Wireless Pers Commun (2018) 101:305–323
https://doi.org/10.1007/s11277-018-5689-2

Cell Selection for Load Balancing in Heterogeneous Networks

Yasin Aghazadeh[1] · Hashem Kalbkhani[1] · Mahrokh G. Shayesteh[1,2] · Vahid Solouk[3]

Published online: 13 April 2018

Abstract The vision of advanced long-term evolution (LTE-A) project is set to ultimate increase of network capacity in heterogeneous networks (HetNets). In HetNets with small cell configuration, a considerable majority of user devices is eventually connected to the macrocell base station (MBS), while small base stations (BSs), such as femtocell access points (FAPs), are still without any user. This results in unbalanced load and reduces the data rate of macrocell user equipment (MUE). In this paper, a method is proposed for load balancing among FAPs, while desired throughput is achieved. The proposed method uses the estimated received signal strength from different BSs and adjusted pilot signals. Under the critical signal to interference plus noise ratio (SINR) condition, a list of candidate FAPs is prepared. The updated candidate list henceforth does not include the least visited FAPs, which in turn leads to lower unnecessary handoffs. Once the BS with the highest number of free RBs and the highest pilot signal power is selected, FAP allocates the RBs with higher SINRs (qualified RBs) to user. In the case of FAP unavailability, the algorithm compels users to connect to the MBS with adequate qualified RBs. The performance of the proposed method was evaluated under a variety of FAPs density, and the number and velocity of users in terms of throughput and Jain's fairness index. The results evidence affordable improvements in the throughput and Jain's index in comparison with other methods.

✉ Hashem Kalbkhani
h.kalbkhani@urmia.ac.ir

Yasin Aghazadeh
st_y.aghazadeh@urmia.ac.ir

Mahrokh G. Shayesteh
m.shayesteh@urmia.ac.ir

Vahid Solouk
v.solouk@it.uut.ac.ir

1 Department of Electrical Engineering, Urmia University, Urmia, Iran

2 Wireless Research Lab, ACRI, Electrical Engineering Department, Sharif University of Technology, Tehran, Iran

3 Department of IT and Computer Engineering, Urmia University of Technology, Urmia, Iran

Keywords Cell selection · Jain's index · Load balancing · Femtocell · Pilot signal

1 Introduction

1.1 Motivations

The exponential growth in wireless technology and smartphones nowadays has led both the users and providers toward deploying extensive wireless services. According to the reports, more than 60% of voice and 90% of data traffic have targets in indoor environments, where macrocells usually lag on efficiency due to spectrum and resource limitation. Hence, providing such coverage and spectrum demand has become a challenge for the operators [1]. Using small cells is a win–win strategy that provides higher data rate, and reliability for users and reduces the amount of traffic on expensive macrocell base station (MBS) for the operator. Using different kinds of base stations (BSs) together in a network makes a multi-tier network known as heterogeneous network (HetNet). Femtocell access point (FAP), a device with low power and small coverage, becomes an inevitable technology due to the increasing demand for high data rates in wireless networks [2]. FAPs connect mobile devices to operator network using residential digital subscriber line or cable broadband connections and can be installed plug and play by users to enhance the cellular network performance at indoor.

In conventional networks, each user is served by the BS which provides the strongest received signal to interference plus noise ratio (SINR), termed as Max-SINR [3]. This is optimal in term of coverage probability for a given SINR, but may results in unequal loads and creating network bottlenecks. MBSs transmit at much higher power than FAPs, thus have larger coverage area. This will diminish the usefulness of FAPs, resulting in very unbalanced load and uneven user experiences [3]. By connecting to the FAP that provides lower SINR but a larger share of resource blocks (RBs), users may get an overall better data rate [3]. Without load balancing, MBSs remain the bottleneck and FAPs are extremely underutilized [4–6]. If load can be distributed more evenly among FAPs, the call success probability could be increased [7].

1.2 Related Works

Some methods proposed for load balancing are based on the channel borrowing from the lightly loaded cells such as hybrid channel assignment [8], channel borrowing without locking [9], and load balancing with selective borrowing [10, 11]. These methods assign most part of the spectrum to the BS with more number of users, i.e., load balancing in spectrum which is different from that of in cells. In our study, each FAP has some spectrum resources, i.e., resource blocks (RBs), and instead of spectrum sharing, users connect to the FAPs with lower number of occupied RBs.

Several methods such as directed retry [12], mobile-assisted call admission algorithms [13], hierarchical macrocell overlay systems [14, 15], cell breathing techniques [16, 17], and biasing methods in HetNets [6, 18] transfer traffic to the lightly-loaded cells. In these methods user association process to the selected cell continues on until the cell is fully loaded. These methods transmit pilot signal with constant power. In this paper, we propose to adjust the pilot signal in proportion to the number of free RBs.

Moreover, there are various studies on investigating different kinds of utility functions to achieve load balancing, such as network-wide proportional fairness [19], network-wide max–min fairness [20], and α-optimal user association [21]. Cost function optimization needs more hardware in FAPs and more complexity too. This paper proposes to push some of measurements and computations to the user equipment unit, which is suitable from the financial aspects and decreases signaling in the network.

The authors in [22, 23] have used user and BS association by gradient projection method in which each user measures the SINR based on pilot signals broadcasted by each BS. However, in practical systems, it is much more difficult to implement multi BS association than single BS association. The idea of leaving blank certain resources of MBS was employed in [4]. Cell selection scheme with a given almost blank sub-frame ratio was proposed in [24] which is based on the SINR without consideration of the loads of BSs.

1.3 Contributions

In this paper, we propose an efficient and practical mobile-assisted cell selection for load balancing in two-tier macrocell-femtocell HetNets. We present a cell selection algorithm to make a decision without radio resource controller and self-organization network to prevent complex computing in FAPs, cost–benefit for operator, and less delay. In this study, two SINR thresholds are used to guarantee the throughput requirement of users. User starts for making a first list of candidate FAPs according to received signal strength (RSS), when received SINR is below the first SINR threshold. Then, user prunes it according to the predicted RSS to make the second list and chooses the target FAP based on the number of free RBs and their SINRs. The proposed pilot signal helps user to provide the third list and finally user selects the RBs with higher SINRs. Simulation results show that users (loads) are distributed fairly among the FAPs, load of MBS is pushed to FAPs, and users have higher throughput compared to the conventional Max-SINR method.

The rest of the paper is organized as follows. Section 2 describes system model used in this study. Section 3 presents the proposed cell selection scheme to achieve load balancing in HetNets. Simulation results are provided in Sect. 4 and finally, Sect. 5 gives the concluding remarks.

2 System Model

One of the important issues in two-tier femtocell–macrocell HetNet is to have trade-off between the spatial reuse of bandwidth and interference in a way that the capacity of network is maximized and load balancing is taken into account. This work considers downlink of the orthogonal frequency-division multiple access (OFDMA)-based two-tier cellular network with fractional frequency reuse (FFR) macrocells.

2.1 Macrocell Tier

FFR is an interference coordination technique used in OFDMA-based wireless networks wherein cells are partitioned into two spatial regions, i.e., inner and outer, with different frequency reuse factors [25]. One cluster of macrocell network and spectrum partitioning into four sub-bands are shown in Fig. 1. It is seen from Fig. 1a that frequency reuse factors for inner and outer regions are equal to one and three, respectively. An omnidirectional

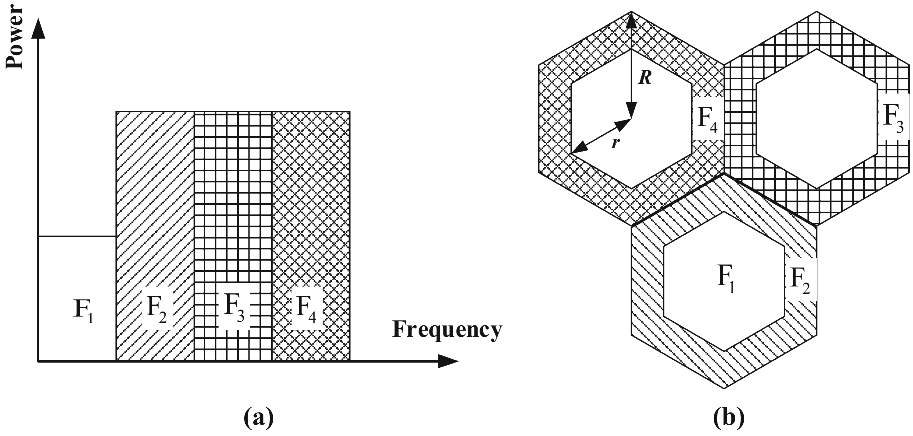

Fig. 1 The layout of cellular network, **a** frequency partitioning into four sub-bands, **b** one cluster of cellular network

antenna is used at the center of the hexagonal macrocell with radius R to serve macrocell user equipments (MUEs) with different transmission powers for the users of inner and outer regions [25, 26]. The central macrocell, that two layers in its coverage is analyzed, is surrounded by six and 12 macrocells in the first layer and the second one of the macrocell network, respectively [27].

It is assumed that the locations of MUEs follow uniform distribution. Therefore, the number of RBs allocated to each region depends on the area of region [28]. As shown in Fig. 1b frequency spectrum is divided into four sub-bands F_1, F_2, F_3, and F_4 each with 25 RBs where the bandwidth of each RB is 180 kHz. In this paper, we analyze the coverage of central macrocell in which the sub-bands F_1 and F_2 are allocated to the MUEs in the inner and outer regions, respectively.

2.2 Femtocell Tier

FAPs are distributed in the coverage area of central macrocell according to the homogeneous spatial Poisson point process (SPPP) with density λ. Each FAP covers a disk area with radius R_f with omnidirectional antenna. Therefore, the average number of FAPs in the central macrocell with area S is $N_{FAP} = \lambda \times S$. To avoid high cross-tier interference from the MBS, it is assumed that FAPs serve users using RBs from the sub-bands F_3 and F_4. In this study, it is assumed that each FAP chooses 12 RBs randomly from 50 available RBs in the sub-bands F_3 and F_4 to serve its users. FAPs operate in hybrid access mode [29] where some of the available RBs are dedicated to the authorized users and the remainings are used to serve the other users.

2.3 Mobility Model

In this paper spatio-temporal parametric stepping (STEPS) model [30] is used to generate the mobility pattern of mobile users. STEPS is a simple parametric mobility model which can cover a large spectrum of human mobility patterns. STEPS makes abstraction of spatio-temporal preferences in human mobility by using a power law to rule the nodes movement.

2.4 Channel Model and SINR

In multi-path wireless channel, transmitted power is affected by three different phenomena; path loss, shadowing, and Rayleigh fading [31]. Path losses of different links are adopted from [32] and presented in Table 1. In this Table, d denotes the distance between the transmitter and receiver, L_w is the wall penetration loss and d_{indoor} is the nearest available distance from FAP to user that is set to 0.5 m. Correlated shadowing has log-normal distribution where its logarithm has normal distribution with mean μ and standard deviation σ [32] and its fluctuations have exponential correlation with correlation distance equal to d_0. Also, flat-fading channel power coefficients have unit mean exponential distribution [1].

Based on the mentioned channel impairments, the received power P_R in the wireless transmission can be calculated as [33]:

$$P_R = P_T H \Psi L^{-1} \tag{1}$$

where P_T, H, Ψ, and L denote the BS transmission power, channel power coefficient, log-normal shadowing, and path loss, respectively. The received SINR of an RB is calculated as

$$\gamma = \frac{P_R}{I_F + I_M + N_0} \tag{2}$$

where I_F and I_M are the received interferences from the interfering co-channel FAPs and MBSs, respectively, and N_0 is the white noise power which is set to -174 dBm/Hz in this study.

Due to mobility of users, by increasing the number of FAPs, the number of handoffs increases as well. Because of the variations of the RSS and small radius of femtocells, several unnecessary handoffs may take place that increase the number of ping-pong handoffs. Ping-pong means that the number of handoffs in a specific time duration exceeds a threshold (e.g., more than one handoff in every 5 s). The ping-pong effect causes some defections including throughput reduction, long handoff delay, and high dropping probability, which all deteriorate the quality of service (QoS).

To overcome the random effects of shadowing and fading, at the receiver, exponential smoothing window is applied to the RSS; then the smoothed RSS ($\bar{P}_R$) is used as the input of recursive least square (RLS) algorithm to predict the next RSS samples ($\hat{\bar{P}}_R$). It is shown that taking advantage of RSS prediction leads to better prevention of frequent handoffs and handoff latency reduction [27].

Table 1 Path losses for different links

Link (Tx/Rx)	Path loss
MBS/outdoor user	$L(\text{dB}) = 15.3 + 37.6 log_{10} d$
MBS/indoor user	$L(\text{dB}) = 15.3 + 37.6 log_{10} d + L_w$
Serving FAP/indoor user	$L(\text{dB}) = 38.46 + 20\log_{10} d + 0.7 d_{indoor}$
Interfering FAP/outdoor user	$L(\text{dB}) = 38.46 + 20\log_{10} d + 0.7 d_{indoor} + L_w$
Interfering FAP/indoor user	$L(\text{dB}) = 38.46 + 20\log_{10} d + 0.7 d_{indoor} + \begin{cases} L_w & \text{if} \quad R_F < d \le 2R_F \\ 2L_w & \text{if} \quad 2R_F \le d \end{cases}$

3 Proposed Cell Selection Algorithm

In this section, we describe the proposed cell selection algorithm to achieve load balancing among FAPs and MBS. In this work, it is assumed that each FAP transmits two types of signal, data signal and pilot signal. Pilot signal should contain information about the status of RBs of FAPs, i.e., number of free RBs. Therefore, we propose that the pilot power of each FAP depends on the number of its free RBs. Hence, the transmitted pilot power of the ith FAP, $P_{T,\,Pilot}^{(i)}$, is calculated as

$$P_{T,Pilot}^{(i)} = P_{T,Pilot}^{\max} \times \frac{N_{RB,\,free}^{(i)}}{N_{RB,\,total}^{(i)}} \tag{3}$$

where $P_{T,Pilot}^{\max}$ is the maximum transmitted pilot power, $N_{RB,\,free}^{(i)}$ denotes the number of free RBs of the ith FAP, and $N_{RB,\,total}^{(i)}$ is the total number of RBs of the ith FAP. In this way, FAP does not transmit pilot signal when all RBs are occupied and pilot signal is transmitted with maximum power when FAP is idle. Therefore, pilot signal is FAPs' tendency to offer service to users. FAP transmits low pilot power when it has low number of free RBs; hence, only users closer to FAP can send connection request to FAP. But, when FAP transmits high pilot power, more users can send connection request to the FAP that results in pushing the load to lower loaded FAPs; which leads to fair user allocation. It should be mentioned that pilot signals suffer from path loss, shadowing, and fading, as data signals.

In the case of dense femtocell deployment, there are large number of FAPs. Users receive high number of pilot signals, which results in considerable increment in the computational complexity and power consumption in user equipment due to great neighbor lists. In the proposed algorithm, user sends connection request when the target BS is chosen, and any signaling does not occur before that. Therefore, the only signaling overhead occurs due to transmission of pilot signal. But, the proposed algorithm prevents unnecessary and redundant handoffs and eliminates the signaling overhead due to those handoffs.

3.1 Cell Selection for Femtocell Users

The block diagram of the proposed cell selection algorithm for femtocell users is shown in Fig. 2. Three lists of candidate FAPs are constructed during the proposed algorithm. It is assumed that users initially connect to the FAP with the highest RSS. RSS is an important parameter to have good QoS; however, the main factor to achieve desired QoS is SINR. Therefore, FUE always check the received SINR from the serving FAP. Since cell selection process takes time, the process of cell selection should be started before decreasing the SINR to the minimum acceptable value. To this end, two different thresholds are considered for SINR. The first SINR threshold γ_{t1} is the minimum acceptable SINR for the service and the second threshold γ_{t2} is higher than γ_{t1}. User stays in the serving FAP until $\gamma_s \geq \gamma_{t2}$ and when γ_s decreases from γ_{t2}, user starts cell selection process.

As mentioned, the main factor in achieving desired QoS is SINR. Therefore, the received SINRs from all FAPs are calculated to construct the first neighbor femtocell list ($NFL^{(1)}$). Since calculation of the received SINR from all FAPs requires huge complexity, we propose to use RSS instead of SINR. Hence, the first step is constructing $NFL^{(1)}$ based

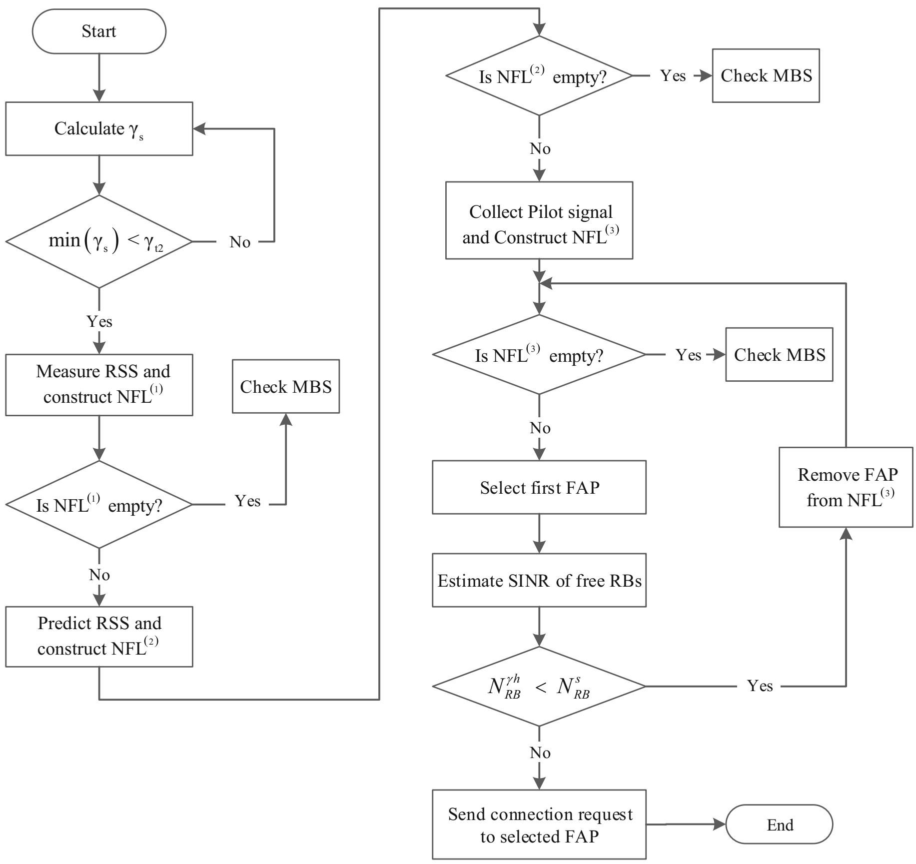

Fig. 2 Proposed cell selection algorithm for femtocell users

on the RSS. The FAPs whose RSSs are greater than P_{TH}, which is the minimum acceptable RSS, are included in $\mathrm{NFL}^{(1)}$, that is,

$$\mathrm{NFL}^{(1)} = \left\{ i : \bar{P}_R^{(i)} > P_{TH} \right\}; \quad i = 1 : N_{FAP} \tag{4}$$

where $\bar{P}_R^{(i)}$ denotes the smoothed RSS from the ith FAP and N_{FAP} is the number of FAPs.

When there is no FAP with RSS higher than P_{TH}, $\mathrm{NFL}^{(1)}$ is empty, which means that user cannot connect to any of neighbor FAPs. In this case, user only can connect to MBS. The process of checking the connection availability to MBS, is indicated with "Check MBS" box in Fig. 2 and described in detail in Fig. 3.

If at least one FAP belongs to $\mathrm{NFL}^{(1)}$, user starts constructing the second neighbor femtocell list ($\mathrm{NFL}^{(2)}$) from the FAPs belonging to the $\mathrm{NFL}^{(1)}$. Due to mobility of users, their residential times in the candidate FAPs of the $\mathrm{NFL}^{(1)}$ are different and vary widely. Therefore, those FAPs in which the user has low residential time, should be removed from the first list. In order to estimate the residential time of user in each FAP, the next RSS samples of FAPs belonging to $\mathrm{NFL}^{(1)}$ are predicted based on some previous RSS samples. $\mathrm{NFL}^{(2)}$ consists of those FAPs from the $\mathrm{NFL}^{(1)}$ whose minimum predicted RSSs are higher than P_{TH}, that is,

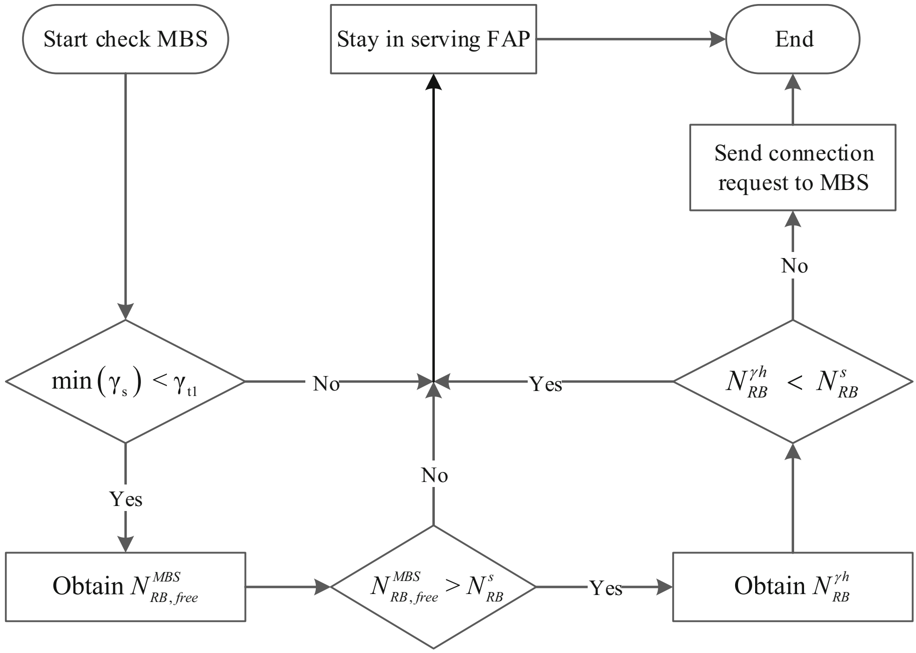

Fig. 3 Proposed connection availability check to MBS

$$\mathrm{NFL}^{(2)} = \left\{ \mathrm{NFL}^{(1)}(j) \ : \min\left(\hat{\bar{P}}_R^{(j)}\right) > P_{TH} \right\}; \quad j = 1 : N_1 \tag{5}$$

where $\hat{\bar{P}}_R^{(j)}$ is the predicted RSS of the jth FAP belonging to $\mathrm{NFL}^{(1)}$ and N_1 denotes the number of FAPs belonging to $\mathrm{NFL}^{(1)}$. Removing the FAPs with low residential time prevents ping-pong effect and decreases signaling cost and complexity.

If $\mathrm{NFL}^{(2)}$ is not empty, user starts to construct the third neighbor femtocell list ($\mathrm{NFL}^{(3)}$); otherwise it checks the connection availability to MBS (Fig. 3). To build $\mathrm{NFL}^{(3)}$, user starts to obtain the information of pilot signal from the FAPs belonging to the $\mathrm{NFL}^{(2)\cdot}$ Since pilot signals also suffer from path loss, shadowing, and fading, the user will grasp the number of free RBs once the received power of the respective pilot signal is above the threshold power γ^p, otherwise user discards the FAP. Those FAPs that their number of free RBs is more than or equal to N_{RB}^s, are included in $\mathrm{NFL}^{(3)}$, i.e.,

$$\mathrm{NFL}^{(3)} = \left\{ \mathrm{NFL}^{(2)}(k) \ : \ P_{R,Pilot}^{(k)} > \gamma^p \ \& \ N_{RB,free}^{(k)} > N_{RB}^s \right\} \quad k = 1 : N_2 \tag{6}$$

where $P_{R,Pilot}^{(k)}$ is the received pilot power from the kth FAP, $N_{RB,free}^{(k)}$ denotes the number of free RBs of the kth FAP belonging to $\mathrm{NFL}^{(2)}$ and N_2 is the number of FAPs in $\mathrm{NFL}^{(2)}$. If $\mathrm{NFL}^{(3)}$ is empty, user checks connection availability to MBS (Fig. 3). Otherwise, the FAPs belonging to $\mathrm{NFL}^{(3)}$ are sorted based on the number of their free RBs in descending order. If some FAPs have the same number of free RBs, FAP with higher pilot power has priority, because it serves lower number of users. It is noted that SINR is the main parameter to achieve the desired QoS. Therefore, we have to make sure about the SINRs of free RBs. To select the target FAP, user chooses the first FAP from $\mathrm{NFL}^{(3)}$ and estimates the SINR in free RBs. If the number of free RBs with SINRs higher than $\gamma_s \left(N_{RB}^{\gamma h}\right)$ is more than N_{RB}^{s}, user sends connection request to the selected FAP; otherwise it removes the FAP from $NFL^{(3)}$ and performs the same steps explained for the first FAP of $NFL^{(3)}$ until it finds the target FAP from the $NFL^{(3)}$. If all FAPs belonging to $NFL^{(3)}$ are checked and none of them is not selected by user, i.e., $NFL^{(3)}$ becomes empty, user checks the connection availability to MBS (Fig. 3).

According to Fig. 3, as long as γ_s is higher than γ_{t1}, user stays in serving FAP and when γ_s becomes lower than γ_{t1}, user obtains the number of free RBs in MBS ($N_{RB,free}^{MBS}$). If $N_{RB,free}^{MBS}$ is less than the demanded RBs by service of user (N_{RB}^{s}), user stays in the serving FAP, otherwise it obtains the number of free RBs that have SINR higher than γ_{t1} ($N_{RB}^{\gamma h}$). If $N_{RB}^{\gamma h}$ is equal to or greater than N_{RB}^{s}, MBS is selected as the target cell and user sends connection request to MBS.

3.2 Cell Selection for Macrocell Users

Due to the resource limitation of MBS, the lowest possible number of users connect to MBS. Therefore, unlike femtocell users, macrocell users can search for available FAP instead of comparing γ_s with γ_{t2}. The process of finding target cell for macrocell users is shown in Fig. 4. Similar to femtocell users, macrocell users construct three lists of candidate neighbor femtocells where the steps are similar to those presented for femtocell users with one exception. When a list is empty, macrocell users start constructing $NFL^{(1)}$, whereas femtocell users check the connection availability to MBS.

As explained in detail, we propose that user equipment performs some operations such as neighbor list creation. When these computations are carried out from BS (FAP or MBS) to the user equipment, the computational cost reduces in the BS, therefore, serving BS only performs signaling operations to connect the user equipment to the target BS. In the case of huge number of users, which is expected to occur in the future, the process of switching connection from one BS to another experiences some delay if the BSs select the target BS for all users.

4 Performance Evaluation

We evaluate the performance of the proposed cell selection approach for different values of user velocity, FAP density, and number of mobile users. Simulation parameters used in this research are shown in Table 2. We performed simulations in MATLAB environment to model physical layer processes. The received power (or interference) is modeled based on (1), and the SINR is determined using (2). We consider frequency-flat fading channel. Based on the bandwidth restriction over the time duration of an RB (around 180 kHz in 1 ms), the variations of channel can be tracked. Hence, we assume that shadowing effect and channel gain in each RB remain constant over an RB duration, but may vary from one RB to another [34]. It is also assumed that all sub-carriers in RB are transmitted with the same power.

Mobility pattern of users were generated using STEPS mobility model. The mobility patterns are generated using one-second sampling time and then the received powers (or interferences) are calculated in the respective locations. We use *Monte Carlo* simulations to consider the randomness of channel power gain, shadowing, and mobility patterns.

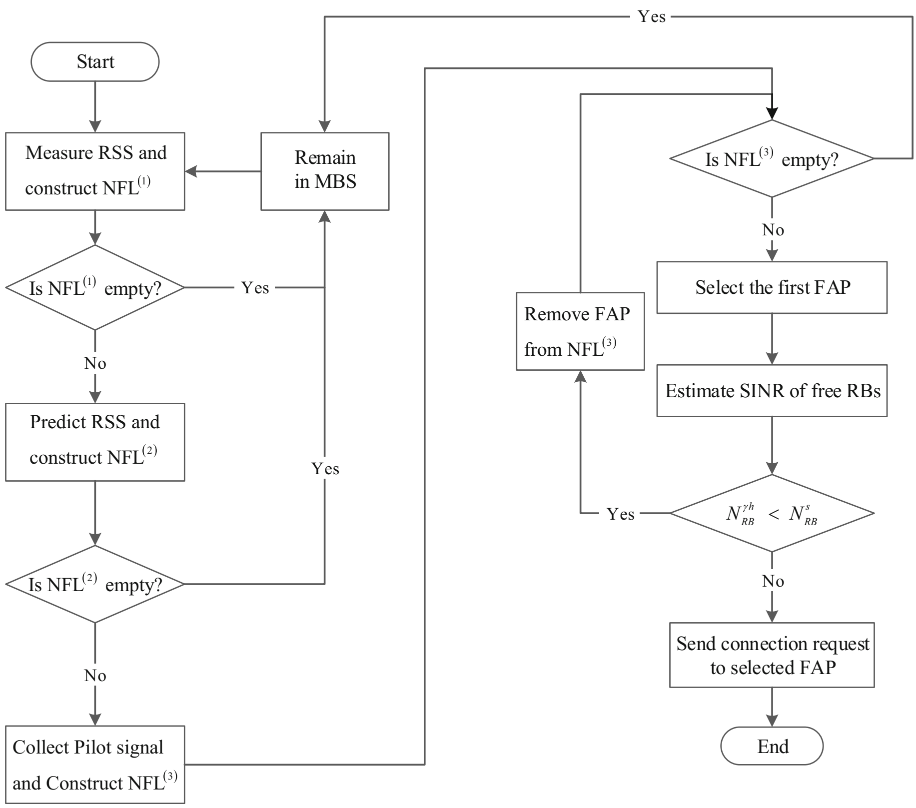

Fig. 4 Proposed cell selection algorithm for macrocell users

Table 2 Simulation parameters used in this paper

Parameter	Value	Parameter	Value
Macrocell radius (R_m)	500 m	Mobile user velocity	2, 5 (km/h)
Inner area radius (r)	350 m	FAP density (λ)	(5, 15) $\times$ 10^{-5}
FAP radius (R_f)	20 m	Number of mobile users	Variable, 50, 100, 150, 200
MBS transmission power for inner region	40 dBm	Threshold for RSS (P_{TH})	− 72 dBm
MBS transmission power for outer region	43 dBm	Maximum pilot power ($P_{t,Pilot}^{\max}$)	15 dBm
FAP transmission power	20 dBm	Number of independent trials for each set of parameters	200
Number of RBs of each FAP	12 RB	Simulation time of each trial	5000 s

Therefore, for each constant value of velocity, number of FAPs, and number of mobile users, simulation of the network is carried out 200 times where each trial includes 5000 user locations.

We carried out our tests and maintained the situation throughout the simulations for three different services, with the option for each user to pick randomly. In addition, the services are assumed to use one, two, and three RBs.

Due to the low transmission power and small coverage region of FAP, 8–12 RBs are considered for FAPs. On the other hand, the number of RBs allocated to the FAP depends on the number of users in its coverage. If the number of users increases, the number of RBs allocated to the FAP should also be increased. In the case of large number of users in the coverage of FAP, if FAP does not have enough RBs to serve the users, some users must connect to other FAPs which results in SINR reduction due to the large distance between the user and FAP. Therefore, efficient number of RBs should be allocated to each FAP. Since FAP has lower computation ability than the MBSs, increasing the number of allocated RBs to the FAP results in more computational complexity and processing delay which in turn reduce the quality of connection. On the other hand, small number of RBs may result in connection lost due to the inability of users in the coverage of FAP in connection to it. Therefore, we consider 12 RBs for each FAP.

The area of macrocell is partitioned into 20×20 grids of size 50 m $\times$ 50 m, therefore, there are 400 zones in the coverage of one macrocell. The residence time of mobile user in each zone is chosen randomly according to uniform distribution in the range (120, 600) s. The maximum pause time of mobile users in each way point is set to 15 s.

4.1 Throughput

Throughput is defined as the number of bits that can be successfully delivered to user within each transmitted symbol in a certain period of time [35]. The throughput is calculated from the Shannon entropy law [36] as follows

$$r = \log_2(1 + \gamma) \quad \text{bits/s/Hz} \tag{7}$$

where γ denotes the SINR. The cumulative distributions of throughputs of mobile users for different FAP densities and mobile user velocities are demonstrated in Fig. 5. For a constant FAP density, as the number of mobile users decreases, the number of users that utilize the same RB decreases accordingly and vice versa. As the number of users that use the specific RB decreases, the interference at the RB also decreases, leading to higher SINR and throughput. This is depicted in Fig. 5. As observed by decreasing the number of mobile users, the average throughput per each RB increases. For example, at the velocity of 2 km/h, FAP density $\lambda = 15 \times 10^{-5}$, and 50 mobile users, 40% of RBs have throughput less than 5 bits/s/Hz. As the number of mobile users increases to 200, around 70% of RBs have throughput less than 5 bits/s/Hz.

Comparing Fig. 5a–c with Fig. 5d indicates that for the same number of mobile users, increasing the FAP density, i.e., increasing the number of FAPs, improves the cell-edge user experience. The reason is that, when the number of FAPs increases, the number of available RBs increases and the number of users that utilize a specific RB, reduces. Therefore, interference in that RB reduces and SINR increases, which results in higher throughput.

4.2 Load Index

The Jain's fairness index has been frequently used to measure the fairness of different resource allocation schemes, which is defined as [37]

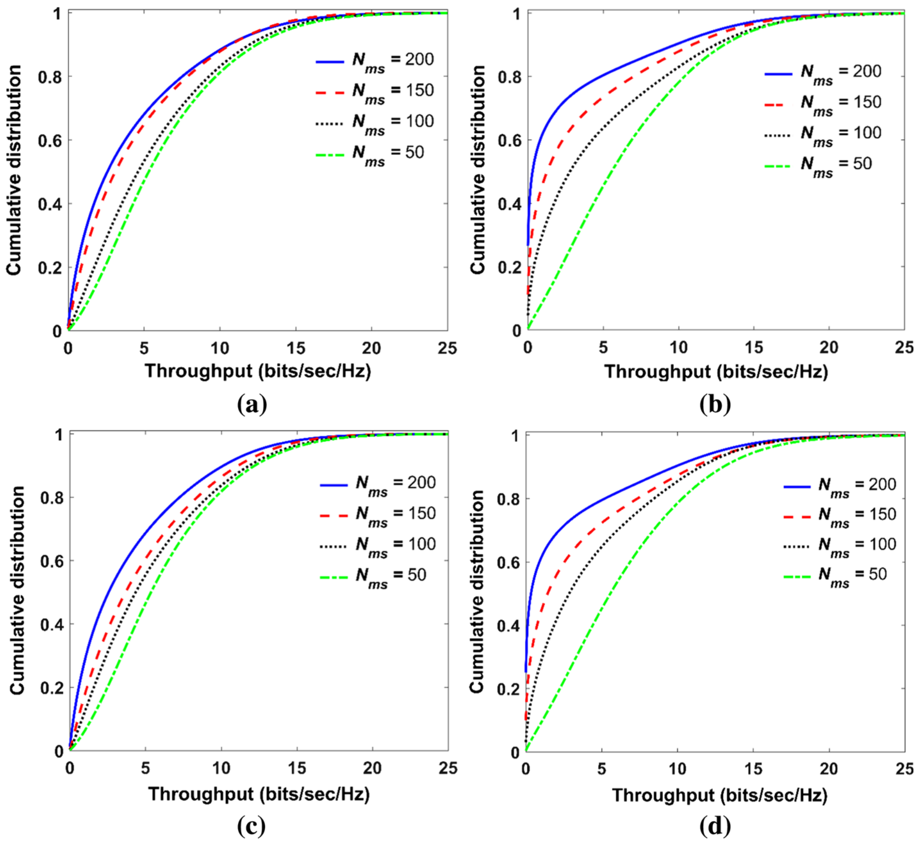

Fig. 5 CDF of throughput for different FAP densities and mobile user velocities. N_{ms} denotes the number of mobile users. **a** velocity = 2 km/h, $\lambda = 15 \times 10^{-5}$, **b** velocity = 2 km/h, $\lambda = 5 \times 10^{-5}$, **c** velocity = 5 km/h, $\lambda = 15 \times 10^{-5}$, **d** velocity = 5 km/h, $\lambda = 5 \times 10^{-5}$

$$I_{Jain} = \frac{\left(\sum_{i=1}^{N_{FAP}} N_{RB}^{(i)}\right)^2}{N_{FAP} \sum_{i=1}^{N_{FAP}} \left(N_{RB}^{(i)}\right)^2} \tag{8}$$

where $N_{RB}^{(i)}$ is the number of RBs that the ith FAP allocates to users. The value of I_{Jain} becomes 1 when all FAPs share an equal load, and it is $1/N_{FAP}$ in the case of extreme imbalance.

Figure 6 shows Jain's fairness index with its standard deviation for different FAP densities and user velocities. Comparing Fig. 6a, b demonstrates that user velocity does not have considerable effect on Jain's fairness index. But, at the constant velocity, the number of mobile users and FAP density considerably affect the performance. It is observed that for the constant FAP density, Jain's fairness index increases by increase in the number of mobile users. For lower number of mobile users, it is possible that some FAPs do not serve any mobile user. Therefore, load balancing between FAPs is degraded. However, as the number of mobile users increases, the probability that more FAPs serve them, increases and as a result, Jain's fairness index increases. Also, it is observed that standard deviations

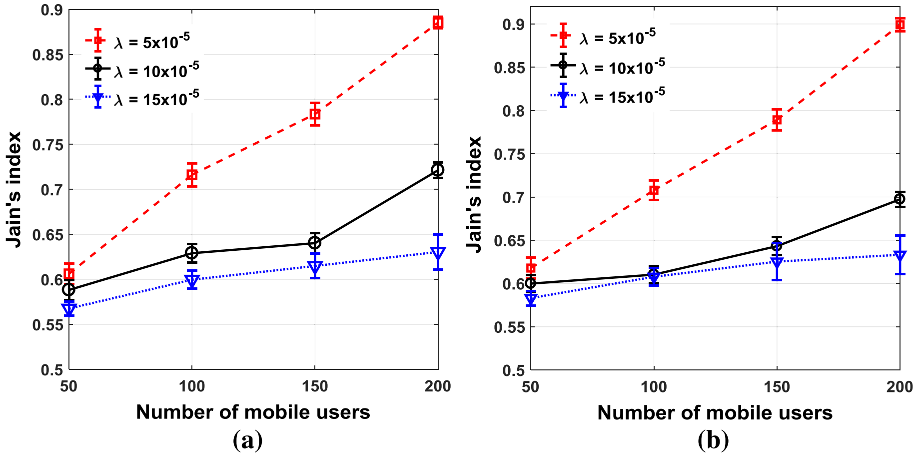

Fig. 6 Jain's fairness index. **a** velocity = 2 km/h, **b** velocity = 5 km/h

of the obtained Jain's index in different trials are small, which means the proposed algorithm efficiently balances the load between base stations in different conditions.

As mentioned, the main purpose of load balancing is to push the load from MBS to FAPs and MBS becomes available for emergency situation. The parameter that can show the effectiveness of load pushing from MBS to FAPs, is the ratio of RBs that FAPs allocate to mobile users to the total number of RBs used by mobile users, which is denoted by "femtocell tier cooperation ratio". In Fig. 7, the femtocell tier cooperation ratio is shown for different user velocities. For the constant FAP density, by increasing the number of mobile users which results in interference increasing in RBs provided by MBS, most of mobile users prefer to connect to FAPs with higher SINRs. It is observed that as Jain's index, the standard deviations of the obtained femtocell cooperation ratio in different trials are small in comparison with means. Therefore, femtocell tier can efficiently serve the users in different conditions.

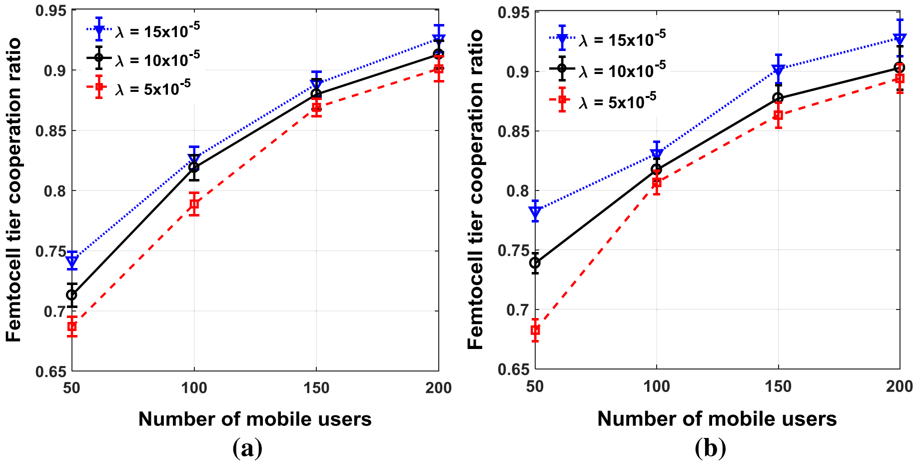

Fig. 7 Femtocell tier cooperation ratio. **a** Velocity = 2 km/h, **b** Velocity = 5 km/h

4.3 Ping-pong Rate

The ping-pong rate of the proposed algorithm in different speeds and FAP densities is given in Fig. 8. Ping-pong rate is defined as the number of ping-pong handoffs to the total number of handoffs. It is observed that ping-pong rate of the proposed method is low and has small standard deviation. Some factors that increase the number of handoffs may cause increase in the ping-pong rate. According to Figs. 2 and 4, in the proposed method by estimation of RSS of target cell for five time steps (which is also ping-pong threshold), ping-pong rate decreases considerably. The most important factor affecting ping-pong rate in the proposed method is prediction error. When the speed of user and number of FAPs increase, the variations of RSS and SINR also increase and prediction becomes a little complicated which increases the prediction error and consequently ping-pong rate increases slightly.

4.4 Performance Comparison

In order to evaluate the performance of our approach, we made the comparisons with the conventional Max-SINR algorithm. In Fig. 9, the performance of both algorithms is presented in terms of throughput, Jain's fairness index, femtocell tier cooperation index, and ping-pong rate. Under the conditions that a service requires multiple RBs, the Max-SINR algorithm uses only the maximum SINR without taking into account the SINRs of other RBs, while the proposed algorithm examines the SINR requirement in all RBs. Such conditions can be seen in Fig. 9(a), where the proposed algorithm incurs higher throughput than what the Max-SINR method offers. The proposed pilot power, which depends on the number of free RBs, results in pushing load to low-loaded FAPs, while Max-SINR uses the same pilot power for all FAPs. Consequently, the proposed algorithm has higher Jain's fairness index as shown in Fig. 9b. For macrocell users, at the first step, γ_s is not compared with γ_{t2} which results in more FAPs association rather than Max-SINR algorithm, leading to femtocell tier cooperation ratio increasing as shown in Fig. 9c. Finally, since the

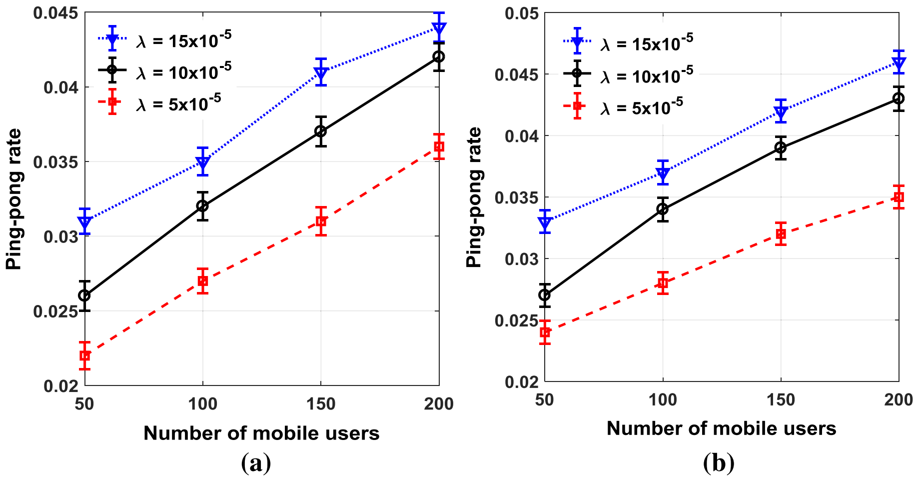

Fig. 8 Ping-pong rate of the proposed method. **a** velocity = 2 km/h, **b** velocity = 5 km/h

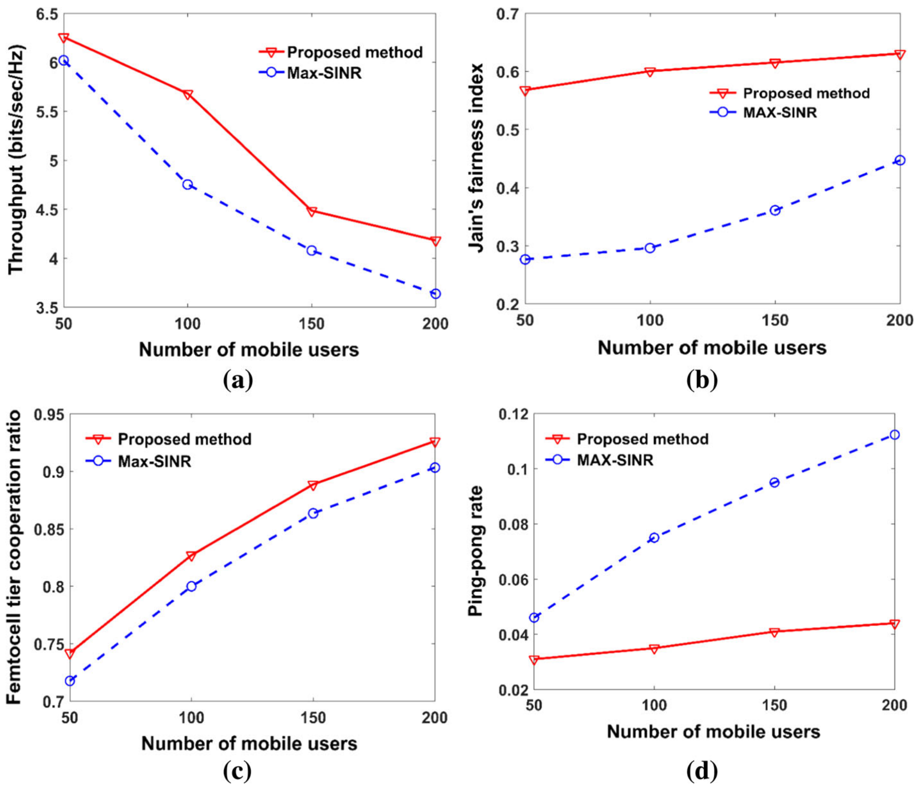

Fig. 9 Performance comparison between the proposed method and MAX-SINR method, velocity = 2 km/h, $\lambda = 15 \times 10^{-5}$. **a** Throughput, **b** Jain's fairness index, **c** Femtocell tier cooperation ratio, **d** Ping-pong rate

proposed method checks the predicted RSS values, it can prevent unnecessary handoffs, and therefore has lower ping-pong rate than the MAX-SINR method as shown in Fig. 9d.

As the main objective throughout the paper, we proposed a load-balancing algorithm for heterogeneous networks, which can be employed in various resource allocation schemes. One of the key indicators of the majority of the resource allocation schemes is known to be user SINR which in turn, is influenced by interference coordination. Therefore, the performance of the resource allocation algorithms with user SINR as a decision metric is definitely dependent upon interference coordination. In part, investigating the impacts of any interference coordination methods is normally performed for resource allocation as they are—whether directly or indirectly—correlated. Hence, we believe that any considerations on interference mitigation/coordination approach determines the performance of resource allocation scheme regardless of whether it has any impact on load-balancing, and such study is barely to reflect any performance detail from our approach.

4.5 Computational Complexity and Synchronization

The complexity of prediction and estimation were solved using recursive least square (RLS) of order $O(F_1^2)$, where F_1 is the length of the used FIR filter. Also, the proposed algorithm is solved by grid search algorithm with the complexity of order $O(L^D)$, where L denotes the number of searches and D is the dimension of variables of optimization algorithm.

In traditional decision method, i.e., max-SINR, the cell with maximum SINR is selected and its complexity is of order $O(n)$, where n is the number of SINRs. In comparison, although max-SINR incurs lower complexity than the proposed method, the proposed method comes with higher throughput, higher Jain's fairness index, and lower ping-pong rate.

According to LTE Release 8 [38], there are three synchronization requirements in LTE: 1) symbol timing acquisition to determine the start of correct symbol; 2) carrier frequency synchronization for mitigating the effect of frequency errors due to Doppler shift and electronics; and 3) sampling clock synchronization.

In user end, two synchronization signals are broadcasted in each cell with every radio frame of 10 ms, namely primary synchronization sequence (PSS) and secondary synchronization sequence (SSS). The information acquired by UE from these signals include radio frame, sub-frame, slot, symbol synchronization. Also, PSS and SSS facilitate UE to identify the carrier frequency and thereby extract the physical layer cell identity (PCI). UE can measure the cell specific reference signals by detecting the synchronization signals and decode the master information block (MIB) on the physical broadcast channel (PBCH) [38].

5 Conclusion

In this paper an efficient cell selection algorithm for load balancing among FAPs in two-tier macrocell-femtocell HetNets was proposed. A new method for power allocation to pilot signal which is based on the number of free RBs was introduced that shows FAP tendency to RB allocation. In this way, the FAPs with more free RBs have priority to allocate RBs to users which leads to load balancing among FAPs. Three lists of neighbor FAPs were constructed based on the RSS, predicted RSS, the number of free RBs extracted from pilot power, respectively. The FAP with more number of free RBs that can provide the throughput requirement of user is selected as the target cell.

Simulation results showed that at the constant FAP density as the number of mobile users increases, the co-channel interference increases which results in throughput reduction. But, since more number of FAPs serve mobile users, Jain's fairness index and femtocell tier cooperation ratio increases. If the number of mobile users remains constant and the number of FAPs increases, the number of RBs with acceptable SINR increases which leads to throughout enhancement and increase in femtocell cooperation ratio. However, Jain's fairness index reduces specially for low number of mobile users. Performance comparison with Max-SINR method demonstrated the efficiency of the proposed algorithm in terms of throughput, Jain's fairness index, femtocell tier cooperation ratio, and ping-pong rate.

In the proposed method, construction of neighbor lists and target cell selection is performed in user equipment, which requires fast processors and results in increment in power consumption and battery life decrement. However, the processors are fasten and battery technology is upgraded every day. Also, in the future some technologies such as energy harvesting can help to have more battery life.

References

1. De La Roche, G., Valcarce, A., López-Pérez, D., & Zhang, J. (2010). Access control mechanisms for femtocells. *Communications Magazine, IEEE, 48*(1), 33–39.
2. Kwon, Y. J., & Cho, D.-H. (2011). Load based cell selection algorithm for faulted handover in indoor femtocell network. In *Vehicular technology conference (VTC Spring), 2011 IEEE 73rd* (pp. 1–5). IEEE.
3. Andrews, J., Singh, S., Ye, Q., Lin, X., & Dhillon, H. (2014). An overview of load balancing in HetNets: Old myths and open problems. *Wireless Communications, IEEE, 21*(2), 18–25.
4. Ye, Q., Al-Shalashy, M., Caramanis, C., & Andrews, J. G. (2013). On/off macrocells and load balancing in heterogeneous cellular networks. In *Global communications conference (GLOBECOM), 2013 IEEE* (pp. 3814–3819). IEEE.
5. Bjerke, B. (2011). LTE-advanced and the evolution of LTE deployments. *Wireless Communications, IEEE, 18*(5), 4–5.
6. Damnjanovic, A., et al. (2011). A survey on 3GPP heterogeneous networks. *Wireless Communications, IEEE, 18*(3), 10–21.
7. Lee, C.-H. (2011). Study of load balance in 3GPP femto-cell network. In *Network operations and management symposium (APNOMS), 2011 13th Asia-Pacific* (pp. 1–4). IEEE.
8. Kahwa, T., & Georganas, N. (1978). A hybrid channel assignment scheme in large-scale, cellular-structured mobile communication systems. *Communications, IEEE Transactions on, 26*(4), 432–438.
9. Jiang, H. & Rappaport Stephen, S. (1994). CBWL: A new channel assignment and sharing method for cellular communication systems. In *IEEE transactions on vehicular technology* (vol. 43(2), pp. 313–322).
10. Das, S. K., Sen, S. K., & Jayaram, R. (1997). A dynamic load balancing strategy for channel assignment using selective borrowing in cellular mobile environment. *Wireless Networks, 3*(5), 333–347.
11. Das, S. K., Sen, S. K., & Jayaram, R. (1998). A novel load balancing scheme for the tele-traffic hot spot problem in cellular networks. *Wireless Networks, 4*(4), 325–340.
12. Eklundh, B. (1986). Channel utilization and blocking probability in a cellular mobile telephone system with directed retry. *IEEE Transactions on Communications, 34*(4), 329–337.
13. Wu, X., Mukherjee, B. & Chan, S.-H. (2000). Maca-an efficient channel allocation scheme in cellular networks. In *Global telecommunications conference, 2000. GLOBECOM'00. IEEE* (vol. 3, pp. 1385–1389). IEEE.
14. Cavalcanti, D., Agrawal, D., Cordeiro, C., Xie, B., & Kumar, A. (2005). Issues in integrating cellular networks WLANs, AND MANETs: A futuristic heterogeneous wireless network. *Wireless Communications, IEEE, 12*(3), 30–41.
15. Yanmaz, E., & Tonguz, O. K. (2004). Dynamic load balancing and sharing performance of integrated wireless networks. *IEEE Journal on Selected Areas in Communications, 22*(5), 862–872.
16. Das, S., Viswanathan, H. & Rittenhouse, G. (2003). Dynamic load balancing through coordinated scheduling in packet data systems. In *INFOCOM 2003. Twenty-Second annual joint conference of the ieee computer and communications. IEEE societies* (vol. 1, pp. 786–796). IEEE.
17. Bejerano, Y., & Han, S.-J. (2009). Cell breathing techniques for load balancing in wireless LANs. *IEEE Transactions on Mobile Computing, 8*(6), 735–749.
18. Khandekar, A., Bhushan, N., Tingfang, J. & Vanghi, V. (2010). LTE-advanced: Heterogeneous networks. In *Wireless conference (EW), 2010 European* (pp. 978–982). IEEE.
19. Ramjee, T. B. L. E. L. R. (2006). Generalized proportional fair scheduling in third generation wireless data networks. In *IEEE INFOCOM* (pp. 1–12).
20. Bejerano, Y., Han, S.-J. & Li, L. E. (2004). Fairness and load balancing in wireless LANs using association control. In *Proceedings of the 10th annual international conference on Mobile computing and networking* (pp. 315–329). ACM.
21. Kim, H., de Veciana, G., Yang, X. & Venkatachalam, M. (2010). Alpha-optimal user association and cell load balancing in wireless networks. In *INFOCOM, 2010 proceedings IEEE* (pp. 1–5). IEEE.
22. Ye, Q., Rong, B., Chen, Y., Al-Shalash, M., Caramanis, C., & Andrews, J. G. (2013). User association for load balancing in heterogeneous cellular networks. *IEEE Transactions on Wireless Communications, 12*(6), 2706–2716.
23. Ye, Q., Rong, B., Chen, Y., Caramanis, C., & Andrews, J. G. (2012). Towards an optimal user association in heterogeneous cellular networks. In *Global communications conference (GLOBECOM), 2012 IEEE* (pp. 4143–4147), IEEE.
24. Oh, J & Han, Y (2012). Cell selection for range expansion with almost blank subframe in heterogeneous networks. In *2012 IEEE 23rd international symposium on personal indoor and mobile radio communications (PIMRC)* (pp. 653–657). IEEE.

25. Novlan, T., Andrews, J. G., Sohn, I., Ganti, R. K., & Ghosh, A. (2010). Comparison of fractional frequency reuse approaches in the OFDMA cellular downlink. In *Global telecommunications conference (GLOBECOM 2010), 2010 IEEE* (pp. 1–5). IEEE.
26. Lee, P., Lee, T., Jeong, J. & Shin, J. (2010). Interference management in LTE femtocell systems using fractional frequency reuse. In *2010 The 12th international conference on advanced communication technology (ICACT)* (vol. 2, pp. 1047–1051). IEEE.
27. Kalbkhani, H., Yousefi, S., & Shayesteh, M. G. (2014). Adaptive handover algorithm in heterogeneous femtocellular networks based on received signal strength and signal-to-interference-plus-noise ratio prediction. *Communications, IET, 8*(17), 3061–3071.
28. Kalbkhani, H., Solouk, V., & Shayesteh, M. (2015). Resource allocation in integrated femto–macrocell networks based on location awareness. *IET Communications, 9*(7), 917–932.
29. Le, L. B., Hoang, D. T., Niyato, D., Hossain, E. & Kim, D. I. (2012). Joint load balancing and admission control in OFDMA-based femtocell networks. In *2012 IEEE international conference on communications (ICC)* (pp. 5135–5139), IEEE.
30. Nguyen, A. D., Sénac, P., Ramiro, V., & Diaz, M. (2011). STEPS-an approach for human mobility modeling networking. *Springer, 2011,* 254–265.
31. Pourmina, M. A., & MirMotahhary, N. (2012). Load balancing algorithm by vertical handover for integrated heterogeneous wireless networks. *EURASIP Journal on Wireless Communications and Networking, 2012*(1), 1–17.
32. Koksal, C. E., Kassab, H. & Balakrishnan, H. (2000). An analysis of short-term fairness in wireless media access protocols (poster session). In *ACM SIGMETRICS Performance Evaluation Review*, (vol. 28, no. 1, pp. 118–119). ACM.
33. Goldsmith, A. (2005). *Wireless communications*. Cambridge: Cambridge University Press.
34. Bălan, I. M., Sas, B., Jansen, T., Moerman, I., Spaey, K., & Demeester, P. (2011). An enhanced weighted performance-based handover parameter optimization algorithm for LTE networks. *EURASIP Journal on Wireless Communications and Networking, 2011*(1), 1–11.
35. Qiu, X., & Chawla, K. (1999). On the performance of adaptive modulation in cellular systems. *IEEE Transactions on Communications, 47*(6), 884–895.
36. Coifman, R. R., & Wickerhauser, M. V. (1992). Entropy-based algorithms for best basis selection. *IEEE Transactions on Information Theory, 38*(2), 713–718.
37. Sediq, A. B., Gohary, R. H., Schoenen, R., & Yanikomeroglu, H. (2013). Optimal tradeoff between sum-rate efficiency and Jain's fairness index in resource allocation. *IEEE Transactions on Wireless Communications, 12*(7), 3496–3509.
38. Innovations, T. (2010). LTE in a Nutshell. In *White paper*

Publisher's Note

Yasin Aghazadeh received the B.Sc. and M.Sc. degrees in Electrical Engineering from Urmia University, Urmia, Iran, in 2011 and 2015, respectively. His research interests lie in the area of wireless communication with special focus on LTE and LTE-advanced heterogeneous networks. He is currently working as radio expert for Communications Regulatory Authority (CRA).

Hashem Kalbkhani received the B.Sc., M.Sc., and Ph.D. degrees from Urmia University, Iran, all in Electrical Engineering. His research interests are cellular networks and signal processing.

Mahrokh G. Shayesteh received the B.Sc. degree from the University of Tehran, Tehran, Iran; the M.Sc. degree from Khajeh Nassir University of Technology, Tehran, Iran; and the Ph.D. degree from Amir Kabir University of Technology, Tehran, Iran, all in Electrical Engineering. She is currently a Professor with the Department of Electrical Engineering, Urmia University, Urmia, Iran. She is also working with the Wireless Research Laboratory, Advanced Communication Research Institute (ACRI), Department of Electrical Engineering, Sharif University of Technology, Tehran, Iran. Her research interests include wireless communications, signal and image processing.

Vahid Solouk is an assistant professor at the department of IT and Computer Engineering, Urmia University of Technology and currently serves as Director of Graduate Campus. He received his Ph.D. degree in Communication and Network Engineering from the department of Communication Engineering, Universiti Putra Malaysia. He is a member of IEEE. His research interests snap specific areas of wireless and mobile communications including mobility management, resource allocation, and channel coding.