Toward Optimal Admission Control and Resource Allocation for LTE-A Femtocell Uplink

Xudong Xiang, *Student Member, IEEE,* Chuang Lin, *Senior Member, IEEE,* Xin Chen, and Xuemin (Sherman) Shen, *Fellow, IEEE*

Abstract—The Third Generation Partnership Project (3GPP) has incorporated femtocell (FC) technology in the Long Term Evolution Advanced (LTE-A) standard to enhance the quality of service of indoor mobile users and extend the coverage area of existing macrocells (MCs). In such two-tier LTE-A MC/FC systems, co-tier and cross-tier interference exists in co-channel deployment, exerting adverse effects on system performance. In this paper, we study Single-Carrier Frequency Division Multiple Access (SC-FDMA) based LTE-A FC uplink. We propose the use of transport-layer data Admission Control (AC) in each Femto User Equipment (FUE) as well as interference-aware Resource Allocation (RA) in each base station to manage the inter-cell interference. We first formulate the problem as a constrained Markov decision problem which aims at maximizing the time average throughput of the entire FC tier subject to queue stability constraint for each FUE. Then we propose a Joint Admission Control and Resource Allocation (JACRA) algorithm to obtain the optimal AC and RA policies. In light of the NP-hardness of the resource allocation subproblem, we further propose an iterative heuristic with polynomial time complexity. Simulation studies show that the proposed JACRA algorithm is throughput optimal, outperforming alternative proportional fair and round robin scheduling schemes. Moreover, the proposed heuristic achieves near-optimal throughput with substantial improvement in computational complexity.

Index Terms—Femtocell, Long Term Evolution-Advanced, admission control, resource allocation, uplink communication.

I. INTRODUCTION

TO meet the requirements of the International Telecommunication Union (ITU) standard for the fourth generation (4G) radio communication, the 3GPP developed LTE-A (a.k.a LTE Release 10) as an enhanced version of LTE Releases 8 and 9. Widely recognized as a promising 4G wireless broadband technology, LTE-A supports improved system capacity and coverage, higher peak data rates, lower latency and reduced operating cost [1]. Particularly, LTE-A incorporates femtocell technology into Universal Mobile Telecommunications System (UMTS), in pursuit of enhancing the Quality of Service (QoS) seen by indoor mobile users and extending the coverage area of existing macrocells (MCs) in indoor environments or at the MC edge.

A femtocell (FC) is a low-power, short-range, and low-cost home base station which provides ubiquitous connectivity to macrocell networks via a broadband backhaul connection such as digital subscriber line or cable modem [2]. From a radio deployment point of view, a FC can either share spectral resources with a MC (known as co-channel deployment), or have dedicated spectral resources. Co-channel deployment is more profitable for mobile operators as it efficiently reuses spectrum. However, it would induce both co-tier interference among FCs and cross-tier interference between the FC and the MC (cf. Sec. II), and is thus far more complex from the technical perspective [3].

In 3GPP LTE-A standard, Single-Carrier Frequency Division Multiple Access (SC-FDMA) has been adopted for the uplink due to its resistance to multi-path fading and low Peak-to-Average Power Ratio (PAPR), among other merits [4], [5]. The low PAPR of SC-FDMA enables User Equipment (UE) to use power more efficiently and hence prolongs UE battery life. In the LTE-A uplink, the system bandwidth is partitioned into multiple subchannels termed as Resource Blocks (RBs). Each subchannel spans 12 consecutive subcarriers in the frequency domain. SC-FDMA requires that subchannels allocated to a single UE must be contiguous within each Transmission Time Interval (TTI), *i.e.*, 1ms as used in LTE-A [5]. This constraint alone makes the subchannel allocation problem hard to solve. Additionally, regarding power allocation, the use of SC-FDMA implies that the transmit power on all subchannels allocated to a UE should be equal, and does not exceed some peak power level [6], [7]. In this work, we consider the allocation of both types of resources (*i.e.*, subchannels and power) for LTE-A femtocell and macrocell uplink.

There have been some great efforts dedicated to investigating subchannel and/or power allocation for *single-tier* LTE uplink. In [4], [5], [8], [9], the authors proved the NP-hardness of the frequency-domain subchannel allocation problem for 3GPP LTE uplink, and proposed approximate/heuristic algorithms to solve the problem in polynomial time. In [10], H. Zhang *et al.* proposed open-loop and closed-loop power control schemes to reduce the average interference in LTE and LTE-A uplink. Joint optimization of subchannel allocation and power control

Xudong Xiang is with the Department of Computer Science and Technology, University of Science and Technology Beijing, Haidian District, Beijing 100083, China (email: xudong.xiang@csnet1.cs.tsinghua.edu.cn).

Chuang Lin is with the Tsinghua National Laboratory for Information Science and Technology, Tsinghua University, Haidian District, Beijing 100084, China (email: chlin@tsinghua.edu.cn).

Xin Chen is with the Computer School, Beijing Information Science and Technology University, Chaoyang District, Beijing 100101, China (email: chenxin@mail.tsinghua.edu.cn).

Xuemin (Sherman) Shen is with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON N2L 3G1, Canada (e-mail: sshen@uwaterloo.ca).

was studied in [6], [7], [11]–[13] for SC-FDMA based uplink. For example, I. C. Wong *et al.* [6] formulated the problem of SC-FDMA resource allocation as a set partitioning problem, and presented a greedy-based heuristic algorithm with low complexity and good performance. In [7], [13], the authors delved into the QoS provisioning and energy efficiency issues of LTE uplink. Different from these works, we study the uplink resource allocation problem in *two-tier* LTE-A MC/FC systems with an additional and unique challenge of handling cross-tier interference.

In two-tier MC/FC uplink, techniques to mitigate Inter-Cell Interference (ICI) have been studied previously for Code Division Multiple Access (CDMA) networks [14]–[16], Orthogonal Frequency Division Multiple Access (OFDMA) systems (*e.g.*, mobile WiMax) [17]–[20] as well as SC-FDMA based LTE/LTE-A systems [21], [22]. Thus far, various solutions have been proposed to alleviate or cancel inter-cell interference such as UE power control [21], [22], hybrid spectrum assignment [23], [24] and Fractional Frequency Reuse (FFR) [25]. In particular, for two-tier SC-FDMA based uplink, M. Morita *et al.* [21] proposed an adaptive power control method for LTE FCs. Using aggregated resource usage information of FCs, the method adaptively changes the UE transmit power to improve the throughput of FCs while maintaining that of MCs. However, it considered UE power control solely without uplink subchannel allocation. In [22], Z. Zheng *et al.* combined power optimization and frequency-domain RB allocation to maximize the uplink throughput of Femto User Equipments (FUEs) without jeopardizing the performance of Macro User Equipments (MUEs). It was assumed that FUE packets were scheduled to a set of RB clusters with fixed cluster selection probabilities. However, such resource allocation scheme cannot fully exploit the gain from multiuser frequency diversity.

In this work, we utilize *FC-tier Admission Control (AC)* and *interference-aware Resource Allocation (RA)* [1] to manage the inter-cell interference of time-slotted LTE-A FC systems. By FC-tier admission control, we mean that each FUE employs a threshold-based rule at the transport layer to admit a certain amount of data into the network layer in each slot. Specifically, if the current network-layer queue backlog of a FUE reaches to some preset upper threshold (measured in number of bits), then all new data arrived at the transport layer will be rejected. Otherwise, the data will be admitted. To achieve interference-aware resource allocation, we assume that each Base Station (BS) has the instantaneous Channel State Information (CSI) of all active UEs at the beginning of each slot. In LTE-A FC system, the uplink CSI can be measured by a BS using the Sounding Reference Signals (SRSs) transmitted by its UEs [5]. We further assume that each BS has the knowledge of the Queue Backlog Information (QBI) of its UEs leveraging the buffer status reporting mechanism [13], [26]. Based on the CSI and the QBI, each BS performs resource allocation according to some selected policy. The principal contributions can be summarized as follows:

- We formulate the admission control and resource alloca-

tion problem as a Constrained Markov Decision Problem (CMDP) whose objective is to maximize the time average throughput of the entire FC tier subject to FUE queue stability constraint and resource allocation constraints intrinsic to SC-FDMA, *e.g.*, contiguous subchannel allocation and constant power allocation.

- We propose a Joint Admission Control and Resource Allocation (JACRA) algorithm to solve the CMDP. Specifically, we first use Lyapunov optimization techniques to transform the CMDP into a *drift-minus-reward* minimization problem. Based on drift analysis, we then decompose the minimization problem into multiple independent admission control and resource allocation subproblems which are to be solved without the statistical information (distributions) of uplink data arrival and channel condition. By rigorous mathematical proof, we also derive the analytical performance bounds on the time average throughput of the entire FC tier and on the FUE queue backlog under JACRA.

- In view of the NP-hardness of the resource allocation subproblem, we propose an iterative heuristic algorithm with polynomial time complexity.

- We conduct extensive simulations to evaluate the achievable performance of the proposed JACRA and heuristic algorithms. Simulation results show that JACRA outperforms alternative Proportional Fair (PF) and Round Robin (RR) scheduling schemes in terms of the time average FC-tier throughput, and the heuristic achieves near-optimal throughput with significant improvement in computational complexity.

The remainder of this paper proceeds as follows. Sec. II provides an overview of the system model; Sec. III presents the CMDP formulation of the admission control and resource allocation problem; In Sec. IV, we take advantage of Lyapunov optimization techniques to design JACRA; In Sec. V, we propose the iterative heuristic algorithm for reducing the computational complexity of the resource allocation subproblem; Sec. VI presents alternative PF and RR schemes for resource allocation; Sec. VII elaborates on the simulation setup; Sec. VIII evaluates the achievable performance of the proposed JACRA and heuristic algorithms; Sec. IX concludes the paper and looks into future research.

II. SYSTEM OVERVIEW

We consider the uplink of a LTE-A FC system that employs SC-FDMA based radio access. The macrocell consists of a set of $\mathcal{I}^m$ MUEs and a MBS located at the MC center. Suppose a set of $\mathcal{J} \triangleq \{1, ..., J\}$ femtocells are *underlaid* within the coverage area of the MC. Each femtocell $j, j \in \mathcal{J}$ consists of a set of $\mathcal{I}_j^f$ FUEs and a FBS j located at the FC center.[2] We assume that each femtocell $j, j \in \mathcal{J}$ is configured in the Closed Subscriber Group (CSG) mode wherein FBS j only serves the set of authorized FUEs $\mathcal{I}_j^f$ on its access control list. Other UEs, regardless of their proximity to FBS j, can only connect to their own associated base stations. In the current

---

[1]We propose resource allocation mechanisms for both FC and MC tiers such that inter-cell interference can be mitigated (cf. Sec. IV-B).

[2]It should be noted that we use the notation j to index both FCs and FBSs in this work.
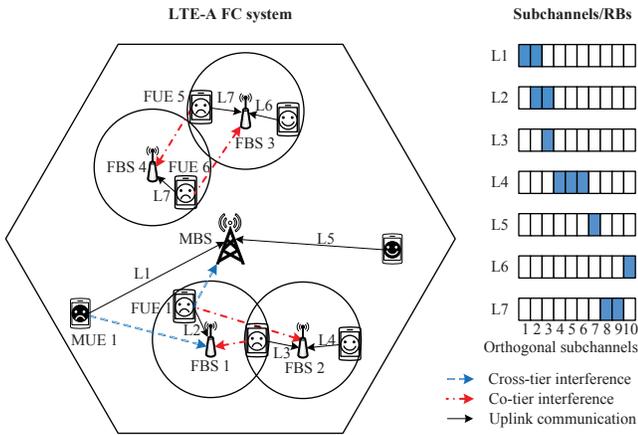
Fig. 1. The architecture of a two-tier LTE-A MC/FC system with co-channel deployment

phase of FC deployment, CSG configurations are expected to be widely used due to practical concerns such as billing, security and contract issues [27]. We use $\mathcal{I}^f \triangleq \cup_j \mathcal{I}_j^f$ to denote the set of all FUEs in the system. The FUEs and MUEs share a set of $\mathcal{K} \triangleq \{1, ..., K\}$ orthogonal subchannels for uplink communications. In practical systems, 3GPP LTE Release 12 specifies $6 \leq K \leq 110$ [28] with the set of allowable configurations being $\{6, 15, 25, 50, 75, 100\}$ [29]. The system model under consideration is illustrated in Fig. 1.

We assume that inter-cell interference in the MC tier is properly managed via FFR [30], [31]. Thus we focus on the co-channel interference within a single MC which can be categorized into the following two types:

- *Cross-tier interference*. This refers to the situation in which the interference source and the victim belong to different network tiers. As illustrated in Fig. 1, MUE 1 and FUE 1 connect to the MBS and FBS 1 via links $L1$ and $L2$, respectively. Since these two links use subchannel 2 simultaneously, such subchannel allocation incurs cross-tier interference at FBS 1 and the MBS.
- *Co-tier interference*. This can happen when proximate FUEs communicate with their associated FBSs via overlapped subchannels. For example, FUE 5 and FUE 6 in Fig. 1 use the same link $L7$ to establish interfered uplink connections to FBS 3 and FBS 4, respectively.

Given the system model illustrated in Fig. 1, our focus is to design uplink admission control and resource allocation policies which maximize the time average throughput of the entire FC tier while guaranteeing FUE queue stability. Tab. I summarizes the key notations used in this work.

III. PROBLEM FORMULATION

We consider a time-slotted system $t \in \mathcal{T} \triangleq \{0, 1, ..., T\}$ where each slot $t, t \in \mathcal{T}$ represents a TTI that has a duration of $\tau = 1$ms as specified in 3GPP LTE-A standard.

A. Uplink Admission Control at FUEs

We consider that in each slot t, FUE $i, i \in \mathcal{I}_j^f$ in FC $j, j \in \mathcal{J}$ receives uplink traffic (measured in bits) from the

TABLE I
SUMMARY OF KEY NOTATIONS

Notation	Meaning
$\mathcal{I}^m$	Set of MUEs in the MC
$\mathcal{I}_j^f, \mathcal{I}^f$	Set of FUEs in FC j and set of all FUEs in the system
$\mathcal{J}$	Set of FCs underlaying the MC with cardinality $|\mathcal{J}| = J$
$\mathcal{K}$	Set of orthogonal subchannels with cardinality $|\mathcal{K}| = K$
$\mathcal{T}$	Set of time slots with cardinality $|\mathcal{T}| = T$ and duration τ
i, j, k, t	Indices of UEs, FCs/FBSs, subchannels and time slots
R, r	Coverage radii of the MC and each FC
$A_{ij}^f(t)$	Amount of data arrived at FUE i in FC j during slot t
$A_i^m(t)$	Amount of data arrived at MUE $i, i \in \mathcal{I}^m$ during slot t
A^{max}	Peak data arrival at a UE in one slot
$Q_{ij}^f(t)$	Queue length of FUE i in FC j at the beginning of slot t
$Q_i^m(t)$	Queue length of MUE $i, i \in \mathcal{I}^m$ at the beginning of slot t
$D_{ij}(t)$	Amount of data admitted by FUE i in FC j in slot t
$D(t)$	Total throughput achieved by all FUEs in slot t
$\bar{D}, \bar{Q}_{ij}$	Time average FC throughput and FUE queue backlog
$X_{ijk}^f(t)$	Binary decision variable on whether to allocate subchannel k to FUE i in slot t
$X_{ik}^m(t)$	Binary decision variable on whether to allocate subchannel k to MUE i in slot t
$\mathcal{K}_{ij}^f(t)$	Set of subchannels allocated to FUE i in FC j in slot t
$\mathcal{K}_i^m(t)$	Set of subchannels allocated to MUE i by the MBS in slot t
$P_{ijk}^f(t)$	Transmit power of FUE $i, i \in \mathcal{I}_j^f$ on subchannel k in slot t
$P_{ik}^m(t)$	Transmit power of MUE i on subchannel k in slot t
P^{max}	Maximum UE transmit power
W	Bandwidth of a subchannel
σ^2	Power spectral density of noise
$C_{ij}^f(t)$	Capacity achieved by FUE $i, i \in \mathcal{I}_j^f$ in FC j during slot t
$C_i^m(t)$	Capacity achieved by MUE $i, i \in \mathcal{I}^m$ during slot t
C^{max}	Maximum achievable capacity of any UE in one slot
$I_{ijk}^f(t)$	ICI received by FUE $i, i \in \mathcal{I}_j^f$ on subchannel k in slot t
$I_{ik}^f(t)$	ICI received by MUE $i, i \in \mathcal{I}^m$ on subchannel k in slot t
$g_{ijk}^f(t)$	CG from FUE $i, i \in \mathcal{I}^f$ to FBS j on subchannel k in slot t
$g_{ijk}^m(t)$	CG from MUE i to FBS j on subchannel k in slot t
$g_{ik}^f(t)$	CG from FUE i to the MBS on subchannel k in slot t
$g_{ik}^m(t)$	CG from MUE i to the MBS on subchannel k in slot t
$S(t)$	System state at the beginning of slot t
V	Lyapunov control parameter satisfying $V > 0$

transport layer according to some i.i.d. random process $A_{ij}^f(t)$, and makes an admission control decision $D_{ij}(t)$ on how much data to admit to the network layer [32], [33]. Here the i.i.d. assumption imposed on uplink FUE traffic is necessary to guarantee the existence of a randomized stationary FC-tier admission control policy (cf. Appendix B), but is not crucial to algorithm performance [34]. Meanwhile, MUE $i, i \in \mathcal{I}^m$ receives an amount of $A_i^m(t)$ data from the transport layer and admits all data to a buffer $Q_i^m(t)$ maintained in the network layer. We do not consider MC-tier admission control because, in general, MUEs have higher priority than FUEs in data transmission [35]. We use $Q_{ij}^f(t)$ to denote the Queue Backlog (QB) of FUE i associated with FBS j in the network layer at the beginning of slot t. Assume that there exists some constant upper bound A^{max} on FUE data arrival such that $A_{ij}^f(t) \leq A^{max}$ for all t. Then we have the following constraint on the admission control decision $D_{ij}(t), \forall t$:

$$D_{ij}(t) \leq A_{ij}^f(t) \leq A^{max}, \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}. \quad (1)$$

Evidently, the maximum amount of data admitted by a FUE cannot exceed the amount of arrived data in every slot t.

FC-tier admission control may help mitigate the co-channel

interference by controlling data injection into FCs. For example, by maintaining the admitted FC-tier traffic load at some lower level, a subset of $\mathcal{K}$ subchannels may be sufficient to transmit all admitted data in each FC. This would reduce the probability that overlapped subchannels are used by FUEs across different FCs for uplink data transmission. As a result, both co-tier and cross-tier (FC to MC) interference can be mitigated. We will obtain the FC-tier admission control policy in Sec. IV-B and validate its effectiveness in mitigating inter-cell interference in Sec. VIII-A.

B. Uplink Subchannel and Power Allocation

The uplink resource allocation takes place in each BS. Specifically, at the beginning of each slot t, each BS collects the current queue backlog information reported from active UEs with non-empty buffers, by means of the buffer status reporting mechanism [13], [26]. Meanwhile, each BS uses the received SRSs to quantize the channel state information of all active UEs, *e.g.*, background noise, channel gains (CGs). We assume that the queue backlog and channel state information is reported back to each BS via the Physical Uplink Control CHannel (PUCCH) without any delay and error. Based on the collected information, each BS makes resource allocation decisions according to the optimal policy obtained using JACRA, and feeds back the decisions to each UE via the Physical Downlink Control CHannel (PDCCH) [13].

In 3GPP LTE-A uplink, the following constraints on subchannel and power allocation need to be satisfied [6], [7]:

C1 (Exclusive subchannel allocation). Each subchannel $k, k \in \mathcal{K}$ can be allocated to at most one FUE/MUE in each slot t. Let $X_{ijk}^f(t)$ and $X_{ik}^m(t)$ be binary subchannel allocation decision variables defined as:

$$X_{ijk}^f(t) = \begin{cases} 1, & \text{Subchannel } k \text{ is assigned to FUE} \\ & i, i \in \mathcal{I}_j^f \text{ by FBS } j \text{ during slot } t, \\ 0, & \text{otherwise.} \end{cases} \quad (2)$$

and

$$X_{ik}^m(t) = \begin{cases} 1, & \text{Subchannel } k \text{ is assigned to MUE} \\ & i, i \in \mathcal{I}^m \text{ by the MBS during slot } t, \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

The following two constraints ensure exclusive allocation of each subchannel in FBS j and in the MBS:

$$\sum_{i \in \mathcal{I}_j^f} X_{ijk}^f(t) \leq 1, \forall j \in \mathcal{J}, \forall k \in \mathcal{K}, \quad (4)$$

$$\sum_{i \in \mathcal{I}^m} X_{ik}^m(t) \leq 1, \forall k \in \mathcal{K}. \quad (5)$$

C2 (Contiguous subchannel allocation). Multiple subchannels allocated to a FUE/MUE must be contiguous. Let $\mathcal{K}_{ij}^f(t)$ and $\mathcal{K}_i^m(t)$ denote the set of subchannels allocated to FUE $i, i \in \mathcal{I}_j^f$ and MUE $i, i \in \mathcal{I}^m$ in slot t, respectively. We have the following two constraints which must be satisfied to ensure contiguous subchannel allocation for each FUE/MUE:

$$|\mathcal{K}_{ij}^f(t)| = \mathbf{ran}\left[\mathcal{K}_{ij}^f(t)\right] + 1, \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}, \quad (6)$$

$$|\mathcal{K}_i^m(t)| = \mathbf{ran}\left[\mathcal{K}_i^m(t)\right] + 1, \forall i \in \mathcal{I}^m, \quad (7)$$

where $|\cdot|$ is the cardinality of a set and the operator $\mathbf{ran}[\cdot]$ denotes the difference between the maximum and the minimum of a set. It can be readily seen that $|\mathcal{K}_{ij}^f(t)| = \sum_{k \in \mathcal{K}} X_{ijk}^f(t)$ and $|\mathcal{K}_i^m(t)| = \sum_{k \in \mathcal{K}} X_{ik}^m(t)$.

To illustrate uplink subchannel allocation in a FBS, consider a FC with three FUEs and accessibility to six subchannels. In each slot t, the FBS performs the uplink subchannel allocation by constructing a binary matrix $\mathbf{H}_{3 \times 6}$ as follows:

$$\mathbf{H} = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{bmatrix}. \quad (8)$$

Each row in $\mathbf{H}$ represents an allocation scheme to a particular FUE while each column represents a subchannel. We can easily verify that each subchannel is exclusively allocated to a FUE in $\mathbf{H}$. However, it is not a valid allocation because the last row indicates that non-contiguous subchannels have been allocated to a FUE.

C3 (Total UE power constraint). The total transmit power of a UE on all allocated subchannels cannot exceed some maximum allowable power level P^{max}.

C4 (Peak power constraint). The transmit power of a UE on each subchannel should be less than some peak power level [6]. We use $\hat{P}_{ijk}^f$ and $\hat{P}_{ik}^m$ to denote the peak transmit power of FUE $i, i \in \mathcal{I}_j^f$ and MUE $i, i \in \mathcal{I}^m$ on subchannel $k, k \in \mathcal{K}$, respectively. In this work, we consider that all UEs use the maximum allowable power level that satisfies both **C3** and **C4** to maximize cell capacity.

C5 (Constant power allocation). The transmit power on all subchannels allocated to a UE should be constant. Although traditional water-filling power allocation algorithms [36] prove to be throughput optimal in OFDM systems, we assume equal power allocation over the subchannels similar to the 3GPP LTE-A standards [10] to preserve the low PAPR of SC-FDMA [6], [22].

Let $P_{ijk}^f(t)$ and $P_{ik}^m(t)$ denote the transmit powers allocated to FUE $i, i \in \mathcal{I}_j^f$ and MUE $i, i \in \mathcal{I}^m$ on subchannel $k, k \in \mathcal{K}$ in slot t, respectively. Based on constraints **C3**$-$**C5**, we have

$$P_{ijk}^f(t) = \begin{cases} \min\left[\hat{P}_{ijk}^f, \dfrac{P^{max}}{|\mathcal{K}_{ij}^f(t)|}\right], & \text{if } X_{ijk}^f(t) = 1, \\ 0, & \text{otherwise,} \end{cases} \quad (9)$$

and

$$P_{ik}^m(t) = \begin{cases} \min\left[\hat{P}_{ik}^m, \dfrac{P^{max}}{|\mathcal{K}_i^m(t)|}\right], & \text{if } X_{ik}^m(t) = 1, \\ 0, & \text{otherwise.} \end{cases} \quad (10)$$

C. System Dynamics

Let $g_{ijk}^f(t)$ and $g_{ijk}^m(t)$ denote the channel gains from FUE $i, i \in \mathcal{I}^f$ and MUE $i, i \in \mathcal{I}^m$ to FBS j on subchannel k in slot t, respectively. The Signal-to-Interference-plus-Noise Ratio (SINR) achieved by FUE $i, i \in \mathcal{I}_j^f$ on subchannel k in

slot t can be written as

$$\Upsilon_{ijk}^f(t) = \frac{P_{ijk}^f(t)g_{ijk}^f(t)}{I_{ijk}^f(t) + \sigma^2}, \qquad (11)$$

$$= \frac{P_{ijk}^f(t)g_{ijk}^f(t)}{\sum_{j'=1,j'\neq j}^J P_{i'j'k}^f(t)g_{i'jk}^f(t) + P_{i''k}^m(t)g_{i''jk}^m(t) + \sigma^2},$$

where σ^2 represents the power of Additive White Gaussian Noise (AWGN), and $I_{ijk}^f(t)$ represents the co-channel interference received by FUE $i, i \in \mathcal{I}_j^f$ on subchannel $k, k \in \mathcal{K}$ in slot t consisting of two components: (1) co-tier interference $\sum_{j'=1,j'\neq j}^J P_{i'j'k}^f(t)g_{i'jk}^f(t)$ from other FCs where the subscript $i', i' \in \mathcal{I}_{j'}^f$, is the index of the FUE that connects to FBS $j', j' \neq j$ on subchannel k; and (2) cross-tier interference $P_{i''k}^m(t)g_{i''jk}^m(t)$ from the MC where the subscript $i'', i'' \in \mathcal{I}^m$ is the index of the MUE that connects to the MBS on subchannel k.

Let $g_{ik}^f(t)$ and $g_{ik}^m(t)$ denote the channel gains from FUE $i, i \in \mathcal{I}^f$ and MUE $i, i \in \mathcal{I}^m$ to the MBS on subchannel k in slot t, respectively. Similarly, we can obtain the SINR achieved by MUE $i, i \in \mathcal{I}^m$ on subchannel k in slot t as

$$\Upsilon_{ik}^m(t) = \frac{P_{ik}^m(t)g_{ik}^m(t)}{I_{ik}^m(t) + \sigma^2}, \qquad (12)$$

$$= \frac{P_{ik}^m(t)g_{ik}^m(t)}{\sum_{j=1}^J P_{\tilde{i}jk}^f(t)g_{\tilde{i}k}^f(t) + \sigma^2},$$

where $I_{ik}^m(t) \triangleq \sum_{j=1}^J P_{\tilde{i}jk}^f(t)g_{\tilde{i}k}^f(t)$ is the cross-tier interference received by MUE $i, i \in \mathcal{I}^m$ on subchannel k in slot t and the subscript $\tilde{i}, \tilde{i} \in \mathcal{I}_j^f$ is the index of the FUE that connects to FBS j on subchannel k. Note that the MUEs may receive only cross-tier interference because we have assumed in Sec. II that inter-cell interference in the MC tier has been properly managed.

Summing together the Shannon capacity of all subchannels allocated to FUE i associated with FBS j, we can write the total uplink capacity (measured in bit/s) of FUE $i, i \in \mathcal{I}_j^f$ in slot t as

$$C_{ij}^f(t) = \sum_{k \in \mathcal{K}_{ij}^f(t)} W \log_2 \left[1 + \Upsilon_{ijk}^f(t)\right], \qquad (13)$$

where W is the bandwidth of a subchannel. Similarly, the total uplink capacity of MUE $i, i \in \mathcal{I}^m$ in slot t can be obtained as

$$C_i^m(t) = \sum_{k \in \mathcal{K}_i^m(t)} W \log_2 \left[1 + \Upsilon_{ik}^m(t)\right]. \qquad (14)$$

We assume that $C_{ij}^f(t), \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}$ and $C_i^m(t), \forall i \in \mathcal{I}^m$ are both upper bounded by some constant C^{max} for all t.

Given the admission control and resource allocation decisions, the FUE and MUE queues evolve over time as follows:

$$Q_{ij}^f(t+1) = \left[Q_{ij}^f(t) - C_{ij}^f(t)\tau\right]^+ + D_{ij}(t), \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J} \qquad (15)$$

$$Q_i^m(t+1) = [Q_i^m(t) - C_i^m(t)\tau]^+ + A_i^m(t), \forall i \in \mathcal{I}^m, \qquad (16)$$

where $[x]^+ \equiv \max[x, 0]$.

Let $\mathbb{S}(t)$ denote the system state at the beginning of slot $t, t \in \mathcal{T}$. We abstract $\mathbb{S}(t)$ as all information that is necessary and sufficient to make the admission control and resource allocation decisions:

$$\mathbb{S}(t) \triangleq [\mathbb{Q}(t); \mathbb{W}(t)] \qquad (17)$$

where $\mathbb{Q}(t) \triangleq [(Q_{ij}^f(t)), (Q_i^m(t))]$ and $\mathbb{W}(t) \triangleq [(A_{ij}^f(t-1)), (g_{ijk}^f(t)), (g_{ik}^m(t))]$ are, respectively, referred to as the *endogenous state* and the *exogenous state* of the LTE-A FC system at time point $t\tau$. The exogenous state $\mathbb{W}(t)$ can be viewed as external random processes that are not subject to the influence of selected control decisions. It should be noted that: (1) $\mathbb{S}(t)$ can only access information that has arrived up to time point $t\tau$, hence the subscript $t-1$ instead of t in $A_{ij}^f(t-1)$; and (2) all channel gains related to co-channel interference are not required for decision-making at each base station, *e.g.*, channel gains $g_{ijk}^m(t)$ and $g_{ik}^f(t)$ (cf. Sec. IV-B).

Based on the analysis above, we can summarize the system dynamics, *i.e.*, the evolvement of the system state $\mathbb{S}(t)$ as follows. At the beginning of slot t, each FBS $j, j \in \mathcal{J}$ makes resource allocation decisions $(\mathcal{K}_{ij}^f(t))_{i \in \mathcal{I}_j^f}$ based on the current queue backlog information $(Q_{ij}^f(t))_{i \in \mathcal{I}_j^f}$ and channel state information $(g_{ijk}^f(t))_{i \in \mathcal{I}_j^f, k \in \mathcal{K}}$. Similarly, the MBS makes resource allocation decisions $(\mathcal{K}_i^m(t))_{i \in \mathcal{I}^m}$ based on $(Q_i^m(t))_{i \in \mathcal{I}^m}$ and $(g_{ik}^f(t))_{i \in \mathcal{I}^m, k \in \mathcal{K}}$. Then, at the end of slot t when the data arrival information $A_{ij}^f(t)$ is available, each FUE $i, i \in \mathcal{I}_j^f$ associated with FBS $j, j \in \mathcal{J}$ makes an admission control decision $D_{ij}(t)$. After all control decisions have been made, the system receives a *one-slot reward* $D(t) \triangleq \sum_{i \in \mathcal{I}_j^f} \sum_{j \in \mathcal{J}} D_{ij}(t)$ and evolves to state $\mathbb{S}(t+1)$. Here $D(t)$ can also be viewed as the total throughput achieved by all FUEs in slot t. We will elaborate on the decision-making process in Sec. IV-B.

D. Control Objective and Problem Statement

For any control algorithm that makes the admission control and resource allocation decisions described in Secs. III-A and III-B, we define the following time-averages:

$$\bar{D} \triangleq \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\{D(t)\} \qquad (18)$$

$$\bar{Q}_{ij} \triangleq \lim_{T \to \infty} \frac{1}{T} \sum_{t=0}^{T-1} \mathbb{E}\left\{Q_{ij}^f(t)\right\}, \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}, \qquad (19)$$

where $\bar{D}$ denotes the time average uplink throughput of the entire FC tier, $\bar{Q}_{ij}$ denotes the time average queue backlog of FUE $i, i \in \mathcal{I}_j^f$ associated with FBS j, and the expectations are with respect to the randomness of the uplink data arrival, channel state and control decisions. Our objective is to design an online control algorithm which maximizes the time average FC-tier throughput subject to FUE queue stability constraints and resource allocation constraints intrinsic to SC-FDMA. We

present the formal statement of the problem (**P1**) as below:

$$\text{Maximize} \quad \bar{D} \tag{20}$$

$$\text{Subject to} \quad (1), (2), (4), (6) \text{ and } (9), \forall t, \tag{21}$$

$$\bar{Q}_{ij} < \infty, \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}, \tag{22}$$

$$\text{Variables} \quad D_{ij}(t), \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}, \forall t \tag{23}$$

$$X_{ijk}^f(t), \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}, \forall k \in \mathcal{K}, \forall t \tag{24}$$

where (21) specifies the admission control and resource allocation constraints in the FC tier and constraint (22) guarantees that all FUEs in the system are *strongly stable*. We stress here that problem **P1** is optimized over the FC-tier admission control decisions $D_{ij}(t)$ and resource allocation decisions $X_{ijk}^f(t)$ on the premise that in each slot t, the MC-tier resource allocation decisions $X_{ik}^m(t)$ are made following the same rule as the FC tier. The two-tier resource allocation problems are coupled by possible cross-tier interference (cf. Sec. IV-B). Although our primary goal is to optimize the throughput of the underlaying FCs, we will also evaluate the performance of the MC by simulations in Sec. VIII.

Problem **P1** can be viewed a CMDP [37] which can usually be solved using traditional Dynamic Programming (DP) based techniques (*e.g.*, [2], [38]). However, this is problematic in our case for the following two reasons: (1) DP-based solutions suffer from the notorious "curse of dimensionality" where the state space $|\mathbb{S}|$ grows exponentially with the number of UEs, as revealed by the definition of $\mathbb{S}(t)$; and (2) the underlying state transition probabilities are unknown due to the absence of the statistical information (distributions) of $\mathbb{W}(t)$. Although learning-based approaches can be applied to estimating such statistics, they would, in general, suffer from a slow rate of convergence. To address these challenges, we take advantage of the recently developed Lyapunov optimization techniques [34] to design a joint admission control and resource allocation algorithm named JACRA in the next section.

IV. JOINT ADMISSION CONTROL AND RESOURCE ALLOCATION ALGORITHM

This section details the design of JACRA. First, we use drift analysis techniques [34] to recast the CMDP into a decomposable *drift-minus-reward minimization* problem. Then, we divide the problem into multiple independent admission control and resource allocation subproblems whose solutions rely solely on the current system state without the statistics of uplink data arrival and channel condition. We will show that each admission control subproblem has a simple solution structure that is based on a threshold criterion. Meanwhile, for each base station, the corresponding resource allocation subproblem is shown to be a Binary Integer Program (BIP) which maximizes the total queue backlog weighted capacity of all subscribing UEs. Finally, we present the analytical performance of JACRA, and comment on its implementation and compatibility issues in practice.

A. Problem Transformation

Let $\boldsymbol{Q}(t) \triangleq (Q_{ij}^f(t))$ denote the concatenated queue backlog of all FUEs in the system. We define the following quadratic Lyapunov function $L(\boldsymbol{Q}(t))$ to measure the uplink congestion state of the FC tier at the beginning of slot t:

$$L(\boldsymbol{Q}(t)) = \frac{1}{2} \sum_{i \in \mathcal{I}_j^f} \sum_{j \in \mathcal{J}} Q_{ij}^f(t)^2. \tag{25}$$

Without loss of generality, we assume that all FUE buffers are empty when $t = 0$ such that $L(\boldsymbol{Q}(0)) = 0$. Intuitively, a larger value of $L(\boldsymbol{Q}(t))$ implies heavier queue congestion in the FC tier. Based on (25), we then define the one-slot conditional Lyapunov drift as

$$\Delta(\boldsymbol{Q}(t)) = \mathbb{E}\{L(\boldsymbol{Q}(t+1)) - L(\boldsymbol{Q}(t))|\boldsymbol{Q}(t)\}, \tag{26}$$

the analysis of which is the key step in designing our stable online control algorithm JACRA.

Subtracting from Eq. (26) the conditional expectation of the FC throughput $D(t)$ in the t-th slot given $\boldsymbol{Q}(t)$ (*i.e.*, a reward function $\mathbb{E}\{D(t)|\boldsymbol{Q}(t)\}$), we obtain the following *drift-minus-reward* term:

$$\Delta(\boldsymbol{Q}(t)) - V\mathbb{E}\{D(t)|\boldsymbol{Q}(t)\}, \tag{27}$$

where $V > 0$ is a tunable parameter that controls the tradeoff between the drift $\Delta(\boldsymbol{Q}(t))$ and the reward $\mathbb{E}\{D(t)|\boldsymbol{Q}(t)\}$. According to the design principle of Lyapunov optimization [34], the admission control and resource allocation decisions should be chosen to minimize an upper bound on Eq. (27) in each slot t. Theorem 1 stated below provides such an upper bound.

Theorem 1. *(Bounding the Drift-Minus-Reward Term) Suppose $A_{ij}^f(t), \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}$ is i.i.d. over slots. Under any control algorithm, the drift-minus-reward term* (27) *has the following upper bound for any values of t, $\boldsymbol{Q}(t)$ and $V \geq 0$:*

$$\Delta(\boldsymbol{Q}(t)) - V\mathbb{E}\{D(t)|\boldsymbol{Q}(t)\} \leq B + \sum_{i \in \mathcal{I}_j^f} \sum_{j \in \mathcal{J}} \mathbb{E}\Big\{\Big[Q_{ij}^f(t)$$
$$- V\Big] D_{ij}(t)|\boldsymbol{Q}(t)\Big\} - \sum_{i \in \mathcal{I}_j^f} \sum_{j \in \mathcal{J}} \tau\mathbb{E}\Big\{Q_{ij}^f(t)C_{ij}^f(t)\Big\}, \tag{28}$$

where $B \triangleq \frac{\sum_{j \in \mathcal{J}} |\mathcal{I}_j^f|[(C^{max}\tau)^2 + (A^{max})^2]}{2}$ is a finite constant.

Proof: See Appendix A. ∎

By Theorem 1, we have transformed the original problem **P1** into minimizing the Right-Hand Side (R.H.S) of Eq. (28) subject to the basic admission control and resource allocation constraints (21). In the next subsection, we design JACRA to achieve this goal.

B. Algorithm Design

Algorithm 1 describes the pseudo-code of JACRA. In each slot t, the algorithm performs the following three control operations: (1) admission control in each FUE (**Lines 8 − 12**); (2) resource allocation in each BS (**Lines 4 − 7**); and (3) queue updating in each UE (**Lines 13 − 20**).

(1) Uplink admission control. Observe that the second term on the R.H.S of Eq. (28) involves the AC decision $D_{ij}(t)$. Considering that each FUE makes AC decisions independently,

1: **procedure** JACRA($\mathcal{I}_j^f, \mathcal{I}^m, \mathcal{J}, \mathcal{K}, T, V$)
2:    **Initialization**: $t \leftarrow 0, \boldsymbol{Q}(0) \leftarrow \boldsymbol{0}, Q_i^m(0) \leftarrow 0, \forall i \in \mathcal{I}^m$.
3:    **while** $t < T$ **do**
4:      **for all** $j \in \mathcal{J}$ **do**
5:        $(\mathcal{K}_{ij}^{f*}(t))_i = $ resource_alloc($(Q_{ij}^f(t))_i, \mathcal{K}, \mathcal{I}_j^f, g_{ijk}^f(t)$);   ▷ RA subproblem **P3** solved in each FBS at the beginning of slot t
6:      **end for**
7:      $(\mathcal{K}_i^{m*}(t))_i = $ resource_alloc($(Q_i^m(t))_i, \mathcal{K}, \mathcal{I}^m, g_{ik}^m(t)$);   ▷ RA subproblem **P4** solved in the MBS at the beginning of slot t
8:      **for all** $j \in \mathcal{J}$ **do**
9:        **for all** $i \in \mathcal{I}_j^f$ **do**
10:          $D_{ij}^*(t) = $ admission_control($Q_{ij}^f(t), A_{ij}^f(t), V$);   ▷ AC subproblem **P2** solved in each FUE at the end of slot t
11:        **end for**
12:      **end for**
13:      **for all** $j \in \mathcal{J}$ **do**
14:        **for all** $i \in \mathcal{I}_j^f$ **do**
15:          $Q_{ij}^f(t+1) = $ queue_updating($Q_{ij}^f(t), \mathcal{K}_{ij}^{f*}(t), D_{ij}^*(t)$);   ▷ Queue updating by Eq. (15) in each FUE at the end of slot t
16:        **end for**
17:      **end for**
18:      **for all** $i \in \mathcal{I}^m$ **do**
19:        $Q_i^m(t+1) = $ queue_updating($Q_i^m(t), \mathcal{K}_i^{m*}(t), A_i^m(t)$);   ▷ Queue updating by Eq. (16) in each MUE at the end of slot t
20:      **end for**
21:      $t \leftarrow t + 1$;
22:    **end while**
23:    **return** $D_{ij}^*(t), \mathcal{K}_{ij}^{f*}(t), \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}, \forall t \in \mathcal{T}$ and $\mathcal{K}_i^{m*}(t), \forall i \in \mathcal{I}^m, \forall t \in \mathcal{T}$.
24: **end procedure**

Algorithm 1: Joint admission control and resource allocation algorithm **JACRA**

we can decouple the minimization of this term into $\sum_{j \in \mathcal{J}} |\mathcal{I}_j^f|$ subproblems (**P2**) as follows

$$\min_{D_{ij}(t)} \left[Q_{ij}^f(t) - V \right] D_{ij}(t) \qquad (29)$$

$$\text{Subject to} \quad D_{ij}(t) \leq A_{ij}^f(t), \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}. \qquad (30)$$

Problem **P2** can be viewed as a min-weight problem which has the following simple *threshold-based* solution:

$$D_{ij}^*(t) = \begin{cases} A_{ij}^f(t), & \text{if } Q_{ij}^f(t) \leq V, \\ 0, & \text{otherwise.} \end{cases} \qquad (31)$$

It is solved by each FUE at the end of each slot t. The computation and storage overhead for implementing this policy in each FUE is negligible due to its simple structure.

Insight: Eq. (31) implies that FUE i associated with FBS j chooses to reject all new data if its current queue backlog $Q_{ij}^f(t)$ reaches to some preset upper threshold V. Otherwise, it would admit all new data. For example, according to Eq. (15), if the queue backlog of a FUE accumulates due to intensive data arrival and/or limited channel capacity and codes, say as a result of severe inter-cell interference, then it is more likely that the newly arrived data will be rejected in avoidance of queue instability. Therefore, we say that each FUE performs *channel-aware* admission control. Furthermore, given the same data arrival and channel conditions, a larger value of V enforces a looser admission control policy, allowing more data to be admitted.

   (2) Uplink resource allocation. The third term on the R.H.S of Eq. (28) involves the subchannel allocation decisions $X_{ijk}^f(t)$ in the FC tier. Since each FBS j allocates resources independently, we can decompose the maximization of this

term into the following $|\mathcal{J}|$ subproblems (**P3**):

$$\max_{X_{ijk}^f(t)} \sum_{i \in \mathcal{I}_j^f} \left\{ Q_{ij}^f(t) \sum_{\mathcal{K}_{ij}^f(t)} \log_2 \left[1 + \frac{P_{ijk}^f(t) g_{ijk}^f(t)}{\sigma^2} \right] \right\} \qquad (32)$$

$$\text{Subject to} \quad (2), (4), (6) \text{ and } (9). \qquad (33)$$

Problem **P3** is solved by each FBS at the beginning of slot t in a distributed fashion. Meanwhile, we assume that the MBS employs the same rule for uplink resource allocation. Then the optimal resource allocation problem (**P4**) for the MBS can be analogously written as

$$\max_{X_{ik}^m(t)} \sum_{i \in \mathcal{I}^m} \left\{ Q_i^m(t) \sum_{\mathcal{K}_i^m(t)} \log_2 \left[1 + \frac{P_{ik}^m(t) g_{ik}^m(t)}{\sigma^2} \right] \right\} \qquad (34)$$

$$\text{Subject to} \quad (3), (5), (7) \text{ and } (10). \qquad (35)$$

We use $(\mathcal{K}_{ij}^{f*}(t))_{i \in \mathcal{I}_j^f}$ and $(\mathcal{K}_i^{m*}(t))_{i \in \mathcal{I}^m}$ to denote the optimal solutions to problems **P3** and **P4** in slot t, respectively.

   Insight: The control objective of problem **P3** implies that each FBS $j, j \in \mathcal{J}$ should choose the resource allocation decisions to maximize the *total weighted capacity* of all subscribed FUEs, where the weight associated with FUE $i, i \in \mathcal{I}_j^f$ is its current queue backlog $Q_{ij}^f(t)$. Intuitively, it is preferable to allocate subchannels with higher quality to the set of most backlogged FUEs to achieve higher system throughput.

As shown in the objective function (32) of problem **P3**, each FBS $j, j \in \mathcal{J}$ does not have the knowledge of the inter-cell interference information $I_{ijk}^f(t)$ at the beginning of each slot t because the derivation of $I_{ijk}^f(t)$ depends upon the resource allocation decisions of other BSs (including the MBS). However, after resource allocation decisions have been made at each BS, $I_{ijk}^f(t)$ becomes available to FBS j, and the actual achievable capacity of FUE $i, i \in \mathcal{I}_j^f$ can be calculated using Eq. (13) wherein the SINR term $\Upsilon_{ijk}^f(t)$ is computed by

Eq. (11). Then the system evolves to state $\mathbb{S}(t+1)$ according to Eq. (15) and the effect of $I_{ijk}^f(t)$ is reflected in each FUE's queue backlog at time point $(t+1)\tau$, *i.e.*, $Q_{ij}^f(t+1)$. Since each FBS uses the queue backlog information reported from active FUEs for making decisions, we say that it performs *interference-aware* resource allocation in each slot t. The analysis above also implies that the FC-tier resource allocation problem **P3** (and hence the original problem **P1**) is coupled with the MC-tier resource allocation problem **P4** by possible cross-tier interference. Similar analysis applies to problem **P4** as well, omitted here for brevity.

Searching optimal solutions to problems **P3** and **P4** is desirable, albeit hard due to their combinatorial nature. Previous studies [4], [5] have proven that the subchannel allocation problem in SC-FDMA based LTE uplink is NP-hard due to the contiguous allocation constraint. Consider the scenario where N UEs are to share K subchannels ($K \geq N$), it has been shown in [4] that the feasible search space has a cardinality of $\sum_{u=0}^{N}\binom{K+u}{2u}\frac{N!}{(N-u)!}$. For a FC where the number of active FUEs is small, exhaustive search for the optimal solution may still apply. For example, the subchannel allocation problem represented by the matrix $\mathbf{H}$ in Sec. III-B has 988 feasible solutions. Suppose checking one feasible solution takes 10^{-9}s [4]. Then the total runtime of exhaustive search is approximately 10^{-6}s ($\ll$ TTI $= 10^{-3}$s). However, for a MC where there are at least tens of active MUEs, say $N = 40, K = 15$ [29], it would require searching through over 1.2×10^9 feasible solutions, resulting in a total runtime of 1.2s ($\gg$ TTI $= 10^{-3}$s). This is unacceptable for online implementation in practical systems. To address this issue, we will propose a heuristic resource allocation algorithm with polynomial computational complexity in Sec. V.

(3) Queue updating. At the end of slot t, the FUE and MUE queues are updated by taking the optimal control decisions $D_{ij}^*(t), \mathcal{K}_{ij}^{f*}(t), \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}$ and $\mathcal{K}_i^{m*}(t), \forall i \in \mathcal{I}^m$ into Eqs. (15) and (16), respectively.

From the analyses above, we can easily verify that the proposed JACRA algorithm minimizes the R.H.S of Eq. (28). In addition, it relies solely on the current system information $\mathbb{S}(t)$ to make control decisions, without the knowledge of the statistics (distribution information) of data arrival and channel state.

C. Analytical Performance

In this subsection, we present the achievable performance of JACRA by mathematical analyses. The following theorem provides deterministic bounds on the FUE queue backlog and the time average FC-tier throughput under JACRA.

Theorem 2. *(Performance of JACRA) Suppose $\mathbf{Q}(0) = \mathbf{0}$. For any control parameter $V > 0$, the queue backlog of FUE i associated with FBS j is upper bounded for all t under the JACRA algorithm as below:*

$$Q_{ij}^f(t) \leq A^{max} + V, \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}. \tag{36}$$

In addition, the time average FC-tier throughput $\bar{D}$ has the

following lower bound:

$$\liminf_{T \to \infty} \sum_{t=0}^{T-1} \mathbb{E}\{D(t)\} \geq \bar{D}^{opt} - \frac{B}{V}, \tag{37}$$

*where B is the same constant defined in Theorem 1 and $\bar{D}^{opt}$ is the optimal value of the objective of problem **P1**.*

Proof: See Appendix B. ∎

Insight: Eq. (36) in Theorem 2 guarantees that all FUE queues in the system are strongly stable since the queue backlog of each FUE is upper bounded by $A^{max} + V$ in every slot t. Thus we have constraint (22) satisfied for all FUEs. Meanwhile, we can also see that the gap between its achieved time average FC-tier throughput and the optimal throughput $\bar{D}^{opt}$ is within B/V, as shown by Eq. (37) in Theorem 2. Evidently, the control parameter $V > 0$ plays a key role in the achievable performance of JACRA. For example, by choosing V to be sufficiently large, the time average FC-tier throughput can approach arbitrarily close to the optimal value $\bar{D}^{opt}$. However, it comes at the cost of increasing the upper bounds on FUE queue backlog. By Little's law, the bound on the response delay of each FUE would increase as well.

V. A Heuristic Subchannel Allocation Algorithm

In this section, we propose an iterative heuristic algorithm with lower computational complexity to solve the resource allocation subproblems **P3** and **P4**. At each iteration, exactly one subchannel is allocated to a single UE. The basic idea of the heuristic is to let the most backlogged UEs enjoy higher priority over the less backlogged UEs in accessing high-quality subchannels. Algorithm 2 describes the pseudo-code of the heuristic designed for MC-tier resource allocation.[3] It comprises of the following four major steps:

Line 2: Initially, the set of subchannels allocated to each MUE $i, i \in \mathcal{I}^m$ is empty, *i.e.*, $\mathcal{K}_i^m(t) = \emptyset$. Let $\mathcal{K}_i^{m,v}(t)$ denote the set of subchannels that are viable to be allocated to MUE i. Let $\bar{\mathcal{K}}$ denote the set of subchannels that are available for allocation. Since all subchannels are available for all MUEs in the beginning, we have $\mathcal{K}_i^{m,v}(t) = \mathcal{K}, \forall i \in \mathcal{I}^m$ and $\bar{\mathcal{K}} = \mathcal{K}$.

Lines $4 - 8$: For all MUEs $i, i \in \mathcal{I}^m$ and for all subchannels that are viable to be allocated to MUE i, *i.e.*, $\mathcal{K}_i^{m,v}(t) \cap \bar{\mathcal{K}}$, the MBS first computes the product $\prod_{i,k}$ of the current queue backlog of MUE i and its achievable capacity on subchannel k, and then chooses the allocation (i^*, k^*) which maximizes $\prod_{i,k}$. In this way, MUEs with larger queue backlog are given priority in accessing high-quality subchannels.

Lines $9 - 13$: After (i^*, k^*) has been determined, subchannel k^* is added to the set of subchannels $\mathcal{K}_{i^*}^m(t)$ allocated to MUE i^*, and is excluded from the available subchannel set $\bar{\mathcal{K}}$. Meanwhile, the viable subchannel set $\mathcal{K}_{i^*}^{m,v}(t)$ for MUE i^* is updated to meet the subchannel adjacency requirement. To prevent less backlogged MUEs from starving, we subtract the total capacity $C_{i^*}^m(t)$ achieved by MUE i^* with current allocation $\mathcal{K}_{i^*}^m$ from $Q_{i^*}^m(t)$ to degrade its priority.

Lines $14 - 15$: The algorithm terminates when the set of available subchannel $\bar{\mathcal{K}}$ becomes empty and returns the

---

[3]The same procedure applies to FC-tier resource allocation as well.

1: **procedure** HEURISTIC-RESOURCE-ALLOCATION($\mathcal{I}^m, \mathcal{K}, t$)
2:    **Initialization**: $\mathcal{K}_i^m(t) = \emptyset, \mathcal{K}_i^{m,v}(t) = \mathcal{K}, \forall i \in \mathcal{I}^m, \bar{\mathcal{K}} = \mathcal{K}$;
3:    **repeat**
4:      **for all** $i \in \mathcal{I}^m$ **do**
5:        **for all** $k \in \mathcal{K}_i^{m,v}(t) \cap \bar{\mathcal{K}}$ **do**

$$\prod_{i,k} = Q_i^m(t)\left[W\log_2\left(1 + \frac{P_{ik}^m(t)g_{ik}^m(t)}{\sigma^2}\right)\right]; \qquad (38)$$

6:        **end for**
7:      **end for**
8:      $(i^*, k^*) = \arg\max_{i,k} \prod_{i,k}$;
9:      $\mathcal{K}_{i^*}^m(t) = \mathcal{K}_{i^*}^m(t) \cup k^*$;
10:     $\mathcal{K}_{i^*}^{m,v}(t) = \{\min(\mathcal{K}_{i^*}^m(t)) - 1, \max(\mathcal{K}_{i^*}^m(t)) + 1\} \cap \bar{\mathcal{K}}$;
11:     $\bar{\mathcal{K}} = \bar{\mathcal{K}} \backslash k^*$;
12:     $C_{i^*}^m(t) = \sum_{k \in \mathcal{K}_{i^*}^m} W\log_2\left(1 + \frac{P_{i^*k}^m(t)g_{i^*k}^m(t)}{\sigma^2}\right)$;
13:     $Q_{i^*}^m(t) = [Q_{i^*}^m(t) - C_{i^*}^m(t)]^+$;   ▷ Priority degradation
14:    **until** $\bar{\mathcal{K}} = \emptyset$
15:    **return** $\mathcal{K}_i^m(t), \forall i \in \mathcal{I}^m$.
16: **end procedure**

Algorithm 2: A heuristic uplink RA algorithm (MC tier)

resource allocation decision $\mathcal{K}_i^m(t)$ for all MUEs $i, i \in \mathcal{I}^m$ in slot t.

A. A Word on Complexity

Since a total number of K subchannels are to be allocated, the maximum number of *major iterations* (**Lines** $4 - 13$) is K. At each major iteration, at most $|\mathcal{I}^m| \times K$ operations are required to search for the allocation (i^*, k^*) that maximizes $\prod_{i,k}$. Therefore, the worst-case time complexity of the proposed heuristic is $\mathcal{O}(NK^2)$ where $N \triangleq |\mathcal{I}^m|$. In practice, the time complexity is lower than this because for each MUE i, the viable subchannel set $\mathcal{K}_i^{m,v}(t)$ has a maximal cardinality of two after the first major iteration (**Line** 10). It matches the time complexity of the greedy based heuristic proposed in [6], and outperforms that of the local ratio technique based approximate algorithm proposed in [8], which is $\mathcal{O}(NK^3)$.

To illustrate the computational complexity of the heuristic, consider the case when $N = 40, K = 15$. The worst-case time complexity is on the order of merely 9×10^3 operations, which is a substantial improvement compared to 1.2×10^9 using the exhaustive search method.

B. Algorithm Implementation and Compatibility

We see from Eqs. (32), (34) and (38) that implementing the proposed algorithms requires the knowledge of UE queue backlog and uplink channel state information at each base station. In current LTE-A standards, this can be achieved utilizing the *existing* Buffer Status Reporting (BSR) [26] and UE sounding procedures [39].

By configuration of a `periodicBSR` timer at the radio resource control protocol layer, the BSR mechanism allows each UE to periodically report its buffer status, in the form of BSR MAC control elements, to the associated BS via the PUCCH. Meanwhile, to obtain uplink CSI, each BS can send SRS requests to its UEs by configuring the `SRS request` field in Downlink Control Information (DCI). In response, each UE receiving the request would transmit the SRS in the last symbol of a subframe via the PUCCH [28]. The optimal resource allocation decisions obtained by Algorithm 2 can be fed back to MUEs via the PDCCH.

TABLE II
ALGORITHM COMPARISON

Algorithm	AC	RA	Problems	Complexity	Solver
JACRA	√	√	**P2, P3, P4**	NP	CPLEX
Heuristic	×	√	-	$\mathcal{O}(NK^2)$	-
PF	×	√	**P5, P6**	NP	CPLEX
RR	×	√	-	$\mathcal{O}(1)$	-

Conclusively, the proposed JACRA and heuristic resource allocation algorithms are compatible with current LTE-A standards.

VI. ALTERNATIVE RESOURCE ALLOCATION SCHEMES

For comparative purposes, we consider two alternative resource allocation schemes for SC-FDMA: (1) proportional fair scheduling scheme [4], [5], [8]; and (2) round robin scheduling scheme [6], [7].

The PF scheduling scheme has been widely used for designing time and Frequency-Domain Packet Scheduling (FDPS) algorithms in wireless systems [4], [5], [8]. It aims at maximizing the system capacity while maintaining proportional fairness among all users [5]. Let $\bar{C}_i^m(t)$ represent the average capacity MUE $i, i \in \mathcal{I}^m$ has received up to time point $t\tau$. Let $C_{ik}^m(t)$ represent the instantaneous capacity for MUE i on subchannel k in slot t. We define $\xi_{ik}^m(t) \triangleq C_{ik}^m(t)/\bar{C}_i^m(t)$ as the *PF metric value* that MUE i has on subchannel k in slot t. Then, for all $t \in \mathcal{T}$, the MC-tier proportional fair FDPS problem (**P5**) can be written as

$$\max_{X_{ik}^m(t)} \sum_{i \in \mathcal{I}^m} \sum_{k \in \mathcal{K}_i^m(t)} X_{ik}^m(t)\xi_{ik}^m(t) \qquad (39)$$

subject to resource allocation constraints (3), (5), (7) and (10). The objective function (39) indicates that the under the PF scheduling scheme, MUEs with higher instantaneous channel capacity $C_{ik}^m(t)$ and lower historical average capacity $\bar{C}_i^m(t)$ stand a better chance of obtaining subchannels. It is easy to see that the hardness of **P5** is the same as that of **P4** due to the contiguous subchannel allocation constraint (7). Similarly, the FC-tier proportional fair FDPS problem (**P6**) in slot t can be written as

$$\max_{X_{ijk}^f(t)} \sum_{i \in \mathcal{I}_j^f} \sum_{k \in \mathcal{K}_{ij}^f(t)} X_{ijk}^f(t)\xi_{ijk}^f(t), \ \forall j \in \mathcal{J}, \qquad (40)$$

subject to resource allocation constraints (2), (4), (6) and (9) where $\xi_{ijk}^f(t)$ denotes the PF metric value that FUE i associated with FBS j has on subchannel k in slot t.

The round robin scheduling scheme assigns an equal number of contiguous subchannels to each UE in circular order, regardless of UE queue backlog and subchannel state [6]. It has fixed time complexity $\mathcal{O}(1)$ that is independent of the number of UEs as well as the number of subchannels.

Tab. II compares the four algorithms in the aspects of involved function blocks, corresponding optimization problems, time complexity and solver used in simulation. Given the hardness of problems **P3**-**P6**, we use the IBM ILOG CPLEX

TABLE III
PATH LOSS MODEL [$P_L(d)$ IN DB]

From \ To	associated FBS	non-associated FBSs	MBS
FUE	$38.46 + 20\log_{10}(d) + 6.9n_{iw}$	$38.46 + 20\log_{10}(r) + 6.9n_{iw} + 0.8d + 20n_{ew}$	$55.3 + 37.6\log_{10}(d)$
MUE		$38.46 + 20\log_{10}(r) + 6.9n_{iw} + 0.8d + 20n_{ew}$	$15.3 + 37.6\log_{10}(d)$

mixed integer optimizer [40] to search for their *optimal* solutions in the C++ API.[4] The optimizer performs dynamic search using linear programming relaxation, branching, cuts, and heuristics. Compared to exhaustive search, the time complexity is significantly reduced (cf. Sec. VIII-B).

VII. SIMULATION SETUP

We simulate a LTE-A FC system where a hexagonal MC is underlaid with $|\mathcal{J}| = 20$ uniformly and randomly distributed FCs. We assume that the system operates at 2 GHz carrier frequency with $|\mathcal{K}| = 15$ subchannels where each subchannel has a bandwidth of $W = 180$ KHz (*i.e.*, 12 subcarriers × 15 KHz spacing) [28], [29]. In the MC, a number of $|\mathcal{I}^m| = 40$ MUEs are uniformly and randomly distributed outdoor within a $R = 500$m radius from the central MBS. Each FC represents a ground-floor apartment in a residential setting and has $|\mathcal{I}_j^f| = 2$ authorized FUEs whose distances from the central FBS are 10m and 20m, respectively. The coverage radius of each FC is assumed to be $r = 20$m. The simulation is carried out for $T = 1000$ consecutive slots.

A. UE Traffic and Power Setup

In 3GPP LTE-A standard, a base station establishes multiple traffic bearers per UE to support diverse QoS demands. Specifically, it defines two main traffic bearer categories [13]: (1) Guaranteed Bit Rate (GBR) bearers established for real-time applications; and (2) non-GBR bearers established for non-real-time applications. Consider the fact that multiple mobile applications with different QoS demands may run simultaneously in a UE, we assume that each FUE (MUE) generates synthetic traffic load which is consisted of: (1) constant rate traffic from GBR applications at rate λ_g^f (λ_g^m) Mbps; and (2) Poisson traffic from non-GBR applications at mean rate λ_n^f (λ_n^m) Mbps. For each FUE, $\boldsymbol{\lambda^f} \triangleq (\lambda_g^f, \lambda_n^f)$ is chosen to be $(20, 40)$. For each MUE, $\boldsymbol{\lambda^m} \triangleq (\lambda_g^m, \lambda_n^m)$ is chosen to be

[4]Problems **P3-P6** need to be transformed into pure BIPs before using the CPLEX mixed integer optimizer. We refer interested readers to refs. [6], [7] for detailed techniques.
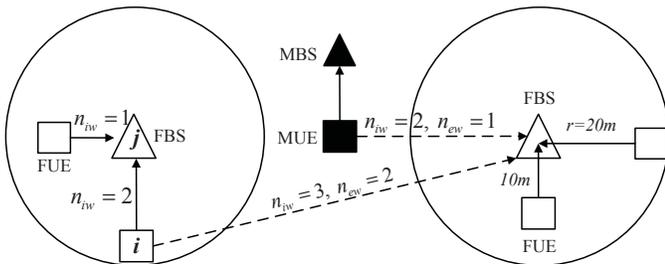
$(2, 4)$. We assume that the size of the uplink buffer in each UE is one megabyte. In case of buffer overflowing, incoming data will be discarded without retransmission.

We assume that the maximum UE transmit power P^{max} is 23 dBm. Each UE is equipped with a single transmit antenna with zero gain [6]. The receive antenna gains at the MBS and the FBSs are set to 14 dBi and 0 dBi, respectively [21].

B. Path Loss Model

The path loss model $P_L(d)$ between a UE and a BS is given in Tab. III [21], [41] where d is the transmitter-to-receiver distance, n_{iw} is the number of internal walls on the radio propagation path and n_{ew} is the number of external walls. Consider the short-range nature of communications inside a FC, we assume that the distance from FUE $i, i \in \mathcal{I}_j^f$ associated with FBS $j, j \in \mathcal{J}$ to any other base station $j' \in \bar{\mathcal{J}} \triangleq \{\mathcal{J} \backslash j\} \cup \{MBS\}$ can be approximated by the distance from FBS j to base station j' [30]. Fig. 2 illustrates the wall penetration setup along different radio propagation paths.

Channel gains are modeled considering both path loss and shadow fading. Specifically, we define $g(t) = 10^{-\frac{P_L(d)+\psi_{dB}}{10}}$ as the channel gain between the UE and the BS where ψ_{dB} is a Gaussian-distributed random variable with mean zero and standard deviation $\delta_{\psi_{dB}} = 6$ dB [41]. The power spectral density of background noise is assumed to be -174 dBm/Hz [2], [7]. Tab. IV summarizes the *default* parameters used in our simulation.

VIII. PERFORMANCE EVALUATION

A. System Performance under JACRA

Fig. 3a plots the average FC-tier throughput $\bar{D}$ vs. V under different settings of $\boldsymbol{\lambda^f}$ and $\boldsymbol{\lambda^m}$. It shows that as the value of V increases from 0 to 1.28 Mbits, $\bar{D}$ converges fast to the optimum for any $(\boldsymbol{\lambda^f}, \boldsymbol{\lambda^m})$ setting, with a declining gap $O(1/V)$ as revealed by Eq. (37) in Theorem 2. This is in consistence with Eq. (31) in Sec. IV-B in the sense that a larger value of V enforces a more relaxed admission control policy, allowing more data to be admitted and transmitted.

Fig. 2. Wall penetration along different radio propagation paths

TABLE IV
DEFAULT SIMULATION PARAMETERS

Parameter	Value	Parameter	Value
J	20	UE buffer size	1 MB
K	15	T	1,000
$|\mathcal{I}_j^f|, \forall j \in \mathcal{J}$	2	τ	TTI (1 ms)
$|\mathcal{I}^m|$	40	W	180 KHz
Carrier frequency	2 GHz	P^{max}	23 dBm
R, r	500, 20 m	BS antenna gains	14, 0 dBi
$\boldsymbol{\lambda^f}$	(20,40) Mbps	σ^2	-174 dBm/Hz
$\boldsymbol{\lambda^m}$	(2,4) Mbps	$\delta_{\psi_{dB}}$	6 dB

(a) Average FC-tier throughput vs. V

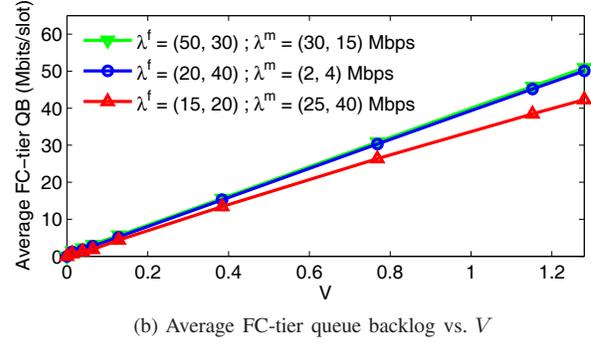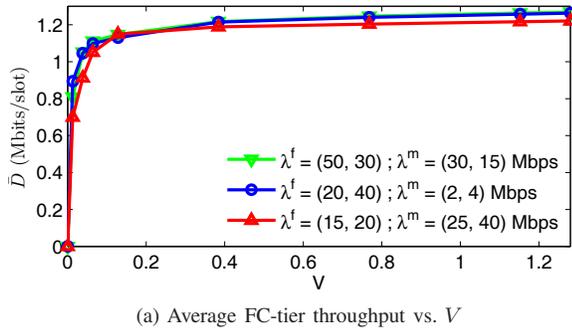(b) Average FC-tier queue backlog vs. V

Fig. 3. Average FC-tier throughput and queue backlog vs. V under **JACRA**
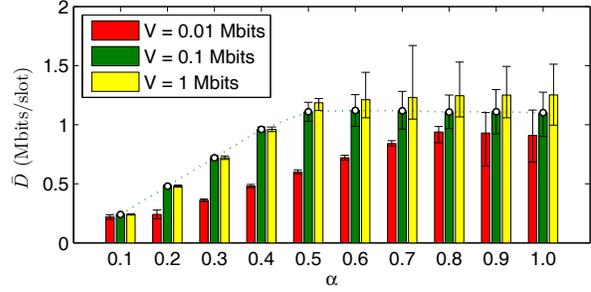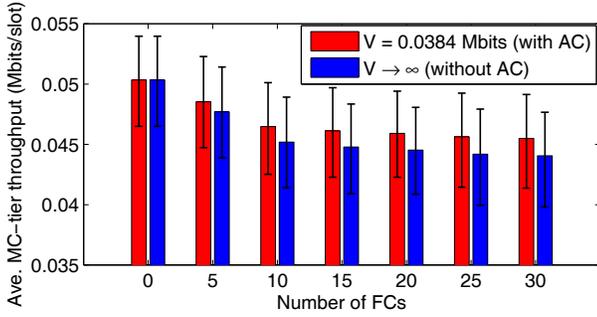

Fig. 4. Average MC-tier throughput at the 95% confidence level vs. number of FCs with/without FC-tier AC

Fig. 5. Average FC-tier throughput at the 95% confidence level vs. α under different admission control polices

Fig. 3b plots the average FC-tier queue backlog vs. V. We can see that the average queue backlog grows linearly in V, as expected by Eq. (36) in Theorem 2. Along with Fig. 3a, this demonstrates the $[O(1/V), O(V)]$ throughput-stability tradeoffs under JACRA. Further, such tradeoffs are not subject to the influence of traffic intensity and composition.

Fig. 4 illustrates the average MC-tier throughput [5] vs. the deployment density of FCs with or without FC-tier admission control. It can been seen that as the FC deployment density grows from 0 to 30, the average MC-tier throughput declines by 12.5% if admission control is not performed in the FC tier (*i.e.*, $V \to \infty$). In contrast, the average MC-tier throughput drops by 9.6% if each FUE performs admission control with a threshold of $V = 0.0384$ Mbits. In both cases, the reduction in the average MC-tier throughput is caused by stronger inter-cell interference in a FC deployment of higher density. However, the enforcement of FC-tier admission control leads to 3.3% extra gain in the average MC-tier throughput. Also, such throughput *gains* increase as the FC deployment density grows. This validates that FC-tier admission control is an effective way of mitigating inter-cell interference in LTE-A FC systems.

Fig. 5 plots the average FC-tier throughput vs. FUE data arrival rate under different admission control polices. In this experiment, we define a scale factor $\alpha \in (0, 1]$ to scale down the data arrival rate of each FUE. For a fixed value of α, it means that the data arrival rate of FUE i associated with FBS j is scaled down to $\alpha \mathbb{E}\{A_{ij}(t)\}, \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}$. We can

observe that for a given admission control policy, say $V = 0.1$ Mbits, the average FC-tier throughput grows *approximately* linearly in $\alpha, \alpha \leq 0.4$ when the system capacity is sufficient enough to transmit all admitted data. Nonetheless, the rate of growth decelerates when $\alpha > 0.4$ as the system becomes saturated.

Fig. 6 plots the average FC/MC-tier throughput vs. the MC radius R. As expected, we see that the average MC-tier throughput drops by 39.0% as the MC radius increases from 0.5 kilometers to 3.5 kilometers. This is because the given set of $|\mathcal{I}^m| = 40$ MUEs are more sparsely distributed within the coverage area of the MC, which results in reduced signal strength received at the MBS. On the contrary, the average FC-tier throughput witnesses an increase, due to mitigated inter-cell interference with a more sparse deployment.
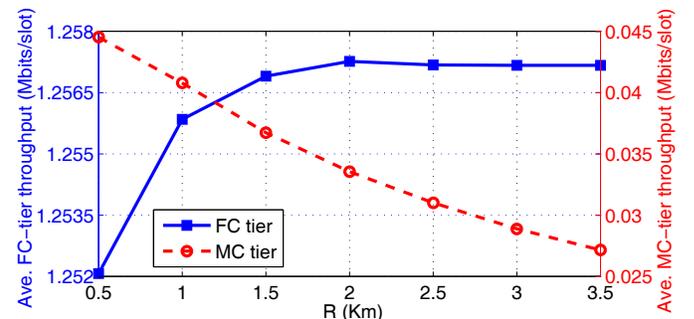
[5]The MC-tier throughput is defined as the total number of bits delivered by all MUEs in a slot, *i.e.*, $\sum_{i \in \mathcal{I}^m} \min[Q_i^m(t), C_i^m(t)]$.

Fig. 6. Average FC/MC-tier throughput vs. MC radius R
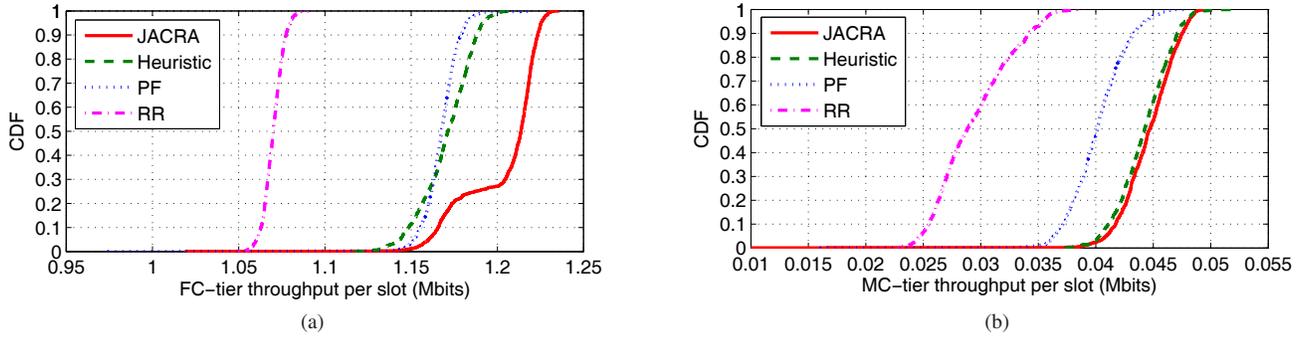
Fig. 7.   CDF of FC/MC-tier throughput per slot under different algorithms

B. Performance Comparison among Algorithms

In the following experiments, we evaluate the achievable performance of the proposed JACRA and heuristic algorithms by comparison with the alternative PF and RR scheduling schemes. We assume that FC-tier admission control is *not* performed at each FUE under JACRA, *i.e.*, $V \rightarrow \infty$, and thus the comparison is with respect to the performance of the resource allocation block.

Fig. 7 plots the Cumulative Distribution Function (CDF) of the throughput achieved per slot in both FC and MC tiers.[6] As expected, we can see that the proposed JACRA algorithm achieves the highest throughput among all algorithms. In the FC tier, JACRA can transmit, on average, 0.036 more megabits per slot than the PF scheduling scheme, and 0.133 more megabits per slot than the RR scheduling scheme, as shown in Fig. 7a. In the MC tier, an average amount of 0.004 and 0.015 more megabits can be transmitted under JACRA when respectively compared to the PF and RR scheduling schemes, as shown in Fig. 7b. Since the RR scheduling scheme performs resource allocation independent of UE queue backlog and subchannel state, we can observe that it achieves the lowest system throughput in both tiers. The PF scheduling scheme outperforms the RR scheduling scheme in that it takes into account the channel state information that each UE has on each subchannel. While it guarantees proportional fairness among UEs, without the knowledge of UE queue backlog, the PF scheduling scheme may assign high-quality subchannels

---

[6]Hereafter, the FC-tier throughput is defined as the total number of bits delivered by all FUEs in a slot, *i.e.*, $\sum_i \sum_j \min[Q_{ij}^f(t), C_{ij}^f(t)]$ instead of $D(t)$, since no admission control is performed at each FUE when $V \rightarrow \infty$.

to a UE which does not have much data to transmit. By means of queue-aware and channel-aware resource allocation, the proposed JACRA algorithm overcomes this drawback and hence achieves higher system throughput. Fig. 7 also shows that the proposed heuristic algorithm outperforms the RR scheduling scheme by transmitting, on average, 0.1 and 0.015 more megabits per slot in the FC and MC tiers, respectively. Compared to the PF scheduling scheme, the heuristic algorithm transmits, on average, 0.004 more megabits per slot in the MC tier, and 0.003 more megabits per slot in the FC tier.

Fig. 8 depicts the CDF of the proportional fair metric value per slot for both FC and MC tiers. The FC-tier PF metric values shown in Fig. 8a are averaged over the number of FCs. It can be seen that the PF scheduling scheme achieves the largest *FC-tier* PF metric value, as maximized in Eq. (40). However, the absolute difference between the *average* PF metric values achieved by any two algorithms does not exceed 0.21 per slot, which is rather small. Interestingly, as shown in Fig. 8b, the average *MC-tier* PF metric values of the proposed JACRA and heuristic algorithms even surpasses that of the PF scheduling scheme by 0.62 and 0.09, respectively. In addition, the average MC-tier PF metric values of the JACRA and heuristic algorithms are, respectively, 0.58 and 0.05 greater than that of the the RR scheduling scheme.

Fig. 9 plots the CDF of the number of iterations that JACRA takes to obtain the optimal allocation for the 1000 slots. For the MC tier, we can see that JACRA converges within 72 iterations for 90% of the slots. Furthermore, our algorithm takes at least 42 iterations and at most 200 iterations to converge. For the FC tier, it shows that JACRA converges much faster within
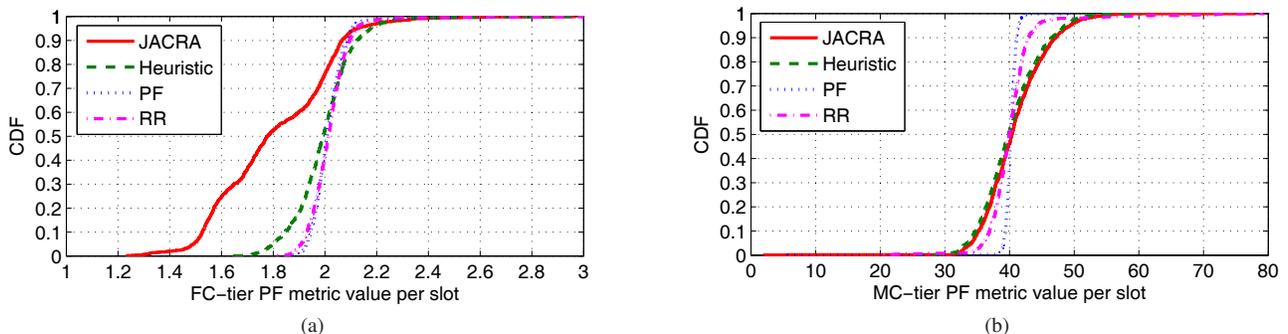

Fig. 8.   CDF of FC/MC-tier proportional fair metric value per slot under different algorithms
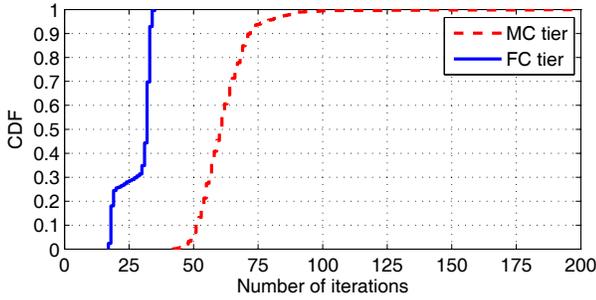
Fig. 9. CDF of the number of iterations to obtain the optimal allocation

17 to 35 iterations, due to a limited number of FUEs in each FC. On average, the proposed JACRA algorithm converges to the optimal allocation in 27 and 56 iterations when applied to FC-tier and MC-tier resource allocation, respectively. The time complexity is thus significantly reduced compared to exhaustive search.

We use the `QueryPerformanceFrequency` function in Windows API to test the rough *average* computation time of the four resource allocation algorithms on a PC with an Intel i3 processor at 3.30 GHz, as shown in Tab. V. It shows that the runtime of the optimal JACRA and PF algorithms is almost on the same order of magnitude as a slot, *i.e.*, 1000 microseconds. This makes these two algorithms infeasible for online implementation in practical LTE-A FC systems. In contrast, the proposed heuristic algorithm has very short computation time that is far less than a slot. From the analyses above, we can safely conclude that the proposed heuristic not only has low computational complexity, but achieves high system throughput with good UE fairness as well.

IX. CONCLUDING REMARKS AND FUTURE RESEARCH

In this paper, we have investigated the uplink admission control and resource allocation problem for LTE-A FC systems with co-channel deployment. We first formulated the problem as a constrained Markov decision problem that aims to maximize the time average throughput of the entire FC tier while maintaining FUE queue stability. Then we proposed a joint admission control and resource allocation algorithm called JACRA to obtain the optimal policies. We demonstrated that the optimal admission control policy has a simple structure that is based on a threshold criterion. Meanwhile, the optimal resource allocation decisions should be chosen to maximize the total queue backlog weighted capacity for each base station. Consider the NP-hardness of the resource allocation subproblem, we proposed an iterative heuristic resource allocation algorithm with polynomial time complexity. Simulation studies have shown that: (1) FC-tier admission control can

TABLE V
COMPARISON OF AVERAGE COMPUTATION TIME (IN MICROSECONDS)

Cell Tier	JACRA	Heuristic	PF	RR
FC-tier	271.3210	0.0380	249.1550	0.0004
MC-tier	2213.9700	0.6645	2038.4600	0.0028

mitigate inter-cell interference, allowing the MC to achieve higher throughput, especially in dense FC deployment scenarios; (2) the proposed JACRA algorithm outperforms baseline proportional fair and round robin scheduling schemes in terms of the time average FC-tier throughput with competitive UE fairness; and (3) the proposed heuristic achieves near-optimal throughput with substantial improvement in computational complexity, and is thus feasible for online implementation in practical LTE-A FC systems.

In our future work, we will address the following issues for *two-tier* LTE-A FC uplink: (1) QoS provisioning for real-time traffic with stringent delay/jitter requirements; and (2) energy-aware resource allocation with reliability guarantees, *e.g.*, via the Hybrid Automatic Repeat Request (HARQ) mechanism.

APPENDIX A
BOUNDING THE DRIFT-MINUS-REWARD TERM

Proof: We first introduce the following Lemma [34].

Lemma 1. *For any nonnegative real numbers a, b and c, there holds* $[\max(a-b,0)+c] \leq a^2 + b^2 + c^2 + 2a(c-b)$.

Squaring both sides of Eq. (15) and using Lemma 1, $\forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}$, we have

$$Q_{ij}^f(t+1)^2 - Q_{ij}^f(t)^2 \leq C_{ij}^f(t)^2\tau^2 + D_{ij}(t)^2 \\ - 2Q_{ij}^f(t)[C_{ij}^f(t)\tau - D_{ij}(t)]. \quad (41)$$

Plugging Eq. (41) into the definition of the Lypunov drift (26) yields

$$\Delta(\boldsymbol{Q}(t)) \leq \frac{1}{2} \sum_{i \in \mathcal{I}_j^f} \sum_{j \in \mathcal{J}} \mathbb{E}\left\{ C_{ij}^f(t)^2\tau^2 + D_{ij}(t)^2 | \boldsymbol{Q}(t) \right\} \quad (42)$$
$$- \sum_{i \in \mathcal{I}_j^f} \sum_{j \in \mathcal{J}} \mathbb{E}\left\{ Q_{ij}^f(t) \left[C_{ij}^f(t)\tau - D_{ij}(t) \right] | \boldsymbol{Q}(t) \right\}.$$

Recall the boundness assumptions $C_{ij}^f(t) \leq C^{max}, \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}, \forall t \in \mathcal{T}$ and $D_{ij}(t) \leq A^{max}, \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}, \forall t \in \mathcal{T}$, it can be readily seen that the first term on the R.H.S. of Eq. (42) is upper bounded by

$$B \triangleq \frac{1}{2} \sum_{j \in \mathcal{J}} |\mathcal{I}_j^f| \left[(C^{max}\tau)^2 + (A^{max})^2 \right]. \quad (43)$$

Substituting Eq. (43) into Eq. (42), subtracting from both sides $V\mathbb{E}\{D(t)|\boldsymbol{Q}(t)\}$ and rearranging terms proves the theorem. ∎

APPENDIX B
PERFORMANCE BOUNDS OF JACRA

Proof: We use mathematical induction to prove Eq. (36) in Theorem 2. Since $\boldsymbol{Q}(0) = \boldsymbol{0}$, Eq. (36) holds for $t = 0$. Now, suppose $Q_{ij}^f(t) \leq A^{max} + V, \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}$ for some $t > 0$. We will show that $Q_{ij}^f(t+1) \leq A^{max} + V$. Consider the following two cases:

- If $Q_{ij}^f(t) > V$, then FUE $i, i \in \mathcal{I}_j^f$ in FC $j, j \in \mathcal{J}$ chooses the AC decision $D_{ij}^*(t) = 0$ according to Eq. (31). Thus, by Eq. (15), we have $Q_{ij}^f(t+1) \leq Q_{ij}^f(t) \leq A^{max} + V$.

- If $Q_{ij}^f(t) \leq V$, by Eq. (31), we know that the optimal AC decision is $D_{ij}^*(t) = A_{ij}^f(t) \leq A^{max}$, *i.e.*, $Q_{ij}^f(t)$ can increase at most A^{max} in slot t. Again, by Eq. (15), we get $Q_{ij}^f(t+1) \leq A^{max} + V$.

Together, this completes the proof of Eq. (36). To prove Eq. (37) in Theorem 2, we first show in the following lemma that there exists a *randomized stationary policy* that achieves the optimal value of objective (20). Under this policy, the admission control and resource allocation decisions are made in every slot t according to some fixed distribution for each state.

Lemma 2. *(Optimality over Randomized Stationary Policy) Suppose $A_{ij}^f(t), \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}$ is an i.i.d. random process. Then there exists a randomized stationary policy π^* that chooses the admission control decision $D_{ij}(t), \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}$ and the resource allocation decision $X_{ijk}^f(t), \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}, \forall k \in \mathcal{K}$ in each slot t independent of $\boldsymbol{Q}(t)$, and yields the following steady-state equations:*

$$\mathbb{E}\{D^{\pi^*}(t)\} = \bar{D}^{opt}, \tag{44}$$

$$\mathbb{E}\{D_{ij}^{\pi^*}(t)\} \leq \mathbb{E}\left\{C_{ij}^{f,\pi^*}(t)\right\}, \forall i \in \mathcal{I}_j^f, \forall j \in \mathcal{J}. \tag{45}$$

Since Lemma 2 can be proven using similar techniques as [34], we omit the detailed proof here for brevity. Based on Lemma 2, we now prove the performance bounds on FUE queue backlog and FC throughput under the JACRA algorithm.

Recall that JACRA is designed to minimize the R.H.S. of Eq. (28). Thus the following inequality holds for *any* other feasible control policy π (inclusive of the optimal randomized stationary policy π^*):

$$\Delta(\boldsymbol{Q}(t)) - V\mathbb{E}\{D(t)|\boldsymbol{Q}(t)\} \leq B - V\mathbb{E}\{D^\pi(t)|\boldsymbol{Q}(t)\} \tag{46}$$
$$- \sum_{i \in \mathcal{I}_j^f} \sum_{j \in \mathcal{J}} \mathbb{E}\left\{Q_{ij}^f(t)\left[C_{ij}^{f,\pi}(t)\tau - D_{ij}^\pi(t)\right]|\boldsymbol{Q}(t)\right\}.$$

Substituting Eqs. (44) and (45) into the R.H.S. of Eq. (46), we have

$$\Delta(\boldsymbol{Q}(t)) - V\mathbb{E}\{D(t)|\boldsymbol{Q}(t)\} \leq B - V\bar{D}^{opt}. \tag{47}$$

Taking expectations over $\boldsymbol{Q}(t)$ on both sides of Eq. (47) yields

$$\mathbb{E}\{L(\boldsymbol{Q}(t+1)) - L(\boldsymbol{Q}(t))\} - V\mathbb{E}\{D(t)\} \leq B - V\bar{D}^{opt}. \tag{48}$$

Summing both sides of inequality (48) over $t \in \{0, ..., T-1\}$, using the fact that $L(\boldsymbol{Q}(0)) = 0, L(\boldsymbol{Q}(t)) \geq 0, \forall t$ and rearranging terms, we get

$$V\sum_{t=0}^{T-1} \mathbb{E}\{D(t)\} \geq (V\bar{D}^{opt} - B)T. \tag{49}$$

Dividing both sides by VT and taking a $\liminf$ as $T \to \infty$ proves Eq. (37) in Theorem 2. $\blacksquare$

ACKNOWLEDGMENT

REFERENCES

[1] A. Ghosh, R. Ratasuk, B. Mondal, N. M. Vedhe, and T. Thomas, "LTE-Advanced: next-generation wireless broadband technology," *IEEE Wireless Communications Magazine*, vol. 17, no. 3, pp. 10–22, June 2010.

[2] X. Xiang, J. Wan, C. Lin, and X. Chen, "A dynamic programming approximation for downlink channel allocation in cognitive femtocell networks," *Computer Networks*, vol. 57, no. 15, pp. 2976–2991, October 2013.

[3] D. López-Pérez, A. Valcarce, G. de la Roche, and J. Zhang, "OFDMA femtocells: A roadmap on interference avoidance," *IEEE Communications Magazine*, vol. 47, no. 9, pp. 41–48, September 2009.

[4] H. Yang, F. Ren, C. Lin, and J. Zhang, "Frequency-domain packet scheduling for 3GPP LTE uplink," in *Proc. of IEEE INFOCOM*, San Diego, CA, USA, March 2010, pp. 1–9.

[5] S.-B. Lee, I. Pefkianakis, A. Meyerson, S. Xu, and S. Lu, "Proportional fair frequency-domain packet scheduling for 3GPP LTE uplink," in *Proc. of IEEE INFOCOM*, Rio de Janeiro, Brazil, April 2009, pp. 2611–2615.

[6] I. C. Wong, O. Oteri, and W. McCoy, "Optimal resource allocation in uplink SC-FDMA systems," *IEEE Transactions on Wireless Communications*, vol. 8, no. 5, pp. 2161–2165, May 2009.

[7] A. Aijaz, X. Chu, and A. H. Aghvami, "Energy efficient design of SC-FDMA based uplink under QoS constraints," *IEEE Wireless Communications Letters*, vol. 3, no. 2, pp. 149–152, April 2014.

[8] F. Ren, Y. Xu, H. Yang, J. Zhang, and C. Lin, "Frequency-domain packet scheduling with stability analysis for 3GPP LTE uplink," *IEEE Transactions on Mobile Computing*, vol. 12, no. 12, pp. 2412–2426, December 2013.

[9] J. Kim, D. Kim, and Y. Han, "Proportional fair scheduling algorithm for SC-FDMA in LTE uplink," in *Proc. of IEEE GLOBECOM*, Anaheim, CA, USA, December 2012, pp. 4816–4820.

[10] H. Zhang, N. Prasad, S. Rangarajan, S. Mekhail, S. Said, and R. Arnott, "Standards-compliant LTE and LTE-A uplink power control," in *Proc. of IEEE ICC*, Ottawa, ON, Canada, June 2012, pp. 5275–5279.

[11] D. J. Dechene and A. Shami, "Energy-aware resource allocation strategies for LTE uplink with synchronous HARQ constraints," *IEEE Transactions on Mobile Computing*, vol. 13, no. 2, pp. 422–433, February 2014.

[12] T. Mert, O. Kaya, and H. A. Çirpan, "Jointly optimal chunk and power allocation in uplink SC-FDMA," in *Proc. of IEEE ICC*, Budapest, Hungary, June 2013, pp. 3393–3397.

[13] M. Kalil, A. Shami, and A. Al-Dweik, "Power-efficient QoS scheduler for LTE uplink," in *Proc. of IEEE ICC*, Budapest, Hungary, June 2013, pp. 6200–6204.

[14] H.-S. Jo, C. Mun, J. Moon, and J.-G. Yook, "Interference mitigation using uplink power control for two-tier femtocell networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 10, pp. 4906–4910, October 2009.

[15] V. Chandrasekhar and J. G. Andrews, "Uplink capacity and interference avoidance for two-tier femtocell networks," *IEEE Transactions on Wireless Communications*, vol. 8, no. 7, pp. 3498–3509, July 2009.

[16] Z. Shi, M. C. Reed, and M. Zhao, "On uplink interference scenarios in two-tier macro and femto co-existing UMTS networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2010, April 2010.

[17] Y. Sun, R. P. Jover, and X. Wang, "Uplink interference mitigation for OFDMA femtocell networks," *IEEE Transactions on Wireless Communications*, vol. 11, no. 2, pp. 614–625, February 2012.

[18] A. T. Gamage and X. S. Shen, "Uplink resource allocation for inter-working of WLAN and OFDMA-based femtocell systems," in *Proc. of IEEE ICC*, Budapest, Hungary, June 2013, pp. 6071–6075.

[19] Z. Zhang, L. Song, Z. Han, and W. Saad, "Coalitional games with overlapping coalitions for interference management in small cell networks," *IEEE Transactions on Wireless Communications*, vol. 13, no. 5, pp. 2659–2669, May 2014.

[20] Z. Feng, J. Deng, L. Song, and Z. Han, "Joint access control and subchannel allocation scheme for OFDMA femtocell network using a truthful mechanism," in *Proc. of ACM PINGEN*, Istanbul, Turkey, August 2012, pp. 39–44.

[21] M. Morita, T. Nobukiyo, and K. Hamabe, "Uplink power control method for LTE femtocells based on resource usage aggregation," in *Proc. of IEEE PIMRC*, Sydney, Australia, September 2012, pp. 442–447.

[22] Z. Zheng, J. Hämäläinen, and Y. Yang, "On uplink power control optimization and distributed resource allocation in femtocell networks," in *Proc. of IEEE VTC Spring*, Yokohama, Japan, 2011, pp. 1–5.

This article has been accepted for publication in a future issue of this journal, but has not been fully edited. Content may change prior to final publication. Citation information: DOI 10.1109/TVT.2014.2351837, IEEE Transactions on Vehicular Technology

15

[23] Y. Bai and L. Chen, "Hybrid spectrum arrangement and interference mitigation for coexistence between LTE macrocellular and femtocell networks," *EURASIP Journal on Wireless Communications and Networking*, vol. 2013, 2013.

[24] W. Liu, X. Chen, and W. Wang, "Price-based uplink resource allocation for hybrid-spectrum femtocell networks," in *Proc. of IEEE WCNC*, Shanghai, China, April 2012, pp. 2665–2670.

[25] A. Dalal, H. Li, and D. P. Agrawal, "Fractional frequency reuse to mitigate interference in self-configuring LTE-femtocells network," in *Proc. of IEEE MASS*, Valencia, Spain, October 2011, pp. 49–54.

[26] 3GPP, "Medium access control (MAC) protocol specification (release 12)," TS 36.321 V12.2.1, Tech. Rep., June 2014.

[27] M. Husso, J. Hämäläinen, R. Jäntti, J. Li, E. Mutafungwa, R. Wichman, Z. Zheng, and A. Wyglinski, "Interference mitigation by practical transmit beamforming methods in closed femtocells," *EURASIP Journal on Wireless Communications and Networking*, vol. 2010, May 2010.

[28] 3GPP, "Physical channels and modulation (release 12)," TS 36.211 V12.0.0, Tech. Rep., December 2013.

[29] ——, "User Equipment (UE) radio transmission and reception (FDD) (release 12)," TS 36.101 V12.3.0, Tech. Rep., March 2014.

[30] 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 Transactions on Wireless Communications*, vol. 12, no. 1, pp. 180–194, January 2013.

[31] F. Jin, R. Zhang, and L. Hanzo, "Fractional frequency reuse aided twin-layer femtocell networks: Analysis, design and optimization," *IEEE Transactions on Communications*, vol. 61, no. 5, pp. 2074–2085, May 2013.

[32] R. Urgaonkar and M. J. Neely, "Opportunistic scheduling with reliability guarantees in cognitive radio networks," *IEEE Transactions on Mobile Computing*, vol. 8, no. 6, pp. 766–777, June 2009.

[33] ——, "Opportunistic cooperation in cognitive femtocell networks," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 607–616, April 2012.

[34] M. J. Neely, *Stochastic network optimization with application to communication and queueing systems*. Morgan & Claypool, 2010.

[35] J. R. Yang and D. I. Kim, "Power control for two-tier femtocell networks using pricing mechanism via emergency message," in *Proc. of IEEE APCC*, Sabah, Malaysia, October 2011, pp. 63–67.

[36] H. Wang, Q. Cui, X. Tao, M. Valkama, and Y. J. Guo, "Optimal cooperative water-filling power allocation for OFDM system," in *Proc. of IEEE WCNC*, Shanghai, China, April 2013, pp. 3742–3747.

[37] E. Altman, *Constrained Markov Decision Processes*. Boca Raton, FL, USA: Chapman and Hall/CRC Press, 1999.

[38] X. Xiang, C. Lin, and X. Chen, "Energy-efficient link selection and transmission scheduling in mobile cloud computing," *IEEE Wireless Communications Letters*, vol. 3, no. 2, pp. 153–156, April 2014.

[39] 3GPP, "Physical layer procedures (release 12)," TS 36.213 V12.2.0, Tech. Rep., June 2014.

[40] *IBM ILOG CPLEX Optimization Studio - CPLEX Users Manual (Version 12 release 4)*, IBM, 2011. [Online]. Available: http://pic.dhe.ibm.com/infocenter/cosinfoc/v12r4/topic/ilog.odms.studio.help/pdf/usrcplex.pdf

[41] 3GPP, "FDD base station (BS) classification (release 11)," TR 25.951 V11.0.0, Tech. Rep., September 2012.

Chuang Lin (IEEE SM'04) is a professor of the Department of Computer Science and Technology, Tsinghua University, Beijing, China. He received the Ph.D. degree in Computer Science from the Tsinghua University in 1994. His current research interests include computer networks, performance evaluation, network security analysis, and Petri net theory and its applications. He has published more than 300 papers in research journals and IEEE conference proceedings in these areas and has published three books. Professor Lin is a member of ACM Council, a senior member of the IEEE and the Chinese Delegate in TC6 of IFIP. He serves as the Technical Program Vice Chair, the 10th IEEE Workshop on Future Trends of Distributed Computing Systems (FTDCS 2004); the General Chair, ACM SIGCOMM Asia workshop 2005; the Associate Editor, IEEE Transactions on Vehicular Technology; the Area Editor, Journal of Computer Networks; and the Area Editor, Journal of Parallel and Distributed Computing.

Xin Chen is a professor in the Computer School of Beijing Information Science and Technology University, Beijing, China. He received the Ph.D degree in Computer Science from Beijing Institute of Technology. In 2006, he received the postdoctoral fellowship in Computer Architecture from Tsinghua University. His current research focuses on performance evaluation of wireless networks. Prof. Chen is a senior member of China Computer Federation (CCF), a member of the CCF Technical Committee of Theoretical Computer Science (TCTCS) and the CCF Technical Committee of Petri Nets (TCPN).

Xuemin (Sherman) Shen (M'97-SM'02-F'09) received the B.Sc. degree from Dalian Maritime University, Dalian, China, in 1982, and the M.Sc. and Ph.D. degrees from Rutgers University, New Brunswick, NJ, USA, in 1987 and 1990, respectively, all in electrical engineering.

He is a Professor and University Research Chair with the Department of Electrical and Computer Engineering, University of Waterloo, Waterloo, ON, Canada. He was the Associate Chair for Graduate Studies from 2004 to 2008. He is a coauthor/editor of six books and has published many papers and book chapters in wireless communications and networks, control, and filtering. His research focuses on resource management in interconnected wireless/wired networks, wireless network security, wireless body area networks, and vehicular ad hoc and sensor networks.

Dr. Shen is a registered Professional Engineer of Ontario, Canada, a Fellow of the Canadian Academy of Engineering, a Fellow of the Engineering Institute of Canada, and a Distinguished Lecturer of the IEEE Vehicular Technology Society and IEEE Communications Society. He served as the Technical Program Committee Chair for IEEE VTC Fall 2010, the Symposia Chair for IEEE ICC 2010, the Tutorial Chair for IEEE VTC Spring 2011 and IEEE ICC 2008, the Technical Program Committee Chair for IEEE GLOBECOM 2007, the General Co-Chair for Chinacom 2007 and QShine 2006, the Chair for the IEEE Communications Society Technical Committee on Wireless Communications, and P2P Communications and Networking. He also serves/served as the Editor-in-Chief for *IEEE Network, Peer-to-Peer Networking and Application*, and *Communications*; a Founding Area Editor for IEEE TRANSACTIONS ON WIRELESS COMMUNICATIONS; an Associate Editor for IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, Computer Networks, and Wireless Networks; and a Guest Editor for the IEEE JOURNAL ON SELECTED AREAS IN COMMUNICATIONS, *IEEE Wireless Communications, IEEE Communications Magazine*, and *Mobile Networks and Applications*.

Xudong Xiang received the B.Sc. (2008) degree in Information Management and Information Systems and the M.Sc. (2011) degree in Management Science and Engineering, both from Beijing Information Science and Technology University (BISTU). Currently, he is a Ph.D candidate in Computer Science in the Department of Computer Science and Technology, University of Science and Technology Beijing (USTB), Beijing, China. Since 2011 fall, he has been doing research in the Next Generation Internet Division of the Tsinghua National Laboratory for Information Science and Technology (TNLIST). His current research interests include optimal control and performance evaluation of computer networks, and resource management in distributed systems. He is a student member of the China Computer Federation (CCF) and the IEEE.