A Novel Load-Aware Cell Association for Simultaneous Network Capacity and User QoS Optimization in Emerging HetNets

Ahmad Asghar, Hasan Farooq and Ali Imran
Department of Electrical and Computer Engineering
University of Oklahoma
Tulsa, Oklahoma 74135, USA
Email: {ahmad.asghar, hasan.farooq, ali.imran}@ou.edu

Abstract—Ultra-dense Heterogeneous networks (UDHN) are emerging as the inevitable approach to cope with the imminent cellular network capacity crunch. However, load imbalance and widely disproportionate SINR distribution between macro and small cells, remains the key hurdle in harnessing the full potential of UDHN. In this paper we address this problem by proposing and analysing a novel load-aware user association methodology that offers a mechanism to simultaneously optimize network capacity, load distribution and coverage. The solution concurrently leverages the three key optimization parameters for Coverage and Capacity Optimization (CCO) and Load Balancing (LB) SON functions i.e. antenna tilts, transmit powers and cell individual offsets (CIOs). The method incorporates exponential-weighting based prioritization of CCO and LB SON functions within the user association process. The results suggest that the proposed approach offers a distribution of load between macro and small cells that yields more gain in terms of both network capacity and user quality of service than conventional max signal strength or max SINR based association methods.

Index Terms—HetNet; Self-organizing networks; CCO; LB; Joint Optimization; User Association; QoS; 5th Generation Cellular Networks.

I. INTRODUCTION

Efficient utilization of resources in emerging cellular networks, vis-a-vis 5G, is the most rapidly growing concern among the telecommunication community today. Despite recent advancements in many physical layer techniques and possible exploitation of new spectrum at higher frequencies, network densification remains the most yielding means to meet the capacity demands of future 5G cellular networks. However, network densification comes with its own set of challenges [1], prime among which is the heavily skewed distribution of load between macro and small cells [2].

Theoretical studies [3] as well as field trials have shown that in a UDHN, macro cells tend to be more heavily loaded owing to the transmit power disparity between macro and small cells. To rectify this imbalance and introduce flexibility into the standard Reference Signal Received Power (RSRP) based user association, 3GPP introduced the cell individual offset (CIO) parameter [4]. CIO introduces a virtual boost to the RSRP of a cell which can be used to forcefully associate users to small cells with lower power and greater available radio resources. This method, however, is far from ideal with several consequences, prime among which is the Signal to Interference and Noise Ratio (SINR) disparity created by such an association. Such CIO induced cell association also fails to accommodate for multiple factors that affect user quality of experience. These factors include SINR from the candidate cell, the effective load generated by the user to be associated, available free resources in the candidate cell, as well as the impact of new user association on interference and hence overall system capacity.

Some aftermaths of CIO induced cell association are illustrated in figures 1a and 1b. Figure 1a plots the user associations with no CIOs, whereas in figure 1b, small cells are assigned 10 dB CIO each. As intended, introduction of CIO does force some UEs to switch from macro cell to small cell thereby, achieving some apparent LB; however, this change in association comes with a caveat, i.e., reduced SINR for small cell users acquired via CIOs, sometimes significantly, due to the lower transmit powers of small cells. This effect can be seen by observing an example UE encircled in red in figures 1a and 1b, whose SINR goes from 17 dB to -20 dB after switching cells. It is worth noting that such arbitrarily set values of CIOs, as is the current practice, do not and cannot ensure that the host small cell has enough surplus Physical Resource Block (PRBs) to offset the 37 dB drop in SINR compared to macro cell. This demonstrates that use of empirically determined CIO values can affect overall resource efficiency in the system negatively, thereby causing the problem that CIOs were introduced to solve in first place. Instead, CIOs need to be determined through a method that considers user traffic demands and current cell loads. Most importantly CIO values should be determined in
conjunction with other two key hard parameters that affect SINR as well as cell association i.e., Tx power and antenna tilts.

A. Related Work

The survey by Liu et. al. [5] provides a comprehensive summary of the work done on user association methodologies in cellular networks especially in conjunction with proposed 5G-enabling technologies including massive MIMO systems, dense HetNets, energy conservation, mmWave, Cloud-RAN and internet of things (IoT). A key observation that can be drawn from this recent survey is the use of maximum RSRP based cell association with CIO by the overwhelming majority of research studies on cell association. This is in part due to the existence of mature standards for parameters and measurements shared with the user equipment (UE) while making the association decision. Later we explain that our proposed resource conscious cell association scheme can also be implemented using existing standard measurements in cellular systems.

Conversely, load and offered cell capacity often provide the underlying motivation in SINR, instead of RSRP, based cell association approaches proposed in literature, such as the one presented in [6]. A similar approach is presented in [7] and [8] which attempt to optimize user association based on offered capacity. Authors in [9] utilize spectral efficiency, a derivative of the SINR, to optimize user association. However SINR based user association is not always the optimal methodology as it can contribute to overloading of macro cells in a HetNet. In this work, we propose a user association methodology that not only takes into account the received power, but also considers cell load modeled as a function of SINR, while performing cell association in order to avoid overloading, and consequently, user quality of experience degradation.

While optimizing user association, the parameters of choice also play a vital role. Many studies employ either antenna tilts [10], Tx power [11], or CIO [12] as the optimization parameters. However, as demonstrated in [13], the impact of cell Tx powers, antenna tilts and CIOs on cell coverage, capacity and load is deeply intertwined. This makes user association optimization in one or two parameters a futile exercise since any change in the third parameter drastically alters the capacity and load scenario of the network.

B. Proposed Approach and Contributions

Optimization of user association has the capability to provide an optimal tradeoff between otherwise conflicting objectives of CCO and LB SON functions. However, along with its dependence on multiple optimization parameters, user association is also affected by the interdependence of load and SINR as highlighted in [2]. The load in a cell for given traffic demand depends on SINR perceived by the users associated with that cell. With poor SINR same traffic demand generates more load on the cell because of low spectral efficiency and hence more PRB consumption. On the other hand, SINR of users associated with a cell also depends on the resource utilization in neighboring cells, thus creating an intertwined chain effect. The contributions and findings of this paper can be summarized as follows:

1. We propose a novel cell association approach that tackles the aforementioned challenges by embedding the goals of CCO and LB SON functions into a single load aware cell association method.
2. We formulate and solve an optimization problem that jointly optimizes soft parameter CIO and hard parameters antenna tilt and Tx power to actuate the proposed cell association method.
3. We also introduce a parameter to set the priority of CCO and LB SON functions for user association without necessitating changes to existing LTE network standards. We also empirically determine the optimal value of this parameter.
4. The analysis and results in this paper call for a shift from signal strength or SINR focused optimization
of cellular networks to a hybrid load-aware optimization that can solve the aforementioned problems in emerging UDHN.

The rest of the paper is organized as follows: a description of system model employed in derivation of the proposed user association is provided in Section II. A description of the proposed approach along with key influencing factors is presented in Section III, while Section IV contains simulation results demonstrating the efficacy of the proposed approach in comparison with coverage and SINR based user association methodologies.

II. SYSTEM MODEL

A. Network and User Specifications

Network topology considers at least one randomly deployed small cell in the coverage area of a macro cell. Frequency reuse of 1 is considered and same frequency spectrum is utilized by macro and small cells. Macro cells use directional antennas with three sectors per site while small cells employ omni-directional antennas. An LTE like OFDMA based system with resources divided into physical resource blocks (PRBs) of fixed bandwidth, is assumed. For conciseness, the downlink direction is chosen for the analysis as this is where most imbalance in coverage of macro cells and small cells occurs. User association is calculated for a snapshot of network user distribution. We also assume that requested user data rate is known which gives a lower-bound on the desired instantaneous user throughput. Desired user throughput can be modelled as a spatio-temporal function of subscriber behavior, subscription level, service request patterns, as well as the applications being used with the help of big data analytics as recently proposed in [14].

B. Parameters and Measurements

We consider the following network information to be known to both UEs and eNodeBs (eNBs).

1) Cell Loads: For an OFDMA based network, we can define instantaneous cell load as the ratio of PRBs occupied in a cell during a Transmission Time Interval (TTI) to total PRBs available in the cell. This information is available as a standard measurement in LTE as "UL/DL total PRB usage" [15] and can be broadcast to the UEs. To define cell load $\eta_c$, for our system model, we first calculate minimum number of PRBs $\eta^c_{ru}$ to be allocated to a user:

$$\eta^c_{ru} = \frac{1}{\omega_B} \left( \frac{\hat{r}_u}{f(\gamma^c_u)} \right)$$

(1)

where $\hat{r}_u$ represents the (desired) throughput of user $u \in U_c$, where $U_c$ is the set of all active users associated with cell $c$ which have requested resources from the cell, $\gamma^c_u$ represents the SINR of user $u$ when associated with cell $c$ and $\omega_B$ is the bandwidth per PRB. $f(\gamma_u)$ denotes the spectral efficiency of the user link for given SINR. If we consider features such as MIMO or coding scheme gains and scheduling gains, $f(\gamma_u)$ can be defined as $f(\gamma_u) := A \log_2 (1 + B \gamma_u)$, where $A$ and $B$ are variables that can be used to model throughput gains (per PRB) achievable from various types of diversity schemes, or losses incurred by signaling overheads, or hardware inefficiencies. For sake of simplicity, without loss in generality, we assume $A = B = 1$. Ratio of the sum of requested PRBs in a cell to the total cell bandwidth $N^c_b$ gives the cell load:

$$\text{Cell Load} = \eta_c = \frac{1}{N^c_b} \left( \frac{1}{\omega_B} \sum_{u \in U_c} \frac{\hat{r}_u}{\log_2 (1 + \gamma^c_u)} \right)$$

(2)

Since there is no limit to the requested PRBs by the users associated with a cell, the range of cell load is $\eta_c \in [0, \infty)$. If cell load exceeds 1, the cell in reality will be fully loaded and oncoming users, for whom there are no more resources left, will face blocking. The value of load $\eta_c$ is therefore referred to as virtual load and $\eta_c > 1$ reflects congestion in cell $c$.

2) Received Power: In LTE networks, downlink Reference Signal Received Power (RSRP) from nearby base stations is continuously monitored by the UEs and reported to the serving eNB for a number of purposes. In the proposed user association method, we use the user received power (RSRP) to calculate coverage probability in the network.

3) Cell Individual Offset: CIOs can be defined as a combination of multiple cell association parameters introduced by the 3GPP in Release 8 E-UTRAN Specifications [4]. More specifically, CIO includes cell hysteresis, cell offsets and event related offsets which are used by the UE to decide association with a cell. CIO information can easily be broadcast by each cell and decoded by the UEs as part of standard operation. For the purpose of this paper we treat CIO as a simple virtual boost in RSRP.

4) Antenna Tilt: As macro cells in the system under consideration use directional antennas, the gain from base station to user $G^u_{t,t}$ is dependent on the 3D antenna gain model. For development purposes, 3GPP has provided a theoretical 3D antenna gain model which is given in [16] as:

$$G^u_{t,t} = 10 \log \left( 1.2 \left[ \frac{\psi^c_u \phi^e_u}{\theta_{tilt}} \right]^2 + \lambda_b \left[ \frac{\phi^c_{azi} \phi^e_{azi}}{\theta_{az}} \right]^2 \right)$$

(3)

where $\psi^c_u$ is the vertical angle between user $c$ and the transmit antenna of cell $c$, $\psi^e_{tilt}$ is the tilt angle of serving cell antenna, $\lambda_b$ and $\lambda_t$ are the weighting factors for horizontal and vertical beam pattern respectively, $\phi^c_{azi}$ is the horizontal angle of user $u$ from cell $c$, $\phi^e_{azi}$ is the azimuth of antenna of cell $c$, and $B_h$ and $B_v$ are horizontal and vertical beamwidths of the transmitter.
antenna of cell $c$. As our variable of interest in (3) is tilt angle and rest of the antenna parameters can be treated as constants, for the sake of conciseness we can simplify (3) using the following substitution:

$$x_u^c = \left(\frac{B_r \lambda_h}{\lambda_v}\right)^2 \left(\frac{\phi_u^c - \phi_{a\pi}^c}{B_h}\right)^2 \tag{4}$$

5) Signal-to-Interference and Noise Ratio: Equation (2) demonstrates the relationship between cell load and user SINR. To develop a load-aware user association methodology, we need to have SINR estimation in order to estimate cell load. Based on the system model, we propose to estimate SINR for user $u$ as a function of antenna tilts and transmit power along with interfering cell loads using the standard exponential pathloss model as:

$$\gamma_u^c = \frac{P_u^c G_u G_u^\prime \delta a (d_u^c)^{-\beta}}{\kappa + \sum_{i \in C_c} \hat{\eta}_i P_i^c G_i G_i^\prime \delta a (d_i^c)^{-\beta}} \tag{5}$$

where $P_i^c$ and $P_u^c$ are the transmit powers of serving cell $c$ and interfering cell $i$, $G_u$ is UE gain, $\delta$ is signal shadowing, $a$ is the pathloss constant, $d_u^c$ and $d_i^c$ represent distance of user $u$ from cell $c$ and $i$, respectively, $\beta$ is the pathloss exponent, and $\kappa$ is the thermal noise power. Here, $\hat{\eta}_i$ denotes actual cell load in a cell such that $\hat{\eta}_i \in [0,1]$ and is used exclusively for SINR calculation. Substituting $x_u^c$ in (3) and replacing $G_u^\prime$ and $G_u$ in (5) with the resulting expression gives:

$$\gamma_u^c = \frac{P_u^c G_u 10^{\alpha_i (\phi_u^c - \phi_{ii}^c) + x_u^c \delta a (d_u^c)^{-\beta}}}{\kappa + \sum_{i \in C_C} \hat{\eta}_i P_i^c G_i 10^{\alpha_i (\phi_u^c - \phi_{ii}^c) + x_i^c \delta a (d_i^c)^{-\beta}}} \tag{6}$$

where $\mu$ is consolidated constant based on fixed antenna characteristics.

III. A New User Association Methodology

The state-of-the-art method of determining cell associations is to use the RSRP measurements along with CIO values as given below:

$$P_{r,u|dBM}^c = P_{r,u|dBM}^c + P_{CIO|dBM}^c \tag{7}$$

where $P_{r,u|dBM}^c$ is the true signal power in dBM received by user $u$ from cell $c$ and $P_{r,u|dBM}^c$ is the received power reported back by user $u$ to cell $c$ in dBM. This value includes the $P_{CIO|dBM}^c$ (the CIO value) of cell $c$ in dBM.

However, as explained earlier, this method overlooks the key impact of user association on cell loads and consequently SINR, and thus results in negative impact on overall capacity and QoS through imbalanced loads and poorer SINR distribution among users (See introduction section and Fig. 1 and 2). To overcome this challenge, we propose to establish user association with cell $j$ not only based on received power but also load in that cell. More specifically, in this load aware cell association method user to be associated with cell $j$ can be determined as:

$$\max_{i \in \mathbb{U}_c} \frac{r_i^c \#_{inter}^c \omega_c \log_2 \left(1 + \gamma_u^c\right)^{\frac{1}{\kappa_{c|1}}} \frac{1}{\rho_{c|1}}} \tag{9}$$

where $\mathbb{U}_{j,t}$ is a set of all active and idle users for whom a scaled product of the RSRP (+CIO) in Watts $P_{r,u|dBM}^c$, and the residual cell capacity for cell $j$ is the highest. $\alpha \in [0,1]$ is a weighting factor introduced to allow trading between the impact of RSRP and cell load measurements in the cell association. Note that in (8), to make new user association decision with a cell we use the virtual cell load which provides a truer picture of effective potential load in the candidate cell by taking into account the users that are already associated with that cell but were not served due to lack of free resources. The time subscripts for $\mathbb{U}_{j,t}$ and $\eta_{c,t-1}$ signify the fact that the user association at time $t$ is calculated using the last known cell loads i.e. before the new set of users are associated.

The set $\mathbb{U}_c$ used in the expression for SINR in (6) represents the set of only active users associated with the cell $c$.
\[
\forall j, t := \left\{ \forall u \in \mathbb{U} \mid j = \arg \max_{c \in C} \left( \frac{1}{\eta_{c, t-1}} \right)^{\alpha} \left( \hat{p}_{r, u, dBm}^{c} \right)^{(1-\alpha)} \right\}
\] (8)

TABLE I: Parameter Settings for Simulation

<table>
<thead>
<tr>
<th>System Parameters</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Base Stations</td>
<td>7</td>
</tr>
<tr>
<td>Sectors per Base Station</td>
<td>3</td>
</tr>
<tr>
<td>Small Cells per Sector</td>
<td>1</td>
</tr>
<tr>
<td>Number of UE per Sector</td>
<td>25</td>
</tr>
<tr>
<td>Transmission frequency</td>
<td>2 GHz</td>
</tr>
<tr>
<td>Transmission Bandwidth</td>
<td>10 MHz</td>
</tr>
<tr>
<td>Network Topology</td>
<td>Hexagonal</td>
</tr>
<tr>
<td>Macro Cell Tx Power</td>
<td>Max: 46 dBm, Min: 40 dBm</td>
</tr>
<tr>
<td>Macro Cell Antenna Tilt</td>
<td>Max: 15°, Min: 0°</td>
</tr>
<tr>
<td>Small Cell Tx Power</td>
<td>Max: 30 dBm, Min: 27 dBm</td>
</tr>
<tr>
<td>Small Cell CIO</td>
<td>Max: 10 dB, Min: 0 dB</td>
</tr>
<tr>
<td>Cellular System Standard</td>
<td>LTE</td>
</tr>
<tr>
<td>Macro Cell Height</td>
<td>25 m</td>
</tr>
<tr>
<td>Small Cell Height</td>
<td>10 m</td>
</tr>
<tr>
<td>Inter-site Distance (Macro)</td>
<td>500 m</td>
</tr>
</tbody>
</table>

We use wrap around model to simulate interference in an infinitely large network thus avoiding boundary effect. To model realistic networks, UEs are distributed non-uniformly in all the sectors such that a fraction of UEs are clustered around randomly located hotspots in each sector. Due to the non-convexity of SINR expression in (6), sequential quadratic programming (SQP) is used to maximize the geometric mean of user throughput since it allows an approximation of a non-convex function as a convex function, while Monte Carlo simulations are used to estimate average performance.

B. Results

In the formulation presented in (8), user association is dependent on 3 features: cell loads at the time of association, RSRP with CIO as reported by the UE, and the user association exponent \( \alpha \). The impact of cell loads and RSRP on user association are obvious from (8); however, the impact of exponent value on user association requires quantitative evaluations of system KPIs. A very relevant KPI is the cell load and its distribution among cells for given total traffic in the network. A lower average cell load and more normal load distribution among cells for given traffic reflects a better performing CCO-LB solution. A comparison for \( \alpha \in [0, 1] \) was done with results showing that for \( \alpha = 0.4375 \), cell load distribution is the closest to a normal distribution. Due to space limitations, figure 2 only presents cell load kernel density distribution for values of \( \alpha \in [0.25, 0.5] \).

Fig. 2: Comparison of Offered Cell Load Distribution for \( \alpha \in [0.25, 0.5] \)

Figure 3 presents a comparison of cell load kernel density distribution of the proposed load-aware user association with Max RSRP and Max SINR based user association methodologies. The results indicate that the
proposed load-aware user association achieves far more balanced load distribution compared to either of the two competing methodologies, despite the fact that Max SINR association scheme provides the best overall SINR distribution, as shown in figure 6. Conversely, the impact of RSRP or SINR centric approaches is clearly evident in terms of uneven load distribution among cells. This is demonstrated in terms of macro and small cell load distributions in figures 4 and 5, which serve to highlight the true capacity gains of the proposed load-aware user association over existing user association approaches by avoiding underloading small cells and overloading macro cells.

![Fig. 3: Comparison of Offered Cell Load Distribution for load-aware vs. Max RSRP and Max SINR user association](image)

Fig. 3: Comparison of Offered Cell Load Distribution for load-aware vs. Max RSRP and Max SINR user association

![Fig. 4: Comparison of Offered Macro Cell Load Distribution for load-aware vs. Max RSRP and Max SINR user association](image)

Fig. 4: Comparison of Offered Macro Cell Load Distribution for load-aware vs. Max RSRP and Max SINR user association

![Fig. 5: Comparison of Offered Small Cell Load Distribution for load-aware vs. Max RSRP and Max SINR user association](image)

Fig. 5: Comparison of Offered Small Cell Load Distribution for load-aware vs. Max RSRP and Max SINR user association

![Fig. 6: Comparison of average SINR CDF for load-aware vs. Max RSRP and Max SINR user association](image)

Fig. 6: Comparison of average SINR CDF for load-aware vs. Max RSRP and Max SINR user association

The even load distribution offered by the proposed load-aware user association methodology also results in gains in terms of user QoS by minimizing the number of users who are unable to achieve their desired throughput due to a lack of physical resources at the serving cell. This is evidenced by the ratio of unsatisfied users in the network and the utilization of physical resources in the network given in figure 7. The results in figure 7 highlight two important points. Firstly, the resource utilization in load-aware user association is higher than the other two schemes due to the lower user SINR which results in more PRBs being utilized to achieve desired user throughput. However, the high user SINR achieved by Max RSRP and Max SINR associations in comparison is inconsequential if there are no physical resources to serve the users at serving cell, as is demonstrated by the ratio of unsatisfied users in figure 7. The ratio of unsatisfied users due to load-aware user association in 7 is significantly lower compared to either Max RSRP and Max SINR association methodologies. This is because the max RSRP and max SINR association methodologies are blind towards network loading while associating users with cells, thus exposing new users to resource unavailability. The load-aware user association avoids this issue by balancing user associations based on RSRP and cell loads, thus forcing users to switch to cells with lower SINR even when a cell with higher RSRP is available.
V. CONCLUSION

In this paper, we presented a novel load-aware user association methodology for capacity optimization with the challenge of load distribution in emerging HetNets for 5G networks taken as the key constraint. The proposed user association methodology exploits joint optimization of CIO, Tx power and antenna tilt to balance loads between all cells, including macro and small cells, in a way that increases overall physical system capacity. Results suggest that for the optimal user association exponent $\alpha$, SINR does not suffer considerably as compared to the RSRP(+CIO) only approach. The proposed association methodology also provides a balanced network loading without any significant depreciation in downlink SINR compared to existing methods. As part of a future study, use of game theoretic techniques to determine optimal $\alpha$ value will be pursued. Moreover, considering the assumption of apriori knowledge of required user throughput, further research can be directed towards incorporation of big data aided knowledge like user mobility prediction to optimally set $\alpha$ values for each cell and UE.

ACKNOWLEDGMENT

This work was supported by NSF Grant No. 1619346. For details about this project see : http://bsonlab.com/ASSH/.

REFERENCES