

RESEARCH

Open Access



A refunding strategy: opportunistic user association with congestion-based pricing in macro-femto hybrid network

Yanjia Qi¹, Hongyu Wang^{1*}, Baoming Li² and Fuliang Chen²

Abstract

Femtocell technology addresses the severe problems of poor network capacity and indoor coverage. Meanwhile, the emergence of high-capacity air interfaces and dense deployment of small cells result in increasingly high backhaul cost in cellular wireless networks. Purchasing on leased lines can guarantee the service provision during busy hours, however, purchased capacity goes to waste in off-peak time. Hybrid mode is the most promising one among all femtocell access modes which allows macro users to associate with adjacent femtocells with idle bandwidth resources. Femto holder (FH) is egoistic and unwilling to share bandwidth with transferred users from macrocells without any compensation, thus the successful implementation of hybrid access becomes a challenging problem. In this paper, we present an economic refunding framework to motivate hybrid access in femtocells. Macro users can opportunistically associate with adjacent femtocells with excess backhaul capacity. FH can receive certain refunding from wireless service provider (WSP) in exchange for traffic offloading. FH employs congestion pricing policy so as to control the cell load in the femtocell. Within this framework, we design a general utility maximization problem for user association that enables macro users to associate with femtocells based on traffic status, cell load, and access price. Dual decomposition is used to obtain an approximate solution. The impact of congestion pricing on the aggregate throughput and load balancing is also analyzed. Extensive simulations show the proposed scheme achieves a remarkable throughput gain compared with that with no compensation and compensation with usage-based pricing policy. Load balancing is substantially improved as well.

Keywords: Heterogeneous network, Backhaul, User association, Congestion pricing, Utility maximization

1 Introduction

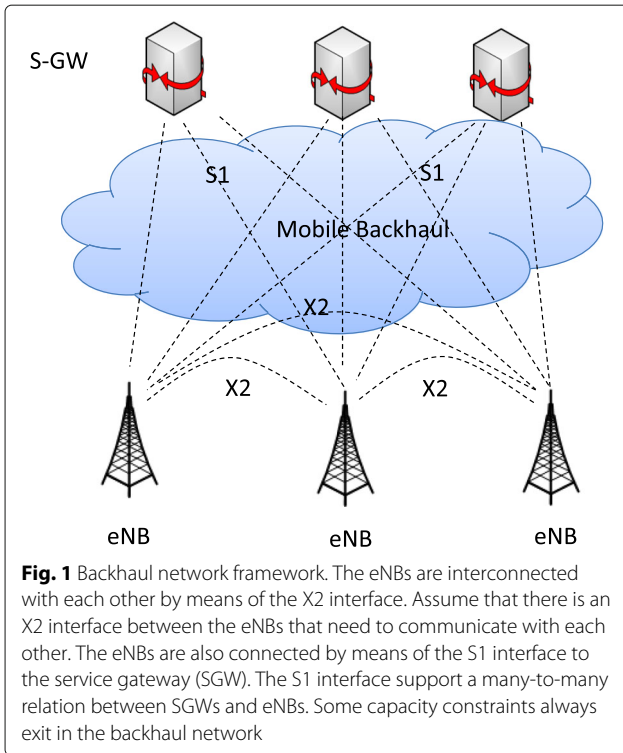
In recent years, there has been a dramatically increase in the number of mobile users and high-speed data services, which places a greater pressure on the conventional cellular network infrastructures. In spite of the necessity for small cells deployed to meet the enormous requirements for traffic data, there are still many technical challenges to be settled. One of the key challenges is to provide extensive backhaul connectivity economically [1]. Backhaul is a term commonly used to describe wired or wireless connectivity between base stations (BSs) and associated mobile switching nodes in a cellular system, as illustrated

in Fig. 1. Wired and wireless technologies have been investigated as backhaul solutions for small cells [2]. For wired backhaul, copper lines and optical fibers are the major mediums, which provide suitable support for voice and other services with low latency and delay. Wireless backhaul solutions incorporate millimeter wave technologies of 60 and 70–80 GHz, microwave technologies between 6 and 60 GHz, and sub 6-GHz radio wave technologies in both licensed and unlicensed bands. The backhaul construction significantly depends on the locations of small cells, the cost of implementing backhaul connections, traffic load intensity of small cells, latency, and target QoS requirement of small cell users and hardware. According to the recent statistics, the number of small cells now deployed has reached up to 13.3 million reported in Small Cell Forum survey [3] and this number is forecasted to

*Correspondence: whyu@dlut.edu.cn

¹School of Information and Communication Engineering, Dalian University of Technology, Linggong Road, 116023 Dalian, China

Full list of author information is available at the end of the article



reach nearly 40 million by 2018 [4]. Such a large backhauling demand is bound to increase the cost substantially. Cost-effective strategies are necessary to relieve the backhauling burden. Therefore, the considerations of backhaul construction and operating costs become extremely crucial in modern communication systems.

Fortunately, various network access modes provide the possibility to relieve the pressure of backhaul cost. Indeed, how to make each user access the appropriate network substantially affects the network performance [5]. Femtocell hybrid access is a promising choice to control user association between macrocells and femtocells [6, 7], rather than the closed access and open access mode which render femtocells fully closed and open to macro users. Hybrid access permits macro users to exploit remaining femtocell resources after each femto user reserves its own capacity. Usually, macrocells and femtocells are controlled by wireless service providers (WSPs) and femto holders (FHs), respectively. FHs are egoistic to share bandwidth with transferred macro users. Incentive mechanisms should be designed from the perspective of economic compensation. Otherwise, FHs do not accept hybrid access mode if they have no benefit from offering own resources to transferred macro users. With the compensation, FHs are willing to share the remaining resources with macro users. Meanwhile, macro users should pay for the used bandwidth from FHs.

Several refunding mechanisms between WSP and FHs are investigated in the past few years. Chen et al. early propose a framework of utility-aware refunding [8], where WSP provides the certain refunding to motivate FHs to open their resource for macro users then FHs decide the resource allocation among femto and macro users. A Stackelberg game is formulated to maximize the utilities for both WSP and FHs. Shih et al. present an economic framework based on the game theoretical analysis [9], where the FHs determine the proportion of femtocell resources they will share with public users, while WSP maximizes its benefit by setting the ratio of the revenue distributed to FHs. Yang et al. show the refunding mechanism for small cell networks with limited-capacity backhaul [10], in which small cell holders receive refunding as incentives to serve guest users with their remaining backhaul capacity. WSP decides individualized refunding and interference constraints to different small cell holders; meanwhile, each small cell holder serves guest users in a best-effort manner while maximizing its own utility. Li et al. show a rate-based pricing framework within which the macro BS provides profit to motivate femto BSs to adopt hybrid access policy and guarantee transmission rates of associated users [11]. Ford et al. study a model where third parties provide backhaul connections and lease out excess capacity to WSP when available [12], presumably at significantly lower costs than guaranteed connections. Multi-leader multi-follower data offloading game is investigated in [13], where macro BSs propose market prices and accordingly small cells determine the traffic volumes they are willing to offload. Shen et al. propose an auction mechanism to establish the hybrid access [14], where femto access points (FAPs) decide their bids independently by maximizing their own utilities. After receiving the bids, the macro BS searches the winner FAP and optimizes the number of offloaded macro users. The compensation is paid by the macro BS to the winner FAP for serving the additional macro users. A price discount strategy for WSP to promote the hybrid access mode of femtocell is developed in which WSP provides a price discount in exchange for the FHs to share part of their resource with macro users [15]. An interference management scheme for the two-tier femtocell networks is studied [16], where the macro BS protects itself by pricing the interference from the femtocell users. Price bargaining between femtocell users and macrocell exists so as to maximize the revenues and protect the QoS requirements. Zhu et al. design an incentive mechanism in which WSP pays the small cell service providers for the shared radio resource [17]. A hierarchical dynamic game framework is proposed in which an evolutionary game is used to model and analyze the service selection of users in the lower lever while a Stackelberg differential game is formulated where WSP and small cell service providers act

as the leader and followers, respectively. A utility gain framework where each femtocell reserves a fraction of resource to macro users and gets a gain from WSP is proposed [18]. A learning mechanism allows both WSP and FH to choose the best strategy to reach a win-win situation. Iosifidis et al. present a market where WSPs lease multiple FAPs and each FAP can concurrently serve traffic from multiple WSPs [19]. An iterative double-auction mechanism is designed to ensure the maximization of differences between offloading benefits of operators and offloading costs of FAPs. Zhang et al. propose an incentive method where macro BS allocates a portion of subchannels to FAP for spurring the FAP to serve macro users [20]. The FAP allocates the subchannels and power to maximize the femtocell network utility and the throughput of the served macro users. Yang et al. propose a bargaining cooperative game where spectrum leasing is used as the incentive mechanism to motivate small cell working as the relays [21]. Macrocell leases some of its dedicated spectrum to the selected relay small cell, and then cooperative bargaining strategy between the relay small cell and the macrocell is formulated to enhance the system spectral efficiency and balance the capacity. In [22], Liu et al. propose an opportunistic user association in multi-service HetNets, where the opportunistic user association is formulated as an optimization problem which can be solved by Nash bargaining solution (NBS).

However, cell load congestion problem in networks will also affect the achieved network performance. Congestion can severely degrade the QoS performance, user's satisfaction, and obtained revenues. Congestion pricing, early proposed in [23], is a promising solution that can help alleviate the problem of congestion. Al-Manthari et al. survey recent congestion pricing techniques for wireless cellular networks [24], which verifies that congestion pricing can reduce congestion and generate higher revenues for network operators. Niu et al. present a congestion pricing model to charge media streaming operators based on the bandwidth-delay product on each overlay link [25]. Khabazian et al. study a mechanism by which the femto and macro capacity resources are jointly priced according to a dynamic pricing-based call admission mechanism [26]. Cheung et al. consider the network selection and data offloading problem in an integrated cellular WiFi system by incorporating the practical considerations [27]. Interactions of the users' congestion-aware network selection decisions across multiple time slots as a non-cooperative network selection game is formulated. When the players repeatedly perform better response updates, the system is guaranteed to converge to a pure Nash equilibrium. Wang et al. solve the optimization problem under the stochastic decision framework and propose a distributed heuristic algorithm to independently and dynamically associate each user with the best BS [28]. By posing a price factor to

the BS evaluation update, users dynamically associate the best BS based on the congestion state.

As a matter of fact, the high fluctuation of traffic load and rate requirement can lead to a waste of provided capacity in some circumstances. For instance, the number of users decreases or users merely need voice service with low-rate requirement in idle hours. Excessively establishing and maintaining small cells will result in the expensive backhaul cost, which can hardly conform to the case of fluctuant traffic. Rather than providing the excessively abundant backhaul capacity to guarantee the peak data rates, WSP should be able to dynamically leverage excess capacity on existing backhaul provided by FHs. The problem is to offload traffic opportunistically when FHs have excess backhaul capacity with the appropriate compensation. Since the capacity will only be purchased when used, the opportunistic capacity can presumably be purchased at a much lower cost than the guaranteed backhaul capacity. Thus, the opportunistic user association can be regarded as a promising method to reduce cost effectively. Meanwhile, FHs will consider the cell load factor to reduce congestion. This observation motivates us to research the performance improvement through dynamic pricing policy. In this paper, we propose an economic compensation framework. Under this framework, FHs provide femtocell and backhaul connections. Traffic can be offloaded opportunistically from macrocells to femtocells. Once the association is implemented, WSP should reimburse FHs for use of backhaul resources. FHs adjust the cell load by congestion pricing policy to guarantee the QoS. The main contributions of the paper are listed as follows:

- 1) We formulate an optimal opportunistic user association problem, in which macro users associate with macrocells or adjacent femtocells with limited backhaul capacity, cell load, and access price. We present a general net utility maximization problem, where the utility is represented by logarithmic utility of throughput minus cost. Cost is measured by price per unit bit rate. Then, we show a dual decomposition method that enables fast computation of global optimal solution in an efficient, distributed manner via augmented Lagrangian techniques.
- 2) We adopt congestion pricing policy to control each cell load. When macro users intend to associate with femtocells, each user will get its own bandwidth to maximize the aggregate utility. Here, the price is not fixed but changes according to the number of users associated with the same femtocell. The more macro users associate with the same femtocell, the higher price per unit bandwidth is. Then, users in congested cells will be impelled to associate with uncrowded femtocells.

- 3) We conduct numerical simulations to evaluate this framework and verify the influence of dynamic price for user association. Results show that when FHs adopt congestion pricing policy, the remarkable throughput gain can be achieved under different congestion levels. Due to dynamic cell load control, the effect of load balancing can also be substantially improved.

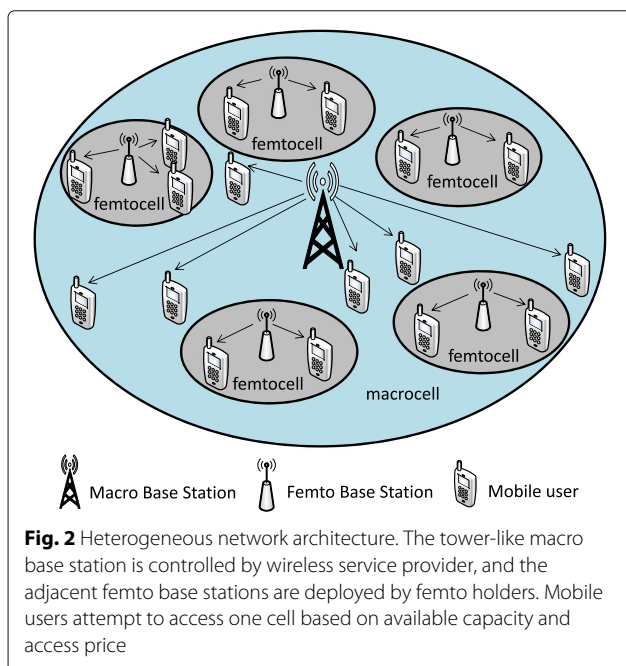
The remainder of this paper is organized as follows. We describe the system model in Section 2. The optimal user association problem and the dual decomposition to solve a net utility maximization problem are proposed in Section 3. In Section 4, extensive simulations are presented along with related discussions, and finally, our work and the outlook are summarized in Section 5.

2 System model

In this section, we describe the system model including the system architecture, interference model, and necessary network constraints. Then, we propose a cell load-based congestion pricing policy where price per bit rate can be adjusted as the cell load changes.

2.1 System architecture

Consider a traditional macrocellular OFDMA network with overlays of several femtocells, as shown in Fig. 2. All subcarriers are orthogonal. There are M BSs including macro BSs (MBSs) and femto BSs (FBSs). We let BS i denote the i th base station, $i = 1, \dots, M$. N mobile users (MUs) uniformly distribute in this area. We let MU j denote the j th mobile user, $j = 1, \dots, N$. $\Psi_{BS}(i)$ is the



set of MUs associated with BS i . Γ_{BS} represents the set of all BSs. Here, we suppose that all the antennas transmit with full power. Thus, the interference suffered by an MU is approximately measured from all BSs except the serving BS. The throughput of one MU is the bandwidth times spectrum efficiency provided by the serving BS $w_{ij} \log(1 + \gamma_{ij})$, where w_{ij} is the bandwidth MU j gets from BS i and γ_{ij} is the SINR of MU j on BS i . The SINR of MU j on BS i is

$$\gamma_{ij} = \frac{P_i H_{ij}}{\sum_{s \in \Gamma_{BS}, s \neq i} P_s H_{sj} + \sigma^2}, \tag{1}$$

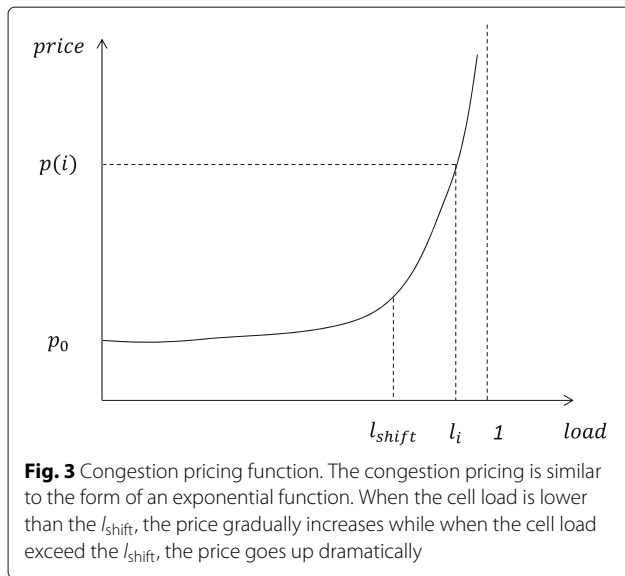
where P_i is transmission power from BS i , Γ_{BS} is the set of BSs, H_{ij} is the channel attenuation coefficient between BS i and MU j , and σ^2 is the thermal noise power. $\sum_{s \in \Gamma_{BS}, s \neq i} P_s H_{sj}$ is the received aggregate interference from all the BSs except the serving BS. In this model, the intra-cell interference can be avoided since there are no overlapped subcarriers for all users served by one cell. Before the bandwidth allocation process, the amount of the subcarriers allocated to one user is uncertain, thus the inter-cell interference is approximately evaluated by the worst case that all BSs generate aggregate interference to the users. Here, we rewrite se_{ij} for short instead of $\log(1 + \gamma_{ij})$. Assume that the attenuation model is slow fading so the channel conditions are fixed through frames.

2.2 Congestion pricing model

We propose a congestion pricing policy in this subsection. The guideline for the definition of this policy is that price changes slowly when the backhaul resource is abundant enough and increases drastically when the backhaul resource becomes scarce. With this pricing policy, resource can be utilized efficiently to benefit load balancing. Three aspects of this pricing policy should be considered:

- 1) The wasted backhaul resource is null regardless of whether the cell is congested or not, which means that bandwidth resource should be fully utilized
- 2) When no congestion occurs, the change of price should be as small as possible to ensure user's fair association
- 3) In case of congestion, the change rate of price should increase faster than that during no congestion period. This faster increasing rate of price can be used to discourage users in associating with heavy-load cell.

In this policy, we let the price be measured by price per bit rate. In Fig. 3, we define l_{shift} as the turning point for the network pricing. When the load is lower than the l_{shift} , the price increases slowly. When the load is higher than the l_{shift} , the price changes rapidly and even dramatically when backhaul resource approaches maximum. We



adopt this variation tendency to describe our pricing policy. When cell load is in a saturated state, the price can be raised to make some users associate with lightly load cell instead.

We show a load-based pricing function that price changes with cell load, which refers to [29].

$$p_i(k) = p_0 \left(\frac{1 - l_{\text{shift}}}{1 - l_i(k)} \right)^n, \quad (2)$$

where the $p_i(k)$ is the price at time k in cell i , p_0 is the initial access price, and $l_i(k)$ is cell load at time k for cell i . Here, $l_i(k)$ is the ratio of actual cell load to cell tolerable maximal load L_{max} . We use parameter n to control the steepness of this function and $n \geq 1$.

3 User association optimization

As mentioned above, an important issue is that how MUs associate with macrocells controlled by WSP or femtocells deployed by FHs when they acquire services within the cellular coverage. We generalize this issue into a net utility maximization problem including network constraints, interference condition, access price, and cell load.

3.1 Optimization formulation

To model the bandwidth constraints, we suppose that the available bandwidth of each BS i is W_i . Let w_{ij} represent the bandwidth BS i allocated to MU j . Thus, the aggregate allocated bandwidth should satisfy the constraint:

$$0 \leq \sum_{j \in \Psi_{\text{BS}(i)}} w_{ij} \leq W_i. \quad (3)$$

We let C_i denote the capacity of BS i . The capacity of FBS is the remaining backhaul resource after each femto

user reserves its own capacity. Thus, the aggregate rate should be less than the capacity upper limit in each cell:

$$0 \leq \sum_{j \in \Psi_{\text{BS}(i)}} se_{ij}w_{ij} \leq C_i. \quad (4)$$

One MU is commonly served by one BS at a time. Thus, a single association constraint should be supplemented.

$$w_{ij} \neq 0 \text{ for only one } i. \quad (5)$$

We adopt logarithmic function as user utility function. Different from linear utility function, logarithmic function can truly reflect the user's satisfaction. Logarithm is concave and has the diminishing growth tendency. This property does not enable to allocate excessive resource to users with excellent channel condition while poor users starve. Therefore, logarithmic function is considered as utility function in particular. In the remainder of this paper, we adopt the natural logarithmic utility function. The aggregate utility can be represented by

$$U(r_{\text{MU}}) = \sum_{i=1}^M \sum_{j=1}^N \ln(se_{ij}w_{ij}). \quad (6)$$

To clarify the backhaul cost that WSP should pay to the FHs, we assume the cost function is represented as follows:

$$C(r_{\text{BS}}) = \sum_{i=1}^M C(r_i) = \sum_{i=1}^M \sum_{j=1}^N p_i se_{ij}w_{ij}, \quad (7)$$

where $C(r_i)$ is the cost that WSP should pay. Once macro users associate with the adjacent femtocells, a positive cost is generated since backhaul resources in femtocell are utilized. Suppose that if macro users associate with macrocells, $C(r_i) = 0$, while $C(r_i) = p_i \sum_{j \in \Psi_{\text{BS}(i)}} w_{ij} se_{ij}$ when macro users associate with adjacent femtocells, where p_i represents price per unit backhaul capacity of each femtocell and this price changes with cell load.

Our goal is to maximize the net utility, which incorporates the MUs' utility and the cost that WSP should pay, with constraints of bandwidth resource and backhaul capacity. Now, we write the user association problem as the optimization:

$$\max_{w_{ij}} U(r_{\text{MU}}) - C(r_{\text{BS}}) \quad (8)$$

$$s.t. \quad 0 \leq \sum_{j \in \Psi_{\text{BS}(i)}} w_{ij} \leq W_i, \quad (9)$$

$$0 \leq \sum_{j \in \Psi_{\text{BS}(i)}} se_{ij}w_{ij} \leq C_i, \quad (10)$$

$$w_{ij} \neq 0 \text{ for only one } i. \quad (11)$$

Then, we will provide the analysis and algorithms for solving optimization problem (8)–(11). We propose a low-complexity distributed algorithm for a large-scale network.

3.2 Dual decomposition algorithm

The optimization (8)–(11) is not convex due to constraint (11). It is unpractical to solve this problem by Karush-Kuhn-Tucker condition. An alternative algorithm is necessary, especially for a large scale network. Fortunately, following [30], we can obtain an approximate solution by dual decomposition method. Traditionally, centralized solution for this convex optimization problem is usually achieved on a central server in the core network. The long computational time and coordination requirement among different tiers result in excessive computational complexity and low reliability. The computational complexity exponentially increases when the network scale is large. An distributed algorithm based on dual decomposition method can overcome this difficulty. First, we neglect the constraint (11), thus the results are the allocated bandwidth from all BSs. Then, among these candidates, the one which offers the largest rate is retained. This truncation method is well-known in network theory and results in few errors [31].

3.2.1 Dual problem

The primal problem in (8)–(11) can be expressed in a Lagrangian formula. Two dual variables are introduced, which are λ^{bw} and λ^{rate} .

$$\begin{aligned}
 P(w_{ij}, \lambda_i^{\text{bw}}, \lambda_i^{\text{rate}}) = & -\sum_{i=1}^M \sum_{j=1}^N \ln(w_{ij}se_{ij}) + \sum_{i=1}^M \sum_{j=1}^N p_i w_{ij}se_{ij} \\
 & + \sum_{i=1}^M \lambda_i^{\text{bw}} \left(\sum_{j \in \Psi_{\text{BS}}(i)} w_{ij} - W_i \right) \\
 & + \sum_{i=1}^M \lambda_i^{\text{rate}} \left(\sum_{j \in \Psi_{\text{BS}}(i)} w_{ij}se_{ij} - C_i \right).
 \end{aligned} \quad (12)$$

The dual problem of (8)–(11) is in regard to a function of variables λ^{bw} and λ^{rate} :

$$\begin{aligned}
 D(\lambda_i^{\text{bw}}, \lambda_i^{\text{rate}}) = & \sum_{i=1}^M \left(\sum_{j \in \Psi_{\text{BS}}(i)} w_{ij} - W_i \right) \lambda_i^{\text{bw}} \\
 & + \sum_{i=1}^M \left(\sum_{j \in \Psi_{\text{BS}}(i)} w_{ij}se_{ij} - C_i \right) \lambda_i^{\text{rate}} \\
 & - \sum_{i=1}^M \sum_{j=1}^N \ln(w_{ij}se_{ij}) + \sum_{i=1}^M \sum_{j=1}^N p_i w_{ij}se_{ij} \\
 \text{s.t. } & \lambda_i^{\text{bw}} > 0, \lambda_i^{\text{rate}} > 0.
 \end{aligned} \quad (13)$$

In a primal problem, both the objective function and all constraints are convex, this satisfies Slater's condition [32]. The well-known weak duality property states that an upper bound to the maximum of the utility is given by

$$\max_{w_{ij}} P(w_{ij}, \lambda_i^{\text{bw}}, \lambda_i^{\text{rate}}) \leq \min_{\lambda_i^{\text{bw}}, \lambda_i^{\text{rate}}} D(\lambda_i^{\text{bw}}, \lambda_i^{\text{rate}}). \quad (14)$$

This bound applies even when the objective function is non-convex. Moreover, $D(\lambda_i^{\text{bw}}, \lambda_i^{\text{rate}})$ is always convex in $\lambda_i^{\text{bw}}, \lambda_i^{\text{rate}}$. Strong duality holds that the maximum value of primal problem equals to the minimum value of its dual problem. Therefore, the primal problem can be solved by its dual problem. By solving the dual optimal $\lambda_i^{\text{bw}*}$ and $\lambda_i^{\text{rate}*}$, the optimal solution w_{ij}^* of the primal problem can be achieved.

3.2.2 Distributed algorithm implementation

The dual problem is solved by the gradient descent method, where lagrange multiplier λ is updated along the opposite direction of the gradient $\nabla D(\lambda)$. The primal and dual problems can be solved in a distributed manner. The iterative process is illustrated in Fig. 4. The k th iteration of gradient descent method is given as follows:

1) MU's side: MUs receive pilot signals from all BSs. Each signal includes the values of λ^{bw} and λ^{rate} which are broadcasted by each BS. The optimal bandwidth which one MU can get from one BS is derived from the first-order partial derivative of w_{ij} at the k th iteration.

$$\begin{aligned}
 \frac{\partial P(w_{ij}(k), \lambda_i^{\text{bw}}(k), \lambda_i^{\text{rate}}(k))}{\partial w_{ij}(k)} = & -\frac{1}{w_{ij}(k)} + p_i(k)se_{ij} \\
 & + \lambda_i^{\text{bw}}(k) + \lambda_i^{\text{rate}}(k)se_{ij} = 0,
 \end{aligned} \quad (15)$$

$$w_{ij}(k) = \frac{1}{\lambda_i^{\text{bw}}(k) + \lambda_i^{\text{rate}}(k)se_{ij} + p_i(k)se_{ij}}. \quad (16)$$

Each MU chooses the optimal serving BS at the k th iteration which satisfies the follows:

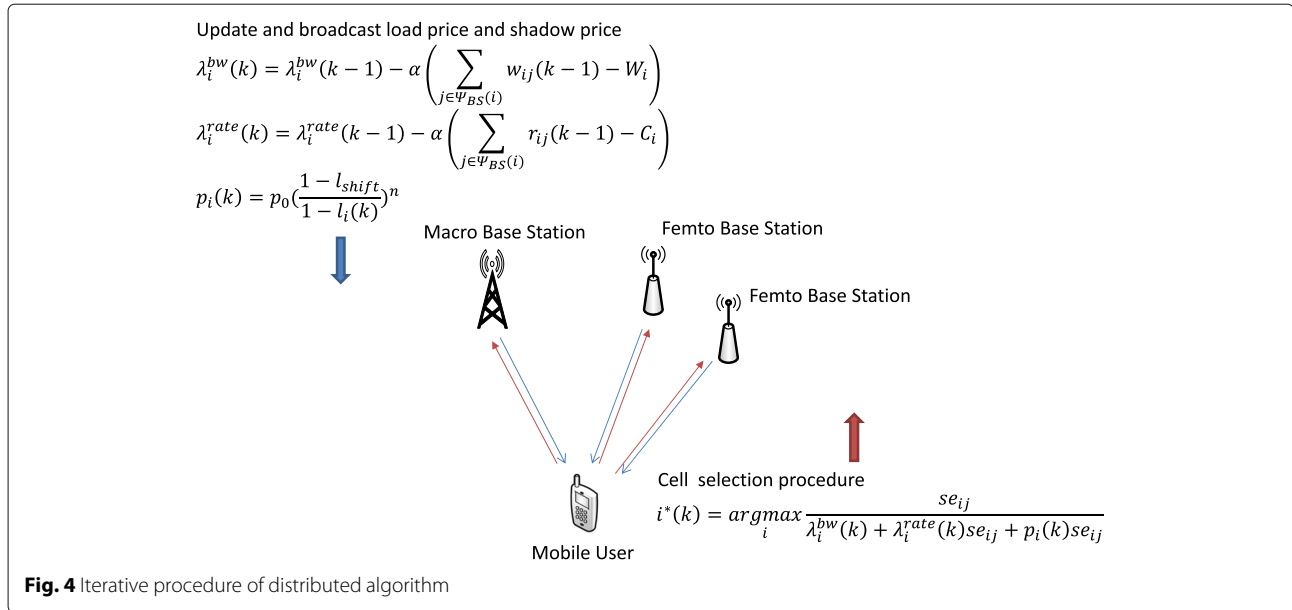
$$i^*(k) = \underset{i}{\operatorname{argmax}} \frac{se_{ij}}{\lambda_i^{\text{bw}}(k) + \lambda_i^{\text{rate}}(k)se_{ij} + p_i(k)se_{ij}}, \quad (17)$$

$$w_{ij}^*(k) = \frac{1}{\lambda_i^{\text{bw}}(k) + \lambda_i^{\text{rate}}(k)se_{ij} + p_i(k)se_{ij}}, \text{ when } i(k)=i^*(k), \quad (18)$$

where $p_i(k)$ is the congestion price which is determined by the cell load of BS i at the k th iteration as shown below:

$$p_i(k) = p_0 \left(\frac{1 - I_{\text{shift}}}{1 - |\Psi_{\text{BS}}(i)|(k)} \right)^n, \quad (19)$$

where $|\Psi_{\text{BS}}(i)|(k)$ is the number of MUs associated with BS i at the k th iteration. In each iteration, a MU may select the different optimal BS which provides maximal rate so cell load may change as the increase of iteration times.



2) BS's side: After each BS receives the demand information from MU's side, the values of λ_i^{bw} and λ_i^{rate} are updated then these two multipliers are announced to MUs in return.

$$\begin{aligned} \lambda_i^{bw}(k+1) &= \lambda_i^{bw}(k) - \alpha \frac{\partial D(\lambda_i^{bw}(k), \lambda_i^{rate}(k))}{\partial \lambda_i^{bw}(k)} \\ &= \lambda_i^{bw}(k) - \alpha \left(\sum_{j \in \Psi_{BS}(i)} w_{ij}(k) - W_i \right), \end{aligned} \tag{20}$$

$$\begin{aligned} \lambda_i^{rate}(k+1) &= \lambda_i^{rate}(k) - \alpha \frac{\partial D(\lambda_i^{bw}(k), \lambda_i^{rate}(k))}{\partial \lambda_i^{rate}(k)} \\ &= \lambda_i^{rate}(k) - \alpha \left(\sum_{j \in \Psi_{BS}(i)} se_{ij} w_{ij}(k) - C_i \right), \end{aligned} \tag{21}$$

where $\alpha > 0$ is a step size and we assume that α remains constant in the process of iterations. After iterations following the above steps, the algorithm can be converged to a sub-optimal solution. In fact, λ_i^{bw} and λ_i^{rate} can be interpreted as the shadow price in economics. If the demand $\sum_{j \in \Psi_{BS}(i)} w_{ij}(k)$ and $\sum_{j \in \Psi_{BS}(i)} se_{ij} w_{ij}(k)$ for BS i exceeds the maximum value, the shadow price will go up. Otherwise, the shadow price will decrease. Thus, when BS i is the congested state, its price will increase and fewer MUs will associate with it, while other lightly load BSs attract more MUs to associate with due to the lower price. In addition, the complexity is

reduced to $\mathcal{O}(M + N)$. In comparison to the complexity $\mathcal{O}(M * N)$ of the centralized method, the distributed method guarantees the convergence fast and effective, especially for a large-scale network.

Since the derivative of $D(\lambda)$ is bounded and this property satisfies the condition of Proposition 6.3.6 in [32], we can confirm that the dual decomposition algorithm converges to a sub-optimal solution.

4 Performance analysis

As the adoption of congestion pricing policy, each cell will change its price according to the load at each iteration, thus MUs select the best serving BSs to associate with. When most MUs associate with the same cell, price will go up even more dramatically when cells are in highly congested state. Due to the lower price, MUs who originally reside in highly load cells are attracted to associate with other lightly load cells. Here, we show some benefits due to the introduction of dynamic pricing policy and related mathematical proofs.

Proposition 1 *The scheme under congestion pricing policy achieves throughput gain in comparison to that under usage-based pricing policy, especially when actual cell load is less than the load threshold.*

Proof Here, we discuss two kinds of cases to prove the throughput gain due to the introduction of congestion pricing policy and then figure out approximate gain value.

Case 1: We consider the single cell case, where all MUs select the same BS to associate with. w_{ij} , se_{ij} , W_i , C_i , and

p_i can be rewritten as w_j , se_j , W , C , and p for short, respectively. Our goal is to explore the relation between bandwidth allocation for each MU and the price that MU is charged.

When the bandwidth and capacity limit are very large, the two constraint conditions in previous optimization problem can be neglected. Then, the optimal bandwidth allocation w_j^* is obtained through the derivation of w_j .

$$\frac{\partial P(w_j)}{\partial w_j} = \frac{1}{w_j} - pse_j = 0 \implies r_j = w_j se_j = \frac{1}{p}. \quad (22)$$

From (22), we see that the allocated bandwidth of MU j is inversely proportional to the price. In other words, when cell load becomes lower, this will make MUs get more bandwidth because of the lower price. However, bandwidth and backhaul resource are not infinite, and therefore, the optimal w_j is about the derivation of w_j , λ^{bw} and λ^{rate} .

$$\frac{\partial P(w_j, \lambda^{bw}, \lambda^{rate})}{\partial w_j} = \frac{1}{w_j + \lambda^{bw} + \lambda^{rate} se_j} = 0, \quad (23)$$

$$\frac{\partial P(w_j, \lambda^{bw}, \lambda^{rate})}{\partial \lambda^{bw}} = \sum_{j=1}^N w_j - W = 0, \quad (24)$$

$$\frac{\partial P(w_j, \lambda^{bw}, \lambda^{rate})}{\partial \lambda^{rate}} = \sum_{j=1}^N w_j se_j - C = 0. \quad (25)$$

From (23)–(25) the optimal resource allocation w_j^* can be obtained. However, the equations are difficult to solve because a large number of MUs result in higher order equations, even if the solution exists. In view of this difficulty, we try to find out the approximate solution to describe the performance improvement. The approximate solution w_j^* is given as iterative recurrence formulas:

$$w_j(k) = \frac{1}{se_j p + \lambda^{bw}(k) + \lambda^{rate}(k) se_j}, \quad (26)$$

where $\lambda^{bw}(k) = \lambda^{bw}(k-1) - \alpha(\sum_{j=1}^N w_j(k-1) - W)$ and $\lambda^{rate}(k) = \lambda^{rate}(k-1) - \alpha(\sum_{j=1}^N w_j(k-1) se_j - C)$ and k is the number of iterations. Initial value $\lambda^{bw}(0)$ and $\lambda^{rate}(0)$ are predefined before the iteration begins. From (26), we can see when actual cell load becomes lower than the cell load threshold, namely the actual cell price decreasing due to lower cell load, the λ^{bw} and λ^{rate} decrease consequently at the $(k-1)$ th iteration and then w_j will go up at the k th iteration. Here, we let an increment of throughput $\Delta_{thr}(k)$ be a difference value at the k th iteration between two pricing policies as below:

$$\begin{aligned} \Delta_{thr}(k) &= \text{throughput}_{con}(k) - \text{throughput}_{use}(k) \\ &= \sum_{j=1}^N se_j (w_{jcon}(k) - w_{juse}(k)) \\ &= \sum_{j=1}^N se_j \left(\frac{1}{\lambda_{con}^{bw}(k) + \lambda_{con}^{rate}(k) se_j + p_{con} se_j} - \frac{1}{\lambda_{use}^{bw}(k) + \lambda_{use}^{rate}(k) se_j + p_{use} se_j} \right) \\ &= \sum_{j=1}^N se_j \left(\frac{p_{use} se_j - p_{con} se_j}{(\lambda_{con}^{bw}(k) + \lambda_{con}^{rate}(k) se_j + p_{con} se_j)(\lambda_{use}^{bw}(k) + \lambda_{use}^{rate}(k) se_j + p_{use} se_j)} \right) \\ &\quad + \frac{\sum_{m=1}^{k-1} (w_{jcon}(m) - w_{juse}(m))}{(\lambda_{con}^{bw}(k) + \lambda_{con}^{rate}(k) se_j + p_{con} se_j)(\lambda_{use}^{bw}(k) + \lambda_{use}^{rate}(k) se_j + p_{use} se_j)} \\ &\quad + \frac{se_j \sum_{m=1}^{k-1} (w_{jcon}(m) - w_{juse}(m))}{(\lambda_{con}^{bw}(k) + \lambda_{con}^{rate}(k) se_j + p_{con} se_j)(\lambda_{use}^{bw}(k) + \lambda_{use}^{rate}(k) se_j + p_{use} se_j)}, \end{aligned} \quad (27)$$

where $p_{con} = p_0 \left(\frac{1-l_{shift}}{1-|\Psi_{BS}|(k)} \right)^n$ and $p_{use} = p_0$. All the formulas on the nominator are greater than zero when $p_{con} < p_{use}$, namely $|\Psi_{BS}| < l_{shift} L_{max}$, the throughput under congestion pricing policy is more than that under usage-based pricing policy. The lower the cell load is, the more the gain is achieved. However, when the optimal solution is reached, the summation of bandwidth or rate allocation approaches the bandwidth or backhaul limit. One MU will reassociate with other lightly load cells if sufficient bandwidth resources are provided for the sake of this throughput increment, which leads to multiple cells case analysis.

Case 2: We consider the multiple cells case, where each MU selects a certain BS to associate with among all MBSs and FBSs. Unlike the single cell case, one MU has many choices because of different positions and spectrum efficiency which makes this case more complicated. According to [31], the multiple cell solution tends to concentrate on dominant single cell. We only need to compare the bandwidth allocation in a certain BS. Then, the total throughput of all MUs is approximately equal to our single cell association problem. The throughput increment is given as below:

$$\begin{aligned} \Delta_{thr}(k) &= \text{throughput}_{con}(k) - \text{throughput}_{use}(k) \\ &= \sum_{j=1}^N se_{ij} (w_{ijcon}(k) - w_{ijuse}(k)) \\ &= \sum_{j=1}^N se_{ij} \left(\frac{1}{\lambda_{icon}^{bw}(k) + \lambda_{icon}^{rate}(k) se_{ij} + p_{con} se_{ij}} - \frac{1}{\lambda_{iuse}^{bw}(k) + \lambda_{iuse}^{rate}(k) se_{ij} + p_{use} se_{ij}} \right) \\ &= \sum_{j=1}^N se_{ij} \left(\frac{p_{use} se_{ij} - p_{con} se_{ij}}{(\lambda_{icon}^{bw}(k) + \lambda_{icon}^{rate}(k) se_{ij} + p_{con} se_{ij})(\lambda_{iuse}^{bw}(k) + \lambda_{iuse}^{rate}(k) se_{ij} + p_{use} se_{ij})} \right) \\ &\quad + \frac{\sum_{m=1}^{k-1} (w_{ijcon}(m) - w_{ijuse}(m))}{(\lambda_{icon}^{bw}(k) + \lambda_{icon}^{rate}(k) se_{ij} + p_{con} se_{ij})(\lambda_{iuse}^{bw}(k) + \lambda_{iuse}^{rate}(k) se_{ij} + p_{use} se_{ij})} \\ &\quad + \frac{se_{ij} \sum_{m=1}^{k-1} (w_{ijcon}(m) - w_{ijuse}(m))}{(\lambda_{icon}^{bw}(k) + \lambda_{icon}^{rate}(k) se_{ij} + p_{con} se_{ij})(\lambda_{iuse}^{bw}(k) + \lambda_{iuse}^{rate}(k) se_{ij} + p_{use} se_{ij})}, \end{aligned} \quad (28)$$

where $p_{\text{con}} = p_0 \left(\frac{1-l_{\text{shift}}}{1-|\Psi_{\text{BS}}(i)|(k)} \right)^n$ and $p_{\text{use}} = p_0$. All formulas on the nominator are greater than zero when $p_{\text{con}} < p_{\text{use}}$, namely $\max(|\Psi_{\text{BS}}(1)|, |\Psi_{\text{BS}}(2)|, \dots, |\Psi_{\text{BS}}(M)|) < l_{\text{shift}}L_{\text{max}}$. Therefore, the total throughput under congestion pricing policy is more than that under usage-based pricing policy. \square

Proposition 2 *The throughput increases monotonously as the parameter n increases ($n \geq 1$).*

Proof As the same analysis method in the proof of Proposition 1, the throughput increment can be given in the form of difference under two different prices. As parameter n increases, the price decreases consequently under the same cell load. Following the proof of Proposition 1, lower price results in higher throughput, and thus the throughput under the congestion pricing policy is more than that under the usage-based pricing policy. \square

Proposition 3 *Under the congestion pricing policy, the cell load tends to be more balancing in comparison to that under the usage-based pricing policy.*

Proof Load balancing is another important criterion in heterogeneous network. Jain fairness index can be used to measure the balance degree of the system [33]. The formula of Jain fairness index is described as follows:

$$JFI = \frac{\left(\sum_{m=1}^M l_i \right)^2}{M \sum_{m=1}^M l_i^2}, \quad (29)$$

where M is the number of cells and l_i is the load of cell i . The balance index is 1 when each cell has the same load and tends to reach $1/M$ when the cell load is severely unbalanced. As shown in the proof of Proposition 1, lower cell load makes bandwidth allocation rise. However, due to bandwidth and backhaul limit, the bandwidth allocation can not increase any more. If sufficient bandwidth resources are provided, a MU will reassociate with other lightly load cells for a larger rate. This switch occurs when $se_{ij}w_{ij}(k) < se_{kj}w_{kj}(k)$, which means that the rate of MU j from BS k is greater than that from BS i . This flexible control property outperforms that of usage-based pricing policy. From Jain fairness index formula, we show the increasing tendency of load balancing as below: if one MU transfers from BS i to BS k , here assuming that cell load in BS k is greater than that in BS i due to lower price, the new cell loads for these two BSs are:

$$l'_i = l_i - 1, \quad l'_k = l_k + 1. \quad (30)$$

The new fairness index value is

$$JFI' = \frac{\left(\sum_{m=1}^M l_m \right)^2}{M \left((l_i - 1)^2 + (l_k + 1)^2 + \sum_{m \neq i,k} l_m^2 \right)} \quad (31)$$

$$= \frac{\left(\sum_{m=1}^M l_m \right)^2}{M \sum_{m=1}^M l_m^2 + 2M(l_i - l_k)}.$$

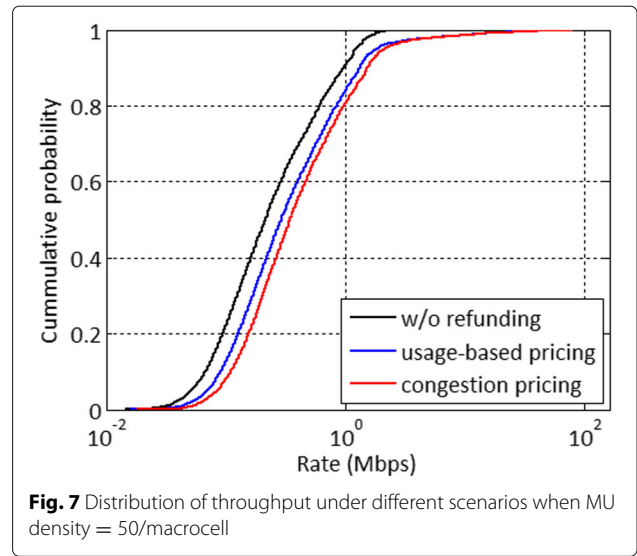
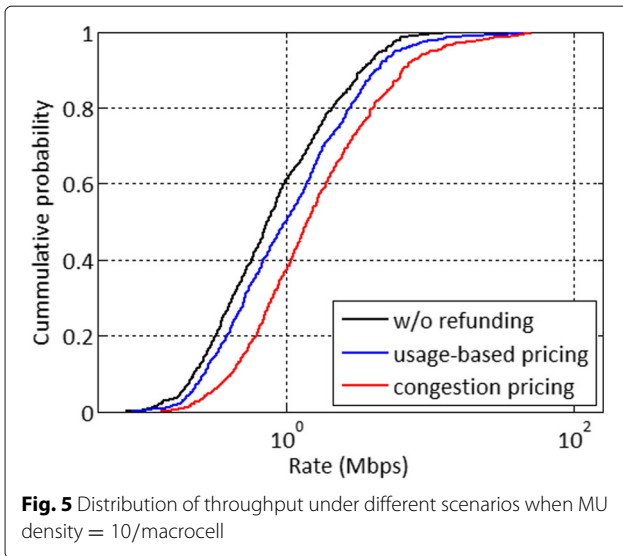
JFI and JFI' differ only in denominators, if and only if $l_i - l_k > 1$, $JFI' > JFI$. Since cell load l_i exceeds l_k , the Jain fairness increases which means cell load tends to be more balancing due to dynamic pricing control. \square

5 Simulation results

We consider a two-tier heterogeneous network with wrap around [34]. Let transmit power of MBS and FBS be 46 and 20 dBm, respectively. Suppose the locations of MBS to be fixed with FBSs uniformly independently distributed around. The density of FBS is 8 per macrocell. MUs locate in space uniformly with the density 10, 30, and 50 per macrocell. In the propagation environment, we use the path loss model $15.3 + 37.6 \log_{10}(d)$ and $35.3 + 37.6 \log_{10}(d)$ for macrocell and femtocell, respectively. We set the lognormal shadowing with a standard deviation to 8 dB. The thermal noise power is -104 dBm. The bandwidth in each cell is 10 MHz, and the backhaul capacity is 50 Mbps. We assume that the throughput is Shannon capacity rate of each MU. All the parameters are shown in Table 1.

Table 1 Simulation parameters

Parameter	Value
Topology	Uniform with wrap around
Total area	1000 m × 1000 m
Antenna pattern	Omni antenna
MU distribution	Uniform, 10, 30, and 50 per macrocell
FBS ditribution	8 per macrocell
MBS Tx power	46 dBm
FBS Tx power	20 dBm
Macrocell pathloss	$15.3 + 37.6 \log_{10}(d)$
Femtocell pathloss	$35.3 + 37.6 \log_{10}(d)$
Bandwidth	10 MHz
Backhaul capacity	50 Mbps
Shadowing	8-dB standard deviation
Thermal noise power	-104 dBm
Carrier frequency	2.1 GHz
Mobile model	Static
Fading	None
Access price	8
l_{shift}	0.8



Figures 5, 6, and 7 compare the throughput CDF under different scenarios with the different number of MUs. Without compensation (labeled without refunding) means there is no relationship between WSP-controlled macrocell and FH-deployed femtocell. FHs are not willing to share even though there are remaining backhaul resource. Therefore, MUs only reside in macrocells without any option. In comparison to the above strategy, usage-based pricing compensation (labeled usage-based pricing) implements the connection between macrocells and adjacent femtocells. FHs receive certain refunding from WSP to open its own backhaul resource for macro users. However, usage-based pricing cannot achieve high throughput due to the possible congestion problem. Our proposed strategy (labeled congestion pricing) can reduce the congestion and achieve high throughput. Table 2

shows the throughput under different number of MUs. We can see that there is a remarkable gain when the number of MUs changes. Cell-edge throughput gets 58.9 and 35.4% gain, respectively, compared with other two scenarios when MU density is 10 per macrocell. The medium rate also gets 44.7 and 34.1% gain. Even when the MU density increases up to 50 per macrocell, cell-edge throughput still gets 125 and 28.6% gain. Medium rate increases significantly as well. The reason is that congestion pricing policy impels macro users to select the best BS which offers abundant bandwidth resource and lower

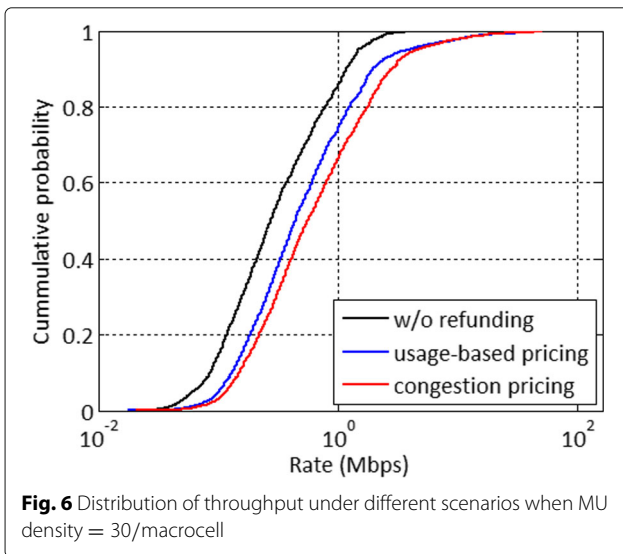


Table 2 The comparison of throughput under different number of MUs ($n = 2$)

Scenario	Without compensation	Usage-based pricing compensation	Congestion pricing compensation
Cell-edge rate (Mbps) (MU density = 10/macrocell)	0.17	0.20	<i>0.27</i>
Cell-edge rate (Mbps) (MU density = 30/macrocell)	0.06	0.11	<i>0.13</i>
Cell-edge rate (Mbps) (MU density = 50/macrocell)	0.04	0.07	<i>0.09</i>
Medium rate (Mbps) (MU density = 10/macrocell)	0.76	0.82	<i>1.10</i>
Medium rate (Mbps) (MU density = 30/macrocell)	0.26	0.39	<i>0.51</i>
Medium rate (Mbps) (MU density = 50/macrocell)	0.17	0.31	<i>0.38</i>

The results of the proposed algorithm are marked in italics to highlight the improvement

Table 3 The comparison of load balancing under different scenarios ($n = 2$)

Scenario	Without compensation	Usage-based pricing compensation	Congestion pricing compensation
JFI (MU density = 10/macrocell)	0.088	0.103	<i>0.268</i>
JFI (MU density = 30/macrocell)	0.098	0.189	<i>0.393</i>
JFI (MU density = 50/macrocell)	0.107	0.254	<i>0.379</i>

The results of the proposed algorithm are marked in italics to highlight the improvement

access price. Comparing to the other two scenarios, our proposed strategy can achieve better dynamic adjustment.

Table 3 shows the comparison of load balancing under different scenarios. Through congestion pricing policy, macro users are attracted to associate with adjacent femtocells, and thus, load balancing is improved reasonably. When the MU density is 10 per macrocell, the JFI increases from 0.088 to 0.268. With the increasing number of MUs, congestion degree becomes serious. When the MU density turns to 50 per macrocell, the JFI is up to 0.379 compared with 0.107 in no compensation scenario and 0.254 in compensation with the usage-based pricing policy.

As shown in Fig. 8, throughput distribution under different n is compared. $n = 0$ represents usage-based pricing compensation scenario because the access price in each cell keeps constant. There is a significant gain when $n = 1$. Indeed, throughput gets even larger as n increases. Table 4 shows the comparison of throughput under different parameters n . We can see a 68.4% gain in cell-edge rate and 60.3% gain in the medium rate when $n = 0$ and $n = 3$, respectively. That larger n leads to much lower

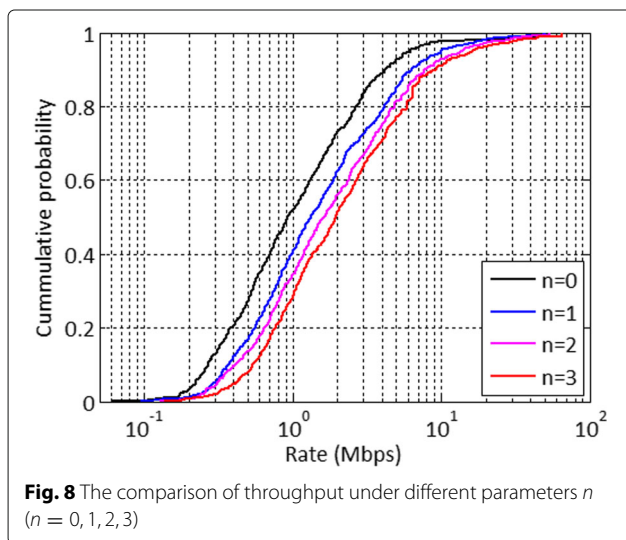


Fig. 8 The comparison of throughput under different parameters n ($n = 0, 1, 2, 3$)

Table 4 The comparison of throughput under different parameters n (MU density = 10/macrocell)

Number	0	1	2	3
Cell-edge rate (Mbps)	0.19	0.22	0.28	0.32
Medium rate (Mbps)	0.78	0.94	1.14	1.25

access price which results in higher throughput than what the smaller one does.

6 Conclusions

In this paper, we present an economic compensation framework between WSP and FHs. Under this framework, WSP pays certain refunding to FHs to implement traffic offloading. Macro users can opportunistically associate with FBS for transmission when there are remaining backhaul resources. We generalize this user association as an utility maximization problem. In the consideration of congestion that occurred in femtocells, each FH adopts the congestion pricing policy to control cell load. To reduce the computation complexity in large-scale networks, a dual decomposition algorithm is presented which incorporates bandwidth, backhaul capacity, and access price. Simulation results show that as the number of MUs increases, our optimization achieves remarkable throughput gains. Load balancing measured by Jain fairness index is also improved drastically. Actually, there are further problems to be investigated. Our work focuses on the interrelation between only one WSP and one type of FH. As a matter of fact, the types of WSP and FH vary widely. Therefore, the interrelation between WSP and each FH becomes more complicated. Our future work is to research multi-WSP-multi-FH problem and design corresponding solutions.

Authors' contributions

All authors contributed to all aspects of the article. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Author details

¹School of Information and Communication Engineering, Dalian University of Technology, Linggong Road, 116023 Dalian, China. ²Space Star Technology Co., Ltd, Zhichun Road, 100086 Beijing, China.

Received: 9 May 2016 Accepted: 11 December 2016

Published online: 03 January 2017

References

1. DC Chen, TQS Quek, M Kountouris, Backhauling in heterogeneous cellular networks: modeling and tradeoffs. *IEEE Trans. Wirel. Commun.* **14**(6), 3194–3206 (2015)
2. U Siddique, H Tabassum, E Hossain, DI Kim, Wireless backhauling of 5G small cells: challenges and solution approaches. *IEEE Wirel. Commun.* **22**(5), 22–31 (2015)
3. 60% of enterprises expect to deploy small cells by end of 2017. <http://www.wireless-mag.com/News/40629/60-of-enterprises-expect-to-deploy-small-cells-by-end-of-2017.aspx>. Accessed 3 Mar 2016

4. Number of small cells to reach 40m by 2018. <http://www.totaltele.com/view.aspx?ID=483099>. Accessed 5 Sept 2013
5. D Liu, L Wang, Y Chen, M ElKashlan, KK Wong, R Schober, L Hanzo, User association in 5G networks: a survey and an outlook. *IEEE Commun. Surv. Tutorials*. **18**(2), 1018–1044 (2016)
6. JG Andrews, H Claussen, M Dohler, S Rangan, MC Reed, Femtocells: past, present, and future. *IEEE J. Sel. Areas Commun.* **30**(3), 497–508 (2012)
7. H-S Jo, P Xia, JG Andrews, Open, closed, and shared access femtocells in the downlink. *EURASIP J. Wirel. Commun. Netw.* **2012**(1), 1–16 (2012)
8. Y Chen, J Zhang, Q Zhang, Utility-aware refunding framework for hybrid access femtocell network. *IEEE Trans. Wirel. Commun.* **11**(5), 1688–1697 (2012)
9. Y Shih, A Pang, M Tsai, C Chai, A rewarding framework for network resource sharing in co-channel hybrid access femtocell networks. *IEEE Trans. Comput.* **64**(11), 3079–3090 (2015)
10. Y Yang, TQS Quek, L Duan, Backhaul-constrained small cell networks: refunding and QoS provisioning. *IEEE Trans. Wirel. Commun.* **13**(9), 5148–5161 (2014)
11. L Li, M Wei, C Xu, Z Zhou, Rate-based pricing framework in hybrid access femtocell networks. *IEEE Commun. Lett.* **19**(9), 1560–1563 (2015)
12. R Ford, C Kim, S Rangan, in *2013 Asilomar Conference on Signals, Systems and Computers*. Opportunistic third-party backhaul for cellular wireless networks (IEEE, Pacific Grove, 2013), pp. 1594–1600
13. L Gao, G Iosifidis, J Huang, L Tassiulas, in *2013 IEEE INFOCOM*. Economics of mobile data offloading (IEEE, Turin, 2013), pp. 3303–3308
14. F Shen, D Li, PH Lin, E Jorswieck, in *2015 IEEE International Conference on Communications (ICC)*. Auction based spectrum sharing for hybrid access in macro-femtocell networks under QoS requirements (IEEE, London, 2015), pp. 3335–3340
15. A Li, X Liao, Z Gao, Y Yang, in *2014 IEEE 80th Vehicular Technology Conference (VTC Fall)*. Price discount strategy for WSP to promote hybrid access in femtocell networks (IEEE, Vancouver, 2014), pp. 1–6
16. Z Liu, L Hao, Y Xia, X Guan, Price bargaining based on the Stackelberg game in two-tier orthogonal frequency division multiple access femtocell networks. *IET Commun.* **9**(1), 133–145 (2015)
17. K Zhu, E Hossain, D Niyato, Pricing, spectrum sharing, and service selection in two-tier small cell networks: a hierarchical dynamic game approach. *IEEE Trans. Mob. Comput.* **13**(8), 1843–1856 (2014)
18. S El-Hammani, K Ibrahim, EH Bouyakhf, in *2015 International Wireless Communications and Mobile Computing Conference (IWCMC)*. Learning gain mechanism to promote the femtocells hybrid access mode (IEEE, Dubrovnik, 2015), pp. 958–963
19. G Iosifidis, L Gao, J Huang, L Tassiulas, A double-auction mechanism for mobile data-offloading markets. *IEEE/ACM Trans. Networking*. **23**(5), 1634–1647 (2015)
20. L Zhang, T Jiang, K Luo, Dynamic spectrum allocation for the downlink of OFDMA-based hybrid-access cognitive femtocell networks. *IEEE Trans. Veh. Technol.* **65**(3), 1772–1781 (2016)
21. C Yang, J Li, M Guizani, Cooperation for spectral and energy efficiency in ultra-dense small cell networks. *IEEE Wirel. Commun.* **23**(1), 64–71 (2016)
22. D Liu, Y Chen, KK Chai, T Zhang, M ElKashlan, Opportunistic user association for multi-service hetnets using Nash bargaining solution. *IEEE Commun. Lett.* **18**(3), 463–466 (2014)
23. T Henderson, J Crowcroft, S Bhatti, Congestion pricing. Paying your way in communication networks. *IEEE Internet Comput.* **5**(5), 85–89 (2001)
24. B Al-Manthari, N Nasser, H Hassanein, Congestion pricing in wireless cellular networks. *IEEE Commun. Surv. Tutorials*. **13**(3), 358–371 (2011)
25. D Niu, B Li, in *2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*. Congestion-aware internet pricing for media streaming (IEEE, Toronto, 2014), pp. 571–576
26. M Khabazian, N Zorba, HS Hassanein, in *2013 IEEE 38th Conference on Local Computer Networks (LCN)*. Can dynamic pricing make femto users and service providers happy? (IEEE, Sydney, 2013), pp. 779–784
27. MH Cheung, R Southwell, J Huang, in *2014 48th Annual Conference on Information Sciences and Systems (CISS)*. Congestion-aware network selection and data offloading, (2014), pp. 1–6
28. L Wang, W Chen, J Li, in *2014 IEEE International Conference on Communications (ICC)*. Congestion aware dynamic user association in heterogeneous cellular network: a stochastic decision approach (IEEE, Sydney, 2014), pp. 2636–2640
29. C Gu, S Zhuang, Y Sun, C Liu, N Sun, in *2010 International Conference on Biomedical Engineering and Computer Science (ICBECS)*. Pricing incentive mechanism supporting load balancing for QoS-enabled networks (IEEE, Wuhan, 2010), pp. 1–4
30. Q Ye, B Rong, Y Chen, M Al-Shalash, C Caramanis, JG Andrews, User association for load balancing in heterogeneous cellular networks. *IEEE Trans. Wirel. Commun.* **12**(6), 2706–2716 (2013)
31. M Pioro, D Medhi, *Routing, Flow and Capacity Design in Communication and Computer Networks*. (Elsevier, San Francisco, 2004)
32. S Boyd, L Vandenberghe, *Convex Optimization*. (Cambridge University Press, New York, 2004)
33. R Jain, D Chiu, W Hawe, A quantitative measure of fairness and discrimination for resource allocation in shared computer systems. Digital Equipment Corporation, Tech. Rep. DEC-TR-301 (1984)
34. 3GPP, Evolved Universal Terrestrial Radio Access (E-UTRA); Further advancements for E-UTRA physical layer aspects (Release 9). TR 36.814, v9.0.0 (2010)

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Immediate publication on acceptance
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► springeropen.com