Wu et al. EURASIP Journal on Wireless Communications and Networking (2015) 2015:117
DOI 10.1186/s13638-015-0280-y

EURASIP Journal on Wireless Communications and Networking
a SpringerOpen Journal

RESEARCH Open Access

Dynamic switching off algorithms for pico base stations in heterogeneous cellular networks

Jie Wu[1], Shi Jin[1*], Lei Jiang[2] and Gang Wang[2]

Abstract

The densification of pico base stations (PBSs) in heterogenous cellular networks (Hetnets) causes redundant energy consumption of the whole network during low traffic periods. In order to address this problem, we take advantage of the traffic load fluctuation over time and area to adapt the necessary resource to the actual traffic demand by introducing sleep mode to the PBSs. To achieve this target, we propose two centralized algorithms (i.e., the heuristic algorithm and the progressive algorithm) in this paper to dynamically switch off the unnecessary PBSs, both of which can track the traffic variation well. We design a utility function of some key factors considered particularly for PBSs in Hetnets. These algorithms rely on the utility function to switch off the redundant PBSs, enabling the working mode of all PBSs be reconfigured periodically. The progressive algorithm is proposed to overcome the inefficiency of the heuristic algorithm. The simulations demonstrate that the execution time of the progressive algorithm is at most one third of that of the heuristic algorithm, which enables the network to respond to the traffic variation more promptly. Besides, the progressive algorithm can switch off more PBSs than the heuristic algorithm while slightly affects the network blocking probability, which indicates that the progressive algorithm has a better potential to save energy. Moreover, simulations also reveal that some key parameters all have nonnegligible influence on the performance of our algorithms. These parameters should be tuned well to trade spare network resource for energy saving.

Keywords: Heterogenous cellular networks (Hetnets); Pico base stations; Dynamic switching off algorithm; Energy saving; Blocking probability

1 Introduction

Nowadays, with the proliferation of smart devices, mobile data traffic is increasing exponentially. It is reported that the number of mobile-connected devices will exceed the world's population in this year. Furthermore, by 2018, mobile network connection speeds will increase twofold and mobile-connected tablets will generate nearly double the traffic generated by the entire global mobile network in 2013 [1]. To satisfy such incremental demand of the future communication, recently, a new framework called heterogenous cellular networks (Hetnets) [2-5] has emerged as a flexible and cost-effective solution. Different from typical macrocells of high power base stations (MBSs) serving as a coverage layer, in Hetnets, small cell access points such as relay nodes, picocell base stations (PBSs), femtocell base stations, and remote radio heads overlaid on macrocells serve as a capacity layer, which bring mobile networks closer to user equipments (UEs), thus enhancing network capacity [6-8].

The densification of small cells strives to excavate the spatial splitting gains, while also increases the energy consumption of the whole network simultaneously. Currently, it has been estimated that the overall energy consumed by information and communications technology (ICT) industry, which includes cellular networks, already constitutes about 2% of global carbon emissions and is projected to increase much further in the coming years [9]. Moreover, it has been revealed in [10] that up to 80% of the energy consumption in a cellular network is attributed to the operations and functionality of the base stations (BSs) in the radio access network while the remaining energy is expended in the switching and core networks.

Since the required network capacity and the number of BSs in a cellular network are typically dimensioned to serve the peak traffic, if all BSs remain active irrespective of traffic load, a tremendous amount of resource will

*Correspondence: jinshi@seu.edu.cn
[1]School of Information Science and Engineering, Southeast University, No. 2 Sipailou, Nanjing 210096, China
Full list of author information is available at the end of the article

be redundant during off-peak times, and energy is inefficiently consumed. Besides, even at the same time, traffic distribution of different network area can be no uniform. Therefore, methods that reduce energy consumption by adapting the network resource to the traffic demand are important research directions. Fortunately, such temporal and spatial traffic load fluctuation gives the opportunity to save energy significantly by switching off the underutilized MBSs or PBSs [11,12].

1.1 Contributions

We consider picocells which are low-power and low-cost cells covered by PBSs and designed to serve a small outdoor area such as hotspot or shaded region. The main objective of this paper is to switch off the redundant PBSs to reduce energy consumption of the network while simultaneously guaranteeing the quality of service (QoS) of UEs. Specifically, the fundamental limit of the achievable QoS of the macrocell UEs (MUEs) and the picocell UEs (PUEs) is given in terms of the required service rate. It is known that BS switching of finding the optimal operation mode of the network system is a difficult combinatorial problem. Since this problem is NP hard and requires high-computational complexity as well as large signaling overhead, there are some works [13-15] considering low complexity algorithms to tackle such problem. These algorithms are all sub-optimal, yet proved to be very useful in the given scenarios. Similarly, in our work, we propose practically implementable algorithms by considering the main characteristics of the Hetnets. The major contributions of this paper are summarized as follows:

1) We design a utility function of PBS considering some key factors particularly for PBSs in Hetnets. These factors within a certain picocell include the total rate of served PUEs, the PBS's traffic load, the number of served PUEs, the number of blocked PUEs, and the received interference signal strength from the nearby cells. The utility function assists the following proposed algorithms to select PBSs in a reasonable order to be tested whether they can be switched off.
2) We first propose a heuristic dynamic switching off (HDSO) algorithm, which tests PBSs to switch off one by one at each step, and is similar to the switching off algorithm proposed in [16]. It turns out that after this algorithm is executed, the number of active PBSs can track the network traffic profile very well. Simulations show that some parameters can significantly influence the performance of HDSO. Besides, its complexity can be prohibitive for large-size network with dense deployment of PBSs and peak traffic time.
3) To overcome the inefficiency of HDSO, we propose a progressive dynamic switching off (PDSO) algorithm (inspired by the works in [17]), by avoiding the unnecessary work of testing. Different from the HDSO algorithm, PDSO is carried out in a round by round manner. Specifically, in one round of testing, this algorithm is based on the utility function to classify the current active PBSs into two groups, then a switching-off process is carried out to test all PBSs in the group with relatively small utilities to be switched off. It intelligently decides whether to launch another round of testing according to the switching off result of the previous round. Simulations demonstrate that the PDSO algorithm is more efficient in switching off the redundant PBSs, for which the complexity of the algorithm is greatly reduced. Moreover, with properly tuned parameters, the PDSO algorithm is verified to have a better potential of energy saving.
4) We propose a PUE transferring algorithm to transfer the PUEs of the PBS which is tested to be switched off to the nearby BSs. As we aim at switching off the PBSs, this algorithm first attempts to transfer the PUE to the nearby MBSs. If this attempt fails, then the nearby PBSs will be considered as the acceptor BSs. Both the acceptor MBSs and PBSs need to reserve some resource blocks (RBs) for subsequent new UEs in order to reduce the network blocking probability. The principle to decide whether to switch off a PBS is that if all its served PUEs can be successfully transferred to the nearby BSs.

1.2 Related work

Different techniques of energy saving have been proposed. The challenge to these techniques is to maintain reliable service coverage and QoS, while simultaneously saving the most energy. In the context of Hetnets, a novel idle mode procedure has been proposed in [18], which allows the femto BS transmissions and associated processing to be switched off completely at all times when the femto BS does not need to support an active call. This is achieved by a low-power sniffer capability in the femto BS and a predetermined threshold of uplink (UL) received power. Besides, threshold-based approaches have also been used in [19,20] to apply sleep mode to PBSs. The basic idea is to switch off the PBSs when their traffic loads are below a certain threshold for a certain period of monitoring time, which means the PBSs' loads can be offloaded by their neighboring BSs. The threshold of either the UL received power or the PBS traffic load is an important metric in the performance of the algorithm, which greatly depends on automatic configuration and fine adjustment of the threshold during operation. In addition, these approaches are derived from the micro perspective (i.e., the cell level); however, from the macro perspective (i.e., the network level), there also exists a fraction threshold of the necessary PBSs needed to remain active. In [17], the authors developed a simple analytical model that allow optimal BS switch-off times to be identified as a function of the daily traffic pattern, in the case where several switch offs per day are permitted (progressively reducing the number of active base stations and the network energy). Some fractions of the BSs needed to be active at certain time instants per day are assumed. Inspired by this work, we propose our PDSO algorithm which adopts the threshold approach from the network perspective.

For the conventional cellular networks, dynamic switching off of BSs has been extensively studied in [11,16,21-26]. A solution proposed in [21] automatically switch off appropriate BSs or sectors, which become the compensation area needed to be covered through the tilting of antennas in the neighboring BSs. A novel energy-efficient cellular access network architecture based on the principle of ecological protocooperation was proposed in [22], which indicates that BSs can cooperatively and dynamically make intelligent decisions based on thresholds for switching between different power modes according to traffic conditions. Similarly, a transmission power increment is required for compensating the areas of switched off BSs. In contrast to these works, the use of coordinated multipoint for MBS switching off without transmission power adjustment from compensating neighboring BSs is presented in [23]. From a game theoretic perspective, the energy efficiency issues in multi-operator mobile networks was studied in [24], where cost-based functions are used to decide the best suitable BSs to remain active. This paper introduces the cost that has to be paid by an operator when its subscribers have to be served by another operator due to the fact that some BSs have been switched off. In [25], the authors proposed a practical switching on/off-based energy saving algorithm that can be realized distributively. The key principle of the algorithm was to switch off a BS one by one that minimally affect the network by using network impact, which takes into account the additional load increments brought to its neighboring BSs. Distance-based approaches have been developed in [11,16]. In [11], the authors used two real datasets (i.e., temporal and spatial) to estimate the energy savings, and they used a greedy algorithm by sequentially switching off the BSs with the minimum distance to its nearest active BSs if the coverage is met. A dynamic switching on/off algorithm where the number of active BSs adapts to the network condition was proposed in [16], which was based on BS traffic load and the position of the associated UEs. This algorithm preferentially tests the BS with larger average distance to its associated UEs to be switched off one by one and terminates when UEs of a BS cannot be accepted by the neighboring BSs. In realistic networks, the distances among BSs in homogeneous cellular network can be easily acquired, but the position of UEs may not be easily and accurately acquired. Besides, the policy that the algorithm just stops when one BS fails to transfer its UEs does not seem reasonable as other BSs with larger average distance may be lowly loaded, thus also have the potential to be switched off. It is important to remark that in the previous works regarding homogeneous cellular network, the decision to switch off BSs was based on either the distance factor or the traffic. However, since we focus on switching off PBSs in Hetnets, the distance between PBS and PUE cannot be accurately or easily obtained due to various reasons, such as technology limits and designing principles. As in some cases, the network side is not aware of the UEs' position, which is also the desire of the users. Moreover, in addition to the load, other factors such as UEs' service rate, the historical blocking probability, and the interference conditions should also be considered. Although BS switching off is primarily designed to reduce network energy consumption, it should be noted that the various approaches must ensure that the QoS in the coverage area is not compromised at all times. Emphasizing on this concern, a progressive BS switching on/off technique has been implemented through the coordination of multiple surrounding BSs in [26], where the main finding shows that the duration of BS sleeping and waking up transients is very short, with no significant reduction of the energy savings achievable with sleep mode approaches.

1.3 Organization

The remainder of the paper is organized as follows. Section 2 presents the system model for the Hetnets. In Section 3, we propose the algorithms and elaborate them in detail. In Section 4, extensive simulations have been performed to investigate the performance of these algorithms, and the simulation results are provided with detailed analysis. Finally, conclusions are drawn in Section 5.

2 System model

Figure 1 shows a system model of Hetnets which consists of a macrocell and multiple picocells and femtocells. The Hetnets considered in this paper consist of M MBSs and P PBSs. We focus on the downlink (DL) transmission scenario based on orthogonal frequency division multiple access (OFDMA), in which the total bandwidth B is divided into N_{RB} RBs with a set $\mathcal{N}$. Co-channel deployment is considered; hence, all RBs (i.e., the scheduled units) are simultaneously allocated by both MBSs and PBSs to their served UEs. Only large-scale fading of the channel model is considered, and small-scale fading is omitted. The subcarriers constituting a single RB are subjected to the same fading, and hence, the channel gain

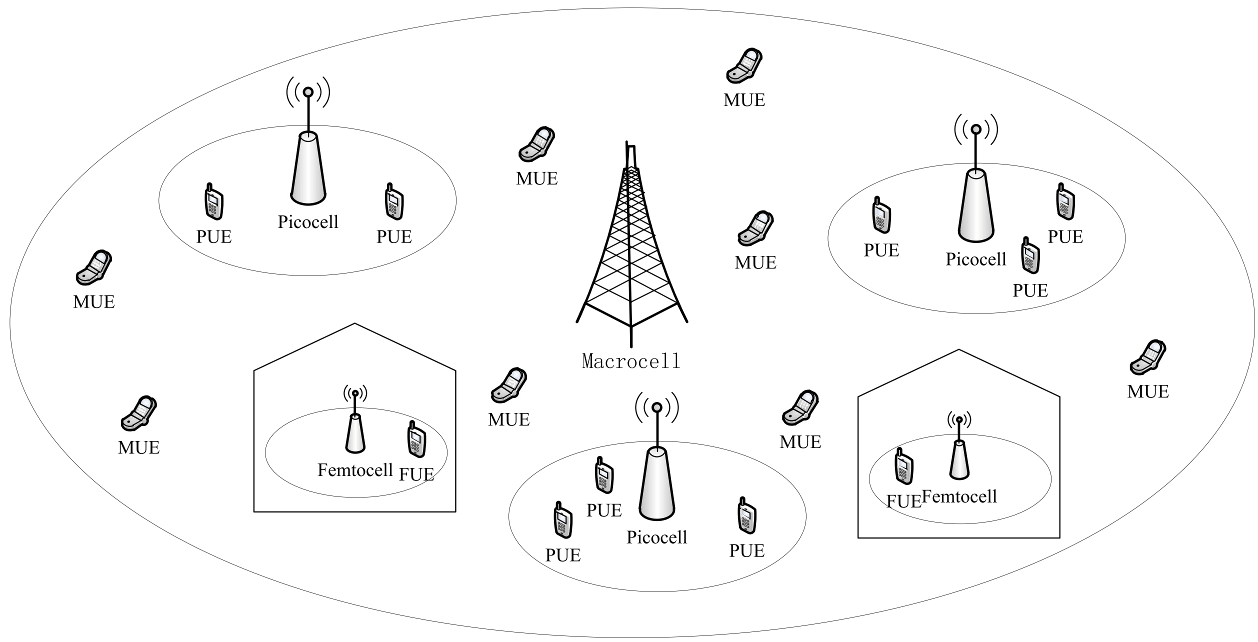

Figure 1 System model of the Hetnets deployments.

on the subcarriers of a single RB is assumed the same. Additionally, the fading is assumed to be independent identically distributed (i.i.d.) across RBs. We generally refer to a MBS or a PBS as a BS and a macrocell or a picocell as a cell. In order to guarantee seamless coverage, MBSs keep active, while PBSs can be switched off to save energy when traffic load is low.

For simplicity, power control is not considered in this paper. Assume that transmission power is uniformly allocated among all RBs for either MBS or PBS, namely, the power on each RB is calculated as $P_m = P_m^{\text{tot}}/N_{\text{RB}}$, $P_m = P_m^{\text{tot}}/N_{\text{RB}}$, where P_m^{tot} is the total transmission power of MBS, and P_p^{tot} is the total transmission power of PBS. We also assume that the unused portion of DL transmission power at the MBS and the PBS is not reallocated to other RBs in order to avoid additional interference toward normal UEs.

2.1 Achievable service rate calculation for UEs

In LTE, the different RBs allocated to a UE can have different modulation and coding schemes (MCSs). However, we do not take the different MCSs into account in this paper and just use the Shannon-Hartlay theorem to calculate the UE throughput for simplicity as in [21]. Assume that MUE k is served by the MBS m_i, $\mathcal{I}_{\text{RB},k}$ is the set of RBs allocated from MBS m_i to MUE k, then the achievable rate of MUE k can be expressed as:

$$R_k^{m_i} = \sum_{n\in\mathcal{I}_{\text{RB},k}} B_{\text{RB}}\log_2\left(1+\text{SINR}_{k,n}^{m_i}\right) \quad (1)$$

where B_{RB} is the bandwidth of one RB, $\text{SINR}_{k,n}^{m_i}$ is the signal-to-interference plus noise ratio (SINR) on the nth RB of MUE k, which can be expressed as:

$$\text{SINR}_{k,n}^{m_i} = \frac{P_m G_{k,n}^{m_i}}{I_{k,n}^{m_i}+\sigma_n^2} \quad (2)$$

where $G_{k,n}^{m_i}$ is the average channel gain from MBS m_i ($'m_i'$ for the ith MBS, $'p_j'$ for the jth PBS) to MUE k on the nth RB, which includes path loss and shadowing. σ_n^2 is the power of additive white Gaussian noise on the nth RB, and $I_{k,n}^{m_i}$ is the interference MUE k receives on the nth RB, which is given by:

$$I_{k,n}^{m_i} = \sum_{i'=1,i'\neq i}^{M}\left(\sum_{k'\in\mathcal{U}_{m_{i'}}}\alpha_{k',n}^{m_{i'}}\right)\cdot P_m G_{k,n}^{m_{i'}} + \sum_{j=1}^{P}\left(\sum_{l\in\mathcal{U}_{p_j}}\alpha_{l,n}^{p_j}\right)\cdot P_p G_{k,n}^{p_j} \quad (3)$$

Note that the first term in the right hand side of (3) is the intra-layer interference from the other MBSs, where $\mathcal{U}_{m_{i'}}$ and $\mathcal{U}_{p_j}$ are the UE sets of the according MBS and PBS, respectively, $\alpha_{k',n}^{m_{i'}} \in \{0,1\}$ is a scheduling indicator which denotes that when MUE k' is associated with MBS $m_{i'}$, if the nth RB is allocated to MUE k', then $\alpha_{k',n}^{m_{i'}}$ is 1, otherwise 0. Since in LTE, every RB in a certain transmission time interval can only allocated to one UE, as for MBS $m_{i'}$, we have $\sum_{k'\in\mathcal{U}_{m_{i'}}} \alpha_{k',n}^{m_{i'}} \leq 1$. When the nth RB of MBS $m_{i'}$ is occupied by UE, MBS $m_{i'}$ will cause interference to MUE k, otherwise no interference. The second term in the right hand side of (3) is the inter-layer interference from all PBSs and detailed explanation is similar as the abovementioned.

Similarly, assume that PUE l is served by PBS p_j, and $\mathcal{I}_{\mathrm{RB},l}$ is the set of RBs allocated from PBS p_j to PUE l, then the rate of PUE l can be expressed as:

$$R_l^{p_j} = \sum_{n\in\mathcal{I}_{\mathrm{RB},l}} B_{\mathrm{RB}}\log_2\left(1+\mathrm{SINR}_{l,n}^{p_j}\right) \tag{4}$$

where $\mathrm{SINR}_{l,n}^{p_j}$ can be expressed as:

$$\mathrm{SINR}_{l,n}^{p_j} = \frac{P_p G_{l,n}^{p_j}}{I_{l,n}^{p_j}+\sigma_n^2} \tag{5}$$

and I_{i,k_p} can be expressed as:

$$I_{l,n}^{p_j} = \sum_{i=1}^{M}\left(\sum_{k\in\mathcal{U}_{m_i}}\alpha_{k,n}^{m_i}\right)\cdot P_m G_{l,n}^{m_i} + \sum_{j'=1,j'\neq j}^{P}\left(\sum_{l'\in\mathcal{U}_{p_{j'}}}\alpha_{l',n}^{p_{j'}}\right)\cdot P_p G_{l,n}^{p_{j'}} \tag{6}$$

2.2 Daily network traffic profile

We assume that the daily traffic profile of the whole Hetnets is the same and repeats periodically, which can be approximated by a sinusoidal-like periodic behavior as follows [27]:

$$\lambda\left(t\right) = \frac{A}{2^b}\left[1+\sin\left(\frac{\pi t}{12}+B\pi\right)\right]^b + C \tag{7}$$

where A denotes the parameter that controls the amplitude of the traffic profile, k denotes the minimum traffic intensity in the network, $b \in \{1,3\}$ is used to modulate the gradient of traffic profile curve (note that with $b = 3$ the curve has steeper slope, and the average traffic is lower), and B means the phase of traffic curve which regulate the position of the peak. Most of the traffic can be simulated through this formula. In this paper, the service arrival is modeled as a Poisson process with intensity $\lambda\left(t\right)$. In (7), we take $A = 19, b = 1, B = -\frac{11}{12}, C = 1$ and draw the traffic curve as shown in Figure 2, assuming that the network has statistic traffic information. This periodic sinusoidal traffic profile in realistic scenario has been provided in [11], which proves to be persuasive in investigating the performance of our following algorithms. Once a service arrives, a UE with a minimum rate requirement is located with certain probability in the network region, and this will be explained later. All UEs remain stationary until the transmission terminates. The transmission duration of each UE follows exponential distribution with mean $1/\mu = 180$s. According to the Little law [28], during the peak time, there are about $\lambda/\mu = 3,600$ UEs in the network.

3 Dynamic PBS switching off algorithms

In this section, we propose two dynamic PBS switching off algorithms for energy saving, which make use of the spatial and temporal traffic load fluctuation in Hetnets. The two proposed algorithms are both centralized which can guarantee a considerable energy saving performance; hence, we assume that there is a centralized controller unit (CCU) that carries out the algorithms. Since the traffic load fluctuates over time, the number of active PBSs should track this fluctuation to make a trade-off between saving energy and satisfying the UEs' service requirement. We divide the traffic period T_{total} into Z equal time intervals T with $Z + 1$ time spots as showed in Figure 3. At each time spot $t^{(i)}$, $i = 0, \ldots, Z$, the CCU carries out the proposed PBS switching off algorithm to determine all PBSs' working mode based on the current network information (the execution time of the algorithm is ignored), and all the reconfigured PBSs' mode keeps constant during the following time interval T. Note that our proposed algorithms should reckon for UEs' service requirement in the current time and the subsequent arriving UEs' service requirement (i.e., guarantee a tolerable blocking probability). The PBSs which can be switched off adjust their mode to sleeping state; the other PBSs remain active to provide service to the current and subsequent arriving UEs in the Hetnets during the time interval T. Note that time interval T which keeps equal just simplifies the resource management of the network. While in realistic scenario, the length of T can vary with $\lambda\left(t\right)$ so as to better track the network traffic variation.

At each time spot $t^{(i)}$, $i = 0, \ldots, Z$, the CCU collects the network information about UE-BS association and RB occupation, based on which the utility of each PBS can then be calculated. The utility is used as a metric which determines whether a certain PBS should be switched off or kept on. The utilities of the current active PBSs are iteratively calculated to keep up to date during the implementation of the proposed algorithms.

3.1 UE-BS association procedure and resource allocation scheme

On the expanded region (ER) of picocells in Hetnets, the power unbalance between DL and UL leads to a mismatch between the DL and UL handover boundaries. Therefore, associating a UE to the BS which provides the strongest DL reference signal receiving power (DL RSRP) may not always be the best strategy and is not an efficient way of network resource usage. Cell range expansion (CRE) is an alternative cell association that has been widely discussed in literature [29]. In CRE, generally, a positive bias is added to the DL RSRPs of PBSs pilot signals at UEs to increase PBSs' DL coverage footprints, thus compensating for the DL/UL mismatch. Although CRE allows more UEs to be associated with PBSs, the UEs in the expended region of

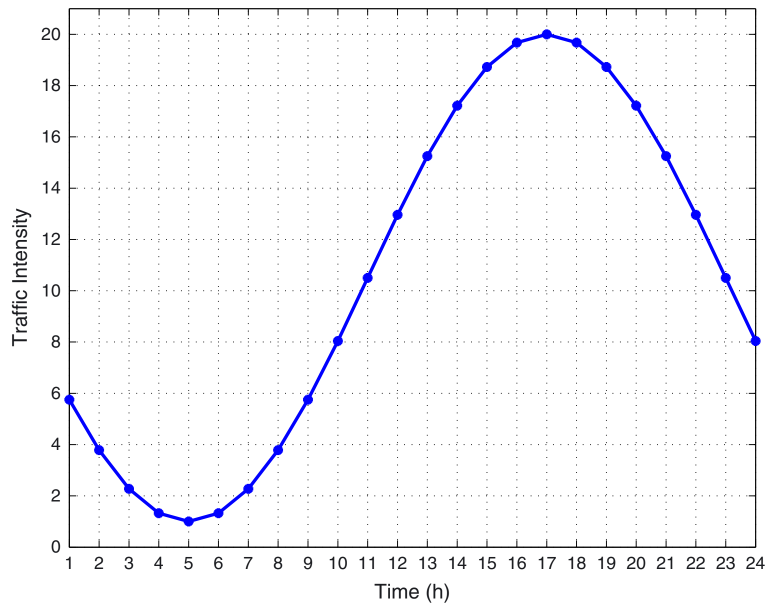

Figure 2 Average daily network traffic variations of UE service arrival rate.

PBSs without using any intercell interference coordination (ICIC) scheme will suffer severe DL cross-layer interference. Therefore, such victim CRE UEs are classified into the protected UE group. In order to improve the service quality of these UEs, a level of radio resource management coordination between the MBSs and the overlaid PBSs is needed. When a new UE enters the Hetnets, the UE-BS association process strives to balance the load of the network, aiming at lowering the network blocking probability. Therefore, the new UE is prone to choose these BSs with relatively lower load and better channel condition (i.e., larger RSRP). We then combine the conventional RSRP association method with CRE and ICIC technique. The path loss-based BS selection procedure is adopted in this paper to realize CRE, namely, associating UEs to BSs with the lowest path loss. If the new UE cannot find a proper MBS or PBS to serve its required rate, then it is blocked. Particularly, a new UE chooses a candidate MBS and a candidate PBS (if there exists one) from the potential serving BSs with the strongest RSRP, respectively, which are denoted as $P_{r,m}$ and $P_{r,p}$. Besides, the corresponding path losses are PL_m and PL_p. When $PL_m < PL_p$, the UE tries to be associated with the candidate MBS. While $PL_m \geq PL_p$, the UE tries to be associated with the candidate PBS. More particularly, if $P_{r,m} < P_{r,p}$, the UE is located in the picocell normal coverage area; otherwise, the UE is located in the picocell extended coverage area. Such association process iteratively continues until the association succeeds or all the candidate BSs cannot serve the UE. The candidate MBS and PBS which cannot afford the required rate will be eliminated from the potential serving cells in the next iteration.

In terms of ICIC technique, practically, the MBSs intelligently stop using or lower transmission power in the spectral and/or temporal resources allocated by PBSs to their CRE UEs [30]. The idea in [30] was that when a

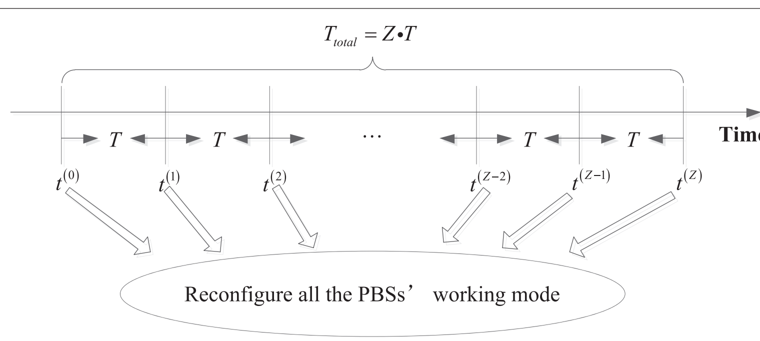

Figure 3 System operation over time.

UE enters or stays within the ER of a PBS, the PBS will inform the MBS about the set of RBs allocated to this CRE UE, and then the MBS will lower its transmission power in the specified RBs so that a desired DL QoS in terms of SINR is guaranteed to this CRE UE. Similarly, a new dynamic frequency distribution strategy was proposed in [31], where the MBSs restrict the transmission power on a part of frequency resources reserved only for CRE UEs operation. The ratio of the bandwidth of the reserved band to that of the whole band was established according to the number of UEs connected to the MBS before CRE and the recent number of CRE UEs. Noh, 2012 [32] proposed a distributed and dynamic cooperative scheme of silencing MBS transmission over part of the system bandwidth, where the silencing fraction of the whole bandwidth was determined online.

Provided that the DL transmission scenario in this paper is based on OFDMA, we adopt the MBS-PBS frequency domain ICIC technique in [30] to protect the victim CRE UEs through dynamic detection. The authors in [30] proposed a method to calculate the maximum power that the MBS can apply in each RB used by the CRE UEs to provide them with the desired QoS. However, in this paper, we focus on energy saving but not the research on ICIC techniques. Therefore, we employ the following reduced scheme for simplicity. When a UE enters the ER of a PBS, the PBS will inform the MBS about the set of RBs allocated to this CRE UE, and then the MBS will silence its transmission in the specified RBs so that a desired DL QoS in terms of required service rate is guaranteed to this CRE UE. When the service for this CRE UE ends, the occupied RBs of the PBS and silenced RBs of the MBS are released and can again be used for subsequent transmission. Cross-layer communication among MBSs and PBSs can be supported through the operator's backhaul and could be periodic or event triggered.

3.2 Utility function for PBS

Contrasting with MBSs, the traffic load of PBSs has more significant fluctuations in space and time due to a number of factors such as user mobility and behavior, as well as the fact that each PBS supports fewer simultaneous UEs. Therefore, when designing the rule of determining which PBS should first be tested to switch off, more factors should be included in the rule, instead of just the PBS traffic load as widely used in literature or the average distance between UEs and PBS as used in [16]. After having UEs in the Hetnets associated with their respective BSs and performing resource allocation as described in Section 3.1, a utility function for each PBS can then be computed. In this paper, the proposed utility function depends on the total rate of served PUEs, the PBS's traffic load, the number of served PUEs, the number of blocked PUEs, and the received interference signal strength from the nearby cells. We do not consider the average distance between a PBS and its associated PUEs, for the accurate positions of PUEs are hard to acquire in realistic scenario. The utility function is expressed as follows:

$$U_p = \frac{\alpha \cdot \frac{R_p}{R_{\max}} + \beta \cdot \frac{L_p}{L_{\max}} + \gamma \frac{W_p}{W_{\max}}}{I_p / I_{\max}} \tag{8}$$

with

$$L_p = \frac{\text{RBs utilized by the PUEs}}{\text{Total RBs available to the PBS}} \tag{9}$$

$$W_p = N_{\text{served},p} * \exp\left(P_{\text{blocked,th}} - \frac{N_{\text{blocked},p}}{N_{\text{served},p} + N_{\text{blocked},p}}\right) \tag{10}$$

where U_p is the utility of PBS p, α, β, γ are the weighting coefficients of the considered factors and satisfy $\alpha + \beta + \gamma = 1$, R_p is the total rate of all PUEs in PBS p, L_p is the traffic load on PBS p defined as [33] (the traffic load on MBS is defined likewise), W_p is the term that considers the number of served UEs $N_{\text{served},p}$ and the number of blocked UEs $N_{\text{blocked},p}$ in PBS p, which takes a similar form to the utility function adopted in [34]. In addition, $P_{\text{blocked},th} = 1\%$ is the blocking probability threshold, indicating the blocking probability that is tolerated in the network. The term I_p is the received interference signal strength from the nearby BSs. Specifically, I_p is once calculated over all RBs assuming that all RBs of nearby BSs are used to account for the worst interference case and then stored for the following utility calculation. The four terms with subscript max are the maximum values of the corresponding terms respectively and are used for normalization. The abovementioned parameters needed to calculate the utility can be obtained from the network information collected by the CCU.

In (8), the utility value is monotonously increasing with the three terms on the numerator, when the sleeping algorithm tries to switch off PBSs, we have to consider the load of the cell (i.e., RB occupation ratio and the number of served UEs) and the throughput of cell, in the interest of switching off those PBSs with relatively lower load and lower cell throughput, and thus trade-off between energy saving and network capacity ensuring. The explanation of W_p is similar to that in reference [34]. The number of the served PUEs, $N_{\text{served},p}$, determines the increase of the PBS utility as long as blocking probability threshold is not exceeded. When the number of blocked PUEs increases, the exponential term in W_p decreases. When the blocking probability threshold is exceeded, the term in the exponential becomes negative, and W_p also decreases. Besides, the interference term on the denominator is used to give more priority to those PBSs which receive relatively strong interference signal from nearby MBSs and other PBSs to be switched off.

It should be noted that the utility in (8) is selected as such because it satisfies the description presented above and integrates the characteristics of Hetnets. Nevertheless, other utility functions and metrics might be used to achieve the same purpose. The algorithms proposed in the next two sections are independent from the utility selected and can be implemented with any appropriate utility.

3.3 Proposed dynamic switching off PBS algorithms

In this section, we provide both heuristic and progressive solutions to dynamically switch off the redundant PBSs for energy saving, which uses the utility function provided in Section 3.2. The effect of the energy-saving algorithms is represented in terms of the number of active PBSs, yet the quantity of the energy saved considering practical power consumption model is an interesting topic for future study. Similar to the switching off algorithm proposed in [16], we first propose the HDSO algorithm, which tests all PBSs in the network one by one to see whether they can be switched off. Unlike the algorithm in [16] which just stops when one BS fails to transfer its UEs due to the unavailable resource the nearby BSs can provide, we consider that there still exists the potential of switching off other active PBSs when just one PBS fails to be switched off. The detailed description of the HDSO algorithm is summarized in Algorithm 1. When the algorithm terminates, the PBSs of set $\mathcal{B}_{\text{off}}$ are switched off, and the remaining PBSs remain active until the next time spot of reconfiguration.

Algorithm 1 Heuristic dynamic switching off PBS.

1: Initialize the set of all PBSs as $\mathcal{B}$, and the set of switched-off PBSs as $\mathcal{B}_{\text{off}} = \emptyset$
2: **while** $\mathcal{B} \neq \emptyset$ **do**
3: Calculate the utility function values of PBSs in set $\mathcal{B}$ based on service state information collected by the CCU, assume p_{off} as the PBS with the minimum utility value
4: Try to transfer the PUEs of PBS p_{off} to the nearby MBSs or PBSs
5: **if** all PUEs of PBS p_{off} can be transferred **then**
6: $\mathcal{B}_{\text{off}} = \mathcal{B}_{\text{off}} + \{p_{\text{off}}\}$
7: Transfer PUEs of PBS p_{off} to the nearby MBSs and PBSs, update the UE association information of the network and the SINR of UEs
8: **else**
9: PBS p_{off} remains on-state and the PUE association state remains the same as before
10: **end if**
11: $\mathcal{B} = \mathcal{B} - \{p_{off}\}$
12: **end while**

The above HDSO algorithm greedily tests all PBSs to be switched off to see the overall energy energy potential and can commendably achieve the energy saving target. In addition, such brute-force approach is straightforward and easy to implement. However, this algorithm needs to traverse all PBSs and all PUEs in the network at any traffic condition, and obviously, such brute-force approach can be inefficient. Note that some PUEs can be transferred several times like playing a football till they finally settle. Besides, the complexity of this algorithm is proportional to the number of the PBSs and the PUEs, which can be very high and causes considerable network signaling overhead. It was mentioned in [20] that the sleep mode mechanism can be used to improve Hetnets energy efficiency at low and medium traffic load only, which indicates that most of PBSs cannot be switched off during peak traffic (this is also verified by our simulation); hence, most part of the traversal is unnecessary. Based on the utility given in (8), the probability that a PBS with relatively large utility can be switched off is smaller than that of a PBS with relatively small utility, which is intuitive and can be easily testified through simulation.

To overcome the shortcoming of the heuristic algorithm, we propose the following expeditious and intelligent PDSO algorithm, which is carried out in a round by round manner and presented in Algorithm 2. This algorithm is based on the utility function to classify the current active PBSs into two groups (one group of PBSs with smaller utility values than the other group), then a switching-off testing process is carried out to test all PBSs in the group with relatively small utilities to be switched off in one round. Upon the termination of this round, the algorithm determines whether to continue to classify the remaining active PBSs and launch another round of testing based on the switching-off result of the previous round. In the same way, the PDSO algorithm tests PBSs as far as possible, and does't stop when one PBS cannot transfer its PUEs successfully. Novelly, PDSO terminates when the predefined condition is met, and the condition can reflect the remaining energy saving potential in the current network. When the algorithm terminates, the PBSs of set $\mathcal{D}$ are switched off, and the remaining PBSs remain active until the next time spot of reconfiguration.

3.4 UE-BS re-association and RB re-allocation during the transferring process

When new UEs arrive in the Hetnets or current PUEs need to be transferred to other BSs, these UEs will select their serving BSs. In Section 3.3, at each time spot $t^{(i)}, i = 0, \ldots, Z$, the principle to decide whether to switch off a PBS is that if all the PUEs associated with this PBS can be successfully transferred to the nearby BSs. The detailed transferring process is elaborated in this section.

Algorithm 2 Progressive dynamic switching off PBS.

1: Allow the PBSs without any PUE association to be switched off, initialize the set of switched-off PBSs as $\mathcal{D}$, the remaining PBSs constitute the set of candidate switched-on PBSs
2: According to the network state information of the current active PBSs collected by the CCU, calculate the utility function values of these PBSs and sort the values incrementally
3: Denote P percentage of the PBSs with relatively small utility values as can testing set $\mathcal{T}$ (the number of elements is N), the remaining PBSs with relatively large utility values constitute the candidate switched-on set $\mathcal{S}$, and initialize the number of switched-off PBSs in this testing iteration as $\text{off}_{\text{count}} = 0$
4: **while** $\mathcal{T} \neq \emptyset$ **do**
5:     Try to transfer the PUEs of PBS with the maximum utility in set $\mathcal{T}$ (denoted as M) to the nearby MBSs or PBSs
6:     **if** all PUEs of PBS M can be transferred **then**
7:         $\mathcal{T} = \mathcal{T} - \{M\}, \mathcal{D} = \mathcal{D} + \{M\}, \text{off}_{\text{count}} = \text{off}_{\text{count}} + 1$
8:         PBS M can be switched off, transfer PUEs of PBS M to the nearby MBSs or PBSs, update the UE association information of the network and the SINR of UEs
9:     **else**
10:         $\mathcal{T} = \mathcal{T} - \{M\}, \mathcal{S} = \mathcal{S} + \{M\}$
11:         PBS M remains on-state and PUE association state remains the same as before, update the elements of sets
12:     **end if**
13: **end while**
14: **if** $\text{off}_{\text{count}}/\text{N} \geq \delta$ **then**
15:     ruturn to step 2
16: **end if**

Assuming the to be transferred PUE as u in PBS p , we define the set of potential acceptor MBSs as $\mathcal{S}_m$, and the set of potential acceptor PBSs as $\mathcal{S}_p$. Particularly, we first consider to transfer PUE u to the potential acceptor MBSs, if these MBSs cannot accept the PUE, then the potential acceptor PBSs will attempt to accept it. We initialize the transmission rate as $R_s = 0$ and the allocated RB set as $\mathcal{R} = \emptyset$. During the transferring process, PUE u chooses the MBS with better channel condition $\bar{G}_u^{m_i}$ (i.e., the average channel gain between MBS m_i and PUE u) and lower load L_{m_i} as well, aiming at balancing the network load. In order to reduce the blocking probability of the subsequent new UEs, the acceptor MBS needs to reserve $\varepsilon_m \cdot N_{\text{RB}}$ RBs for the new UEs, where $\varepsilon_m \in (0, 1]$ is the bandwidth protection margin of each MBS. This leaves some spare bandwidth at each MBS to decrease the blocking probability of subsequent UEs entering the network and larger ε_m means less PBSs being switched off. Therefore, the number of available RBs of each acceptor MBS is restricted to $(1 - \varepsilon_m) \cdot N_{RB}$ in the execution phase of the PUE transfering algorithm. If all the potential acceptor MBSs cannot accept the PUE, the potential acceptor PBSs will be considered. PUE u chooses the PBS with better channel condition $\bar{G}_u^{p_j}$ and higher load L_{p_j}, which aims at switching off as many PBSs as possible and avoiding the possible repeated transferring of a single PUE. For the same purpose, the acceptor PBS needs to reserve $\varepsilon_p \in (0, 1]$ bandwidth for the subsequent new UEs. The bandwidth protection margin is consistent with the idea of system load threshold used in reference [25]. The detailed transferring process is presented in Algorithm 3. When the process ends, if the candidate BS set $\mathcal{S}_m \neq \emptyset$ or $\mathcal{S}_p \neq \emptyset$, PUE u successfully reselects a proper serving BS and the RBs in $\mathcal{R}$ are allocated to it. Otherwise, there is no BS that has enough resource for PUE u. That is PUE u can't be transferred to other BSs, and the PBS p will stay active in the following time interval.

Algorithm 3 Transferring PUE.

Initialize the to be transferred PUE as u, the set of potential acceptor MBSs as $\mathcal{S}_m$, and the set of potential acceptor PBSs as $\mathcal{S}_p$
while $\mathcal{S}_m \neq \emptyset$ **do**
    $i^* = \arg \max_{m_i \in \mathcal{S}_m} \bar{G}_u^{m_i} / L_{m_i}, R_s = 0, \mathcal{R} = \emptyset$
    while $R_s < R_{\min}$ and $\sum_{k \in \mathcal{U}_{m_{i^*}}} \sum_{n \in \mathcal{N}} \alpha_{k,n}^{m_{i^*}} < N_{\text{RB}}$ **do**
        $n^* = \arg \max_{n \in \mathcal{N}} \left(1 - \sum_{k \in \mathcal{U}_{m_{i^*}}} \alpha_{k,n}^{m_{i^*}} \right) G_{u,n}^{m_{i^*}}$
        $\mathcal{R} = \mathcal{R} + \{n^*\}, R_s = R_s + B_{\text{RB}} \log_2 \left(1 + \text{SINR}_{u,n^*}^{m_{i^*}} \right)$
    end while
    if $R_s >= R_{\min}$ and $L_{m_{i^*}} \leq (1 - \varepsilon_m)$ **then**
        break
    else
        $\mathcal{S}_m = \mathcal{S}_m - \{m_{i^*}\}$
    end if
end while
if $\mathcal{S}_m = \emptyset$ **then**
    while $\mathcal{S}_p \neq \emptyset$ **do**
        $j^* = \arg \max_{p_j \in \mathcal{S}_p} \bar{G}_u^{p_j} / L_{p_j}, R_s = 0, \mathcal{R} = \emptyset$
        while $\sum_{l \in \mathcal{U}_{p_{j^*}}} \sum_{n \in \mathcal{N}} \alpha_{l,n}^{p_{j^*}} < N_{RB}$ **do**
            $n^* = \arg \max_{n \in \mathcal{N}} \left(1 - \sum_{l \in \mathcal{U}_{p_{j^*}}} \alpha_{l,n}^{p_{j^*}} \right) G_{u,n}^{p_{j^*}}$
            $\mathcal{R} = \mathcal{R} + \{n^*\}, R_s = R_s + B_{\text{RB}} \log_2 \left(1 + \text{SINR}_{u,n^*}^{p_{j^*}} \right)$
        end while
        if $R_s >= R_{\min}$ and $L_{p_{j^*}} \leq \left(1 - \varepsilon_p \right)$ **then**
            break
        else
            $\mathcal{S}_p = \mathcal{S}_p - \left\{ p_{j^*} \right\}$
        end if
    end while
end if

After the working modes of all PBSs are determined, all MBSs and active PBSs need not to reserve any frequency resource when serving the newly arriving UEs during the time interval, the UE association algorithm is similar to the UE transferring algorithm expect two points. Firstly, UEs choose the serving BSs with higher channel quality and lower traffic load as well, so as to balance the traffic load of the network to reduce system blocking probability. Secondly, there is no protection margin in UE association algorithm (i.e., $\varepsilon_m = 0$, $\varepsilon_p = 0$).

4 Performance evaluation

In this section, extensive numeric simulation is done to validate our analysis and evaluate the performance of our proposed algorithms in terms of energy saving and blocking probability. Next, we present the simulation scenario, the results and analysis of our algorithms.

4.1 Simulation scenario

The evaluation is performed through numerical simulations, and Table 1 gives the major simulation parameters [35]. We employ a 19-hexagonal macrocell model with three sectors per MBS, denoting the antenna downtilt of MBS, the horizontal and vertical lobe width of MBS as φ, $\phi_{3\text{ dB}}$ and $\theta_{3\text{ dB}}$, respectively. We assume that four PBSs are randomly deployed within each sector with a uniform distribution. We consider the cluster distribution for UE deployment configuration, where two thirds of the UEs are randomly located within 40 m from the PBSs with a uniform distribution (i.e., forming a hotspot area), and the location of the remaining one third of the UEs are assigned randomly across the network with a uniform distribution. It should be noted that the UE-BS association is decided in a centralized manner in our simulation. However, in realistic network, this process happens in a distributed way and involves the UE and the potential serving BSs in the vicinity of this UE. Wraparound is applied to eliminate the network edge effect and generate accurate simulation results.

As for PDSO, we generally take the threshold and the percentage as $\delta = 0.5, P = 0.5$ for simplicity, which is referred to as fixed PDSO (F-PDSO). Considering that adaptive thresholds and percentages matching the traffic variation can further enhance the efficiency of the algorithm in the realistic scenario; hence, we can have the adaptive version of PDSO (A-PDSO). In the following simulation, except for the one with respect to (w.r.t.) the traffic profile (Figures 4, 5, 6), we take PDSO as F-PDSO. For the adaptive version in the following simulation, we correspondingly take both the thresholds and the percentages in the same manner as the the ratios of switched-off PBSs when the F-PDSO algorithm is applied w.r.t. all traffic intensities. Note that these thresholds and percentages can be obtained by other methods in reality, such as the historical traffic trace about the network resource utilization as in [11] and the anticipative necessary network resource based on some traffic prediction techniques. We take the weighting coefficients in utility function as $\alpha = 0.3, \beta = 0.6, \gamma = 0.1$ by considering the importance of the corresponding factors.

4.2 Simulation results

In the following figures, we provide the results of our algorithms and the analysis of the performance using a realistic daily traffic profile.

First, we investigate the performance of the proposed algorithms w.r.t. varying traffic intensities of a single day. Figure 4 shows that the number of active PBSs tracks the variation of the statistic traffic profile very well. Figure 5 shows the average blocking probability of the proposed

Table 1 Simulation parameters

Parameters	Macrocell setting	Picocell setting
Cellular layout	19 cell sites, 3 sectors per site	4 picocells randomly distributed per sector
System Bandwidth	10 MHz	10 MHz
Number of RBs	50	50
Path loss model	$L = 128.1 + 37.6 \log 10(R)$, R in km	$L = 140.7 + 36.7 \log 10(R)$, R in km
Shadowing standard deviation	8 dB	10 dB
Antenna pattern	3D pattern	Omni-directional
Antenna gain	14 dBi	5 dBi
$\phi_{3\text{ dB}}, \theta_{3\text{ dB}}$	70°, 10°	N/A
Thermal noise density	-174 dBm/Hz	-174 dBm/Hz
BS transmission power	46 dBm	30 dBm
UE distribution radius	289 m	40 m
UE's rate requirement	400 Kbps	400 Kbps
Minimum distance requirement	Macro-pico: 75 m, Macro-UE: 35 m	Pico-pico: 40 m, Pico-UE: 10 m

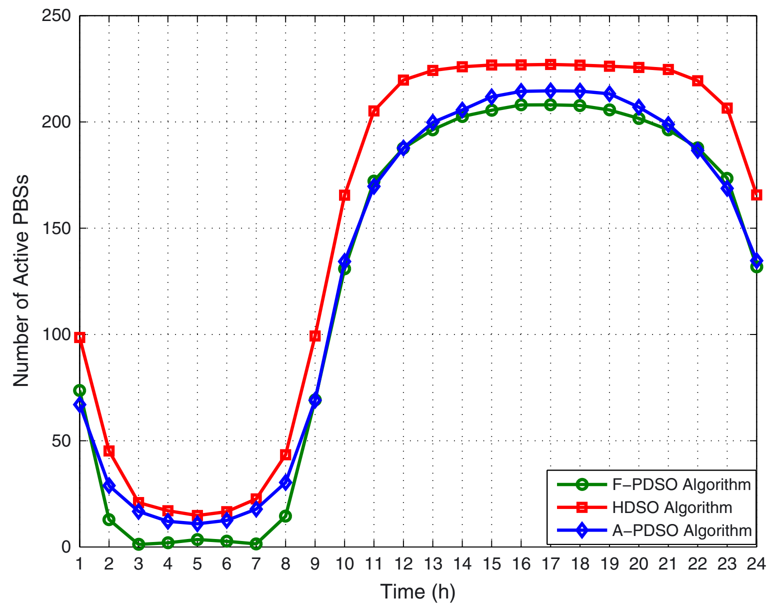

Figure 4 The number of active PBSs versus time, $\varepsilon_{\mathrm{m}} = 0.6$, $\varepsilon_{\mathrm{p}} = 0.9$, $\varphi = 13°$, $T = 3$ min.

algorithms. The F-PDSO algorithm can switch off averagely 25 more PBSs than the HDSO algorithm, while having a slight impact on the network blocking probability which can be observed from Figure 5. The more significant energy-saving effect is explained through the following analysis. Since the F-PDSO algorithm classifies the current active PBSs into two different groups and handles each group on its own characteristics, which guarantees that PBSs with similar characteristics can be handled (i.e., switched on or off) fairly, and gives more opportunity to some PBSs with relatively small utility to be switched off. However, the HDSO algorithm only tests PBSs one by one, there can be close-by PBSs all with low load, some PBSs act as acceptor PBSs to take over the transferred UEs which can be far from the respective PBSs and occupy more RBs than before. This phenomenon can increase the originally low-loaded acceptors' load and make them unable to be switched off.

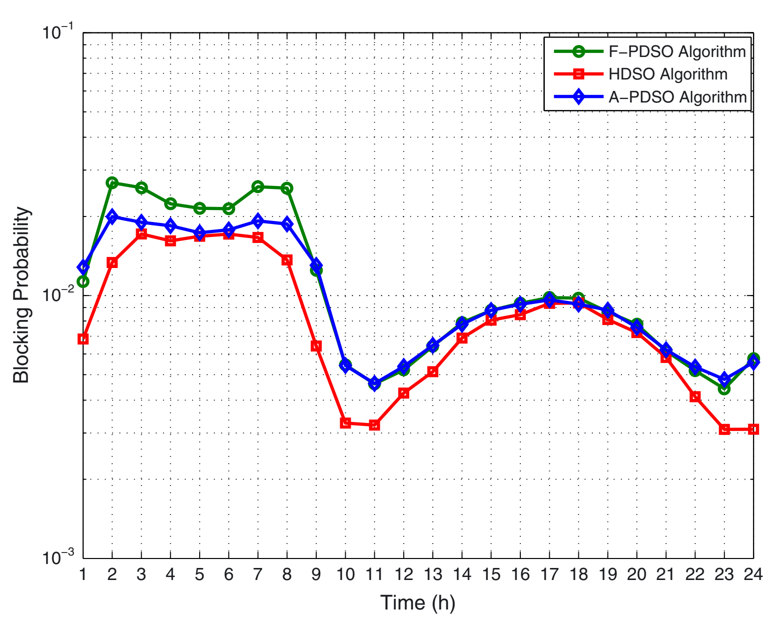

Figure 5 The blocking probability versus time, $\varepsilon_{\mathrm{m}} = 0.6$, $\varepsilon_{\mathrm{p}} = 0.9$, $\varphi = 13°$, $T = 3$ min.

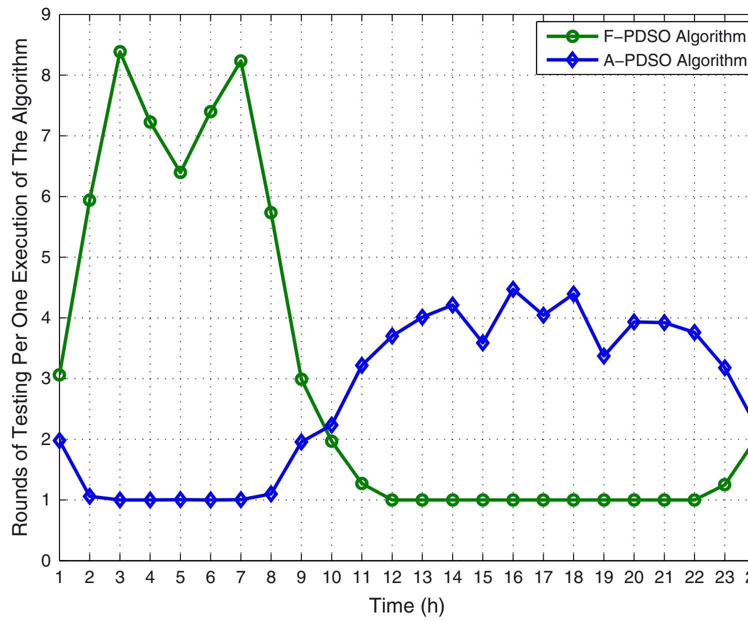

Figure 6 The rounds of testing per one execution of the proposed algorithms versus time, $\varepsilon_m = 0.6$, $\varepsilon_p = 0.9$, $\varphi = 13°$, $T = 3$ min.

Besides, the A-PDSO algorithm can keep more PBSs active than the fixed version during the bottom times and peak times. In order to explain this, we present the rounds of testing per one execution of the algorithms in Figure 6. According to the simulation result of F-PDSO algorithm, the percentages of switched-off PBSs during these two times are about 99% and 9%, respectively. In the first case, the A-PDSO tests 99% of the current active PBSs in one round aggressively, while the result of switching off may not be that enough to launch another round of testing (i.e., the termination condition of the A-PDSO algorithm can be more easily met, compared to that of F-PDSO). This is verified in Figure 6, as we can observe that during low traffic times, the A-PDSO only needs one round of testing. Therefore, more PBSs are left active at the end of the A-PDSO algorithm. However, the F-PDSO algorithm launches about seven rounds of testing to further switch off more PBSs. In the second case, the A-PDSO algorithm tests 9% of the current active PBSs in one round of testing conservatively. From Figure 6, we can see that about four rounds of testing are executed; hence, the A-PDSO algorithm can test at most 40% of the total PBSs. However, the F-PDSO algorithm just launches one round of testing and test 50% of the total PBSs, which is prone to switch off more PBSs. Therefore, as the process of the A-PDSO algorithm moves on, one round of testing only considers a small percentage of the current active PBSs, which gives more opportunity for the remaining PBSs to keep active. Now we can conclude that in extreme cases of the network traffic (i.e., bottom times and peak times), the A-PDSO algorithm is prone to keep more PBSs active so as to guarantee the QoS of the network. Moreover, the thresholds and the percentages used in PDSO have a significant influence on the performance of the algorithm.

In Figure 5, from 5:00 to 17:00, the same variation trend of blocking probability can be observed about the proposed algorithms, and it is not monotonous with the traffic variation. The average blocking probability is around 1%, which can satisfy the blocking probability constrain of general communication service. It can be observed that from 5:00 to 10:00, the blocking probability is decreasing with the incremental traffic, and the rationale can be explained as follows. When the traffic increases from the trough to the moderate intensity, more PBSs need to be active to provide service; thus, the resource is sufficient to serve the UEs, and the probability that an UE cannot find an appropriate BS to serve is relatively small. Therefore, the overall blocking probability is decreasing. But when the traffic continues to increase to the peak, the blocking probability is increasing contrarily even with more active PBSs. This is due to that the interference among UEs is becoming significant with more UEs served in the network, which results in decreasing the UEs' spectrum efficiency and makes BSs provide more RBs to UEs to satisfy their rate requirement. However, the frequency resource is limited, some newly arriving UEs needing more RBs can be blocked, and the increasing active PBSs cannot compensate the decreasing QoS of the network due to the decreasing UEs' spectrum efficiency. Hence, the blocking probability is increasing with the traffic curve until the peak. The explanation to the variation trend of the other two periods is the same as above owing to the

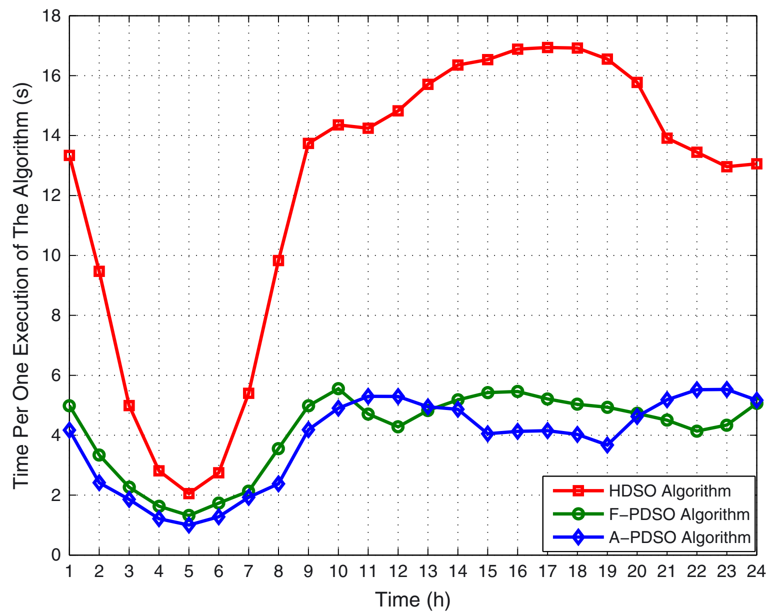

Figure 7 The time per one execution of the proposed algorithms versus time, $\varepsilon_{\mathrm{m}} = 0.6$, $\varepsilon_{\mathrm{p}} = 0.9$, $\varphi = 13°$, $T = 3$ min.

symmetrical variation of the sinusoidal-like traffic curve. Moreover, the average blocking probability during the peak traffic is lower than that during low traffic due to the densification of active PBSs, which provides sufficient network capacity. However, during low traffic, some UEs enter the macrocells' edge where no PBS exists cannot obtain the minimum service rate, and thus become easily blocked.

Figure 7 depicts the remarkable difference of the time per one execution of the proposed algorithms. Averagely, the PDSO algorithm takes at most one third time of the HDSO algorithm, which indicates that PDSO can reduce the time for the CCU making decision to a great extent and may also involve less network signaling overhead. Therefore, PDSO can quickly reconfigure the working mode of all PBSs to accommodate the traffic. Since

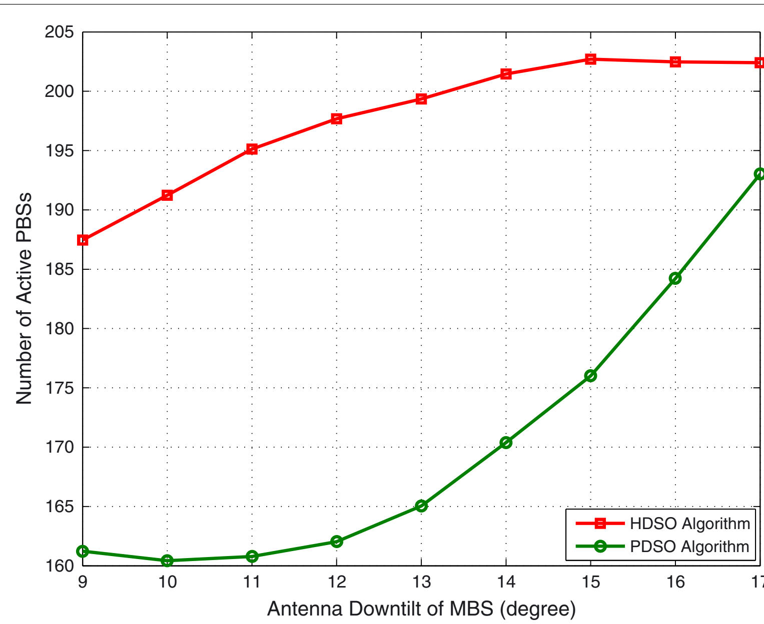

Figure 8 The number of active PBSs versus antenna downtilt of the MBS, $\varepsilon_{\mathrm{m}} = 0.6$, $\varepsilon_{\mathrm{p}} = 0.9$, $\lambda\,(t) = 10$, $T = 3$ min.

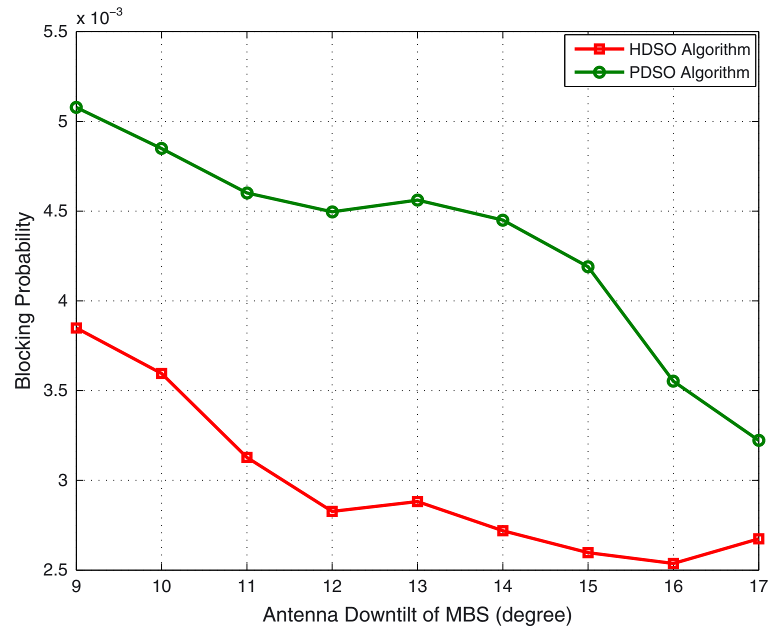

Figure 9 The blocking probability versus antenna downtilt of the MBS, $\varepsilon_m = 0.6$, $\varepsilon_p = 0.9$, $\lambda(t) = 10$, $T = 3$ min.

the PDSO algorithm avoids the unnecessary traversal of some PBSs and can adaptively decide when to stop the switching off process, a large amount of time can thus be saved. The phenomenon that the execution time of A-PDSO is slightly less than that of F-PDSO during most time reveals the adaptive version of PDSO algorithm can further quicken the decision-making of the CCU. As most PBSs without any association or being lightly loaded can be quickly judged whether they can be switched off or not during trough traffic, the difference between the PDSO and HDSO becomes the least. Note that the execution time is a simulation result and can vary among different simulation environments; however, the relative difference between the proposed two algorithms

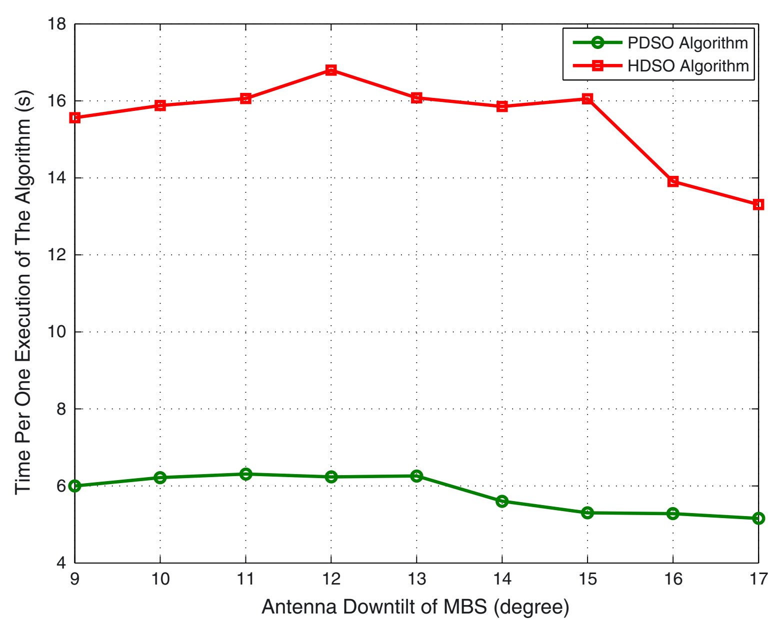

Figure 10 The time per one execution of the proposed algorithms versus antenna downtilt of the MBS, $\varepsilon_m = 0.6$, $\varepsilon_p = 0.9$, $\lambda(t) = 10$, $T = 3$ min.

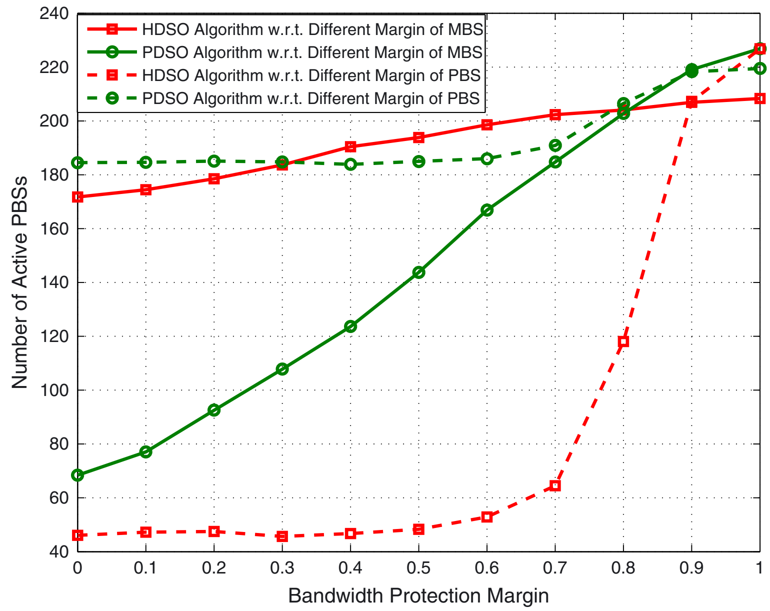

Figure 11 The number of active PBSs versus bandwidth protection margin, $\lambda(t) = 10$, $\varphi = 13°$, $T = 3$ min.

remains the same when they are simulated in the same environment.

Figures 8, 9, and 10 show the impact of antenna downtilt of the MBS on the performance of the proposed algorithms. It can be observed that the number of active PBSs increases with the downtilt, which is due to that when the MBSs reduce their coverage area, the emerging coverage holes need to be compensated by the nearby PBSs to provide constant QoS of the network. These PBSs' load increases, therefore they have less opportunity to be switched off. The difference should be noted is that the impact on the PDSO algorithm is relatively more significant than the HDSO algorithm, indicating that an appropriate smaller downtilt can lead to a better performance enhancement of PDSO algorithm compared to HDSO algorithm. Figure 9 shows the influence of the MBS antenna downtilt on the network blocking probability. From Figure 10, we can also observe that PDSO

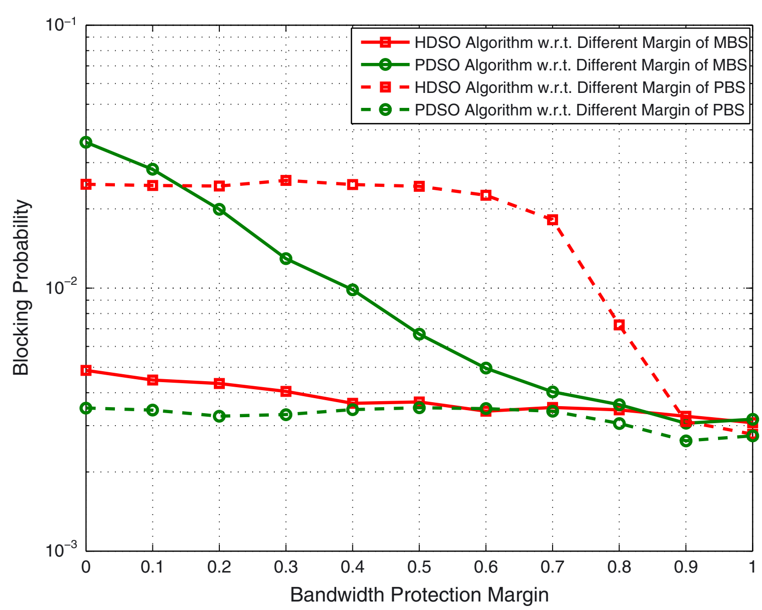

Figure 12 The blocking probability versus bandwidth protection margin, $\lambda(t) = 10$, $\varphi = 13°$, $T = 3$ min.

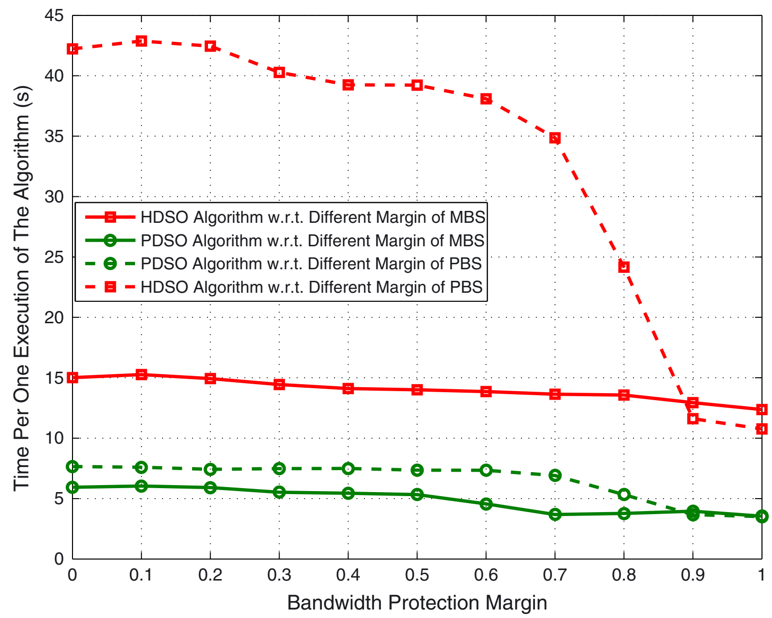

Figure 13 The time per one execution of the proposed algorithms versus bandwidth protection margin, $\lambda(t) = 10$, $\varphi = 13°$, $T = 3$ min.

algorithm exceeds HDSO algorithm in terms of algorithm complexity and that the downtilt has indistinctive impact on the execution time of each algorithm. As the downtilt increases, more PBSs are determined that their PUEs cannot be accepted by MBSs, which accelerates the execution of the two algorithms to some degree.

Figures 11, 12, and 13 show the impact of bandwidth protection margin of MBS and PBS on the performance of the proposed algorithms, respectively. As for the margin of MBS, the number of active PBSs of the two proposed algorithms both increases with the margin, and this is obvious. When the margin is low, the PDSO algorithm can switch off more PBSs and show more potential of saving energy while resulting in slight QoS deterioration observed from Figure 12. When $\varepsilon_{ma} \geq 0.4$, the PDSO algorithm can achieve the target that the blocking probability is maintained below 1%. In Figure 11, as the margin increases, the energy-saving difference

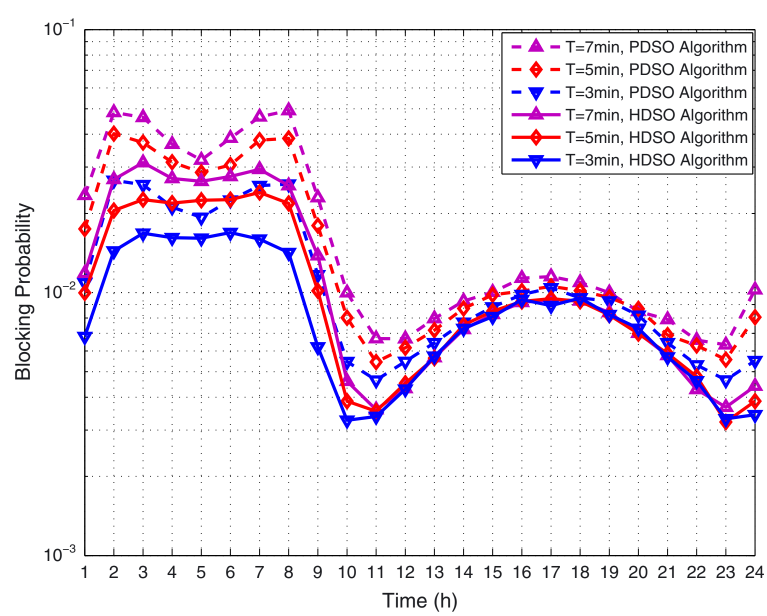

Figure 14 The blocking probability of the proposed algorithms under different time interval T, $\varepsilon_{\mathrm{m}} = 0.6$, $\varepsilon_{\mathrm{p}} = 0.9$, $\varphi = 13°$.

between the two algorithms reduces, and in extreme case, the energy-saving effects of the two algorithms become approximately the same. As for the margin of PBS, similar trends can be observed. However, the PDSO algorithm is more robust than the HDSO algorithm, and in extreme case, the effects of the two algorithms become approximately the same. An interesting phenomenon is that when $\varepsilon_{\text{pi}} < 0.9$, the HDSO algorithm can switch off more PBSs than the PDSO algorithm. This is due to the termination condition of PDSO algorithm which is $\text{off}_{\text{count}}/N < \delta$. For the minimum distance restriction among PBSs, the PUEs of the switched-off PBSs are more likely transfered to the nearby MBSs than the nearby PBSs. However,the available RBs that MBSs can provide are limited, hence, it's likely that the PUEs of the switched-off PBSs can't be transferred, and these PBSs need to keep active. The termination condition can be easily met as we set δ=0.5, and the PDSO algorithm terminates to test the remaining PBSs which may have the potential to be switched off. However, the HDSO algorithm tests all PBSs which exploit the full potential of energy saving, which is at the expense of needing much more execution time. When $\varepsilon_{\text{pi}} \geq 0.9$, the two algorithms show similar performance. The blocking probability is greatly related to the number of active PBSs, which can be observed from Figure 12. Figure 13 presents the time per one execution of the proposed algorithms. It can be observed that the time is decreasing with all bandwidth protection margins to some degree. Specifically, the time of the HDSO algorithm is much more sensitive to a different margin of PBS than the other three cases. From these results, we can see that the bandwidth protection margins have significant impacts on the performance of the proposed algorithms and should be tuned well in realistic scenarios.

Figure 14 shows the blocking probability of our proposed algorithms under different time interval T. We can see that the blocking probability varies in the same trend under different time intervals but increases with increasing time interval. The difference among different intervals is significant in the night zone of traffic profile, while unobvious during the peak traffic time. This is because most of the PBSs are switched off during the night zone leading to inadequate resource. During peak traffic time, what is important is that this result indicates that the sufficient resource for UEs benefiting from the densification of active PBSs provides more time for some PBSs to be switched off, while has slight influence on the blocking probability. The result implies that the proposed algorithms can track the variation of the traffic better when shortening the time interval of carrying out the algorithms, which yet increases the overhead of signaling due to the frequent reconfiguration. Hence, a tradeoff can again be made between the network signaling overhead and the QoS.

5 Conclusions

In this paper, two PBS switching off algorithms have been proposed to dynamically reconfigure PBSs' working mode for energy saving in Hetnets. With our algorithms, network energy consumption can be significantly reduced. The simulation results show that the number of active PBSs tracks the variation of traffic well while ensuring an acceptable blocking probability. Particularly, the PDSO algorithm can lead to a better effect of energy saving and slightly deteriorates the blocking probability. What is important is that the complexity of the PDSO algorithm is much lower than that of the HDSO algorithm, enabling the network respond to the traffic variation more promptly, which is demonstrated by the simulations that the execution time of the PDSO algorithm is at most one third of the HDSO algorithm. Additionally, simulations imply that the antenna downtilt of the MBS, the bandwidth protection margin of MBS and PBS, and the time interval T all have nonnegligible influence on the performance of our algorithms. As the downtilt increases, more PBSs remain active to guarantee the QoS of the network, which accelerates the execution of the algorithms to some degree. In terms of the bandwidth protection margin, the performance of the PDSO algorithm is more sensitive to the margin of MBS, and the performance of the HDSO algorithm is more sensitive to the margin of PBS, compared to the other two cases. The number of active PBSs of the algorithms increases with the two margins due to less available capacity to accept the transferred PUEs. The blocking probability increases with the time interval more significantly during low traffic times, indicating that the active PBSs set should change frequently in these times. The unobvious difference during peak traffic times indicates that sufficient resource for UEs benefiting from the densification of active PBSs provides more time for some PBSs to be switched off. Hence there exist tradeoffs that can be made to achieve a preferable system performance and energy saving.

Competing interests
The authors declare that they have no competing interests.

Acknowledgements
This work was supported by the National Natural Science Foundation of China under Grant 61222102, the National Basic Research Program of China (973 Program 2012CB316004), Research Project of Jiangsu Province under Grant BK2012021, the International Science & Technology Cooperation Program of China under Grant 2014DFT10300, the National Science and Technology Major Project of China under Grant 2013ZX03001032-004, and NEC Research Fund.

Author details
[1]School of Information Science and Engineering, Southeast University, No. 2 Sipailou, Nanjing 210096, China. [2]NEC Laboratories China, 11F, Bldg A, Innovation Plaza, Tsinghua Science Park, No. 1 Zhongguancun East Road, Haidian District, Beijing 100084, China.

Received: 11 July 2014 Accepted: 9 February 2015
Published online: 24 April 2015

References

1. Cisco: Visual networking index. White paper at Cisco.com (2014)
2. A Damnjanovic, J Montojo, Y Wei, T Ji, T Luo, M Vajapeyam, T Yoo, O Song, D Malladi, A survey on 3GPP heterogeneous networks. IEEE Wireless Commun. **18**(3), 10–21 (2011)
3. V Chandrasekhar, JG Andrews, A Gatherer, Femtocell networks: a survey. IEEE Commun. Mag. **46**(9), 59–67 (2008)
4. M Yavuz, F Meshkati, S Nanda, A Pokhariyal, N Johnson, B Raghothaman, A Richardson, Interference management and performance analysis of UMTS/HSPA+ femtocells. IEEE Commun. Mag. **47**(9), 102–109 (2009)
5. S-M Cheng, S-Y Lien, F-S Chu, K-C Chen, On exploiting cognitive radio to mitigate interference in macro/femto heterogeneous networks. IEEE Wireless Commun. **18**(3), 40–47 (2011)
6. D López-Pérez, I Guvenc, G De La Roche, M Kountouris, TQ Quek, J Zhang, Enhanced intercell interference coordination challenges in heterogeneous networks. IEEE Wireless Commun. **18**(3), 22–30 (2011)
7. J Sangiamwong, Y Saito, N Miki, T Abe, S Nagata, Y Okumura, in *11th European Wireless Conference 2011-Sustainable Wireless Technologies (European Wireless)*. Investigation on cell selection methods associated with inter-cell interference coordination in heterogeneous networks for LTE-advanced downlink (VDE, 2011)
8. R Madan, J Borran, A Sampath, N Bhushan, A Khandekar, T Ji, Cell association and interference coordination in heterogeneous LTE-A cellular networks. IEEE J. Selected Areas Commun. **28**(9), 1479–1489 (2010)
9. G Rittenhouse, S Goyal, DT Neilson, S Samuel, in *IEEE Technical Symposium at ITU Telecom World (ITU TW)*. Sustainable telecommunications, (2011)
10. F Richter, AJ Fehske, GP Fettweis, in *IEEE 70th Vehicular Technology Conference Fall (VTC 2009-Fall)*. Energy efficiency aspects of base station deployment strategies for cellular networks, (2009)
11. E Oh, B Krishnamachari, X Liu, Z Niu, Toward dynamic energy-efficient operation of cellular network infrastructure. IEEE Commun. Mag. **49**(6), 56–61 (2011)
12. I Ashraf, F Boccardi, L Ho, Sleep mode techniques for small cell deployments. IEEE Commun. Mag. **49**(8), 72–79 (2011)
13. J Gong, Z Niu, A dynamic programming approach for base station sleeping in cellular networks. IEICE Trans. Commun. **95**(2), 551–562 (2012)
14. S Cai, L Xiao, H Yang, J Wang, S Zhou, in *22nd Wireless and Optical Communication Conference (WOCC)*. A cross-layer optimization of the joint macro- and picocell deployment with sleep mode for green communications, (2013)
15. Y Shi, J Zhang, KB Letaief, Group sparse beamforming for green Cloud-RAN. IEEE Trans. Wireless Commun. **13**(5), 2809–2823 (2014)
16. A Bousia, E Kartsakli, L Alonso, C Verikoukis, in *IEEE Global Communications Conference (GLOBECOM)*. Dynamic energy efficient distance-aware base station switch on/off scheme for LTE-advanced, (2012)
17. MA Marsan, L Chiaraviglio, D Ciullo, M Meo, in *IEEE 4th International Conference on Communications and Electronics (ICCE)*. Multiple daily base station switch-offs in cellular networks, (2012)
18. H Claussen, I Ashraf, LT Ho, Dynamic idle mode procedures for femtocells. Bell Labs Tech. J. **15**(2), 95–116 (2010)
19. L Falconetti, P Frenger, H Kallin, T Rimhagen, in *IEEE Online Conference on Green Communications (GreenCom)*. Energy efficiency in heterogeneous networks, (2012)
20. AA Abdulkafi, TS Kiong, D Chieng, A Ting, J Koh, Energy efficiency improvements in heterogeneous network through traffic load balancing and sleep mode mechanisms. Wireless Pers. Commun. **75**(4), 2151–2164 (2014)
21. O Blume, H Eckhardt, S Klein, E Kuehn, WM Wajda, Energy savings in mobile networks based on adaptation to traffic statistics. Bell Labs Tech. J. **15**(2), 77–94 (2010)
22. MF Hossain, KS Munasinghe, A Jamalipour, in *IEEE Wireless Communications and Networking Conference (WCNC)*. An eco-inspired energy efficient access network architecture for next generation cellular systems, (2011)
23. G Cili, H Yanikomeroglu, FR Yu, in *IEEE International Conference on Communications (ICC)*. Cell switch off technique combined with coordinated multi-point (CoMP) transmission for energy efficiency in beyond-LTE cellular networks, (2012)
24. A Bousia, E Kartsakli, A Antonopoulos, L Alonso, C Verikoukis, in *IEEE International Conference on Communications (ICC)*. Game theoretic approach for switching off base stations in multi-operator environments, (2013)
25. E Oh, K Son, B Krishnamachari, Dynamic base station switching-on/off strategies for green cellular networks. IEEE Trans. Wireless Commun. **12**(5), 2126–2136 (2013)
26. A Conte, A Feki, L Chiaraviglio, D Ciullo, M Meo, MA Marsan, Cell wilting and blossoming for energy efficiency. IEEE Wireless Commun. **18**(5), 50–57 (2011)
27. MA Marsan, M Meos, in *Proc. of GreenMetrics Workshop in Conjunction with ACM SIGMETRICS Conf.* Energy efficient management of two cellular access networks, (2009)
28. JD Little, A proof for the queuing formula: $L = \lambda$ w. Oper. Res. **9**(3), 383–387 (1961)
29. D López-Pérez, X Chu, I Guvenc, On the expanded region of picocells in heterogeneous networks. IEEE J. Selected Topics Signal Process. **6**(3), 281–294 (2012)
30. D López-Pérez, X Chu, in *IEEE 20th International Conference on Computer Communications and Networks (ICCCN)*. Inter-cell interference coordination for expanded region picocells in heterogeneous networks, (2011)
31. Y Li, X Xu, J Han, W Liu, in *IEEE 78th Vehicular Technology Conference (VTC Fall)*. Coordinated inter-cell interference management for expanded region picocells in heterogeneous network, (2013)
32. W Noh, W Shin, C Shin, K Jang, H-h Choi, in *IEEE Wireless Communications and Networking Conference (WCNC)*. Distributed frequency resource control for intercell interference control in heterogeneous networks, (2012)
33. MW Arshad, A Vastberg, T Edler, in *IEEE Wireless Communications and Networking Conference (WCNC)*. Energy efficiency improvement through pico base stations for a green field operator, (2012)
34. E Yaacoub, in *19th International Conference on Telecommunications (ICT)*. Performance study of the implementation of green communications in LTE networks, (2012)
35. Evolved universal terrestrial radio access (E-UTRA); Further advancements for E-UTRA physical layer aspects. in *3GPP Tech. Rep., http://www.3gpp.org/DynaReport/36814.htm?device=desktop*, (2010)