac{r_{\text{under},i}}{N_{\text{under}}+1} - \sum_{i=1}^{N_{\text{under}}} \frac{r_{\text{under},i}}{N_{\text{under}}}}_{\triangleq \Omega} - \frac{r_{\text{over},N_{\text{over}}}}{N_{\text{over}}} > 0. \quad (8)
$$

If $N_{\text{under}} \to \infty$, then $\Omega \to 0$, leading to (8) becoming false. Therefore, we conclude that (8) is most easily satisfied if N_{under} is small and N_{over} is large, which is exactly the case when users are heavily clustered. Note that for the special case when the underloaded base station is initially empty, i.e., $N_{\text{under}} = 0$, (8) reduces to

$$
r_{\text{under},1} > \frac{r_{\text{over},N_{\text{over}}}}{N_{\text{over}}}, \quad (9)
$$

which is more easily satisfied when $N_{\text{over}} \to \infty$.

Relationship between Sum Rate and Fairness: We can obtain mathematical insight into how an increase in sum rate may affect JFI. Using the JFI definition [8], the difference in fairness before and after a user is offloaded is

$$
\frac{\left(A + \sum_{i=1}^{N_{\text{over}}-1} \frac{r_{\text{over},i}}{N_{\text{over}}-1} + \sum_{i=1}^{N_{\text{under}}+1} \frac{r_{\text{under},i}}{N_{\text{under}}+1} \right)^2}{N\left(B + \sum_{i=1}^{N_{\text{over}}-1} (\frac{r_{\text{over},i}}{N_{\text{over}}-1})^2 + \sum_{i=1}^{N_{\text{under}}+1} (\frac{r_{\text{under},i}}{N_{\text{under}}+1})^2 \right)}
$$

$$
- \frac{\left(A + \sum_{i=1}^{N_{\text{over}}} \frac{r_{\text{over},i}}{N_{\text{over}}} + \sum_{i=1}^{N_{\text{under}}} \frac{r_{\text{under},i}}{N_{\text{under}}} \right)^2}{N\left(B + \sum_{i=1}^{N_{\text{over}}} (\frac{r_{\text{over},i}}{N_{\text{over}}})^2 + \sum_{i=1}^{N_{\text{under}}} (\frac{r_{\text{under},i}}{N_{\text{under}}})^2 \right)}, \quad (10)
$$

where A is the sum of rates for the other users, and B is the sum of their squared rates.

If users are very clustered, i.e., N_{under} is small and N_{over} is large, we have already established that sum rate tends to increase. Further, that increase is almost solely attributed to the offloaded user, since if N_{over} is large we assume that the $N_{\text{over}}-1$ users of the offloading base station experience similar rates as before. Thus, we can approximate (10) as

$$
\frac{(A' + \epsilon)^2}{N(B' + \epsilon^2)} - \frac{(A')^2}{N(B')}, \quad (11)
$$

where A' and B' are new constants and ϵ is the increase in sum rate. If we set (11) to 0 to study under which conditions it will be positive or negative, the numerator can be reduced to

$$
2A'B' + \epsilon(B' - (A')^2) = 0. \quad (12)
$$

Since $(A')^2$ is a square of sums, and B' is a sum of squares, $(B' - (A')^2) \leq 0$. Therefore, when ϵ is large, which tends to be the case when users are initially highly clustered, fairness tends to decrease since (12) will become a '$<$' inequality. Conversely, as ϵ becomes smaller, fairness tends to increase.

V. SIMULATION RESULTS

We simulate a network with one macro in the centre of a 1 km $\times$ 1 km region, with 4 small cells spaced at a radius 250 m around the macro. Each small cell could be a pico or a femto and this is randomly generated in each realization. Bias values are 7 dB and 13 dB for picos and femtos respectively, while transmit powers are 40 dBm, 30 dBm and 20 dBm in order of decreasing tiers. Pathloss exponent is $\alpha = 3$, log-normal shadowing variance σ_s^2 is 6 dB, noise power σ_n^2 is -174 dBm/Hz, and the transmission bandwidth is 20 MHz. Users are randomly distributed according to a Thomas cluster process[1] centered at the macro with intensity function [11]

$$
\lambda(\mathbf{x}) = \frac{\bar{c}}{2\pi\sigma^2} \exp\left(-\frac{\|\mathbf{x}\|^2}{2\sigma^2} \right), \quad (13)
$$

where the average number of users is $\bar{c} = 500$, σ^2 is the cluster variance, $\mathbf{x}$ is the position vector of the user relative to the parent point, and $\|\cdot\|$ denotes Euclidean norm. The maximum rate for each user is limited to 2 bits/s/Hz, and we set $\theta = 0.1$. To calculate the Voronoi cell area x_i, we divide the region into a 100×100 grid and determine how many grid elements each Voronoi cell contains.

Sum rate improvement: We first compare the sum rate performance of our proposed algorithm with minimum distance association (where users associate with the closest base station) and the dynamic range heuristic proposed in [1], which aims to select the best number of users to connect to picos. Since the system model in [1] is different, for fair comparison we have adapted the heuristic to our system model, such that

---

[1]Our NBI definition is independent of the user distribution. Thomas cluster process is adopted as an example of a clustered process. Other clustered processes display similar performance. Uniformly distributed processes such Point Poisson Process display minor improvements in sum rate and NBI, and thus their results are not shown.
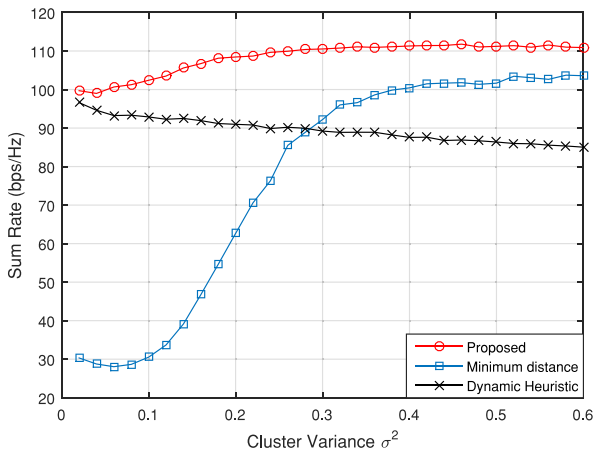
Fig. 1. Sum rate of three user association schemes - minimum distance, dynamic heuristic and proposed algorithm with varying Thomas cluster variance.
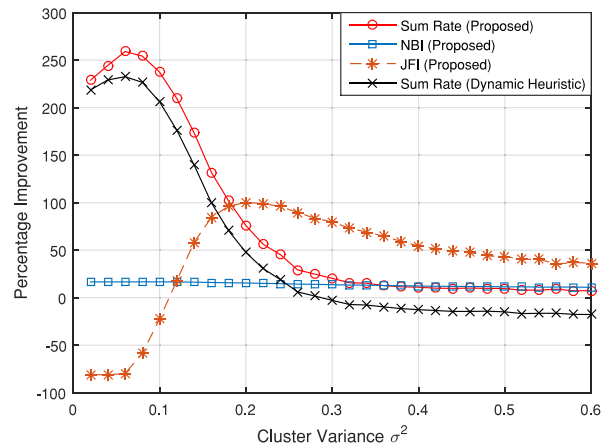

Fig. 2. Percentage improvement in sum rate using proposed algorithm compared to conventional minimum distance association and dynamic heuristic with increasing Thomas cluster variance. Percentage improvements in NBI and JFI with proposed algorithm are also shown.

each user compares the received powers from its nearest pico and the macro, and at most $\lfloor \theta N \rfloor$ users can be reassociated.

From Fig. 1 we observe that our proposed algorithm drastically improves the sum rate compared to conventional minimum distance association, and slightly outperforms the dynamic range heuristic for all cluster variances. Interestingly, while [1] shows that its dynamic range heuristic is very close to optimal association, this claim is only valid for their system model where a user as the option of connecting to one macro or one pico. Since our system model contains multiple small cells for the user to choose from, our proposed algorithm outperforms the dynamic range heuristic as it takes load balancing into account by deciding which small cells should receive which reassociated user.

Average improvement: Fig. 2 shows the average improvement of our proposed algorithm over conventional minimum distance association and dynamic range heuristic after 100 realizations for varying σ^2. User locations, channel fading and small cell types are all varied for each realization. Our algorithm always improves NBI, but when users are heavily clustered (small σ^2), sum rate is drastically improved despite decreasing JFI. As users become more uniform, sum rate improvement decreases, at which point both NBI and JFI will increase, and continue to do so even as users become more uniformly located. This is due to the fact that larger initial clustering leads to offloaded users gaining more from associating with a less loaded base station. For more uniform user distribution (larger σ^2), offloaded users will benefit much less, if at all, hence maintaining sum rate. The observations from our analysis are also verified by these results, as the largest increases in sum rate occur when clustering is high (small σ^2) as suggested by (8), which also approximately coincides with the largest decreases in fairness. *This behaviour suggests that using the NBI as an indicator is most beneficial when users are heavily clustered.* As users become more uniformly located, fairness begins to increase from our algorithm, as higher rates experienced by users associated with the underloaded cell are brought down closer to those of clustered users. This behaviour is consistent with that described by (12), as a reduced sum rate improvement (i.e., smaller ϵ) leads to increasing fairness.

VI. CONCLUSION

In this letter we have proposed a metric which quantifies the amount of balance in cellular HetNets. We have described how it is conceptually different to user fairness, and have shown how a user association algorithm can use this as a means to improve sum rate and fairness. Our NBI can be exploited to increase sum rate despite decreasing fairness when users are heavily clustered, and maintain sum rate while increasing fairness when users are more uniform. Future work can explore directly incorporating NBI in optimizing user association.

REFERENCES

[1] S. Corroy, L. Falconetti, and R. Mathar, "Dynamic cell association for downlink sum rate maximization in multi-cell heterogeneous networks," in *Proc. IEEE ICC*, Ottawa, ON, Canada, Jun. 2012, pp. 2457–2461.

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

[3] D. Liu *et al.*, "User association in 5G networks: A survey and an outlook," *IEEE Commun. Surveys Tuts.*, vol. 18, no. 2, pp. 1018–1044, 2nd Quart., 2016.

[4] T. Zhou *et al.*, "QoS-aware user association for load balancing in heterogeneous cellular networks," in *Proc. VTC-Fall*, Vancouver, BC, Canada, Sep. 2014, pp. 1–5.

[5] M. Mirahsan, R. Schoenen, and H. Yanikomeroglu, "HetHetNets: Heterogeneous traffic distribution in heterogeneous wireless cellular networks," *IEEE J. Sel. Areas Commun.*, vol. 33, no. 10, pp. 2252–2265, Oct. 2015.

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

[7] H. Kim, G. de Veciana, X. Yang, and M. Venkatachalam, "Distributed α-optimal user association and cell load balancing in wireless networks," *IEEE/ACM Trans. Netw.*, vol. 20, no. 1, pp. 177–190, Feb. 2012.

[8] T. Lan, D. Kao, M. Chiang, and A. Sabharwal, "An axiomatic theory of fairness in network resource allocation," in *Proc. IEEE INFOCOM*, San Diego, CA, USA, Mar. 2010, pp. 1–9.

[9] S. Singh, H. S. Dhillon, and J. G. Andrews, "Offloading in heterogeneous networks: Modeling, analysis, and design insights," *IEEE Trans. Wireless Commun.*, vol. 12, no. 5, pp. 2484–2497, May 2013.

[10] A. Okabe, B. N. Boots, and K. Sugihara, *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*. New York, NY, USA: Wiley, 1992.

[11] R. K. Ganti and M. Haenggi, "Regularity, interference, and capacity of large ad hoc networks," in *Proc. Asilomar Conf. Signals Syst. Comput.*, Pacific Grove, CA, USA, Oct./Nov. 2006, pp. 3–7.