

Network Configuration for Two-tier Macro-Femto Systems with Hybrid Access

Binglai Niu and Vincent W.S. Wong, *Senior Member, IEEE*

Abstract—In this paper, we study uplink network configuration in a two-tier macro-femto heterogeneous system with hybrid access control. We consider a system where one macro base station and a cluster of adjacent femto base stations together serve a number of mobile users. In this system, base stations and users make decisions in various network configuration processes with different optimization objectives. Such decision making processes are usually correlated and an efficient mechanism is needed to coordinate the decision makers. In this work, we propose a five-stage network configuration mechanism where access control, resource allocation and power management are performed sequentially at the base stations and users, respectively. We show that this mechanism provides incentive for the FBSs to operate at the hybrid access mode. We model the configuration mechanism as a multi-stage decision making process and formulate a multi-level optimization problem. We analyze the problem in a bottom-up manner, and propose efficient algorithms to solve the optimization problem in each level sequentially. Simulation results show that the proposed network configuration mechanism achieves a higher system utility than configuration mechanisms with topology-based hybrid access or closed access.

Index Terms—Network configuration, macro-femto heterogeneous systems, hybrid access, multi-level optimization.

I. INTRODUCTION

THE increasing demand of mobile data service has triggered the evolution of cellular architecture from single-tier networks towards multi-tier heterogeneous networks [2]. Recently, there is an emerging trend of deploying low cost and low power base stations overlay the existing cellular infrastructure to extend the coverage area and increase the system capacity. In particular, in urban residential area, installing femto base stations (FBSs) in houses and apartments has been considered as an efficient approach to improve the indoor signal quality [3]. FBSs are inexpensive low power base stations referred as home nodeB or home eNodeB in the Third Generation Partnership Project (3GPP). FBSs are installed by users and can serve a small number of subscribers at the operator's licensed spectrum [4]. By deploying FBSs, home users' devices can experience better indoor signal quality and

longer battery life due to the close proximity. The operator can also offload the data traffic at the macro base stations (MBSs) to improve the system capacity [5]. Despite the promise, the deployment of FBSs also introduces technical challenges, such as interference management between the macrocell and femtocells and adaptive configuration of the system parameters [6]. Several approaches have been proposed to address the interference issue in heterogeneous networks, such as frequency planning, orthogonal resource allocation, and power control [7].

Recent attention on two-tier heterogeneous systems has been drawn to designing efficient network configuration mechanisms (i.e., resource allocation, power management) with access control at the FBSs [8]–[20]. Since FBSs are user deployed, typically they are configured to operate in closed access mode and only serve authorized subscribers, referred as femtocell users (FUs). However, it has been shown that allowing non-subscribers, i.e., macrocell users (MUs), to access FBSs can improve the overall system performance when the MUs experience poor signal quality from the MBS [8]. As a result, open access and hybrid access modes have been proposed to deal with this issue. In open access mode, MUs can access the FBSs freely and share the same resources with FUs. Thus, the available radio resources can be allocated in a flexible manner. The overall system performance can be improved at the expense of performance degradation at some FUs due to the altruistic resource sharing. In hybrid access mode, FBSs may grant access to MUs with limited resources while serving their subscribed FUs with higher priority [9]. This access mode aims to seek a balance between improving the overall system performance and preserving the quality of service (QoS) at the FUs, and has become a popular choice in designing novel network configuration mechanisms [10].

The emergence of open and hybrid access modes bring new challenges to designing network configuration mechanisms for heterogeneous networks. First, since an FBS is owned by its FUs, incentives should be provided to motivate the operation of open or hybrid access mode. In [11], a spectrum leasing mechanism has been proposed, where the MBS lease the spectrum to the FBSs, and the FBSs further lease their spectrum to nearby MUs. The mechanism is formulated as a Stackelberg game and the optimal spectrum leasing prices are determined via equilibrium analysis. Similar approaches have been employed in [12] and [13] with different optimization objectives. An access permission based incentive mechanism has been proposed in [14], where the macrocell service provider purchases access permissions from multiple competing femtocell service providers. In [15], the authors

Manuscript received on Mar 3, 2014; revised on August 21, 2014, February 14, 2015; and accepted on March 29, 2015. This work was supported by the Natural Sciences and Engineering Research Council (NSERC) of Canada. Part of this paper was presented at the *IEEE Global Communications Conference (GlobeCom'13)*, Atlanta, GA, Dec. 2013 [1]. The review of this paper was coordinated by Dr. Berk Canberk.

Copyright (c) 2015 IEEE. Personal use of this material is permitted. However, permission to use this material for any other purposes must be obtained from the IEEE by sending a request to pubs-permissions@ieee.org.

B. Niu and V. W. S. Wong are with the Department of Electrical and Computer Engineering, The University of British Columbia, Vancouver, BC, Canada, V6T 1Z4, e-mail: {bniu,vincentw}@ece.ubc.ca.

propose a spectrum-sharing rewarding framework to encourage hybrid access at FBSs, where femtocell owners determine the amount of resources shared with public users and the operator determines the ratio of revenue distribution to femtocell owners. The aforementioned pricing schemes are developed based on specific business models, which may not be available in the existing systems. In practice, an MU may pay access fee to the service provider instead of making separate payments to FBSs. Therefore, it is interesting to study how to encourage the FBSs to adopt hybrid access without changing the business model, which motivates the research in this paper.

In addition to providing incentives for using open or hybrid access mode at the FBSs, an efficient access control scheme is needed to determine the set of MUs to be served at each FBS. Several access control schemes have been proposed in recent works. In [16], FBSs grant access to the MUs who generate significant interference to the FUs in the uplink, so that the total number of interferers is reduced below a certain threshold. In [17], an MU is selected to be associated with its geographically closest FBS if the distance to this FBS is smaller than a threshold. However, these schemes are either heuristic or designed independently from other network configuration processes. In practice, the access control decisions at FBSs may affect other network configuration processes such as resource allocation at the MBS and power management at the users. For instance, when allocating channels, the MBS may need the user association decisions from the FBSs to optimize the overall network performance. Therefore, optimization of access control, resource allocation and power management should be jointly considered when designing the network configuration mechanism for heterogeneous systems. In [18], joint admission control and resource allocation has been studied using the theory of semi-Markov decision process, and a distributed power adaptation algorithm is proposed to reduce energy consumption at femtocells. In [19], the authors propose optimal access control, subcarriers allocation and power control algorithms in an orthogonal frequency division multiplexing (OFDM) macrocell/femtocell network assuming hybrid access mode at the FBSs. However, the incentive for private owners to use hybrid access is not discussed in the aforementioned work. The work in [20] proposed an on-demand waterfilling type channel allocation scheme with the consideration of user grouping for MBS and FBSs. The aforementioned joint optimization problems only consider the scenario where open access mode is used at the FBSs, while leaving the hybrid access control scenario unexplored. In our previous work [1], we proposed a resource management mechanism with hybrid access control for a macro-femto system. However, the mechanism is designed for a simple system with one MBS and one FBS, which limits its application in practice.

In this paper, we study uplink network configuration of a two-tier macro-femto system, where an MBS and a cluster of adjacent FBSs together serve mobile users with different QoS requirements. Different from existing works, we consider base stations and users are individual decision makers with different optimization objectives. We aim to design an efficient mechanism which provides incentive for using hybrid access at the FBSs, and jointly optimizes the access control, resource

allocation and power management processes. The major contributions are summarized as follows:

- We propose a five-stage network configuration mechanism for a two-tier macro-femto system. In this mechanism, FBSs make access control decisions at the beginning. Then, the scheduler at the operator allocates channels to the base stations adaptively according to the access control decisions. Next, the MBS and FBSs allocate their available resources to their associated users. Finally, the users determine their transmission power. We show that this mechanism provides incentives for the FBSs to operate in hybrid access mode without changing the business model.
- We model the decision making process in the proposed network configuration mechanism as a multi-level optimization problem and design efficient algorithms to find the optimal solution. In this problem, each level corresponds to a stage in the configuration mechanism, and the optimization in one level requires the knowledge of the optimality of the following levels. We analyze the multi-level optimization problem in a bottom-up manner, and develop efficient algorithms to find the solution for each level.
- We evaluate the performance of the proposed configuration mechanism under various scenarios. Simulation results show that our proposed mechanism achieves a higher system utility and throughput compared to the mechanisms with closed access scheme or a topology-based access control with orthogonal channel allocation scheme at the FBSs. The results also show that our proposed mechanism can provide incentives for using hybrid access mode at the FBSs.

The rest of this paper is organized as follows. In Section II, we describe the system model and network configuration mechanism, and formulate the multi-level optimization problem. In Section III, we analyze the multi-level optimization problem and design the network configuration mechanism based on the algorithm developed in each optimization level. Simulation results are presented in Section IV, and conclusions are drawn in Section V.

II. SYSTEM MODEL AND PROBLEM FORMULATION

We consider the uplink of a two-tier macro-femto wireless system, where an MBS and a cluster of $J > 1$ FBSs (which are close to each other) together serve a number of mobile users as shown in Fig. 1. This model represents the cellular network in a residential area (e.g., in a building) where FBSs are installed by home users to improve the service quality. In such networks, each FBS has a dedicated group of subscribed FUs, and the MBS has a number of subscribed MUs that are also close to the FBSs. The FBSs connect to the MBS and the operator's core network via wired backhaul links. We assume the capacity of the backhaul link for each FBS is large enough to handle the data transmission for its associated users. A central scheduler located at the operator's core network is responsible for allocating the available spectrum resources to all the base stations. We denote the set of base stations

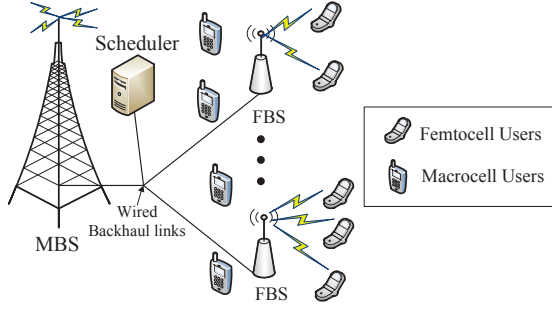


Fig. 1. Two-tier macro-femto system model: the MBS has a set of subscribed MUs and each FBS has a set of subscribed FUs.

as $\mathcal{I} = \{0, 1, \dots, J\}$, where 0 indicates the MBS. For base station $j \in \mathcal{I}$, we define the set of associated MUs as \mathcal{S}_j^m , and the set of associated FUs as \mathcal{S}_j^f . Thus, the set of all users associated with base station j is denoted as $\mathcal{S}_j = \mathcal{S}_j^m \cup \mathcal{S}_j^f$. The set of MUs in the system is denoted as $\mathcal{S}^M = \bigcup_{j \in \mathcal{I}} \mathcal{S}_j^m$. A list of key notations is provided in Table I.

We consider a time slotted system, where the time frame is divided into slots of equal length. We denote the set of available channels as $\mathcal{N} = \{1, 2, \dots, N\}$, where each channel has a bandwidth of B . We assume a frequency flat block fading wireless channel model [21], where for any of the N channels, the channel fading between a user and its base station remains constant during a time slot and is independent and identically distributed (i.i.d.) in different time slots with zero mean and unit variance complex Gaussian distribution. Since the MUs are close to the FBSs, to avoid the cross-tier interference (from the MUs to the FBSs and from the FUs to the MBS), the scheduler allocates orthogonal channels to the macrocell tier and the femtocell tier. However, in the femtocell tier, the same channel may be reused by multiple femtocells under the condition that the maximum aggregate interference over that channel at each FBS is less than a predefined threshold θ . Users associated with the same base station may share the available resources using time division multiple access, and at most one user in a femtocell can access a particular channel at a time slot.

Each FBS can serve at most L users and is configured to operate in the hybrid access mode. In this access mode, the FBS may allow some MUs to access it but reserves a proportion of the available resources for its own FUs. To facilitate the access control process, we introduce an access ratio $0 \leq \epsilon_j < 1$, which denotes the maximum proportion of available resources that FBS $j \in \mathcal{I} \setminus \{0\}$ can allocate to the MUs associated with it. Note that $\epsilon_j = 0$ corresponds to closed access mode where the FBSs only serve their own FUs. We assume users' devices have the capability to sense the signal quality from nearby base stations and may send requests to access the FBS that provides the best signal quality. We consider each user requests a service (e.g., video chatting) that requires a minimum average data rate (over a transmission period) to guarantee its QoS. We assume the total amount of resources in the system is sufficient to satisfy the QoS requirements for all users. We consider there is a network configuration phase to setup system parameters prior to users'

TABLE I
LIST OF KEY NOTATIONS

a_{ij}	Indicator for association between MU i and FBS j .
\mathbf{a}_j	User association profile at FBS j .
\mathcal{A}_j	Feasible set of association profiles at base station j .
$\tilde{\mathcal{A}}_j$	Set of association profiles with which FBS j obtains more resources than the maximum amount it needs.
B	Bandwidth of each channel.
$c_i(\cdot)$	Function that characterizes the energy consumption of user i 's device.
d	Distance between MBS and the region center where MUs and FBSs are located in the simulation.
$f_i(\cdot)$	Satisfaction function of user i defined over $[0, +\infty)$.
g_{ij}	Average channel gain from user i to base station j due to path loss.
h_{ij}	Instantaneous small scale channel fading from user i to base station j .
\mathcal{I}	Set of base stations including MBS and FBSs.
\mathcal{I}^k	Set of base stations to which channel k has been allocated.
J	Total number of FBSs in the system.
L	Maximum number of users an FBS can serve.
\mathcal{N}	Set of available channels for the system.
N	Total number of channels for the system.
n_j	Number of channels allocated to base station j .
n_j^c	Optimal number of channels allocated to base station j under closed access.
P_i	User i 's transmission power.
r_{ij}	Indicator for access request from MU i to FBS j .
$\bar{R}_{ij}, \tilde{R}_{ij}$	Average data rate and approximate data rate from user i to base station j .
\mathcal{S}^M	Set of MUs in the system.
$\mathcal{S}_j^m, \mathcal{S}_j^f$	Set of MUs and set of FUs associated with base station j , respectively.
T	Number of time slots in one transmission period.
u_{ij}	Utility of user i associated with base station j .
Δu_{ij}	Utility increment of user i associated with FBS j given one additional time slot.
U_j	Total utility at base station j .
ΔU_j	Base station j 's utility increment given one additional channel.
w_{ij}	The proportion of time that user i is allocated one channel from base station j .
w_{i0}^c	Optimal resource sharing factor for MU i under closed access.
x_j^k	Channel allocation indicator of channel k for base station j .
$\delta(k, \mathbf{x}^k)$	Interference condition indicator for channel k given channel allocation profile \mathbf{x}^k .
$\Delta \varphi_j$	Additional channels that base station j can obtain from MBS.
$\Phi_i(\cdot)$	Satisfaction function of user i defined over $[R_i^{\min}, +\infty)$.
ϵ_j	Access ratio for FBS j .
θ	Interference threshold at FBSs.
β	Weighting factor in utility function.
σ^2	Additive white Gaussian noise power.
ξ_j	Maximum interference allowed at base station j .

data transmission, which includes access control, resource allocation, and power management at the base stations and users. Such network configuration is performed periodically, i.e., every T time slots.

A. Utility Function

In this work, we define utility functions to characterize the objectives at the users and base stations. In practice, a wireless user prefers receiving high QoS with low energy consumption. Therefore, we define the utility function for user $i \in \mathcal{S}_j$ who

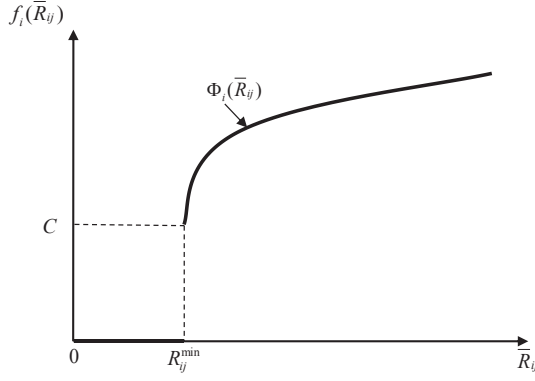


Fig. 2. User's satisfactory function $f_i(\bar{R}_{ij})$ with respect to the average data rate \bar{R}_{ij} . R_i^{\min} is the minimum rate requirement, C is a constant, and $\Phi_i(\cdot)$ is a concave function defined in $[R_i^{\min}, +\infty)$.

is associated with base station $j \in \mathcal{I}$ as

$$u_{ij} = f_i(\bar{R}_{ij}) - c_i(\bar{P}_i), \quad (1)$$

where $f_i(\cdot)$ characterizes the satisfaction of user i with respect to the average data rate over one transmission period (\bar{R}_{ij}), and $c_i(\cdot)$ is a function of the average transmission power (\bar{P}_i) that characterizes the energy consumption of user i 's device.

According to [22], the satisfaction of a user is an increasing function of the data rate, which has a decreasing marginal improvement as the data rate increases. Such property can be modeled using concave functions [23] [24]. In this work, to characterize users' satisfaction under different QoS requirements, we consider a satisfaction function $f_i(\cdot)$ as shown in Fig. 2. For user i associated with base station j , we have

$$f_i(\bar{R}_{ij}) = \begin{cases} \Phi_i(\bar{R}_{ij}), & \text{if } \bar{R}_{ij} \geq R_i^{\min}, \\ 0, & \text{otherwise,} \end{cases} \quad (2)$$

where $\Phi_i(\cdot)$ is a concave function defined in $[R_i^{\min}, \infty)$, and R_i^{\min} is the minimum average data rate required to guarantee the basic QoS for user i . To simplify the analysis, we assume users have the same satisfaction at their QoS threshold, where $\Phi_i(R_i^{\min}) = C, \forall i \in \mathcal{S}_j, \forall j \in \mathcal{I}$ and C is a constant that guarantees positive utility for users. Equation (2) implies that user i gains nothing when its data rate is less than the basic QoS requirement. Once the basic QoS is guaranteed, the user's satisfaction becomes a concave function of the data rate. Equation (2) can be used to characterize the user's satisfaction when running delay sensitive applications such as video chatting. Note that user's satisfaction function may vary with respect to different applications. In this paper, our analysis and results can be applied to satisfaction function defined in (2) with any concave function $\Phi_i(\cdot)$.

We assume each user chooses a fixed transmission power from $[0, P_{\max}]$ to optimize their utilities during the transmission phase (T time slots), that is $\bar{P}_i = P_i$ for $i \in \mathcal{S}_j, j \in \mathcal{I}$. We propose to use

$$c_i(P_i) = \beta w_{ij} P_i, \quad (3)$$

where β is a constant weighting factor and $w_{ij} \in \{\frac{1}{T}, \frac{2}{T}, \dots, 1\}$ is the resource sharing variable, which is defined as the proportion of time that user $i \in \mathcal{S}_j$ is granted access to base station $j \in \mathcal{I}$ during the transmission phase. β is

used to characterize the importance of the energy consumption in the utility function. It can be seen that when β is small such that $c_i(P_i) < f_i(\bar{R}_{ij})$, users care more about data rate rather than their energy consumption. However, when β is sufficiently large such that $c_i(P_i) > f_i(\bar{R}_{ij})$, energy consumption becomes more critical when making the transmission decisions. Therefore, we can achieve different optimization objectives by adjusting the weighting factor β . Note that in the configuration phase, the actual data rate of a user for the transmission phase cannot be achieved since the channel gain in the future T time slots is unknown. Therefore, we use an approximate data rate to calculate the utility when optimizing the decisions. Based on the channel capacity in [25], we approximate the average data rate for user i who communicates with base station j using the average channel gain between them, where the approximate data rate is

$$\begin{aligned} \tilde{R}_{ij} &= w_{ij} B \log_2 \left(1 + \frac{\mathbb{E}[g_{ij}|h_{ij}|^2] P_i}{\sigma^2 + \xi_j} \right) \\ &= w_{ij} B \log_2 \left(1 + \frac{g_{ij} P_i}{\sigma^2 + \xi_j} \right), \end{aligned} \quad (4)$$

where g_{ij} is the average channel gain from user i to base station j , h_{ij} is the corresponding small scale fading that follows a complex Gaussian distribution with zero mean and unit variance, σ^2 is the additive white Gaussian noise power, and ξ_j is the maximum interference allowed at base station j . Here, we use the fact that $\mathbb{E}[|h_{ij}|^2] = 1$ for complex Gaussian variable h_{ij} with unit variance. Note that in (4) we use the maximum allowed interference instead of the actual interference, since the actual interference experienced at a base station is unknown prior to the transmission phase. Since orthogonal channels are allocated to the macrocell tier and femtocell tier, there is no interference at the MBS. Therefore, we have

$$\xi_j = \begin{cases} \theta, & \text{if } j \in \mathcal{I} \setminus \{0\}, \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

Thus, the approximate data rate of user i to base station j is a function of the resource sharing variable w_{ij} and the transmission power P_i . In the remaining of the work, we use the approximate data rate to calculate the utility for each user when designing the network configuration mechanism.

We assume the MBS aims to optimize the total utility of the MUs. Each FBS allocates the resource to maximize the total utility of its own FUs, while guaranteeing the QoS of all its associated users (FUs and MUs). Since users receive wireless service from the same provider, and the wireless spectrum is limited, the operator may make channel allocation decisions to optimize the performance of the whole network, i.e., to maximize the system throughput [17] or total utility [26]. We consider the operator's objective at the scheduler is to optimize the total utility of all users in the system. Note that the proposed configuration mechanism and corresponding analysis in this work can also be applied to other utility functions with proper adjustment.

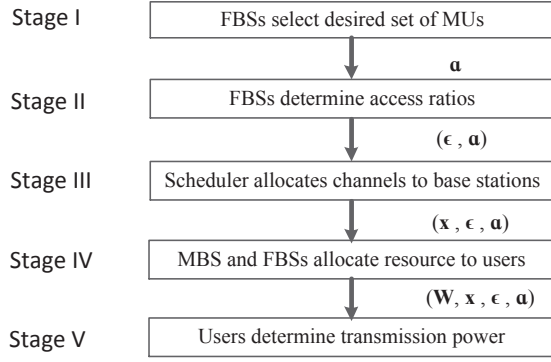


Fig. 3. Five-stage network configuration mechanism: Stages I and II correspond access control process, Stages II and IV are the resource allocation process. Stage V is the power management process.

B. Network Configuration Mechanism

In this work, we aim to design a network configuration mechanism that includes access control, resource allocation, and power management between base stations and users. We consider the FBSs perform access control at the beginning and the scheduler allocates the channels to the FBSs adaptively based on their access control decisions. This is to encourage the FBSs to use hybrid access mode to serve MUs who receive low QoS from the MBS. Specifically, if an FBS grants access to some MUs, it may obtain more channels and thus may improve its own utility. Note that to determine the transmission power that optimizes the utility, a user may require the knowledge of which base station it is associated with and the amount of resource it is granted from that base station. Therefore, it is reasonable to consider power management after the resource allocation process.

Based on the above discussion, we propose a network configuration mechanism that follows the procedure as shown in Fig. 3, which consists of five stages. The first two stages correspond to the access control process at the FBSs. In Stage I, after receiving the access requests from the MUs, each FBS j ($j \in \mathcal{I} \setminus \{0\}$) determines its association profile $\mathbf{a}_j = (a_{ij}, i \in \mathcal{S}^M)$, where $a_{ij} = 1$ indicates that MU i is granted access to FBS j . Note that each MU can only be associated with one base station, which leads to $\sum_{j \in \mathcal{I} \setminus \{0\}} a_{ij} \leq 1$. Then, FBS j determines its access ratio ϵ_j in Stage II. After the access control process, the FBSs send the access control decisions (including the association decision profile $\mathbf{A} = \{\mathbf{a}_j, j \in \mathcal{I} \setminus \{0\}\}$ and the access ratios $\epsilon = (\epsilon_j, \forall j \in \mathcal{I} \setminus \{0\})$ to the scheduler. In Stage III, the scheduler performs channel allocation. We define $x_j^k \in \{0, 1\}$ as the indicator whether channel k is allocated to base station j ($x_j^k = 1$) or not ($x_j^k = 0$). The scheduler determines the set of channel allocation profiles $\mathbf{x} = (\mathbf{x}_j, \forall j \in \mathcal{I})$, where $\mathbf{x}_j = (x_j^k, \forall k \in \mathcal{N})$ denotes the channel allocation profile at base station j . Then, the MBS and the FBSs allocate their available resources (channels and time slots) to their associated users in Stage IV, i.e., base station $j \in \mathcal{I}$ determines its resource allocation profile $\mathbf{w}_j = (w_{ij}, \forall i \in \mathcal{S}_j)$. Finally, each user i determines its transmission power P_i based on the available resource from its associated base station.

The proposed network configuration mechanism can be viewed as a multi-stage decision making process, where base

stations and users make decisions sequentially. Our objective is to design efficient decision making strategies for the base stations and users in order to optimize their own utilities.

C. Multi-level Optimization Problem

In the proposed network configuration mechanism, base stations and users have different objectives. Therefore, instead of formulating a single utility maximization problem which only targets at one objective, in this paper, we adopt a multi-level optimization approach [27] and formulate the network configuration process as a five-level optimization problem. The multi-level optimization approach models each decision making stage as an optimization problem with a specific objective function. It finds the optimal decision at one stage by considering the decisions made in the previous stages and the decisions to be made in the following stages. We define r_{ij} as the indicator whether FBS j receives a request from MU i , where $r_{ij} = 1$ indicates the request is received and $r_{ij} = 0$ otherwise. We define the set of feasible association decisions for FBS $j \in \mathcal{I} \setminus \{0\}$ as $\mathcal{A}_j \triangleq \{\mathbf{a}_j \mid a_{ij} \leq r_{ij}, \forall i \in \mathcal{S}^M, \sum_{i \in \mathcal{S}^M} a_{ij} \leq L - |\mathcal{S}_j^f|\}$, where $|\cdot|$ is the cardinality of the set. We further denote $\mathbf{P}_j = (P_i, i \in \mathcal{S}_j)$ for $j \in \mathcal{I} \setminus \{0\}$, and define $\epsilon_0 = 0$ and $\mathbf{a}_0 = (a_{i0}, i \in \mathcal{S}_0)$. We define U_j as the utility of base station $j \in \mathcal{I}$. As mentioned in Section II-A, the utility of MBS is the total utility of MUs, which is $U_0 = \sum_{i \in \mathcal{S}^M} u_{i0}$. The utility of FBS $j \in \mathcal{I} \setminus \{0\}$ is the total utility of its FUs, which is $U_j = \sum_{i \in \mathcal{S}_j^f} u_{ij}$. Note that a user's utility depends on its transmission power P_i and the amount of allocated resources w_{ij} , while the determination of P_i and w_{ij} is also related to the access control decision \mathbf{a}_j, ϵ_j and the channel allocation decision \mathbf{x}_j . Therefore, the utility at each base station $j \in \mathcal{I}$ is a function of the variables $\mathbf{a}_j, \epsilon_j, \mathbf{x}_j, \mathbf{w}_j$ and \mathbf{P}_j . In the following, we use $U_j(\mathbf{a}_j, \epsilon_j, \mathbf{x}_j, \mathbf{w}_j, \mathbf{P}_j)$ as the utility function at base station j during the problem formulation. Then, the multi-level optimization problem, denoted as problem \mathbb{P} , can be represented as

Level I : For $j \in \mathcal{I} \setminus \{0\}$,

maximize $U_j(\mathbf{a}_j, \epsilon_j, \mathbf{x}_j, \mathbf{w}_j, \mathbf{P}_j)$

subject to $\mathbf{a}_j \in \mathcal{A}_j$,

where ϵ_j ($j \in \mathcal{I} \setminus \{0\}$) is the solution of

Level II : For $j \in \mathcal{I} \setminus \{0\}$,

maximize $U_j(\mathbf{a}_j^*, \epsilon_j, \mathbf{x}_j, \mathbf{w}_j, \mathbf{P}_j)$

subject to $\epsilon_j^{\min} \leq \epsilon_j < 1$,

where \mathbf{x}_j ($j \in \mathcal{I} \setminus \{0\}$) is the solution of

Level III :

maximize $\sum_{j \in \mathcal{I}} U_j(\mathbf{a}_j^*, \epsilon_j^*, \mathbf{x}_j, \mathbf{w}_j, \mathbf{P}_j)$

subject to $n_j^{\min} \leq \sum_{k \in \mathcal{N}} x_j^k \leq n_j^{\max}, \quad \forall j \in \mathcal{I},$
 $x_j^k \sum_{l \in \mathcal{I} \setminus \{j, 0\}} x_l^k \max_{i \in \mathcal{S}_l} \{g_{il} P_{\max}\} \leq \theta,$
 $\quad \forall k \in \mathcal{N}, j \in \mathcal{I} \setminus \{0\},$

where \mathbf{w}_j ($j \in \mathcal{I}$) is the solution of

Level IV : For $j \in \mathcal{I}$,

maximize $U_j(\mathbf{a}_j^*, \epsilon_j^*, \mathbf{x}_j^*, \mathbf{w}_j, \mathbf{P}_j)$

subject to $\sum_{i \in \mathcal{S}_j} w_{ij} \leq \sum_{k \in \mathcal{N}} x_j^k,$
 $w_{ij} \geq w_{ij}^{\min}, \quad \forall i \in \mathcal{S}_j,$

where \mathbf{P}_j ($j \in \mathcal{I}$) is the solution of

Level V : For $i \in \mathcal{S}_j, j \in \mathcal{I}$,
 maximize $u_{ij}(w_{ij}^*, P_i)$
 subject to $P_i \in [0, P_{\max}]$,

where $a_j^*, \epsilon_j^*, \mathbf{x}_j^*$ and w_{ij}^* represent the optimal decisions at the corresponding optimization level. In the above formulation, at each level, we need to solve an optimization problem taking into account the parameter(s) obtained from the upper level(s) and the optimality of problem(s) in the following level(s). For example, in Level II, FBS j determines its access ratio ϵ_j based on the user association decision a_j obtained in Level I and the other optimal decisions (i.e., \mathbf{x}_j) obtained from lower levels. The determination of access ratio in this stage is important since ϵ_j works as an indicator for how many channels should be allocated to FBS j in order to guarantee the minimum QoS requirement(s) for the MU(s) accepted by FBS j . We require ϵ_j be no smaller than a certain value, denoted as ϵ_j^{\min} . It also indicates how much resources the FBS would like to share with its associated MUs. In Level III, n_j^{\min} and n_j^{\max} denote the minimum and maximum number of channels that can be allocated to FBS j , respectively. Given the association decisions from the FBSs, n_j^{\min} and n_j^{\max} should be properly selected to guarantee the feasibility of optimization in the following levels. The channel allocation decision at this level also satisfies the maximum allowed interference constraint at each FBS. In Level IV, w_{ij}^{\min} represents the minimum resource that user i requires from base station j to guarantee its QoS requirement. The determination of the parameters (ϵ_j^{\min} , n_j^{\min} , n_j^{\max} , and w_{ij}^{\min}) is discussed in Section III. The solution for each optimization problem corresponds to the optimal decision(s) for the decision maker(s) at that stage. Thus, our objective becomes designing efficient algorithms to find the optimal solution for each optimization level.

III. ANALYSIS OF THE MULTI-LEVEL OPTIMIZATION PROBLEM

In this section, we analyze the multi-level optimization problem and design efficient algorithms to achieve the optimal solution at each level. The multi-level optimization problem can be analyzed by a bottom-up approach [27], which follows the order from the lowest to the highest level sequentially.

A. Power Management at Users

We first consider the last level of problem \mathbb{P} . Our objective is to find the optimal transmission power for each user $i \in \mathcal{S}_j$ given the resource sharing variable w_{ij} from its associated base station $j \in \mathcal{I}$. In this level, the utility of a user becomes a function of the transmission power. The corresponding optimization problem is

$$\begin{aligned} & \text{maximize}_{P_i} u_{ij}(w_{ij}, P_i) \\ & \text{subject to } P_i \in [0, P_{\max}]. \end{aligned} \quad (6)$$

We analyze problem (6) with respect to different values of w_{ij} . Note that the minimum amount of resources that user i requires from base station j to guarantee its QoS can be calculated as $w_{ij}^{\min} = R_i^{\min} / (B \log_2(1 + \frac{g_{ij} P_{\max}}{\sigma^2 + \xi_j}))$. First, when $w_{ij} < w_{ij}^{\min}$, according to equations (1), (2) and (3), we have

$u_{ij}(P_i) = -\beta w_{ij} P_i$, and the optimal power is $P_i^* = 0$ in this scenario. Next, when $w_{ij} \geq w_{ij}^{\min}$, we define $P_i^{\min} \leq P_{\max}$ as user i 's minimum power that satisfies its QoS requirement, which is given by $P_i^{\min} = \frac{\sigma^2 + \xi_j}{g_{ij}} (2^{R_i^{\min}/(w_{ij} B)} - 1)$. In this scenario, to solve problem (6), we only need to consider the feasible region $[P_i^{\min}, P_{\max}]$. For $P_i \in [P_i^{\min}, P_{\max}]$, we have $u_{ij}(P_i) = \Phi_i(\tilde{R}_{ij}) - \beta w_{ij} P_i$ and

$$\begin{aligned} \frac{du_{ij}}{dP_i} &= \frac{d\Phi_i}{d\tilde{R}_{ij}} \frac{d\tilde{R}_{ij}}{dP_i} - \beta w_{ij} \\ &= w_{ij} \left(\frac{d\Phi_i}{d\tilde{R}_{ij}} \frac{Bg_{ij}}{\ln 2(\sigma^2 + \xi_j + g_{ij} P_i)} - \beta \right). \end{aligned} \quad (7)$$

Note that $\Phi_i(\cdot)$ is a concave increasing function of \tilde{R}_{ij} , which implies that $\frac{d\Phi_i}{d\tilde{R}_{ij}}$ is a decreasing function of \tilde{R}_{ij} (or $\frac{d^2\Phi_i}{d\tilde{R}_{ij}^2} < 0$). Since \tilde{R}_{ij} is increasing with P_i , $\frac{d\Phi_i}{d\tilde{R}_{ij}}$ is a decreasing function of P_i . From (7), we conclude that $\frac{du_{ij}}{dP_i}$ is also decreasing with P_i in (P_i^{\min}, P_{\max}) . Based on the decreasing property of $\frac{du_{ij}}{dP_i}$, if $\frac{du_{ij}}{dP_i} \Big|_{P_i=P_i^{\min}} \leq 0$, we have $\frac{du_{ij}}{dP_i} < 0$ for $P_i \in (P_i^{\min}, P_{\max})$. Therefore, the optimal power in this case is $P_i^* = P_i^{\min}$. On the other hand, if $\frac{du_{ij}}{dP_i} \Big|_{P_i=P_i^{\min}} > 0$, since $\frac{du_{ij}}{dP_i} \Big|_{P_i \rightarrow \infty} = -\beta w_{ij} < 0$ and $\frac{du_{ij}}{dP_i}$ is decreasing with P_i , equation $\frac{du_{ij}}{dP_i} = 0$ has a unique root $P_i = \tilde{P}_i$ in (P_i^{\min}, ∞) , where we have $\frac{du_{ij}}{dP_i} > 0$ in $(P_i^{\min}, \tilde{P}_i)$ and $\frac{du_{ij}}{dP_i} < 0$ in (\tilde{P}_i, ∞) . Therefore, in this case, the optimal power is $P_i^* = \min(\tilde{P}_i, P_{\max})$. In summary, the optimal solution to problem (6) can be represented as

$$\begin{aligned} & P_i^* \\ &= \begin{cases} P_i^{\min} & \text{if } \frac{du_{ij}}{dP_i} \Big|_{P_i=P_i^{\min}} \leq 0 \\ \min(\tilde{P}_i, P_{\max}) & \text{if } \frac{du_{ij}}{dP_i} \Big|_{P_i=P_i^{\min}} > 0 \end{cases} \\ & \quad \text{otherwise,} \end{cases} \quad (8)$$

where indicator function $\mathbf{I}_{\mathcal{X}} = 1$ if \mathcal{X} is true, and is equal to zero otherwise. It can be seen that given the resource sharing factor w_{ij} from base station $j \in \mathcal{I}$, the optimal power management decision for user $i \in \mathcal{S}_j$ is provided in (8).

B. Resource Allocation at Base Stations

Level IV of problem \mathbb{P} corresponds to resource allocation at the base stations. As mentioned in Section II, the channel gain between a user and a base station is i.i.d.. Therefore, when optimizing the resource allocation decisions, we only determine the resource sharing variables w_{ij} , but do not consider which channel or time slots are allocated to each user. After determining the resource sharing variables, the number of time slots for each user is fixed. The set of time slots for each user can be determined by applying a scheduling algorithm, such as round robin scheduling algorithm or random scheduling algorithm. We first consider resource allocation at FBS j given its access control decision (a_j and ϵ_j) and the channel allocation decision \mathbf{x}_j from the scheduler. Note that we require

the FBS's decision guarantee the QoS requirements of all its associated users, and we assume such decision is feasible. The optimal resource allocation decision $\mathbf{w}_j = (w_{ij}, \forall i \in \mathcal{S}_j)$ that maximizes FBS j 's utility can be found by solving the following problem:

$$\begin{aligned} & \underset{w_{ij}, i \in \mathcal{S}_j}{\text{maximize}} && \sum_{i \in \mathcal{S}_j^f} u_{ij}(w_{ij}, P_i^*(w_{ij})) \\ & \text{subject to} && \sum_{i \in \mathcal{S}_j^m} w_{ij} \leq \epsilon_j \sum_{k \in \mathcal{N}} x_j^k, \\ & && \sum_{i \in \mathcal{S}_j^f} w_{ij} + \sum_{i \in \mathcal{S}_j^m} w_{ij} \leq \sum_{k \in \mathcal{N}} x_j^k, \\ & && w_{ij} \geq w_{ij}^{\min}, \quad \forall i \in \mathcal{S}_j, \\ & && w_{ij} \in \left\{ \frac{1}{T}, \frac{2}{T}, \dots, 1 \right\}, \quad \forall i \in \mathcal{S}_j, \end{aligned} \quad (9)$$

where $P_i^*(w_{ij})$ is the optimal transmission power for user i given w_{ij} . The first constraint indicates that the FBS allocates at most ϵ_j of the total resource to the MUs. The second constraint implies that the total amount of resources allocated to users should not exceed the total number of channels available at the FBS. The third constraint guarantees the QoS of all users associated with this FBS. The last constraint implies that a user can only obtain an integer number of time slots and cannot access more than one channel simultaneously. Since $P_i^*(w_{ij})$ also depends on w_{ij} , from (7) and (8), it can be seen that problem (9) is a nonlinear discrete optimization problem, which is hard to solve in general. In the rest of this section, we design an efficient algorithm to achieve the optimal solution.

Since the objective function in problem (9) is only related to FUs associated with FBS j , problem (9) can be solved in two steps, where we first determine the resource sharing variables for MUs associated with FBS j , and then we determine the resource sharing variables for FUs. Intuitively, to maximize FUs' total utility, the FBS should allocate as much resources as possible to its FUs, or equivalently, to minimize the amount of resources allocated to the MUs while satisfying their QoS constraints. Therefore, in the first step, the optimal resource sharing variables for the MUs can be found by solving the following subproblem:

$$\begin{aligned} & \underset{w_{ij}, i \in \mathcal{S}_j^m}{\text{minimize}} && \sum_{i \in \mathcal{S}_j^m} w_{ij} \\ & \text{subject to} && \sum_{i \in \mathcal{S}_j^m} w_{ij} \leq \epsilon_j \sum_{k \in \mathcal{N}} x_j^k, \\ & && w_{ij} \geq w_{ij}^{\min}, \quad \forall i \in \mathcal{S}_j^m, \\ & && w_{ij} \in \left\{ \frac{1}{T}, \frac{2}{T}, \dots, 1 \right\}, \quad \forall i \in \mathcal{S}_j^m. \end{aligned} \quad (10)$$

In the second step, with the solution to subproblem (10), w_{ij}^* , $\forall i \in \mathcal{S}_j^m$, problem (9) is transformed into the following subproblem:

$$\begin{aligned} & \underset{w_{ij}, i \in \mathcal{S}_j^f}{\text{maximize}} && \sum_{i \in \mathcal{S}_j^f} u_{ij}(w_{ij}, P_i^*(w_{ij})) \\ & \text{subject to} && \sum_{i \in \mathcal{S}_j^f} w_{ij} \leq \sum_{k \in \mathcal{N}} x_j^k - \sum_{i \in \mathcal{S}_j^m} w_{ij}^*, \\ & && w_{ij} \geq w_{ij}^{\min}, \quad \forall i \in \mathcal{S}_j^f, \\ & && w_{ij} \in \left\{ \frac{1}{T}, \frac{2}{T}, \dots, 1 \right\}, \quad \forall i \in \mathcal{S}_j^f. \end{aligned} \quad (11)$$

By solving subproblem (11), we obtain the optimal resource sharing variables for the FUs. It can be seen that the solu-

Algorithm 1 Optimal resource allocation algorithm at FBS j with fixed user association

-
- 1: **Basic resource allocation:**
 - 2: Calculate the minimum amount of resource required by each user, $w_{ij}^{\min}, \forall i \in \mathcal{S}_j$.
 - 3: Set $w_{ij} := w_{ij}^{\min}, \forall i \in \mathcal{S}_j$.
 - 4: **Remaining resource allocation:**
 - 5: Determine the total remaining time slots $T_R := T(\sum_{k \in \mathcal{N}} x_j^k - \sum_{i \in \mathcal{S}_j} w_{ij}^{\min})$.
 - 6: Calculate the initial value of $\Delta u_{ij}(w_{ij} + 1/T), \forall i \in \mathcal{S}_j^f$.
 - 7: **Repeat**
 - 8: Find the set of FUs, \mathcal{S}_t , that satisfies $w_{ij} < 1, \Delta u_{ij}(w_{ij} + 1/T) > 0, \forall i \in \mathcal{S}_t$.
 - 9: **if** $\mathcal{S}_t \neq \emptyset$
 - 10: Find user $i \in \mathcal{S}_t$ that has the largest $\Delta u_{ij}(w_{ij} + 1/T)$ and set $w_{ij} := w_{ij} + 1/T, T_R := T_R - 1$.
 - 11: Update $\Delta u_{ij}(w_{ij} + 1/T)$ for user i .
 - 12: **endif**
 - 13: **Until** $T_R = 0$ or $\mathcal{S}_t = \emptyset$.
-

tions to subproblems (10) and (11) constitutes the solution to problem (9). We define $\Delta u_{ij}(w_{ij}) \triangleq u_{ij}(w_{ij}, P_i^*(w_{ij})) - u_{ij}(w_{ij} - \frac{1}{T}, P_i^*(w_{ij} - \frac{1}{T}))$ as the changes of utility when one time slot is assigned to user i . Based on the previous discussion, we propose a two-step greedy allocation algorithm to find the optimal solution for (9), as shown in Algorithm 1. In the first step, we allocate the minimum number of time slots ($w_{ij}^{\min}T$) for each user to satisfy their QoS constraints. Intuitively, this solves subproblem (10). In the second step, we allocate the remaining time slots one by one, where each time slot is allocated to the FU who has the largest utility increment ($\Delta u_{ij}(w_{ij})$). With Algorithm 1, the following result can be obtained.

Theorem 1: The resource allocation profile \mathbf{w}_j obtained using Algorithm 1 constitutes the optimal solution to problem (9).

The proof of Theorem 1 is provided in Appendix A. For the MBS, the resource allocation algorithm is similar to that at the FBS, except that the MBS only serves MUs. This algorithm can be obtained by replacing the terms FU by MU, and \mathcal{S}_j^f by \mathcal{S}_0 , respectively, in Algorithm 1. Thus, we have found the desired strategy for FBSs and MBS in Level IV. Note that we can apply other objective functions to realize alternative optimization objectives. For example, we can use $\sum_{i \in \mathcal{S}_j} \log(R_{ij})$ to achieve proportional fairness of data rate among the users. The optimality of the proposed algorithm is still valid as long as the objective function is concave with respect to the data rate.

C. Channel Allocation at MBS

Next, we consider the channel allocation problem in Level III of problem \mathbb{P} . In this level, the scheduler allocates the available channels (N) to each base station in order to maximize the total system utility. The channel allocation decision should guarantee the feasibility of the decisions in the following levels, which is to guarantee the QoS requirement for each user in the system. To achieve this, we require that the number of channels allocated to FBS $j \in \mathcal{I}$, defined as $n_j \triangleq \sum_{k \in \mathcal{N}} x_j^k$, satisfies $n_j^{\min} \leq n_j \leq n_j^{\max}$. We define n_j^{\min} and n_j^{\max} based

on the access control decisions at base station j as follows. First, since we assume each user can only access one channel at a time, the total number of channels allocated to a base station should not exceed the number of its associated users. Therefore, it is reasonable to define $n_j^{\max} = |\mathcal{S}_j|$. For base station j which only serves its own subscribed users (either MUs or FUs), n_j^{\min} is defined as $n_j^{\min} = \lceil \sum_{i \in \mathcal{S}_j} w_{ij}^{\min} \rceil$, where $\lceil \cdot \rceil$ is the ceiling function. For FBS j , which serves both FUs and MUs, to guarantee the QoS of all its associated users, n_j^{\min} should satisfy

$$\epsilon_j n_j^{\min} \geq \sum_{i \in \mathcal{S}_j^m} w_{ij}^{\min}, \quad \text{and} \quad (1 - \epsilon_j) n_j^{\min} \geq \sum_{i \in \mathcal{S}_j^f} w_{ij}^{\min}. \quad (12)$$

That is $n_j^{\min} \geq \max(\lceil \frac{1}{\epsilon_j} \sum_{i \in \mathcal{S}_j^m} w_{ij}^{\min} \rceil, \lceil \frac{1}{1 - \epsilon_j} \sum_{i \in \mathcal{S}_j^f} w_{ij}^{\min} \rceil)$. Note that if FBS j intentionally selects a small ϵ_j that is close to zero, n_j^{\min} may become a large number, which makes the allocation problem infeasible. To prevent such behavior, we set an upper bound for n_j^{\min} as follows. We denote n_j^c as the optimal number of channels allocated to base station j when all base stations use closed access mode, and represent the corresponding optimal resource sharing variable for MU $i \in \mathcal{S}^M$ (associated with MBS) as w_{i0}^c . Then, for FBS j who accepts to serve the set of MUs \mathcal{S}_j^m , the minimum number of channels to be allocated satisfies

$$n_j^{\min} \leq n_j^c + \lfloor \sum_{i \in \mathcal{S}_j^m} w_{i0}^c \rfloor. \quad (13)$$

Inequality (13) implies that if FBS j accepts to serve MUs, it is guaranteed to obtain up to $\lfloor \sum_{i \in \mathcal{S}_j^m} w_{i0}^c \rfloor$ additional channels from the scheduler (depending on the selected access ratio). Note that allocating these additional channels to the FBS does not reduce the number of channels allocated to other base stations when the set of MUs switch from accessing the MBS to accessing the FBS.

Based on the previous discussion, we determine n_j^{\min} as

$$n_j^{\min} = \begin{cases} \min\{\max(\lceil \frac{1}{\epsilon_j} \sum_{i \in \mathcal{S}_j^m} w_{ij}^{\min} \rceil, \lceil \frac{1}{1 - \epsilon_j} \sum_{i \in \mathcal{S}_j^f} w_{ij}^{\min} \rceil), \\ n_j^c + \lfloor \sum_{i \in \mathcal{S}_j^m} w_{i0}^c \rfloor\}, & \text{if } \epsilon_j > 0, \\ \lceil \sum_{i \in \mathcal{S}_j} w_{ij}^{\min} \rceil, & \text{otherwise.} \end{cases} \quad (14)$$

With the above definition, the optimal channel allocation strategy that maximizes the system utility can be found by solving the following problem

$$\begin{aligned} & \underset{\mathbf{x}_j, j \in \mathcal{I}}{\text{maximize}} \quad \sum_{j \in \mathcal{I}} \sum_{i \in \mathcal{S}_j} u_{ij}((w_{ij}^*(\mathbf{x}_j), P_i^*(w_{ij}^*(\mathbf{x}_j))) \\ & \text{subject to} \quad n_j^{\min} \leq \sum_{k \in \mathcal{N}} x_j^k \leq n_j^{\max}, \quad \forall j \in \mathcal{I}, \\ & \quad x_j^k \sum_{i \in \mathcal{S}_j} x_i^k \max_{i \in \mathcal{S}_i} \{g_{ij} P_{\max}\} \leq \theta, \\ & \quad \quad \quad \forall k \in \mathcal{N}, j \in \mathcal{I} \setminus \{0\}, \\ & \quad x_j^k \in \{0, 1\}, \quad \forall k \in \mathcal{N}, j \in \mathcal{I}, \end{aligned} \quad (15)$$

where $P_i^*(w_{ij}^*(\mathbf{x}_j))$ and $w_{ij}^*(\mathbf{x}_j)$ can be found for a given \mathbf{x}_j using algorithms in Sections III-A and III-B, respectively. Note that x_j^k takes integer values and the objective function in (15)

does not have an explicit form with respect to x_j^k . Therefore, it is difficult to find the optimal solution. Similar to Algorithm 1, we propose a greedy allocation algorithm to find a suboptimal solution.

We define $\Delta U_j(n_j) \triangleq U_j(n_j) - U_j(n_j - 1)$ as the utility increment by assigning one additional channel to base station j . We define $\delta(k, \mathbf{x}^k) \in \{0, 1\}$ as the indicator whether the interference constraint for all FBSs over channel k is satisfied or not, given the channel allocation profile $\mathbf{x}^k = (x_j^k, j \in \mathcal{I} \setminus \{0\})$. We further denote $\mathbf{x}^k|_j$ as the channel allocation profile obtained by setting $x_j^k = 1$ from \mathbf{x}^k , and $\delta(k, \mathbf{x}^k|_j)$ indicates whether the interference constraints are satisfied when allocating channel k to FBS j . We determine the value of $\delta(k, \mathbf{x}^k|_j)$ as follows. At the beginning of the channel allocation process, we determine a $J \times J$ matrix Γ^f , where $\Gamma_{i,j}^f$ represents the maximum interference from users associated with FBS i to FBS j . That is, $\Gamma_{i,j}^f = \max_{u \in \mathcal{S}_i} \{P_{u g_{uj}}\}$. Note that Γ^f characterizes the interference between any pair of femtocells and $\Gamma_{i,j}^f$ is different from $\Gamma_{j,i}^f$. We also maintain a $J \times N$ interference matrix Γ^c , where $\Gamma_{j,k}^c$ represents the aggregate interference experienced at FBS j on channel k . To determine $\delta(k, \mathbf{x}^k|_j)$ when allocating a channel k to an FBS j , we first find the set of FBSs that have already been allocated channel k , denoted as \mathcal{I}^k . Then, we calculate the temporary aggregate interference for FBS $i \in \mathcal{I}^k$ over channel k assuming channel k is allocated to FBS j , which is $\Gamma_{i,k}^{c'} = \Gamma_{i,k}^c + \Gamma_{j,i}^f$, where we add the new interference generated from users associated with FBS j ($\Gamma_{j,i}^f$) to existing interference at FBS $i \in \mathcal{I}^k$ ($\Gamma_{i,k}^c$). We also calculate the maximum aggregate interference experienced at FBS j from other femtocells on channel k , which is $\Gamma_{j,k}^{c'} = \sum_{i \in \mathcal{I}^k} \Gamma_{i,j}^f$. Then, the interference condition indicator can be determined as

$$\delta(k, \mathbf{x}^k|_j) = \begin{cases} 1, & \text{if } \Gamma_{i,k}^{c'} < \theta, \forall i \in \mathcal{I}^k \cup \{j\}, \\ 0, & \text{otherwise.} \end{cases} \quad (16)$$

Once channel k is allocated to FBS j , we update $\Gamma_{i,k}^c$ and $\Gamma_{j,i}^f$ $\forall i, j \in \mathcal{I}$ accordingly. By introducing the indicator function $\delta(k, \mathbf{x}^k|_j)$, the proposed greedy channel allocation algorithm is shown in Algorithm 2. The main idea of Algorithm 2 is to allocate the channels one by one, where each channel is allocated to as many base stations as possible under the interference constraints. The available channels are indexed from 1 to N . Since MBS and FBSs use different set of channels, we allocate the channels to MBS starting from index N in a decreasing order, while allocating channels to FBSs starting from index 1 in an increasing order.

Algorithm 2 contains two main steps. In the first step, we allocate the minimum number of channels to each base station so that $n_j = n_j^{\min}, \forall j \in \mathcal{I}$. Specifically, we allocate n_0^{\min} channels, which indexed from N to $N - n_0^{\min} + 1$, to the MBS in Line 3. Then, we allocate the minimum number of channels to FBSs in Lines 4 to 14 starting from channel indexed at 1, where for each channel, we try to allocate it to as many FBSs as possible by checking the interference constraint for each FBS (from Lines 8 to 12). After allocating the minimum number of channels, in the second step, we allocate the remaining

Algorithm 2 Greedy channel allocation algorithm

```

1: Basic channel allocation:
2: Calculate  $n_j^{\min}$  and  $n_j^{\max}$ ,  $\forall j \in \mathcal{I}$ .
3: Set  $x_0^k := 1, \forall k = N - n_0^{\min} + 1, \dots, N$ .
4: Set  $k := 1$ .
5: Find the set of FBSs  $\mathcal{I}_b$  that satisfies  $n_j < n_j^{\min}, \forall j \in \mathcal{I}_b$ .
6: while(  $\mathcal{I}_b \neq \emptyset$  )
7:   Set  $k := k + 1$ .
8:   for each FBS  $j \in \mathcal{I}_b$ 
9:     if  $\delta(k, \mathbf{x}^k|_j) = 1$ 
10:      Set  $x_j^k := 1, n_j := n_j + 1$ .
11:     endif
12:   endfor
13:   Find the set of FBSs  $\mathcal{I}_b$  that satisfies  $n_j < n_j^{\min}, \forall j \in \mathcal{I}_b$ .
14: endwhile
15: Remaining channel allocation:
16: Set the starting point of remaining channels for FBS and MBS
    $n_{\text{FBS}} := 2, n_{\text{MBS}} := N - n_0^{\min}$ .
17: Calculate the initial value of  $\Delta U_j(n_j + 1), \forall j \in \mathcal{I}$  based on
   Algorithm 1.
18: Find the base station set  $\mathcal{I}'_b$  that satisfies  $n_j < n_j^{\max}$  and
    $\Delta U_j(n_j + 1) > 0$ .
19: while(  $\mathcal{I}'_b \neq \emptyset$  and  $n_{\text{FBS}} < n_{\text{MBS}}$  )
20:   Find base station  $j \in \mathcal{I}'_b$  that has the largest  $\Delta U_j(n_j + 1)$ .
21:   if  $j = 0$  and  $n_{\text{MBS}} > k$ 
22:     Set  $x_0^{n_{\text{MBS}}} := 1, n_0 := n_0 + 1, n_{\text{MBS}} := n_{\text{MBS}} - 1$ .
23:     Update  $\Delta U_0(n_0 + 1)$  by computing  $U_0(n_0 + 1)$  and  $U_0(n_0)$ 
       according to Algorithm 1.
24:   else
25:     for FBS  $j$  in decreasing order with respect to  $\Delta U_j(n_j + 1)$ 
26:       if  $\delta(n_{\text{FBS}}, \mathbf{x}^{n_{\text{FBS}}}|_j) = 1$  and  $x_j^{n_{\text{FBS}}} = 0$ 
27:         Set  $x_j^{n_{\text{FBS}}} := 1, n_j := n_j + 1$ .
28:         Update  $\Delta U_j(n_j + 1)$  by computing  $U_j(n_j + 1)$  and
            $U_j(n_j)$  according to Algorithm 1.
29:       endif
30:     endfor
31:   endif
32:   Set  $n_{\text{FBS}} := n_{\text{FBS}} + 1$ .
33:   Find the base station set  $\mathcal{I}'_b$  that satisfies  $n_j < n_j^{\max}$  and
      $\Delta U_j(n_j + 1) > 0, \forall j \in \mathcal{I}'_b$ .
34: endwhile

```

channels one by one to the MBS and FBSs, respectively, from Lines 16 to 33. We introduce two variables n_{FBS} and n_{MBS} to denote the index of channels to be allocated to FBSs and MBS, respectively. We initialize $n_{\text{FBS}} = 2$ in Line 16. This is because in the basic allocation process we only consider those FBSs whose minimum requirements are not satisfied, and the channels indexed from 2 to k may still be allocated to some FBSs without violating the interference constraints. We initialize $n_{\text{MBS}} = N - n_0^{\min}$ and allocate channels to MBS following a decreasing order of channel index. Then, in each iteration, we find the base station with the largest utility increment assuming one additional channel is allocated to it in Line 20. If it is the MBS, we allocate channel n_{MBS} to MBS in Lines 22 to 23. Otherwise, we allocate channel n_{FBS} to as many FBSs as possible under the interference constraints from Lines 25 to 30 according to the order of utility increment at the FBSs. The allocation terminates when no more channels are available ($n_{\text{FBS}} \geq n_{\text{MBS}}$) or all the base stations have been fully allocated ($n_j = n_j^{\max}, \forall j \in \mathcal{I}$).

Note that the aforementioned closed access scenario (where $\epsilon_j = 0, \forall j \in \mathcal{I}$) is a special case of the hybrid access scenario

and the values of n_j^c ($\forall j \in \mathcal{I}$) and w_{ij}^c ($\forall i \in \mathcal{S}_j, j \in \mathcal{I}$) can be determined by applying Algorithm 2 at the beginning of the network configuration (before the access control stages). We denote $\mathcal{N}_{\text{FBS}}^c = \{1, \dots, n_{\text{FBS}}^c\}$ and $\mathcal{N}_{\text{MBS}}^c = \{n_{\text{MBS}}^c, \dots, N\}$ as the set of channels allocated to FBSs and the MBS in the closed access scenario, respectively. Then, in the hybrid access scenario, we only need to determine the additional channels that should be reallocated from the MBS to FBSs who accept to serve MUs, denoted as N_a , where $N_a = \sum_{j \in \mathcal{I} \setminus \{0\}} (n_j^{\min} - n_j^c)$ according to (14). Then, we allocate these N_a channels to FBSs using Algorithm 2 by setting $k = n_{\text{FBS}}^c + 1$, $n_{\text{FBS}} = n_{\text{FBS}}^c + 1$ and $n_{\text{MBS}} = n_{\text{MBS}}^c + N_a$ in Line 4 and Line 16, respectively. Thus, the desired channel allocation strategy is provided in Algorithm 2.

D. Access Control at FBSs

In this subsection, we study the first two levels of problem \mathbb{P} , and find the desired access control decisions for FBSs assuming the resource allocation and power management algorithms proposed in previous sections are adopted. We first determine the access ratio ϵ_j at FBS $j \in \mathcal{I} \setminus \{0\}$. Obviously, if FBS j does not serve any MU, its access ratio ϵ_j should be set to zero. If FBS j serves both FUs and MUs, according to (12), given the set of MUs associated with FBS j (\mathcal{S}_j^m), a smaller access ratio ϵ_j may result in a larger number of channels allocated to this FBS (n_j), since the channel allocation should satisfy $\epsilon_j n_j \geq \sum_{i \in \mathcal{S}_j^m} w_{ij}^{\min}$. Moreover, a smaller access ratio means sharing fewer resources with the MUs. Therefore, in order to maximize the utility, FBS j prefers selecting the minimum achievable access ratio. However, an FBS should also guarantee the minimum QoS requirements for its associated MUs. According to (13), the maximum number of channels that FBS j can be guaranteed is $n_j^c + \lfloor \sum_{i \in \mathcal{S}_j^m} w_{i0}^c \rfloor$. Thus, the minimum access ratio for FBS j to guarantee its associated MUs' QoS requirements is $\epsilon_j^{\min} = \frac{\sum_{i \in \mathcal{S}_j^m} w_{ij}^{\min}}{n_j^c + \lfloor \sum_{i \in \mathcal{S}_j^m} w_{i0}^c \rfloor}$. Note that n_j^c and w_{i0}^c can be calculated by the scheduler and broadcast to the FBSs and MUs at the beginning of the configuration period. In summary, the desired strategy for base station $j \in \mathcal{I}$ in Level II is

$$\epsilon_j^* = \begin{cases} 0, & \text{if } \mathcal{S}_j^m = \emptyset, \\ \epsilon_j^{\min}, & \text{otherwise.} \end{cases} \quad (17)$$

Finally, we study the user association decision at FBS j when multiple access requests are received from the MUs. As mentioned before, the FBSs are selfish and they select the desired MUs to maximize their own utilities. For FBS j , the optimal user association decision can be obtained by solving the following problem

$$\begin{aligned} & \underset{\mathbf{a}_j}{\text{maximize}} && U_j \\ & \text{subject to} && \mathbf{a}_j \in \mathcal{A}_j, \end{aligned} \quad (18)$$

where $U_j = \sum_{k \in \mathcal{S}_j^f} u_{kj}(w_{kj}^*(\mathbf{x}_j^*(\epsilon_j^*(\mathbf{a}_j), \mathbf{a}_j)), P_k^*(w_{kj}^*(\mathbf{x}_j^*(\epsilon_j^*(\mathbf{a}_j), \mathbf{a}_j))))$ and \mathcal{A}_j is the set of feasible association profiles defined in Section II-C. $P_k^*(w_{kj}^*(\mathbf{x}_j^*(\epsilon_j^*(\mathbf{a}_j), \mathbf{a}_j))), w_{kj}^*(\mathbf{x}_j^*(\epsilon_j^*(\mathbf{a}_j), \mathbf{a}_j)), \mathbf{x}_j^*(\epsilon_j^*(\mathbf{a}_j), \mathbf{a}_j)$ and $\epsilon_j^*(\mathbf{a}_j)$ are obtained using strategies proposed in Sections III-A, III-B, and III-C,

and equation (17), respectively, for a given association profile \mathbf{a}_j . Note that the FBS prefers obtaining more resources from the MBS to serve its FUs. The optimal values of \mathbf{a}_j in (18) can be found by searching among the association profiles in \mathcal{A}_j and selecting the one that brings the largest amount of additional resources to the FBS. Specifically, for each association profile, i.e., \mathbf{a}_j , the guaranteed number of channels allocated to FBS j is $n_j = n_j^c + \lfloor \sum_{i \in \mathcal{S}^M} a_{ij} w_{i0}^c \rfloor$, and the additional resources FBS j obtains by accepting MUs can be represented as $\Delta\varphi_j(\mathbf{a}_j) \triangleq \lfloor \sum_{i \in \mathcal{S}^M} a_{ij} w_{i0}^c \rfloor - \sum_{i \in \mathcal{S}^M} a_{ij} w_{ij}^{\min}$, where the first term is the additional channels obtained from the MBS and the second term is the amount of resources to be allocated to the MUs. Therefore, the association profile that maximizes $\Delta\varphi_j(\mathbf{a}_j)$ at FBS j is $\arg \max_{\mathbf{a}_j \in \mathcal{A}_j} \Delta\varphi_j(\mathbf{a}_j)$. Note that it is not necessary for the FBS j to obtain more than $|\mathcal{S}_j^f|$ channels to serve its FUs, since we assume each FU can only access one channel at a time. We denote $\mathcal{A}_j' = \{\mathbf{a}_j \mid \mathbf{a}_j \in \mathcal{A}_j, n_j^c + \Delta\varphi_j(\mathbf{a}_j) > |\mathcal{S}_j^f|\}$. It can be seen that if $\mathcal{A}_j' \neq \emptyset$, accepting any association profile in \mathcal{A}_j' maximizes the total utility of FUs at FBS j , since the resources available for the FUs are sufficient to guarantee that each FU obtains one channel. In this case, we choose the profile that contains the minimum number of MUs as the association decision. On the other hand, if $\mathcal{A}_j' = \emptyset$, the optimal association decision can be selected as the association profile that gives the largest $\Delta\varphi_j(\mathbf{a}_j)$. Therefore, the optimal user association decision for FBS j can be represented as

$$\mathbf{a}_j^* = \begin{cases} \arg \min_{\mathbf{a}_j \in \mathcal{A}_j'} \sum_{i \in \mathcal{S}^M} a_{ij}, & \text{if } \mathcal{A}_j' \neq \emptyset, \\ \arg \max_{\mathbf{a}_j \in \mathcal{A}_j} \Delta\varphi_j(\mathbf{a}_j), & \text{otherwise.} \end{cases} \quad (19)$$

E. Procedures of the Proposed Mechanism

We have derived the desired strategies for base stations and users in each network configuration stage. The proposed network configuration mechanism can be summarized as follows. (1) *Initialization*: The scheduler collects global information of the network, and calculates the resource allocated to the FBSs and MUs using Algorithms 2 and 1, respectively, assuming closed access is used at all FBSs. Then, it sends the information n_j^c and $w_{i0}^c, \forall i \in \mathcal{S}^M$ to each FBS $j \in \mathcal{I} \setminus \{0\}$. The scheduler also obtains the channel allocation decisions in the closed access mode, where the sets of channels allocated to the FBSs and the MBS are $\{1, \dots, n_{\text{FBS}}^c\}$ and $\{n_{\text{MBS}}^c, \dots, N\}$, respectively.

(2) *Access Control Process*: Each MU sends an access request to the FBS from which the MU can obtain the largest utility. Then, each FBS j determines its association decision according to equations (19) and (17), and sends the information of \mathbf{a}_j and ϵ_j to the scheduler.

(3) *Resource Allocation Process*: The scheduler first determines the number of additional channels (N_a) to be reallocated from MBS to FBSs based on the resource allocation decisions in closed access mode (obtained in the initialization step) and the access control decisions. Next, the scheduler calculates the minimum number of channels required at each base stations according to (14) and allocates the additional channels N_a one by one using Algorithm 2 by setting $k = n_{\text{FBS}}^c + 1$,

$n_{\text{FBS}} = n_{\text{FBS}}^c + 1$ and $n_{\text{MBS}} = n_{\text{MBS}}^c + N_a$ in Line 4 and Line 16, respectively. Then, the MBS and FBSs allocate their available resources to each associated user according to the results obtained using Algorithm 1 during the channel allocation process.

(4) *Power Management Process*: Each user determines its optimal power according to equation (8) based on the resources allocated by its associated base station.

Note that the proposed mechanism explores channel reuse among femtocells, which significantly improves the system performance. Moreover, the proposed mechanism has a nice property, it encourages an FBS to use hybrid access control in order to improve the total utility of its FUs. On one hand, an FBS may obtain additional resources to serve its own FUs by accepting to serve nearby MUs, as discussed in Section III-D. On the other hand, an MU who communicates with the MBS with poor signal quality can improve its utility by accessing the nearby FBS without additional cost (i.e., extra payment). This type of incentive is important since our model considers FBSs are installed by selfish private users who may not share their services for free. Without such incentive, FBSs may refuse to serve MUs and always use closed access mode.

F. Complexity of the Proposed Algorithms

In the proposed configuration mechanism, the optimization in Levels 2 and 5 takes constant time, as can be seen from (17) and (8), respectively. From (19), the computational complexity of Level 1 is proportional to the cardinality of the feasible set $|\mathcal{A}_j|$. Therefore, the major computation complexity originates from Algorithms 1 and 2. Both algorithms use iterative approaches. In the following, we provide an analysis of these two algorithms with respect to the number of iterations. In Algorithm 1, the basic allocation step from Line 2 to Line 3 can be computed in constant time. Therefore, the complexity mainly depends on the loop from Line 7 to Line 13. Note that in each iteration, we allocate one time slot to a user. There are $T_R = T(n_j - \sum_{i \in \mathcal{S}_j} w_{ij}^{\min})$ iterations in this loop. Within each iteration, Line 10 searches for the user with the largest utility increment, which can be implemented using binary search. The average number of searching iterations for this step is $\log(|\mathcal{S}_j|)$. Therefore, the complexity of Algorithm 1 with respect to the number of iterations is $T_R \log(|\mathcal{S}_j|)$, and each iteration takes constant time for computation. Algorithm 2 also contains two steps, and the major complexity depends on the second step from Line 17 to Line 33. We denote the number of remaining channels after the basic allocation as N_R . In the second step, we allocate the channels to each base station one by one iteratively. For simplicity, we assume all FBSs have the same number of subscribed FUs. In this case, the maximum number of iterations for the outer loop is N_R . Within each iteration for channel allocation, we check all base stations one by one. Each time we need to invoke Algorithm 1 to update the utility increment with one additional channel, which further takes $T \log(|\mathcal{S}_j|)$ iterations. Thus, the complexity with respect to the number of iterations in this worst case is $N_R(|\mathcal{I}|)T \log(|\mathcal{S}_j|)$. It can be seen that the complexity increases linearly with the number of remaining

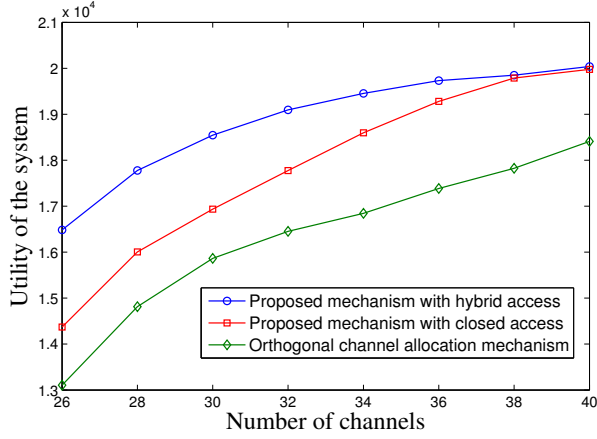


Fig. 4. Total utility of the system versus total number of channels N , with $r = 40$ m and $d = 150$ m.

channels, the number of FBSs, and the number of time slots in the transmission phase.

IV. PERFORMANCE EVALUATION

In this section, we evaluate the performance of the proposed network configuration mechanism using simulation. We consider a system that consists of an MBS and 10 FBSs which are close to each other. The FBSs are randomly located within a circular region whose radius is r . The distance between the MBS and the region center is d . The radius of each femtocell is 10 m, and four FUs are randomly distributed within each femtocell. The MBS serves 15 MUs who are randomly distributed in the region with the constraint that at least 8 of them are indoor users who are located within at least one femtocell. The maximum number of users an FBS can serve is $L = 8$. There are N channels available for the system, each with a bandwidth of 180 kHz. The wireless channel model follows [28], where the path loss exponent between a user and the MBS (or an FBS) is 4 (or 3), and we choose the wall penetration loss as 8 dB. The users' maximum transmission power is $P_{\max} = 250$ mW and the noise power is -120 dBm. The interference threshold θ is -90 dBm. Each time slot is 1 ms and one transmission period consists of $T = 1000$ time slots. We consider two types of applications, the MUs are requesting regular video chatting with the minimum rate requirement (R_i^{\min}) of 256 kbps. The FUs are requesting high definition video chatting with minimum rate requirement of 400 kbps. Similar to [24], we implement the user satisfaction function with $\Phi_i(\bar{R}_{ij}) = \alpha \ln(1 + (\bar{R}_{ij} - R_i^{\min})) + C$, where $\alpha = 200$ and $C = 50$. We choose $\beta = 0.1$. In this section, the utility of a user is calculated using the actual average data rate, which is obtained by averaging the instantaneous data rate over the transmission phase (1000 time slots). We adopt a random scheduling algorithm to determine the set of time slots for each user based on the resource sharing variables. Specifically, at each time slot, FBS j randomly selects n_j users from the scheduling set, which initially contains all the users to be served. Once the number of time slots allocated to a user reaches the desired number, the user is removed from the scheduling set. We simulate the proposed configuration mechanism and compare its performance with configuration

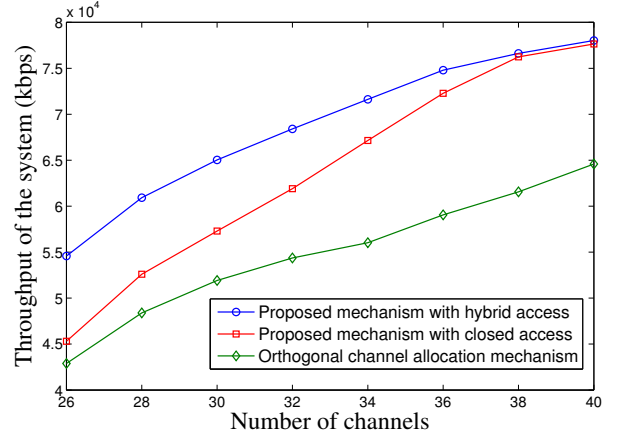


Fig. 5. System throughput versus total number of channels N , with $r = 40$ m and $d = 150$ m.

mechanism with closed access and configuration mechanism with topology-based access introduced in [17], respectively. We first evaluate the performance of the aforementioned configuration mechanisms with respect to different number of channels N , and the results are averaged over 50 simulation runs with $r = 40$ m and $d = 150$ m. Fig. 4 shows the total system utility achieved by different network configuration mechanisms, and Fig. 5 shows their corresponding system throughput. It can be seen that the proposed mechanism achieves the largest total utility and the highest throughput among the three mechanisms, and the performance gap between the proposed mechanism and the other two mechanisms becomes smaller as the number of channels increases. The reason is as follows: Using the proposed mechanism, when the number of channels is small, these MUs who receive low service quality may switch to a nearby femtocell to improve their utility, which may also improve the total utility (or throughput) of the FUs in that cell. Similarly, the mechanism with topology-based access can also improve the system utility (or throughput) by accepting MUs at the FBSs. However, this approach allocate orthogonal channels to each FBS, which neglect the benefit of channel reuse among the FBSs. In addition, since the access control is purely based on topology information, it is possible that an FBS may reject the requests of some MUs and miss the opportunity to improve its own performance. It is also possible that an FBS accepts to serve some MUs but does not obtain additional channels from the MBS, which may degrade the utility of FUs due to resource sharing. Therefore, the performance of the mechanism with topology-based access control is not as good as the proposed mechanism. As the number of channels increases, some FBSs or MUs may obtain sufficient resources, and access control between these MUs and FBSs are not necessary. Thus, the performance gap between the proposed mechanism and the other mechanisms becomes smaller.

In Table II, we show the average running time of the proposed mechanism and the topology-based mechanism with respect to different number of channels using MATLAB. As the number of channels increases, the average running time of these two mechanisms increases, since the computational complexity increases due to the increased number of variables.

TABLE II
AVERAGE RUNNING TIME VERSUS NUMBER OF CHANNELS

N	26	28	30	32	34	36	38	40
Proposed mechanism	6.33 sec	6.36 sec	6.95 sec	7.09 sec	7.13 sec	7.23 sec	7.48 sec	7.63 sec
Topology-based mechanism	9.00 sec	10.22 sec	10.93 sec	11.33 sec	11.62 sec	11.87 sec	12.13 sec	12.70 sec

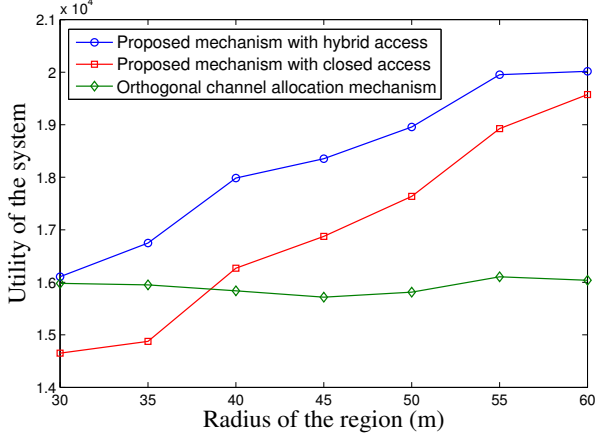


Fig. 6. Total utility of the system versus radius of the region, with $N = 30$, and $d = 150$ m.

It is implied in Table II that these two mechanisms have similar computational complexity. However, the proposed mechanism achieves better total utility and system throughput, as shown in Figs. 4 and 5.

Next, we evaluate the performance of the three mechanisms with respect to different radius of the circular area r . It is shown in Fig. 6 that the system utility of the proposed mechanism and the closed access mechanism increase as the radius of the region increases. This is because as the radius increases, the average distance between two randomly distributed FBSs becomes larger, which increases the possibility of channel reuse among the FBSs. The topology-based mechanism remains almost constant as r changes. This is because this mechanism uses an orthogonal channel allocation scheme without the consideration of interference among the femtocells, and the decision is not affected by the radius of the circular region. Therefore, when the radius is small, the proposed mechanism with hybrid access achieves a similar performance to the topology-based mechanism since channel reuse rarely happens due to the close proximity of the FBSs, as the radius increases, the proposed mechanism achieves significant utility improvement and the performance gap between the two mechanism becomes larger.

Then, we adjust the distance between the MBS and the considered region center, and evaluate the performance of the configuration mechanisms with respect to different values of d . Fig. 7 shows that the system utility decreases as d increases for the proposed mechanism and the closed access mechanism. However, the utility decrement step size of the proposed mechanism becomes smaller when d increases from 100 m to 140 m and then increases thereafter. This is because when d is small, the signal quality from the MBS to the MUs are good enough to satisfy their QoS requirements and accessing the MBS may achieve a larger utility for each MU. As d increases, the signal quality from the MBS to the MUs

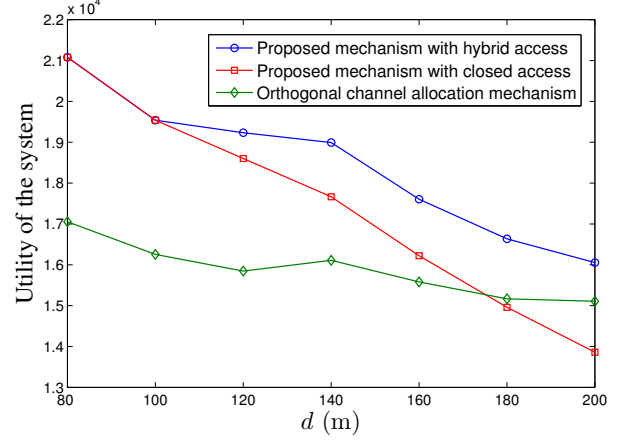


Fig. 7. Total utility of the system versus d , with $N = 30$ and $r = 40$ m.

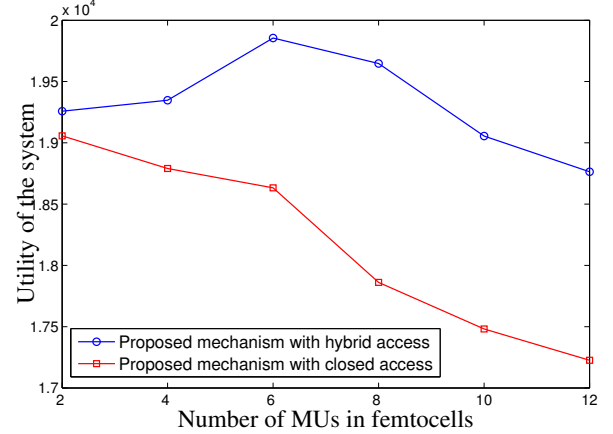


Fig. 8. Total utility versus number of MUs moved into the femtocell region with $N = 30$, $d = 150$ m and $r = 40$ m.

degrades accordingly, which decreases the total utility. When d is greater than a certain value, i.e., 100 m, the signal quality from the MBS to the MUs is poor and the MUs may request to access their nearby FBSs to improve their utilities. With the proposed mechanism, the FBSs can accept to serve the MUs who experience severe channel fading from the MBS, which improves both the utilities of MUs and FUs when d varies from 100 m to 140 m. After that, FBSs may not accept more MUs, and the total utility decrease linearly as d increases. A similar trend can be seen for the topology-based mechanism.

In Fig. 8, we adjust the locations of MUs and evaluate their impact on the performance of the proposed mechanism. We randomly position MUs in a ring area and the average distance between MU to the region center is 80 m. We then move a number of MUs into the circular region where the FBSs are located. It is shown in Fig. 8 that the system utility of the proposed mechanism first increases and then decreases as more MUs move into femtocells, and the performance gap between the proposed mechanism and the closed access

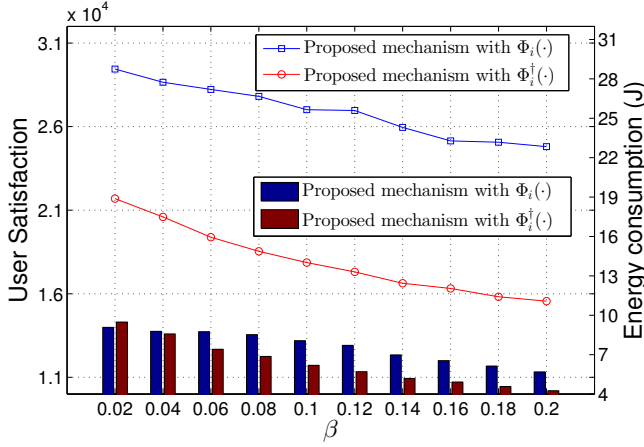


Fig. 9. Energy consumption and user satisfaction versus β , with $N = 28$, $d = 120$ m and $r = 40$ m.

mechanism becomes larger. The reason is that as the MU moves into the femtocells, their minimum resource demand from the MBS becomes larger due to the wall penetration loss of signal, which decreases their utility in the closed access mode. However, as MUs move closer to the FBSs, they require less resources from the FBSs to satisfy their minimum rate requirements. In this case, the proposed mechanism improves the performance of MUs and FUs by redistributing the channels among the MBS and FBSs, which results in an increase of the performance gap. As more MUs move into the femtocells, once the utility improvement with hybrid access does not compensate the utility decrease (for these MUs), the system utility starts to decrease. Nevertheless, the proposed mechanism achieves superior performance than the mechanism with closed access mode.

Finally, we evaluate the performance of the proposed mechanism with respect to different values of the weighting factor β . We select $\beta \leq C/P_{\max}$ to guarantee positive utility of each user in the system. We also implement another user satisfaction function [29] as $\Phi_i^dagger(\tilde{R}_{ij}) = \alpha(1 - \exp(-(\tilde{R}_{ij} - R_i^{\min})/R_i^{\min})) + C$. Fig. 9 shows the users' satisfaction and the energy consumption decrease as β increases for both scenarios. The reason is that when β is small, improving users' satisfaction with respect to data rate is more important than saving energy, and users may use large transmission power to achieve high data rates. However, as β increases, the energy consumption component in the utility function becomes more critical. To maximize the utility, users may consider reducing transmission power to save energy. Therefore, we can adjust the parameter β accordingly to balance the trade-off between users' satisfaction and energy consumption, and to realize different optimization objectives. Note that this result is valid for other types of concave satisfaction functions as well, since the proposed mechanism is shown to achieve optimal resource allocation solution with any concave satisfaction function.

V. CONCLUSION

In this paper, we studied uplink network configuration in a two-tier macro-femto system with hybrid access control. To motivate hybrid access at the FBSs and allocate resource efficiently, we proposed a network configuration mechanism

which consists of access control, channel allocation and power management processes. We modeled the configuration mechanism as a multi-stage decision making process, where base stations and users make decisions sequentially to optimize their own utilities. To find the optimal strategy at each stage, we formulated the sequential decision making process as a multi-level optimization problem, and analyzed each optimization level in a bottom-up manner. We proposed efficient algorithms to solve the optimization problems sequentially. Simulation results showed that the proposed network configuration mechanism achieves higher system utility than configuration mechanism with topology-based hybrid access or closed access, especially when the number of available channels in the system is small.

APPENDIX

A. Proof of Theorem 1

According to the discussion in Section III-B, we only need to prove that Algorithm 1 solves subproblems (10) and (11). First, it is clear that the solution to subproblem (10) is to allocate the minimum amount of resource to the MUs that guarantees their QoS constraints, which corresponds to Lines 1 and 2 in Algorithm 1.

Next, we show that Algorithm 1 also solves subproblem (11) optimally. We denote $\Delta\tilde{u}_{ij}(k) = \Delta u_{ij}(\frac{k}{T})$ as the utility increment when FU i obtains its k th time slot from the FBS. It can be shown that $\Delta\tilde{u}_{ij}(k)$ is a non-decreasing function of k , which gives

$$\Delta\tilde{u}_{ij}(k) \geq \Delta\tilde{u}_{ij}(k+1) \quad \text{if } w_{ij}^{\min}T < k < T. \quad (20)$$

The proof of (20) is shown in Appendix B. We denote set $\mathcal{K} = \{1, 2, \dots, T\}$, and define W_{ij}^k as an indicator such that $W_{ij}^k = 1$ indicates user i obtains its k th time slot ($W_{ij}^k = 0$ otherwise). Then, subproblem (11) is equivalent to the following problem

$$\begin{aligned} & \underset{W_{ij}^k, i \in S_j^f, k \in \mathcal{K}}{\text{maximize}} && \sum_{i \in S_j^f} \sum_{k \in \mathcal{K}} W_{ij}^k \Delta\tilde{u}_{ij}(k) \\ & \text{subject to} && \sum_{i \in S_j^f} \sum_{k \in \mathcal{K}} W_{ij}^k \leq T \left(\sum_{k \in \mathcal{N}} x_j^k - \sum_{i \in S_j^m} w_{ij} \right), \\ & && W_{ij}^k = 1, \quad \forall k \leq w_{ij}^{\min}T, \quad \forall i \in S_j^f, \\ & && \prod_{t=1}^k W_{ij}^t = W_{ij}^k, \quad \forall k \in \mathcal{K}, \quad \forall i \in S_j^f. \end{aligned} \quad (21)$$

The first constraint indicates the total number of time slots allocated to the FUs is no larger than $T(\sum_{k \in \mathcal{N}} x_j^k - \sum_{i \in S_j^m} w_{ij})$. The second constraint indicates that user i should obtain at least $w_{ij}^{\min}T$ time slots from the FBS. The third constraint implies that if user i obtains its k th time slot from the FBS ($W_{ij}^k = 1$), it should also obtain all the previous time slots ($W_{ij}^t = 1, \forall t \leq k$). We define the set $\Delta\mathcal{U}_j = \{\Delta\tilde{u}_{ij}(k), \forall k \in \{w_{ij}^{\min}T + 1, \dots, T\}, \forall i \in S_j^f\}$ and denote $\Delta\mathcal{U}'_j$ as the set of the largest T_R (defined in Algorithm 1) elements from $\Delta\mathcal{U}_j$. With property (20), it can be shown that if $\Delta\tilde{u}_{ij}(k) \in \Delta\mathcal{U}'_j$, then $\Delta\tilde{u}_{ij}(t) \in \Delta\mathcal{U}'_j, \forall t < k$. Therefore, the corresponding values $W_{ij}^k = 1$ associated with the positive elements $\Delta\tilde{u}_{ij}(k)$ in $\Delta\mathcal{U}'_j$ constitute the solution

to (21). Note that in Algorithm 1, for each remaining time slot after basic allocation, we allocate it to the user with the largest positive $\Delta u_{ij}(w_{ij} + 1/T)$. Based on the non-increasing property of $\Delta u_{ij}(w_{ij}) = \Delta \tilde{u}_{ij}(w_{ij}T)$, Algorithm 1 exactly finds the positive elements among the first largest T_R elements in set $\Delta \mathcal{U}_j$. Therefore, Algorithm 1 also solves problem (21), which implies that it solves subproblem (11). This completes the proof.

B. Proof of (20)

According to the definition of $\Delta \tilde{u}_{ij}(k)$, to prove property (20), we only need to prove that $u_{ij}(w_{ij}, P_i^*(w_{ij}))$ is a concave function with respect to w_{ij} in $(w_{ij}^{\min}, 1)$ (where we relax the integer constraint on w_{ij}). We define $\phi(w_{ij}) = \frac{du_{ij}}{dw_{ij}}$ and show that $\frac{d\phi}{dw_{ij}} < 0, \forall w_{ij} \in (w_{ij}^{\min}, 1)$. Note that according to (8) in Section III-A, the optimal power P_i^* is a function of w_{ij} , which may be either P_i^{\min} , P_{\max} or \tilde{P}_i . We define $\mathcal{W}_1, \mathcal{W}_2$ and \mathcal{W}_3 as the intervals of w_{ij} where P_i^* is P_i^{\min} , P_{\max} and \tilde{P}_i , respectively. We show that for $w_{ij} \in \mathcal{W}_i, i \in \{1, 2, 3\}$, we always have $\frac{d\phi}{dw_{ij}} < 0$.

First, if $\mathcal{W}_1 \neq \emptyset$, we have $P_i^* = P_i^{\min}$ and $\tilde{R}_{ij} = R_i^{\min}$ for $w_{ij} \in \mathcal{W}_1$. In this case, $u_{ij} = C - \beta w_{ij} \frac{\sigma^2 + \xi_j}{g_{ij}} \left(2^{R_i^{\min}/(w_{ij}B)} - 1 \right)$. Then, we have $\phi(w_{ij}) = -\frac{\beta(\sigma^2 + \xi_j)}{g_{ij}} 2^{R_i^{\min}/(w_{ij}B)} (1 - \frac{R_i^{\min} \ln 2}{B w_{ij}}) + 1$, and $\frac{d\phi}{dw_{ij}} = -\frac{\beta(\sigma^2 + \xi_j)}{g_{ij}} \frac{(R_i^{\min})^2 (\ln 2)^2}{B^2 w_{ij}^3} 2^{\frac{R_i^{\min}}{B w_{ij}}} < 0$. Next, if $\mathcal{W}_2 \neq \emptyset$, we have $P_i^* = P_{\max}$, and $\tilde{R}_{ij} = w_{ij}B \log_2(1 + \frac{g_{ij} P_{\max}}{\sigma^2 + \xi_j})$ for $w_{ij} \in \mathcal{W}_2$. Then, we have $u_{ij} = \Phi_i(\tilde{R}_{ij}) - \beta w_{ij} P_{\max}$ and $\phi(w_{ij}) = \frac{d\Phi_i}{d\tilde{R}_{ij}} \cdot B \log_2(1 + \frac{g_{ij} P_{\max}}{\sigma^2 + \xi_j}) - \beta P_{\max}$. Since $\Phi_i(\cdot)$ is a concave function of \tilde{R}_{ij} , $\frac{d\Phi_i}{d\tilde{R}_{ij}}$ decreases with \tilde{R}_{ij} . Note that in this case, \tilde{R}_{ij} is an increasing function of w_{ij} . Therefore, $\phi(w_{ij})$ decreases with w_{ij} , which gives $\frac{d\phi}{dw_{ij}} < 0$. Finally, if $\mathcal{W}_3 \neq \emptyset$, we have $P_i^* = \tilde{P}_i$ for $w_{ij} \in \mathcal{W}_3$, where \tilde{P}_i satisfies $\frac{d\Phi_i}{d\tilde{R}_{ij}} \frac{d\tilde{R}_{ij}}{dP_i} - \beta w_{ij} = 0$. Then, we have

$$\frac{d\Phi_i}{d\tilde{R}_{ij}} \bigg|_{P_i=\tilde{P}_i} \frac{B}{\ln 2(\tilde{P}_i + (\sigma^2 + \xi_j)/g_{ij})} - \beta = 0. \quad (22)$$

In this case, we first show that \tilde{P}_i decreases with w_{ij} . Assume w'_{ij} and \tilde{P}'_i also satisfy (22) and $w'_{ij} > w_{ij}$. If $\tilde{P}'_i \geq \tilde{P}_i$, we have $\tilde{R}'_{ij} > \tilde{R}_{ij}$. Since $\frac{d\Phi_i}{d\tilde{R}_{ij}}$ decreases with \tilde{R}_{ij} , we have

$$\begin{aligned} & \frac{d\Phi_i}{d\tilde{R}_{ij}} \bigg|_{P_i=\tilde{P}_i} \frac{B}{\ln 2(\tilde{P}_i + (\sigma^2 + \xi_j)/g_{ij})} - \beta \\ & > \frac{d\Phi_i}{d\tilde{R}_{ij}} \bigg|_{P_i=\tilde{P}'_i} \frac{B}{\ln 2(\tilde{P}'_i + (\sigma^2 + \xi_j)/g_{ij})} - \beta = 0, \end{aligned} \quad (23)$$

which contradicts (22). Therefore, for $w'_{ij} > w_{ij}$, we must have $\tilde{P}'_i < \tilde{P}_i$, which implies that $\frac{d\tilde{P}_i}{dw_{ij}} < 0$.

We further have

$$\begin{aligned} & \phi(w_{ij}) \\ &= \frac{d\Phi_i}{d\tilde{R}_{ij}} \bigg|_{P_i=\tilde{P}_i} \left(B \log_2 \left(1 + \frac{g_{ij} \tilde{P}_i}{\sigma^2 + \xi_j} \right) \right. \\ & \quad \left. + w_{ij} \frac{B}{\ln 2(\tilde{P}_i + (\sigma^2 + \xi_j)/g_{ij})} \cdot \frac{d\tilde{P}_i}{dw_{ij}} \right) \\ & \quad - \beta \left(\tilde{P}_i + w_{ij} \frac{d\tilde{P}_i}{dw_{ij}} \right) \\ &= \beta \ln 2 \left(\tilde{P}_i + \frac{\sigma^2 + \xi_j}{g_{ij}} \right) \log_2 \left(1 + \frac{g_{ij} \tilde{P}_i}{\sigma^2 + \xi_j} \right) - \beta \tilde{P}_i. \end{aligned} \quad (24)$$

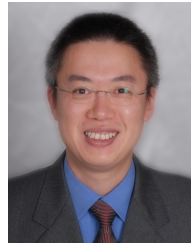
From (24), we have $\frac{d\phi}{dw_{ij}} = \frac{d\phi}{d\tilde{P}_i} \frac{d\tilde{P}_i}{dw_{ij}} = \beta \ln 2 \log_2(1 + \frac{g_{ij} \tilde{P}_i}{\sigma^2 + \xi_j}) \frac{d\tilde{P}_i}{dw_{ij}} < 0$ according to $\frac{d\tilde{P}_i}{dw_{ij}} < 0$.

In summary, we have shown that $\frac{d\phi}{dw_{ij}} < 0$ for $w_{ij} \in \mathcal{W}_1 \cup \mathcal{W}_2 \cup \mathcal{W}_3$. Note that $(w_{ij}^{\min}, 1) \subseteq \mathcal{W}_1 \cup \mathcal{W}_2 \cup \mathcal{W}_3$, and $\phi(w_{ij})$ is a continuous function over $(w_{ij}^{\min}, 1)$. Therefore, we conclude that $\frac{d\phi}{dw_{ij}} < 0, \forall w_{ij} \in (w_{ij}^{\min}, 1)$, which implies $u_{ij}(w_{ij}, P_i^*(w_{ij}))$ is a concave function of w_{ij} in $(w_{ij}^{\min}, 1)$. This completes the proof.

REFERENCES

- [1] B. Niu and V. W. S. Wong, "The design of resource management mechanism with hybrid access in a marco-femto system," in *Proc. of IEEE GLOBECOM*, Atlanta, USA, Dec. 2013.
- [2] J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3GPP heterogeneous networks," *IEEE Wireless Commun. Mag.*, vol. 18, no. 3, pp. 10–21, June 2011.
- [3] A. Barbieri, A. Damnjanovic, T. Ji, J. Montojo, Y. Wei, D. P. Malladi, O. Song, and G. Horn, "LTE femtocells: System design and performance analysis," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 586–594, Apr. 2012.
- [4] 3GPP TS 22.220, "Service requirements for home node B (HNB) and home enode B (HeNB) (Release 11)," Sept. 2012.
- [5] V. Chandrasekhar, J. Andrews, and A. Gatherer, "Femtocell networks: A survey," *IEEE Commun. Mag.*, vol. 46, no. 9, pp. 59–67, Sept. 2008.
- [6] J. G. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. C. Reed, "Femtocells: Past, present, and future," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 497–508, Apr. 2012.
- [7] T. Zahir, K. Arshad, A. Nakata, and K. Moessner, "Interference management in femtocells," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 1, pp. 293–311, First Quarter 2013.
- [8] A. Golaup, M. Mustapha, and L. B. Patanapongpibul, "Femtocell access control strategy in UMTS and LTE," *IEEE Commun. Mag.*, vol. 47, no. 9, pp. 117–123, Sept. 2009.
- [9] G. de la Roche, A. Valcarce, D. Lopez-Perez, and J. Zhang, "Access control mechanisms for femtocells," *IEEE Commun. Mag.*, vol. 48, no. 1, pp. 33–39, Jan. 2010.
- [10] R. Balakrishnan and B. Canberk, "Traffic-aware QoS provisioning and admission control in OFDMA hybrid small cells," *IEEE Trans. on Vehi. Technol.*, vol. 63, no. 2, pp. 802–810, Feb. 2014.
- [11] F. Pantisano, M. Bennis, W. Saad, and M. Debbah, "Spectrum leasing as an incentive towards uplink macrocell and femtocell cooperation," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 617–630, Apr. 2012.
- [12] Y. Yi, J. Zhang, Q. Zhang, and T. Jiang, "Spectrum leasing to femto service provider with hybrid access," in *Proc. of IEEE INFOCOM*, Orlando, FL, Apr. 2012.
- [13] S. Yun, Y. Yi, D.-H. Cho, and J. Mo, "Open or close: On the sharing of femtocells," in *Proc. of IEEE INFOCOM*, Shanghai, China, Apr. 2011.
- [14] Y. Chen, J. Zhang, and Q. Zhang, "Incentive mechanism for hybrid access in femtocell network with traffic uncertainty," in *Proc. of IEEE ICC*, Budapest, Hungary, June 2013.

- [15] C. H. Chai, Y. Y. Shih, and A. C. Pang, "A spectrum-sharing rewarding framework for co-channel hybrid access femtocell networks," in *Proc. of IEEE INFOCOM*, Turin, Italy, Apr. 2013.
- [16] P. Xia, V. Chandrasekhar, and J. G. Andrews, "Open vs. closed access femtocells in the uplink," *IEEE Trans. on Wireless Commun.*, vol. 9, no. 12, pp. 3798–3809, Dec. 2010.
- [17] W. C. Cheung, T. Q. S. Quek, and M. Kountouris, "Throughput optimization, spectrum allocation, and access control in two-tier femtocell networks," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 561–574, Apr. 2012.
- [18] L. B. Le, D. Niyato, E. Hossain, D. I. Kim, and D. T. Hoang, "QoS-aware and energy-efficient resource management in OFDMA femtocells," *IEEE Trans. on Wireless Commun.*, vol. 12, no. 1, pp. 180–194, Jan. 2013.
- [19] R. Estrada, A. Jarray, H. Otrok, Z. Dziong, and H. Barada, "Energy-efficient resource-allocation model for OFDMA macrocell/femtocell networks," *IEEE Trans. on Vehi. Technol.*, vol. 62, no. 7, pp. 3429–3437, Sept. 2013.
- [20] C. H. Ko and H. Y. Wei, "On-demand resource-sharing mechanism design in two-tier OFDMA femtocell networks," *IEEE Trans. on Veh. Technol.*, vol. 60, no. 3, pp. 1059–1071, Mar. 2011.
- [21] E. Biglieri, J. Proakis, and S. Shamai, "Fading channels: Information-theoretic and communications aspects," *IEEE Trans. on Inf. Theory*, vol. 44, no. 6, pp. 2619–2692, Oct. 1998.
- [22] S. Shenker, "Fundamental design issues for the future internet," *IEEE J. Select. Areas Commun.*, vol. 13, no. 7, pp. 1176–1188, Sept. 1995.
- [23] Z. Jiang, Y. Ge, and Y. Li, "Max-utility wireless resource management for best-effort traffic," *IEEE Trans. on Wireless Commun.*, vol. 4, no. 1, pp. 100–111, Jan. 2005.
- [24] S. Bashar and Z. Ding, "Admission control and resource allocation in a heterogeneous OFDMA wireless network," *IEEE Trans. on Wireless Commun.*, vol. 8, no. 8, pp. 4200–4210, Aug. 2009.
- [25] D. Tse and P. Viswanath, *Fundamentals of Wireless Communication*. Cambridge University Press, 2005.
- [26] D. T. Ngo, L. B. Le, and T. Le-Ngoc, "Distributed Pareto-optimal power control for utility maximization in femtocell network," *IEEE Trans. on Wireless Commun.*, vol. 11, no. 10, pp. 3434–3446, Oct. 2012.
- [27] A. Migdalas, P. M. Pardalos, and P. Varbrand, *Multilevel Optimization: Algorithms and Applications*. Springer, 1998.
- [28] ITU-R Std. Recommendation ITU-R M.1225, "Guidelines for evaluation of radio transmission technologies for IMT-2000," Feb. 1997.
- [29] N. Enderle and X. Lagrange, "User satisfaction models and scheduling algorithms for packet-switched services in UMTS," in *Proc. of IEEE VTC 2003-Spring*, Jeju, South Korea, Mar. 2003.



Vincent W.S. Wong (SM'07) received the B.Sc. degree from the University of Manitoba, Winnipeg, MB, Canada, in 1994, the M.A.Sc. degree from the University of Waterloo, Waterloo, ON, Canada, in 1996, and the Ph.D. degree from the University of British Columbia (UBC), Vancouver, BC, Canada, in 2000. From 2000 to 2001, he worked as a systems engineer at PMC-Sierra Inc. He joined the Department of Electrical and Computer Engineering at UBC in 2002 and is currently a Professor. His research areas include protocol design, optimization, and resource management of communication networks, with applications to wireless networks, smart grid, and the Internet. Dr. Wong is an Editor of the *IEEE Transactions on Communications*. He has served on the editorial boards of *IEEE Transactions on Vehicular Technology* and *Journal of Communications and Networks*. He has served as a Technical Program Co-chair of *IEEE SmartGridComm'14*, as well as a Symposium Co-chair of *IEEE SmartGridComm'13* and *IEEE Globecom'13*. He received the 2014 UBCs Killam Faculty Research Fellowship. He is the Chair of the IEEE Communications Society Emerging Technical Sub-Committee on Smart Grid Communications and IEEE Vancouver Joint Communications Chapter.



Binglai Niu (S'11) received the B.S. degree in electronics engineering from Fudan University, Shanghai, China, in 2008 and the M.Sc. degree in electrical engineering from the University of Alberta, Edmonton, Alberta, Canada, in 2010. He is currently working towards his Ph.D. degree at the University of British Columbia, Vancouver, British Columbia, Canada. His research interests include optimization, game theory, resource allocation and interference management in wireless networks, and incentive mechanism design for cooperative communications.