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/JIOT.2015.2496161, IEEE Internet of Things Journal

Probabilistic Analysis on QoS Provisioning for Internet of Things in LTE-A Heterogeneous Networks with Partial Spectrum Usage

Ran Zhang, Miao Wang, Xuemin (Sherman) Shen, and Liang-liang Xie

***Abstract*—This paper investigates QoS provisioning for Internet of Things (IoT) in LTE-Advanced heterogeneous networks (HetNets) with partial spectrum usage (PSU). In HetNets, the IoT users with ubiquitous mobility support or low-rate services requirement can connect with macrocells (MCells), while femtocells (FCells) with PSU mechanism can be deployed to serve the IoT users requiring high-data-rate transmissions within small coverage. Despite the great potentials of HetNets in supporting various IoT applications, the following challenges exist: i) how to depict the unplanned random behaviors of the IoT-oriented FCells and cope with the randomness in user QoS provisioning, and ii) how to model the interplay of resource allocation between MCells and FCells under PSU mechanism. In this work, the Stochastic Geometry theory is first exploited to statistically analyze how the unplanned random behaviors of the IoT-oriented FCells impact the user performance, considering the user QoS requirements and FCell PSU policy. Particularly, to satisfy the QoS requirements of different IoT user types, the concept of effective bandwidth (EB) is leveraged to provide the users with probabilistic QoS guarantee, and a heuristic algorithm named QA-EB algorithm is proposed to make the EB determination tractable. Then, the interplay of resource allocation between the MCells and FCells is formulated into a two-level Stackelberg game, where the two parties try to maximize their own utilities through optimizing the macro-controlled interference price and the femto-controlled PSU policy. A backward induction method is proposed to achieve the Stackelberg equilibrium. Finally, extensive simulations are conducted to corroborate the derived SINR and ergodic throughput performance of different user types and demonstrate the Stackelberg equilibrium under varying user QoS requirements and spectrum aggregation capabilities.**

***Index Terms*—Internet of things (IoTs), femtocells, heterogeneous networks, partial spectrum usage, stochastic geometry, Stackelberg game.**

I. Introduction

As one of the most promising communication paradigms in recent decades, the Internet of Things (IoT) has expanded the scope of the conventional Internet by providing a pervasive network to intelligently interconnect and manage billions of physical objects embedded with sensing, computing and communication capabilities. Through IoT, connected devices can communicate with Internet for desired services or exchange information with other devices. To accommodate the enormous amount of IoT traffic with different Quality of Service (QoS) requirements, the Long-term evolution advanced (LTE-A) system [1] has become a key enabler by providing ubiquitous, reliable and high-data-rate communications. Particularly, the LTE-A standard has specified the heterogeneous networks (HetNets) technology [2] [3] for service-differentiated communications. In HetNets, macrocells (MCells) can provide ubiquitous mobility support and low-rate services; while in data-intensive areas, femtocells (FCells) can be deployed overlaid with MCells to significantly boost the transmission rates within a small coverage range.

Although the HetNets offer a capacity surge for LTE-A systems to support IoT, the extensive overlaid deployment of FCells has caused considerable co-channel interference between MCells and FCells as well as among FCells themselves, thus degrading the overall system performance. There have been extensive research works devoted to mitigating the co-channel interference in HetNets [4]–[8], which mainly focus on aborative resource coordination between MCell base stations (MBSs) and FCell base stations (FBSs) within a single operator carrier. On the other hand, instead of seeking for optimal interference mitigation scheme in one single carrier, LTE-A has offered a new promising technology, Carrier Aggregation (CA) [9], to better coordinate the HetNets interference. As demonstrated in [10] [11], CA empowers concurrent utilization of multiple carriers so that IoT devices can transmit on a wider bandwidth with higher rates. With CA, MBSs can operate on a whole set of carriers, while FBSs can dynamically choose a subset of carriers based on the interference intensity in different carriers [12]. In this way, both the macro-femto and inter-femto interference can be more effectively suppressed compared to single-carrier interference mitigation. In this paper, the carrier-level spectrum management in HetNets is referred to as partial spectrum usage (PSU).

PSU in HetNets has been studied in a few works [13] [14]. In [13], Cao *et al.* studied the optimal PSU factor for FCells to improve both the energy and spectrum efficiency. In [14], Garcia *et al.* proposed an autonomous carrier selection strategy where newly plugged FBSs avoid interfering the nearby BSs by choosing the proper subset of carriers in a cognitive manner. However, there are still two main challenges in HetNets with PSU mechanism to properly support IoT applications.

R. Zhang, M. Wang, X. Shen, and L. Xie are with the Department of Electrical and Computer Engineering, University of Waterloo, 200 University Avenue West, Waterloo, Ontario, Canada, N2L 3G1 (e-mail: {r62zhang, m59wang, sshen, llxie}@uwaterloo.ca).

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/JIOT.2015.2496161, IEEE Internet of Things Journal

First, the unplanned random behaviors of IoT-oriented F-Cells make it difficult to appropriately evaluate the user performance and further provide QoS guarantee. As IoT is a pervasive network connecting various kinds of devices, its generated traffic has much randomness due to human activities, power instabilities, environment changes, etc. As a result, the FCells deployed to support IoT applications may (dis)appear anytime at anyplace. In addition, IoT-oriented FCells are usually installed by end users and are connected to operators' core network via private Internet provider [2], causing relatively large signaling latency for exchanging resource management information between MCells and FCells. Consequently, it is difficult for MBSs to monitor the random behaviors of FCells and directly manage the interference from FCells in a timely manner. Thus, it is important to model the FCell randomness and investigate how it affects the HetNets performance.

Some recent works [15]–[17] have investigated this issue by exploiting stochastic geometry (SG) [18] [19] to capture the FCell randomness and provide tractable interference modeling and throughput evaluation. In [15], Zhong *et al.* considered hybrid-access FCells in a single-carrier scenario to study the throughput for both FCell subscribers and nonsubscribers. In [16], Zhang *et al.* studied a multi-carrier scenario and exploited SG to derive the user ergodic throughput for LTE and LTE-A users. In [17], Lin *et al.* proposed a generalized SG-based framework where the multi-carrier multi-flow users are analyzed in HetNets. Most of the existing works exploiting SG assume that every user is assigned with the same portion of bandwidth disregarding the users' quality of service (QoS) requirements. Moreover, the PSU mechanism and CA capabilities (i.e., the number of carriers a user device can aggregate concurrently) are not considered either. Therefore, to converge service-differentiated IoT applications into LTE-A HetNets, it is desirable to incorporate the FCell randomness into performance analysis of HetNets with PSU and meanwhile consider the IoT users' QoS requirements and CA capabilities.

Second, the interplay of resource allocation (RA) between the MBSs and FBSs under LTE-A PSU mechanism has not been well studied. Although an MBS cannot directly control the RA of FBSs deployed to serve IoT, it can influence the RA decisions indirectly through price control [20] [21]. For instance, in [20], Duan *et al.* designed a game-theory-based price control strategy where MBSs influence the FBS behaviors by determining the user service prices for MCells and FCells. In [21], Bu *et al.* proposed to set an interference price for MBSs over FBSs based on the interference from FBSs, which is considered by FBSs into RA to maximize their own utilities. Through such interaction, the interference between MCells and FCells can be effectively coordinated. However, most of the existing works have not considered the PSU mechanism, and usually assume either an isolated MCell scenario or no interference from neighboring MCells. The assumptions do not hold in LTE-A systems where the MCell frequency reuse factor equals to 1, i.e., one MCell shares the same spectrum with its neighboring MCells, making the interference from other MCells is nonneglectable. Therefore, a new interaction strategy considering the inter-macro interference under the PSU mechanism is indispensable for LTE-A HetNets.

In this paper, we investigate the QoS provisioning issue for IoT in LTE-A HetNets with PSU mechanism. The IoT users that require ubiquitous mobility support or low-rate services connect to MBSs, while FBSs are deployed to serve the IoT users that require high-data-rate transmissions. Hybrid-access FCells are considered where a subset of carriers are reserved for the FCell subscribers (FSs), i.e., the FCell IoT users subscribed to FCells, while another disjoint subset is open to provide paid services to FCell nonsubscribers (FNSs), i.e., the FCell IoT users not subscribed to FCells. Two challenging issues are addressed: i) under PSU, both the FCell random behaviors and the inter-macro interference are deliberately modeled and incorporated into the performance analysis framework, with considering the IoT users' QoS requirements and CA capabilities; and ii) the interplay between MBSs and FBSs is formulated into correlated utility maximization problems to determine the optimal RA decisions. Specifically, our contributions are four-fold,

- We first model the locational randomness of MBSs, FBSs and IoT users into Poisson Point Processes (PPPes) [22]. Stochastic Geometry theory is exploited to obtain the signal-to-interference-plus-noise (SINR) distributions and ergodic throughput (measures long-term average user throughput) for different user types in each carrier. The derivation considers PSU mechanism, user CA capabilities and configurable user bandwidth.
- To satisfy the QoS requirements of different IoT user types with appropriate bandwidth assignment, the concept *effective bandwidth* [23] is leveraged to provide a unified bandwidth for each user type based on the derived SINR distributions. With the derived effective bandwidth, IoT users are provided with probabilistic QoS guarantee. Particularly, to make the decision process for effective bandwidth practical and tractable, an heuristic algorithm named QA-EB algorithm is proposed based on iterations.
- The interplay of RA between MCells and FCells is formulated into a two-level Stackelberg game. In the game, based on the aggregate throughput of FBSs, MBSs first impose an interference-related price upon FBSs, and FBSs adjust their PSU policy accordingly. A backward induction method is proposed to achieve the Stackelberg equilibrium (i.e., optimal price and PSU policy) and to show how the price and PSU policy are tuned to maximize the utilities of both parties.
- Finally, simulation results validate our analytical ones, and the Stackelberg equilibrium is demonstrated under different IoT user QoS requirements and CA capabilities.

The remainder of the paper is organized as follows. Section II describes the system model. The stochastic-geometry-based analytical framework is proposed to evaluate the performance for different user types in Section III, and the Stackelberg game is formulated and analyzed in Section IV. Simulation results are presented in Section V. Finally, Section VI con-

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/JIOT.2015.2496161, IEEE Internet of Things Journal

cludes the paper.

II. System Model

In this section, the PPP-based HetNets layout is first presented. The bandwidth access mechanisms and physical channel model are then introduced, followed by the interaction model between macro and femto cells. The main notations are listed in Table I.

TABLE I: Notation Table

Parameters	Values
λ^{MBS} (λ^{FBS})	The density of the MBSs (FBSs)
λ^{MU}	The density of the MUs
λ^{FS} (λ^{FNS})	The density of the FSs (FNSs)
P^{MBS} (P^{FBS})	MBS (FBS) power spectrum density
H	Fast fading channel gain
N	Total number of carriers
n_{agg}	User CA capability
n^{res} (n^{open})	The number of carriers for FS (FNSs) access
W_{PRB}	PRB bandwidth in carrier i
r^{MU}	The throughput requirement of MUs
r^{F}	The throughput requirement of FSs and FNSs
W_i^{MU}	The effective bandwidth assigned to each MU in carrier i
W_i^{FS} (W_i^{FNS})	The effective bandwidth assigned to each FS (FNS) in carrier i
R_i^{MU}	Ergodic throughput of the MUs
R_i^{FS} (R_i^{FNS})	Ergodic throughput of the FSs (FNSs)
$Q_{s\|i}^{MU}$	User service probability of one MU
$Q_{s\|i}^{FS}$ ($Q_{s\|i}^{FNS}$)	User service probability of one FS (FNS)
$\theta_i^{MBS,usa}$	Bandwidth usage probability of one MBS
$\theta_i^{FBS,usa}$	Bandwidth usage probability of one FBS
y	Interference price from the MBS to each FBS
g^{MU} (g^{FNS})	Unit profit of MU (FNS) services
g^{FS}	Unit cost of FS services

A. Network Deployment

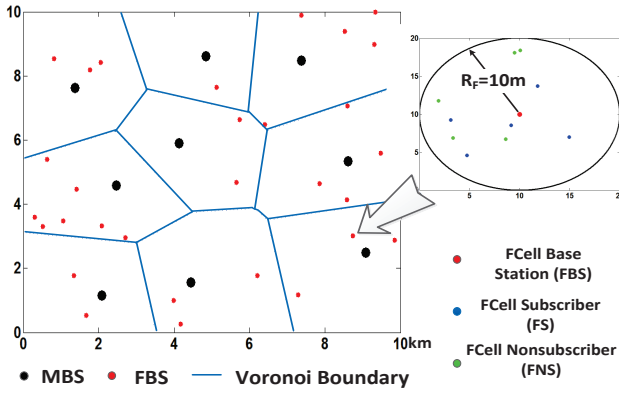


Fig. 1: The network layout of HetNets. Voronoi cells formed by 9 MBSs are uniformly located in a $10 \times 10km^2$ area.

As shown in Fig. 1, we consider an arbitrary region A with area $|A|$, where the MBSs and FBSs are deployed as homogeneous PPPes with density measure λ^{MBS} and λ^{FBS}, respectively. In other words, MBSs (or FBSs) are uniformly distributed within A with the total number following a Poisson distribution - $Poisson(\lambda^{MBS}|A|)$ (or $Poisson(\lambda^{FBS}|A|)$). Due to much smaller transmission powers, FCell coverage is much smaller than MCell coverage. The sets of MBSs and FBSs are denoted as Φ^{MBS} and Φ^{FBS}, respectively.

The MCell users (MUs) are distributed within A following a homogeneous PPP with density λ^{MU}. Each MU connects to its nearest MBS for service. Under such an association policy, the actual coverage of an MBS becomes a Voronoi cell [24] where any point in a Voronoi cell has a shorter distance to the associated MBS than to other MBSs. The FSs (or FNSs) are distributed as a homogeneous PPP with density λ^{FS} (or λ^{FNS}), in a disk coverage of FCell with radius R_F.

B. Bandwidth Allocation Mechanisms

The system bandwidth consists of N carriers. Each carrier i ($\in \{1, ..., N\}$) is further divided into P_i orthogonal Physical Resource Blocks (PRBs[1]), each with bandwidth W_{PRB}. A PRB is the minimum bandwidth allocation unit in LTE-A systems, as shown in Fig. 2.

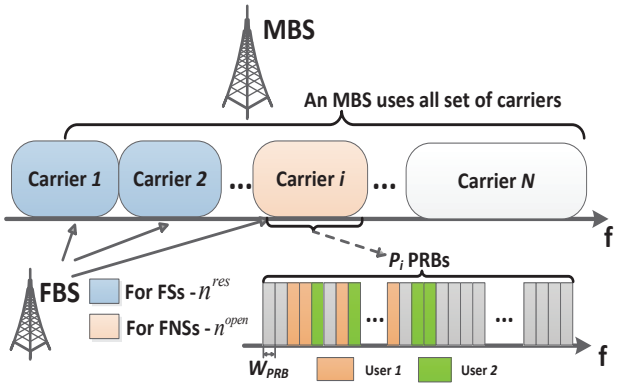


Fig. 2: Bandwidth structure of HetNets under PSU. In carrier i, the PRBs are all orthogonal. One PRB can only be assigned to one user within 1 subframe which is 1ms long, while one user can occupy several PRBs concurrently. The PRBs assigned to one user can be contiguous or not [25].

All users are assumed to have a CA capability n_{agg} indicating that a user can transmit on n_{agg} carriers simultaneously. One MU requires a minimum throughput r_u^{MU}, while one FS or FNS enjoying high-speed services requires minimum throughput r_u^F ($r_u^F > r_u^{MU}$). To provide the users with probabilistic guarantee on the throughput requirement, each MU, FS and FNS is assigned with effective bandwidth W_i^{MU}, W_i^{FS} and W_i^{FNS}, respectively in carrier i, such that

$$\begin{aligned} &\Pr\left(\sum_{i=1}^{N} R_i^{MU}(W_i^{MU}) < r_u^{MU}\right) < e \ll 1, \\ &\Pr\left(\sum_{i=1}^{N} R_i^{T}(W_i^{T}) < r_u^{F}\right) < e \ll 1,\ T \in \{FS, FNS\}, \end{aligned} \tag{1}$$

where $R_i^T(W_i^T)$ denotes the ergodic throughput that a type-T user can get from carrier i given effective bandwidth W_i^T, and e denotes a small positive value much smaller than 1. Eq. (1) means that the total ergodic rate of all carriers for a user

[1]LTE-A is built upon the Orthogonal Frequency Division Multiple Access (OFDMA) technology, and a PRB consists of 12 contiguous OFDMA subcarriers.

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/JIOT.2015.2496161, IEEE Internet of Things Journal

should be smaller than its required throughput with a very small probability.

For the access mechanism, each MBS operates on all N carriers to serve MUs, while each hybrid-access FBS randomly and independently chooses n^{res} carriers to serve the FSs and n^{open} disjoint carriers for open access, satisfying $n^{res}+n^{open}\leq N$. For one MBS or FBS, denote the number of type-T users choosing carrier i for transmission as N_i^T and the maximum number of users that can be concurrently served in carrier i as $N_i^{T,ser}=\lfloor P_i\cdot W_{PRB}/W_i^T\rfloor,\ T\in\{MU,FS,FNS\}$. In each subframe, if $N_i^T\leq N_i^{T,ser}$, the system will randomly choose N_i^T pieces of W_i^T bandwidth for type-T users; otherwise, time-sharing scheduling is adopted to randomly select $N_i^{T,ser}$ type-T users to transmit. In this way, each type-T user can be served with equal long-term time-proportion within one carrier.

C. Physical Channel Model

The path loss and fast fading effects are considered in this paper. The shadowing effects are not included as [26] has proved that the shadowing can be well approximated by the randomness of the Poisson distributed BS locations. This is a strong justification that the distribution of MBSs can be modelled as a PPP.

We consider that the power spectrum densities (PSD) of MBSs and FBSs are fixed in carrier i and denoted as P_i^{MBS} and P_i^{FBS}, respectively. For a user, its received PSD in carrier i from an MBS (or FBS) B with a distance of D_B is

$$P_i^r=P_i^T H {D_B}^{-\alpha_i},\ B\in\Phi^T,\ T\in\{MBS,FBS\} \tag{2}$$

where H is the fast fading channel gain and α_i is the path loss exponent. The fast fading of the useful signal is considered as Rayleigh fading, so the fast fading channel gain follows an exponential probability density function[2] (pdf), i.e., Exp(μ). For simplicity, we set μ as 1. The fast fading of the interference signals is considered as generally distributed.

D. Economic Interaction between Macro and Femto Cells

The objective of both parties is to maximize their own utilities, which are expressed as the weighted summations of multiple parts of profits. Each MBS charges MUs' services with unit price g^{MU}/bit. Meanwhile, to preserve the MU performance from the interference by FBSs, MBSs impose an interference unit price y_i over FBSs for interfering the carrier i. For analytical simplicity, y_i's are set to be equal for all carriers and denoted as y in the rest of the paper. Besides, an upper bound y^{max} is imposed on y to avoid overcharging, which is reasonable in practical. Therefore, there are two parts of profits for one MBS: the profits from MU services and the profits from charging all FBSs within its coverage.

Bearing the interference price y, each FBS will optimize the subsets of carriers assigned to FSs and FNSs, i.e., n^{res} and n^{open}, considering the effective bandwidth of all the user types and CA capabilities. One FBS pays unit price g^{FS}/bit for the FSs' services and can gain profits with unit price g^{FNS}/bit from FNSs. Therefore, the total utility of one FBS is the three-fold: profits from FNSs, service payment for FSs, and the interference cost charged by the MBSs.

[2]If a random variable Ra is Rayleigh distributed, then its power Ra^2 is exponentially distributed with parameter μ.

III. Probabilistic Analysis on User Performance for HetNets with PSU

In this section, the Stochastic Geometry is exploited to model the HetNets interference with PSU, considering user QoS requirements and CA capabilities. Both the FCell randomness and the multi-macro interference are included. Specifically, the SINR distributions and ergodic rates for each type of users in each carrier are first derived (Subsection III-A and III-B), and then the effective bandwidth is finalized according to the user QoS requirement (Subsection III-C).

A. SINR Distributions and User Ergodic Rates

The SINR distribution of an MU in carrier i is derived first. The probability that $SINR_i^{MU}$ is larger than a threshold β is

$$\begin{aligned}&\mathrm{P}(SINR_i^{MU}>\beta)=\mathrm{P}(\tfrac{P_i^{MBS}HD_{B_0}{}^{-\alpha_i}}{I_i^{MBS}+I_i^{FBS}+n_0}>\beta)\\&=\mathrm{P}(H>\tfrac{\beta(I_i^{MBS}+I_i^{FBS}+n_0)D_{B_0}{}^{\alpha_i}}{P_i^{MBS}}),\\&where\ I_i^{MBS}=\sum_{B\in\Phi_i^{MBS}\backslash B_0}P_i^{MBS}H_i^{MBS}D_B{}^{-\alpha_i}\\&I_i^{FBS}=\sum_{F\in\Phi_i^{FBS}}P_i^{FBS}H_i^{FBS}D_F{}^{-\alpha_i}.\end{aligned} \tag{3}$$

In Eq. (3), B_0 is the associated MBS of the considered MU. Notation Φ_i^{MBS} (Φ_i^{FBS}) denotes the set of MBSs (FBSs) that use the same PRBs with the considered MU in carrier i; I_i^{MBS} (I_i^{MBS}) denotes the interference PSD from Φ_i^{MBS} (Φ_i^{FBS}); and H_i^{MBS} (H_i^{FBS}) denotes the fast fading channel gain between the considered user to the MBSs (FBSs). As $H\sim\mathrm{Exp}(1)$, we have $\mathrm{P}(H>h)=e^{-h}$. Then

$$\begin{aligned}&\mathrm{P}(H>\tfrac{(I_i^{MBS}+I_i^{FBS}+n_0)D_{B_0}{}^{\alpha_i}\beta}{P_i^{MBS}})\\&=E_{\Phi_i^{MBS}\backslash B_0,\Phi_i^{FBS},H_i^{MBS},H_i^{FBS},D_{B_0}}\\&\qquad[\exp(-\tfrac{(I_i^{MBS}+I_i^{FBS}+n_0)D_{B_0}^{-\alpha_i}\beta}{P_i^{MBS}})],\end{aligned} \tag{4}$$

where $E[\cdot]$ denotes the expectation and $\Phi_i^{MBS}\backslash B_0$ is the set Φ_i^{MBS} excluding MBS B_0. As randomness exists in $\Phi_i^{MBS}\backslash B_0$, Φ_i^{FBS}, H_i^{MBS}, H_i^{FBS} and D_{B_0}, $\mathrm{P}(SINR_i^{MU}>\beta)$ should be an expectation over all these items. Proposition 1 gives the derived SINR distribution of one MU in carrier i.

Proposition 1: In the HetNets described in the system model, given the effective bandwidth of all the user types (i.e., W_i^{MU}, W_i^{FS}, and W_i^{FNS}), the probability that the SINR of

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/JIOT.2015.2496161, IEEE Internet of Things Journal

one MU in carrier i is larger than a threshold β is given as

$$\begin{aligned}&\mathrm{P}(SINR_i^{MU} > \beta)\\&= \int_0^{+\infty} 2\pi\lambda^{MBS} d e^{-\pi\lambda^{MBS}d^2} e^{-n_0 d^{\alpha_i}\beta/P_i^{MBS}}\\&\quad\cdot \exp\{-2\pi\lambda_i^{MBS}\eta(d, H_i^{MBS}, \beta)\}\\&\quad\cdot \exp\{-2\pi\theta_i^{FBS}\lambda^{FBS,usa}\epsilon(d, H_i^{FBS}, \beta, A)\}d(d),\\&\text{where}\\&\eta(d, H_i^{MBS}, \beta) = -\tfrac{1}{2}d^2 + \tfrac{1}{2}d^2 E_{H_i^{MBS}}\{e^{-\beta H_i^{MBS}} +\\&\qquad(\beta H_i^{MBS})^{2/\alpha_i}\left[\Gamma(1-\tfrac{2}{\alpha_i}, 0) - \Gamma(1-\tfrac{2}{\alpha_i}, \beta H_i^{MBS})\right]\},\\&\epsilon(d, H_i^{FBS}, \beta, A) = \tfrac{1}{2}d^2\Gamma(1-\tfrac{2}{\alpha_i}, 0)E_{H_i^{FBS}}\left[(A\beta H_i^{FBS})^{2/\alpha_i}\right],\\&\Gamma(s,t) = \textstyle\int_t^{+\infty} x^{s-1}e^{-x}dx, \text{ and } A = P_i^{FBS}/P_i^{MBS}.\end{aligned} \tag{5}$$

Please refer to **Appendix A** for the detailed derivations of Proposition 1. One interesting observation in Proposition 1 is that the SINR distribution is not related to the effective bandwidth of any user type. This is because fixed transmission PSDs are considered for both MBSs and FBSs in the system model. Therefore, the ergodic throughput of the MUs in carrier i can be calculated as

$$\begin{aligned}R_i^{MU} &= E_{SINR_i^{MU}}\left[\mathrm{Q}_{s|i}^{MU} W_i^{MU}\log(1+SINR_i^{MU})\right]\\&= \tfrac{\mathrm{Q}_{s|i}^{MU} W_i^{MU}}{ln2}\int_0^{+\infty}\mathrm{P}(\ln(1+SINR_i^{MU}) > t)dt\\&\overset{\beta=e^t-1}{=} \tfrac{\mathrm{Q}_{s|i}^{MU} W_i^{MU}}{ln2}\int_0^{+\infty}\tfrac{1}{1+\beta}\mathrm{P}(SINR_i^{MU} > \beta)d\beta,\end{aligned} \tag{6}$$

where $\mathrm{Q}_{s|i}^{MU}$ denotes the service probability that an MU can be scheduled to have W_i^{MU} bandwidth conditioning on that it selects carrier i. Variables θ_i^{MBS}, θ_i^{FBS} and $\mathrm{Q}_{s|i}^{MU}$ are closely related to the PSU policy and calculated in Subsection III-B.

Similarly as MUs, the SINR distribution of FSs is given in Proposition 2.

Proposition 1: In the HetNets described in the system model, the probability that the SINR of one FS in carrier i is larger than a threshold β can be expressed as

$$\begin{aligned}&\mathrm{P}(SINR_i^{FS} > \beta) = \textstyle\int_0^{R_F}\tfrac{2d}{R_F^2}e^{-n_0 d^{\alpha_i}\beta/P_i^{FBS}}Fd(d)\\&where\quad F = \exp\{-2\pi\theta_i^{FBS,usa}\lambda^{FBS}\tau(d, H_i^{FBS}, \beta)\}\\&\qquad\qquad\cdot\exp\{-2\pi\theta_i^{MBS,usa}\lambda^{MBS}\rho(d, H_i^{MBS}, \beta, B)\},\\&\tau(d, H_i^{FBS}, \beta) = \tfrac{d^2}{2}\Gamma(1-\tfrac{2}{\alpha_i})E_{H_i^{FBS}}\left[(\beta H_i^{FBS})^{\frac{2}{\alpha_i}}\right],\\&\rho(d, H_i^{MBS}, \beta, B) = \tau(d, BH_i^{MBS}, \beta),\\&B = P_i^{MBS}/P_i^{FBS}.\end{aligned} \tag{7}$$

Please refer to **Appendix B** for the detailed derivations of Proposition 2. Therefore, the ergodic throughput of FSs in carrier i, R_i^{FS}, is calculated similarly as Eq. (6),

$$R_i^{FS} = \frac{\mathrm{Q}_{s|i}^{FS} W_i^{FS}}{ln2}\int_0^{+\infty}\frac{1}{1+\beta}\mathrm{P}(SINR_i^{FS} > \beta)d\beta, \tag{8}$$

where $\mathrm{Q}_{s|i}^{FS}$ denotes the user service probability of one FS in carrier i given that it uses carrier i. Following the same procedure, the SINR distribution of FNS in carrier i, $P(SINR_i^{FNS} > \beta)$, and ergodic throughput R_i^{FNS} can be calculated similarly as Eq. (7) and (8) where the superscript *"FS"* is replaced with *"FNS"*.

B. User Service Probability and Bandwidth Usage Probability

In this subsection, we calculate the user service probabilities ($\mathrm{Q}_{s|i}^{MU}$, $\mathrm{Q}_{s|i}^{FS}$ and $\mathrm{Q}_{s|i}^{FNS}$) that one user can be served by an MBS or FBS, and the bandwidth usage probability ($\theta_i^{MBS,usa}$ and $\theta_i^{FBS,usa}$) that a portion of effective bandwidth in one carrier is occupied by any user in one MBS or FBS. All the probabilities are closely related to the PSU policy (n^{res} and n^{open}) and the user CA capabilities (n_{agg}) and are conditioned on that the user selects carrier i to transmit.

To calculate the MBS-related probabilities $\mathrm{Q}_{s|i}^{MU}$ and $\theta_i^{MBS,usa}$, the number of MUs in one MCell needs to be calculated first. Denote the number of MUs in a Voronoi cell and the cell size as N^{MU} and S, respectively. As the MUs are distributed as a PPP with density λ^{MU}, the number of MUs in one Voronoi cell with area S (denoted as N^{MU}) follows Poisson($\lambda^{MU}S$). Thus, we have

$$\mathrm{P}(N^{MU} = k|S) = \frac{(\lambda^{MU}S)^k e^{-\lambda^{MU}S}}{k!},\ k = 0, 1, \dots \tag{9}$$

And

$$\mathrm{P}(N^{MU} = k) = \int_0^{+\infty} P(N^{MU} = k|S)f(S)dS, \tag{10}$$

where $f(S)$ is the pdf of S. As indicated in [27], a simple but accurate enough approximation of $f(S)$ is given as,

$$f(S) = \frac{343}{15}\sqrt{\frac{7}{2\pi}}(S\lambda^{MBS})^{\frac{5}{2}}e^{-\frac{7}{2}S\lambda^{MBS}}\lambda^{MBS}. \tag{11}$$

By substituting Eq. (11) into Eq. (10), the distribution of N^{MU} can be obtained. As one MU can access any carrier and concurrently transmit on n_{agg} carriers, the probability that one MU chooses carrier i is n_{agg}/N. Then the probability that there are totally k MUs in one cell among which l MUs choose carrier i is denoted as $P_{l,k|i}^{MU}$ and calculated as

$$\mathrm{P}_{l,k|i}^{MU} = C_k^l(\frac{n_{agg}}{N})^l(1-\frac{n_{agg}}{N})^{k-l}\mathrm{P}(N^{MU} = k). \tag{12}$$

Given l MUs choose carrier i, the probability that one MU can have bandwidth from carrier i is

$$\min\{1, \frac{P_i W_{PRB}/W_i^{MU}}{l}\} = \min\{1, \frac{P_i W_{PRB}}{W_i^{MU} l}\}. \tag{13}$$

Then user service probability $\mathrm{Q}_{s|i}^{MU}$ can be achieved by averaging Eq. (13) over $\mathrm{P}_{l,k|i}^{MU}$,

$$\mathrm{Q}_{s|i}^{MU} = \sum_{k=1}^{\infty}\sum_{l=1}^{k}\min\{1, \frac{P_i W_{PRB}}{W_i^{MU} l}\}P_{l,k|i}^{MU}. \tag{14}$$

Similarly, the bandwidth usage probability $\theta_i^{MBS,usa}$ is calculated as

$$\theta_i^{MBS,usa} = \sum_{k=1}^{\infty}\sum_{l=1}^{k}\min\{1, \frac{W_i^{MU} l}{P_i W_{PRB}}\}P_{l,k|i}^{MU}. \tag{15}$$

Then, the FBS-related probabilities (i.e., $\mathrm{Q}_{s|i}^{FS}$, $\mathrm{Q}_{s|i}^{FNS}$ and

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/JIOT.2015.2496161, IEEE Internet of Things Journal

$\theta_i^{FBS,usa}$) can be calculated. As the area of one FBS coverage is πR_F^2, the total number of FSs in one FCell (denoted as N^{FS}) is Poisson distributed with $\lambda^{FS}\pi R_F^2$. Thus, we have

$$\mathrm{P}(N^{FS}=k)=\frac{(\lambda^{FS}\pi R_F^2)^k}{k!}e^{-\lambda^{FS}\pi R_F^2}. \tag{16}$$

Similarly as the calculation of $Q_{s|i}^{MU}$, $\mathrm{Q}_{s|i}^{FS}$ is calculated as

$$\begin{aligned} &Q_{s|i}^{FS}= \\ &\begin{cases} \sum\limits_{k=1}^{\infty}\min\{1,\frac{P_iW_{PRB}}{kW_i^{FS}}\}P(N^{FS}=k), & if\ n_{agg}\geq n^{res} \\ \sum\limits_{k=1}^{\infty}\sum\limits_{l=1}^{k}\min\{1,\frac{P_iW_{PRB}}{lW_i^{FS}}\}P_{l,k|i}^{FS}, & otherwise, \end{cases} \\ &where\ \ \mathrm{P}_{l,k|i}^{FS}=C_k^l(\tfrac{n_{agg}}{n^{res}})^l(1-\tfrac{n_{agg}}{n^{res}})^{k-l}\mathrm{P}(N^{FS}=k). \end{aligned} \tag{17}$$

The probability $Q_{s|i}^{FNS}$ is calculated exactly the same way as that of $Q_{s|i}^{FS}$. For $\theta_i^{FBS,usa}$, the probabilities that carrier i is assigned to FSs and FNSs are n^{res}/N and n^{open}/N, respectively. For either possibility, the bandwidth usage probability is calculated similarly with Eq. (15), i.e.,

$$\begin{aligned} &\theta_i^{FBS}=\tfrac{n^{res}}{N}P_1+\tfrac{n^{open}}{N}P_2, \\ &where \\ &P_1= \\ &\begin{cases} \sum\limits_{k=1}^{\infty}\min\{1,\frac{kW_i^{FS}}{P_iW_{PRB}}\}P(N^{FS}=k), & if\ n_{agg}\geq n^{res} \\ \sum\limits_{k=1}^{\infty}\sum\limits_{l=1}^{k}\min\{1,\frac{lW_i^{FS}}{P_iW_{PRB}}\}P_{l,k|i}^{FS\,\prime}, & otherwise, \end{cases} \end{aligned} \tag{18}$$

The probability P_2 is FNS-related and can be calculated similarly with P_1.

C. QoS-Aware Effective Bandwidth: Formulation and Algorithm

The effective bandwidth for each type of user is finalized based on the derived user SINR distributions. For analytical simplicity, the intra-band contiguous CA [25] is considered where the radio characteristics of all carriers are the same, so each carrier contribute equal portion of throughput for each user. Thus for MUs, the minimum throughput requirement on carrier i is r_u^{MU}/n_{agg}. According to Eq. (1), W_i^{MU} should be determined such that

$$\mathrm{P}(Q_{s|i}^{MU}W_i^{MU}\log(1+SINR_i^{MU})<r_u^{MU}/n_{agg})<e\ll 1, \tag{19}$$

which can be rearranged as

$$\mathrm{P}(SINR_i^{MU}<2^{r_u^{MU}/(n_{agg}Q_{s|i}^{MU}W_i^{MU})}-1)<e. \tag{20}$$

If β is equal to $2^{r_u^{MU}/(n^{agg}Q_{s|i}^{MU}W_i^{MU})}-1$, Eq. (39) can be leveraged to achieve the value range of W_i^{MU}. As one PRB is the minimum bandwidth allocation unit in LTE-A systems, W_i^{MU} should be an integral multiple of W_{PRB}. Then W_i^{MU} can be finalized as the product of W_{PRB} and an integer value denoted as m_i^{MU}. The physical meaning of m_i^{MU} is the minimum number of PRBs in carrier i that can satisfy Eq. (20), i.e.,

$$\begin{aligned} &m_i^{MU}=\min_m\ m \\ &s.t.\ \ W_i^{MU}=mW_{PRB}; \\ &\quad 0\leq m\leq P_i,\ m\in\mathbb{Z}^+; \\ &\quad Eq.\ (20). \end{aligned} \tag{21}$$

According to Eq. (20), W_i^{MU} is closely related to $Q_{s|i}^{MU}$ and $\theta_i^{MBS,usa}$ which are further determined by the PSU policy and CA capabilities. Therefore, W_i^{MU} is jointly determined by the PSU policy, user QoS requirements and CA capabilities. For the integer values of FSs (or FNSs), denoted as m_i^{FS} (or m_i^{FNS}), the derivation is the same as that of MUs except the minimum throughput requirements in carrier i, which is $r_u^F/\min\{n^{res},n_{agg}\}$ (or $r_u^F/\min\{n^{open},n_{agg}\}$).

It can be seen from Eq. (21) that for any type of users, the optimization problem to calculate the effective bandwidth is constrained integer non-convex. In addition, the determination of its effective bandwidth is highly dependent on the effective bandwidth of the other types, which is because the constraint on its SINR distribution (i.e., the third constraint of the optimization problem) is closely related to the effective bandwidth of the other types. Therefore, it is infeasible to obtain the optimum in polynomial time.

To make the proposed strategy tractable and practical, a heuristic algorithm, referred to as QoS-aware effective bandwidth (QA-EB) algorithm in this paper, is proposed. The basic idea of the QA-EB algorithm is to augment m_i^{MU}, m_i^{FS} and m^{FNS} step by step according to a specified priority. Each of the above three variables starts from 1 with augmentation step 1. Each time when one variable increases 1, the algorithm checks whether the SINR constraints of the user types with higher priority are satisfied. If the constraints are satisfied, the variable of the user type with the next lower priority is augmented; otherwise the variable of the user type with the highest priority and unsatisfied SINR constraint is augmented. The algorithm stops when the SINR constraints of all the user types are satisfied. Note that The priority of user types can be determined according to the vendor/operator's preference, and different priority assignments may lead to different effective bandwidth sets $\{m_i^{MU},m_i^{FS},m_i^{FNS}\}$. In this paper, the MUs and FNSs are given the highest and lowest priority, respectively. The reason is that the MUs are more sensitive to the change of effective bandwidth of FSs and FNSs, which has been validated through simulations in Section V.

Remark: According to Eq. 1, if the throughput requirement for one user (i.e., r_u^{MU} and r_u^F) is higher, the effective bandwidth (i.e., W_i^{MU}, W_i^{FS} and W_i^{FNS}) will be higher. But according to Eq. (14) and (17), increasing the user effective bandwidth will decrease the user service probability when the system bandwidth is saturated, which may result in a decrease instead in the average user throughput. Correspondingly, the average user packet delay may be higher due to lower buffer service rate in user equipment. Therefore, it is important to study the tradeoff between packet delay, time-average user bandwidth and average user throughput. Quantitative analysis on the tradeoff related to user packet delay relies on extra

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/JIOT.2015.2496161, IEEE Internet of Things Journal

Algorithm 1 QA-EB Algorithm

/* **Initialization** */
Define auxiliary variables
$A_1 := r_u^{MU}/(n_{agg}Q_{s|i}^{MU}W_i^{MU})$,
$A_2 := r_u^{FS}/(\min\{n_{agg}, n^{res}\}Q_{s|i}^{FS}W_i^{FS})$,
$A_3 := r_u^{FNS}/(\min\{n_{agg}, n^{open}\}Q_{s|i}^{FNS}W_i^{FNS})$;
Define events E_1, E_2, E_3 as
$E_1 := \mathrm{P}(SINR_i^{MU} < 2^{A_1} - 1) < e$,
$E_2 := \mathrm{P}(SINR_i^{FS} < 2^{A_2} - 1) < e$,
$E_3 := \mathrm{P}(SINR_i^{FNS} < 2^{A_3} - 1) < e$;
$m_i^{MU} \leftarrow 1, m_i^{FS} \leftarrow 1, m_i^{FNS} \leftarrow 1$;
/***Loop Augmentation***/
EndFlag←FALSE;
while EndFlag == FALSE **do**
  while E_1 is TRUE and EndFlag == FALSE **do**
    if E_2 is FALSE **then**
      $m_i^{FS} \leftarrow m_i^{FS} + 1$;
    end if
    while E_1 is TRUE and E_2 is TRUE and EndFlag == FALSE **do**
      if E_3 is TRUE **then**
        EndFlag ← TRUE;
      else
        $m_i^{FNS} \leftarrow m_i^{FNS} + 1$;
      end if
    end while
  end while
  if E_1 is FALSE **then**
    $m_i^{MU} \leftarrow m_i^{MU} + 1$;
  end if
end while

mathematical tools such as Queueing theory, which is beyond the scope of this work but will be explored in our future research.

IV. Two-Level Stackelberg Game Between Macro and Femto Cells

In this section, we model the interaction between MCells and FCells into a Stackelberg game and propose a backward induction method to determine the optimal interference price y and PSU policy.

A. Game Formulation

The interaction is formulated as a two-level Stackelberg game, jointly considering the utility maximization of both MBSs and FBSs. In the first level, each MBS, as the game leader, imposes an interference-related price y upon the FBS throughput according to the interference from FBSs. In the second level, each FBS, as a follower, decides the PSU policy (i.e., n^{res} and n^{open}) based on the imposed price y, user QoS requirements and user CA capabilities. The utilities are expressed to be the total weighted profits as follows.

1) MBS Level Game: For each MBS, its total utility is composed of two parts: the service profits from MUs and the profits from the interference charge on FBSs. To calculate either part, it is required to have i) the average number of MUs and FBSs in one MCell (denoted as $\overline{N^{MU}}$ and $\overline{N^{FBS}}$, respectively), and ii) the average number of FSs and FNSs in one FCell (denoted as $\overline{N^{FS}}$ and $\overline{N^{FNS}}$, respectively). Based on Eq. (10), $\overline{N^{MU}}$ is calculated as,

$$\overline{N^{MU}} = \sum_{k=1}^{\infty} k \cdot \mathrm{P}(N^{MU} = k). \tag{22}$$

Variable $\overline{N^{FBS}}$ can be calculated similarly. Based on Eq. (16), $\overline{N^{FS}}$ and $\overline{N^{FNS}}$ can also be achieved similarly as Eq. (22). Then, the utility of one MBS is given as,

$$\begin{aligned} &U^{MBS} = n_{agg}R_i^{MU} \cdot \overline{N^{MU}}g^{MU} \\ &+\omega^{MBS}y\overline{N^{FBS}}\Big[n_aR_i^{FS}\overline{N^{FS}} + n_bR_i^{FNS}\overline{N^{FNS}}\Big], \\ &where\ n_a = \min\{n_{agg}, n^{res}\},\ n_b = \min\{n_{agg}, n^{open}\} \end{aligned} \tag{23}$$

Here, the total throughput from all FBSs in one MCell is used to represent the interference caused by FBSs as the FBS-part interference is hard to extract from MU report in realistic implementation. ω^{MU} is the weight of interference charge over service profits. R_i^{MU} and R_i^{FS} are given in Eq. (6) and (8). As aforementioned, the MBSs can only influence the resource allocation of FBSs indirectly through price control. Therefore, one MBS can only optimize the imposed interference-related price y to maximize its own total utility:

$$\max_{0 \le y \le y^{max}} U^{MBS}. \tag{24}$$

2) FBS Level Game: For each FBS, it needs to pay g^{FS}/bit for FS services, and can gain g^{FNS}/bit for FNS services. Thus, its total utility can be expressed as:

$$\begin{aligned} &U^{FBS} = -g^{FS}n_aR_i^{FS}\overline{N^{FS}} + g^{FNS}n_bR_i^{FNS}\overline{N^{FNS}} \\ &-w^{FBS}y\Big[n_aR_i^{FS}\overline{N^{FS}} + n_bR_i^{FNS}\overline{N^{FNS}}\Big], \end{aligned} \tag{25}$$

where n_a and n_b are given in Eq. (23). Variable ω^{FBS} is the weight of interference cost over the profits. Given the interference price y imposed by the MBSs, as the user QoS requirements and CA capabilities are known, the PSU policy alone can determine the effective bandwidth $W_i^T, T \in \{MU, FS, FNS\}$ and further determine the FBS utility U^{FBS}. Therefore, one FBS only needs to optimize n^{res} and n^{open} to maximize its own utility, i.e.,

$$\begin{aligned} &\max_{n^{res}, n^{open}} \ U^{FBS} \\ &s.t.\ \ n^{res} + n^{open} \le N, \ \ n^{res} \text{ and } n^{open} \in \mathbb{Z}^+. \end{aligned} \tag{26}$$

B. Analysis of the Proposed Game

Tradeoffs exist in this game. On one hand, if one MBS hopes to improve its MU performance to gain more profits from MU services, it needs to increase y to lower interference from FBSs. As a result, the throughput from FBSs will be reduced, resulting in a reduction in MBS gains from interference charge. On the other hand, one FBS can increase its utiliy by opening more carriers for FNSs, however, it needs to pay more for the increased throughput due to the interference-related price y. Therefore, MBSs need to optimize y and FBSs need to optimize the PSU policy (i.e., n^{res} and n^{open}) to achieve their own maximum utilities, i.e., to obtain the Stackelberg equilibrium.

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/JIOT.2015.2496161, IEEE Internet of Things Journal

To achieve the Stackelberg equilibrium, a backward induction method is utilized to analyze the proposed game, which captures the dependence of FBS decisions on MBS decisions. The followers of the game, i.e., the FBSs, are analyzed first. Given the imposed interference price y, the optimal PSU policy (n^{res} and n^{open}) can be achieved by solving optimization problem (26). The primary challenge of solving (26) is that the exact value of y is unknown, which means the optimal (n^{res}, n^{open}) combination is not fixed and should be a function of y. In other words, the goal of solving (26) is to find a mapping between different value intervals of y and the corresponding optimal (n^{res}, n^{open}) combinations. For a given y value, the general method to obtain the optimal (n^{res}, n^{open}) combination is the classic branch and bound algorithm [28], since (26) is typical integer nonlinear optimization. But as the backward induction method potentially needs to know the optimal (n^{res}, n^{open}) combinations for all the y values in $[0, y^{max}]$, the computation workload can be huge when N is large. Fortunately, it is specified in the LTE-A standard [25] that at most 5 carriers can be aggregated in one system, i.e., $N \leq 5$. Therefore, there are at most 15 feasible (n^{res}, n^{open}) combinations for problem (26). By comparing the values of U^{FBS} under each combination within the interval $[0, y^{max}]$, the optimal (n^{res}, n^{open}) combination with the corresponding y value interval can be easily determined, as denoted below.

$$\begin{gathered} \{n_{opt}^{res}(Y_s), n_{opt}^{open}(Y_s)\}, \\ where \quad Y_s \subset [0, y^{max}], \\ \bigcup_s Y_s = [0, y^{max}], \quad and \quad Y_{s_1} \bigcap Y_{s_2} = \emptyset, \\ \forall s, s_1, s_2 \in \{1, 2, \cdots, S\}. \end{gathered} \tag{27}$$

In Eq. (27), S is the total number of y value intervals that correspond to a different optimal (n^{res}, n^{open}) combination compared to its adjacent value interval.

The game for MBSs is then analyzed. As the optimal (n^{res}, n^{open} is different for different y value intervals Y_s, the utility maximization problem for MBSs (24) can be decomposed into a series of sub-optimization problems as follows.

$$\max_{y \in Y_s} \quad U_s^{MBS}, \quad s \in \{1, 2, \cdots, S\}. \tag{28}$$

Denote the optimal value for U_s^{MBS} and the corresponding optimal y as $U_{s,opt}^{MBS}$ and $y_{s,opt}$, respectively, then the optimal solution of the original problem (24) (denoted as y_{opt}) can be determined as

$$\begin{gathered} y_{opt} = y_{s^*,opt}, \\ where \quad s^* = \arg\max_s \quad (U_{s,opt}^{MBS}), \\ s \in \{1, 2, \cdots, S\}. \end{gathered} \tag{29}$$

Consequently, the optimal (n^{res}, n^{open}) combination is finalized as $\{n_{opt}^{res}(Y_{s^*}), n_{opt}^{open}(Y_{s^*})\}$.

In summary, the backward induction method obtains the Stackelberg equilibrium in two steps. It first solves the utility maximization problem of the game followers (i.e., the FBSs) by finding a mapping between a set of y value intervals and a set of corresponding optimal (n^{res}, n^{open}) combinations. With the mapping, the utility maximization problem of game leaders (i.e., MBSs) is decomposed into a series of sub-problems with different y value intervals; by comparing the optimal utility values of each sub-problem, the optimal y for the original problem can be finalized. In this manner, the stackelberg equilibrium is determined.

V. Simulation Results

In this section, *Monte Carlo* simulation results are presented to i) validate our analytical results and ii) demonstrate the optimal PSU policy and interference price under different user QoS requirements and CA capabilities.

A. Simulation Setup

Simulation setup of starts with an area of $20 \times 20km^2$ with $\lambda^{MBS} = 0.5/(\pi 500^2)/m^2$. The homogeneous-carrier case is considered where P_i, b_i, and α_i are identical for $\forall i$. The PSD of each MBS (FBS) is the same for every PRB in each carrier. The detailed parameter settings are presented in Table II. With this setting, the average number of BSs is 255 and that of MUs is 7650. Thus, the boundary effect can be neglected by such a large-scale network. Furthermore, each presented result is averaged over 1000 runs.

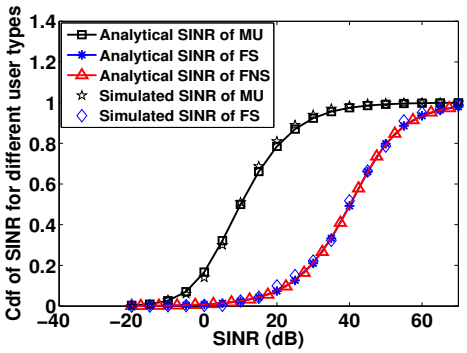
TABLE II: Simulation Parameters

Parameters	Values
Coverage radius of FCell, R_F	10m
FBS density, λ^{FBS}	$10\lambda^{MBS}$
MU density, λ^{MU}	$30\lambda^{MBS}$
FS density, λ^{FS}	$3/(\pi R_F^2)/m^2$
FNS density, λ^{FNS}	$2/(\pi R_F^2)/m^2$
MBS PSD, P^{MBS}	-23.5dBm
FBS PSD, P^{FBS}	-49.5dBm
Noise PSD, n_0	$-174dBm$
Fast fading of interference	Rayleigh fading
Total number of carriers, N	5
The number of PRBs per carrier, P_i	100
PRB bandwidth in carrier i, W_{PRB}	$180kHz$
Path loss component, α_i	4
User CA capability, n_{agg}	$1 \sim 5$
MU required throughput, r^{MU}	$320 \sim 460kbps$
FS (FNS) required throughput, r^F	$5 \sim 23Mbps$
QoS violation probability, e	0.05
Unit profit of MU services, g^{MU}	10
Unit cost of FS services, g^{FS}	1.5
Unit profit of FNS services, g^{FNS}	3
(y^{max},ω^{MBS},ω^{FBS})	(10,0.01,2)

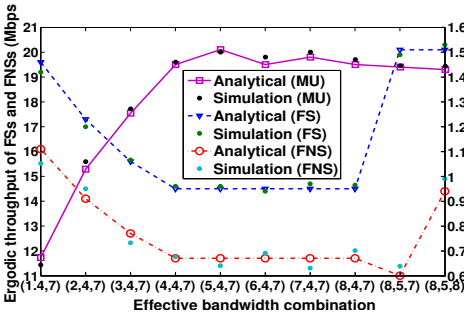
B. Numerical and Simulation Results

We first corroborate our analytical results on user SINR distributions and ergodic throughput. In Fig. 3(a), the cdfs of single-carrier SINR are given when the effective bandwidth of MUs, FSs and FNSs are 1,4,7, respectively. It can be observed that the SINR performance of FCell users is much better than that of MUs since MUs generally have a much longer distance to MBSs than FCell users to FBSs. Besides, FSs and FNSs have the same SINR performance. This is because of the same FBS PSD for both user types and the random bandwidth access mechanism, resulting in the same strength of average useful signal and interference.

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/JIOT.2015.2496161, IEEE Internet of Things Journal



(a) (W_i^{MU},W_i^{FS},W_i^{FNS}) are set to (1,4,7) PRBs.

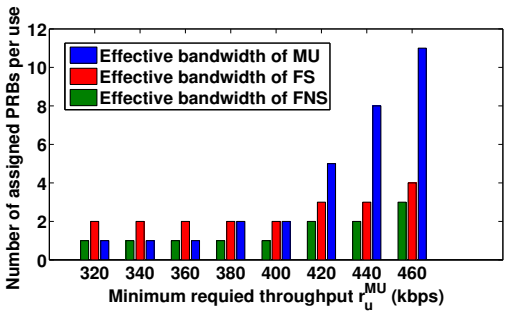


(b) Total ergodic throughput per user.

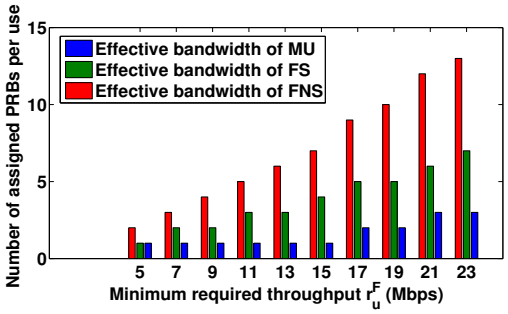
Fig. 3: User SINR and ergodic throughput performance in HetNets. Default values: $n_{agg} = 2$, $n^{res} = 3$, $n^{open} = 1$, $\lambda^{MU} = 30\lambda^{MBS}$, $\lambda^{FBS} = 10\lambda^{MBS}$.

In Fig. 3(b), the ergodic throughput of FCell users is significantly higher ($\sim$ 10 times) than that of MUs under different effective bandwidth combinations due to much better SINR performance. It can be further observed that when each MU is assigned with more PRBs per carrier, the MU ergodic throughput first increases and then remains stable. This can be explained as follows: when W_i^{MU} is small, increasing W_i^{MU} will bring each MU more bandwidth without increasing the interference intensity from other MBSs very much. Thus, the ergodic throughput increases. However, if W_i^{MU} keeps increasing, the service probability of each MU ($Q_{s|i}^{MU}$) will drop considerably, which counterbalances the performance gain brought by wider bandwidth. So the ergodic throughput becomes stable. In addition, FSs have a higher throughput than FNSs since FSs can concurrently transmit on 2 carriers ($n_{agg} = 2$ and $n^{res} = 2$) compared to 1 carrier for FNSs. When W_i^{MU} increases, the ergodic throughput of FSs and FNSs both decreases first and then becomes stable. This is because with larger W_i^{MU}, the interference from MBSs first increases and then remains unchanged since the bandwidth usage probability $\theta_i^{MBS,usa}$ has reached its maximum, i.e., 1. Moreover, when W_i^{FS} increases, e.g., from (8,4,7) to (8,5,7), the throughput of FNS decreases, which is because the interference perceived by FNSs from FBSs increases.

Fig. 4 shows how the effective bandwidth of different user types is decided with the user throughput requirements given QoS violation probability e and CA capabilities. The effective bandwidth is represented by the number of PRBs assigned to



(a) r_u^{MU} changes while $r_u^F = 5Mbps$.



(b) r_u^F changes while $r_u^{MU} = 360kbps$.

Fig. 4: Effective bandwidth with different minimum throughput requirements. Default values: same with Fig. 3.

each user. It can be seen that as the throughput requirements increase, users need to be assigned with more PRBs to satisfy the maximum QoS violation probability. Besides, MUs are more sensitive to the throughput increase than FSs and FNSs: W_i^{MU} increases 10 times to satisfy only 44% increase of r^{MU} while W_i^{FS} or W_i^{FNS} increases 7 times to satisfy 360% increase of r^F. The reason is as follows. In an MBS, there are more MUs selecting the same carrier than FSs (or FNSs) do in an FBS, resulting in that the bandwidth in one carrier is more likely to be saturated in an MBS than in an FBS. Thus, increasing W_i^{MU} will more likely reduce the user service probability of MUs, making the MU throughput increase harder than FSs and FNSs. Furthermore, it can be observed in Fig. 4(a) that even if r^F is not changed, the effective bandwidth of FSs and FNSs still increases to supplement the throughput loss due to increased interference from MBSs. The similar phenomenon is also observed in Fig. 4(b).

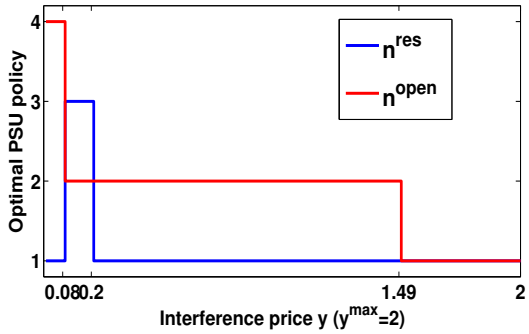


Fig. 5: The optimal PSU policy of FBS (n^{res} and n^{open}) given different y when $n_{agg} = 4$,$r^{MU} = 400kbps$, and $r^F = 15Mbps$.

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/JIOT.2015.2496161, IEEE Internet of Things Journal

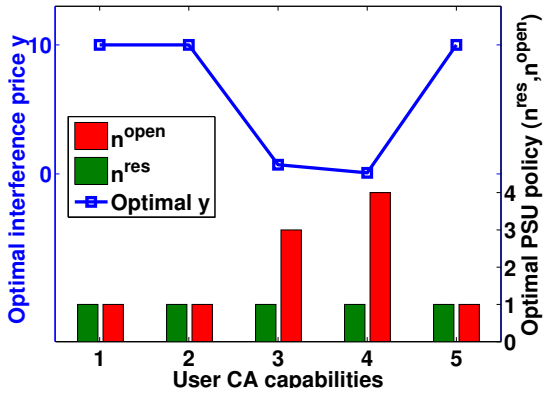


Fig. 6: The optimal interference price and PSU policy for different CA capabilities, when $r^{MU} = 400kbps$ and $r^{F} = 15Mbps$.

Finally, we show how the optimal PSU policy and interference price y is determined with given user throughput requirements and CA capabilities. As mentioned in Subsection IV-B, to maximize the utilities of both parties, the FBSs first offer MBSs with the knowledge of optimal PSU policies to maximize the utility of FBS given different y values. Then the MBS choose the optimal y to maximize its own utility. Fig. 5 shows the mapping between optimal PSU policy and y. It can be seen that the optimal PSU policy changes when y reaches the point 0.08, 0.2 and 1.49. As y increases, the total number of carriers selected by an FBS decreases in order to reduce the interference cost charged by MBSs.

The optimal interference price and PSU policy for different CA capabilities is evaluated in Fig. 6. When $n_{agg} = 3, 4$, the optimal interference price is 0.71 and 0.08, respectively. This implies that for FBSs, the increase of utility due to enlarging the number of carriers for FNS surpasses the resultant interference cost charged by MBSs. When $n = 1, 2$ and 5, the situation is adverse where the FBSs have to reduce the number of open-access carriers to the minimum to avoid the relatively excessive interference charges from MBSs.

VI. Conclusion

In this paper, we have studied the QoS provisioning for IoT in LTE-A HetNets with PSU mechanism. Specifically, the Stochastic geometry has been leveraged to consider the random behaviors of IoT-oriented FCells and inter-macro interference into performance analysis under PSU mechanism. Then, the concept of effective bandwidth has been applied to decide the user bandwidth with considering the user QoS requirements and CA capabilities. Furthermore, the interaction between MBSs and FBSs has been modelled into a Stackelberg game to maximize the utilities of both parties. Finally, *Monte Carlo* simulations have been performed to verify our analytical results and demonstrate the decision process of effective bandwidth as well as the optimal PSU policy and interference price. The research outcomes should shed some light on how to optimally coordinate the resource utilization in HetNets among different operator bands, which is a future trend for the cellular systems. For the future work, we intend to remove the upper bound on the interference price by introducing operator-competition mechanism where multiple operators managing different bands will contend to offer the best price to users.

Appendix A

We first solve the expectation over D_{B_0} - the distance between the serving MBS and the considered MU. Since one MU always chooses the nearest MBS to connect, D_{B_0} implies that there is no MBSs within the distance D_{B_0} to the considered MU. As the MBSs are distributed as PPP with λ^{MBS}, thus

$$\mathrm{P}(D_{B_0} > d) = \mathrm{P}(\text{No MBSs within } \pi d^2) = e^{-\lambda^{MBS}\pi d^2}. \quad (30)$$

Then, the pdf of D_{B_0} is given as

$$f_{D_{B_0}}(d) = 2\pi\lambda^{MBS} d e^{-\lambda^{MBS}\pi d^2},\ d \in (0, +\infty). \quad (31)$$

Thus, $\mathrm{P}(SINR_i^{MU} > \beta)$ can be calculated as

$$\begin{aligned}
&\mathrm{P}(SINR_i^{MU} > \beta) \\
&= E_{\Phi_i^{MBS}\setminus B_0, \Phi_i^{FBS}, H_i^{MBS}, H_i^{FBS}} \Big[\int_0^{+\infty} 2\pi\lambda^{MBS} \\
&\quad \cdot d e^{-\pi\lambda^{MBS} d^2} \exp[- \tfrac{(I_i^{MBS} + I_i^{FBS} + n_0) d^{\alpha_i}\beta}{P_i^{MBS}}] d(d) \Big] \\
&= \int_0^{+\infty} 2\pi\lambda^{MBS} d e^{-\pi\lambda^{MBS} d^2} e^{-n_0 d^{\alpha_i}\beta / P_i^{MBS}} F d(d) \\
&where\ F = F_1 \cdot F_2, \\
&F_1 = \mathrm{E}_{\Phi_i^{MBS}\setminus B_0, H_i^{MBS}} \left[\exp(- \tfrac{I_i^{MBS} d^{\alpha_i}\beta}{P_i^{MBS}}) \right], \\
&F_2 = \cdot \mathrm{E}_{\Phi_i^{FBS}, H_i^{FBS}} \left[\exp(- \tfrac{I_i^{FBS} d^{\alpha_i}\beta}{P_i^{MBS}}) \right].
\end{aligned} \quad (32)$$

Then, we calculate F_1 expected over $\Phi_i^{MBS}\setminus B_0$. As Φ_i^{MBS} denotes the set of MBSs that use the same PRBs in carrier i with the considered MU, Φ_i^{MBS} can be viewed as a thinning of the original PPP Φ^{MBS} with probability $\theta_i^{MBS,usa}$. Here, $\theta_i^{MBS,usa}$ is the bandwidth usage probability that one piece of W_i^{MU} bandwidth is occupied by any MU in one MBS. In other words, Φ_i^{MBS} is a homogenous PPP with density $\lambda_i^{MBS} = \theta_i^{MBS,usa}\lambda^{MBS}$. The calculation of $\theta_i^{MBS,usa}$ is shown in Subsection III-B. Define a set $\tilde{\Phi}_i^{MBS}$ as

$$\tilde{\Phi}_i^{MBS} \triangleq \{D_B : B \in \Phi_i^{MBS}\setminus B_0\}. \quad (33)$$

According to *Slivnyak's Theorem* [18], if Φ_i^{MBS} is a PPP, $\Phi_i^{MBS}\setminus B_0$ is also a PPP with the same density as Φ_i^{MBS}. And $\tilde{\Phi}_i^{MBS}$ is an inhomogeneous PPP with density function

$$\lambda(x) = 2\pi\lambda_i^{MBS} x,\ 0 \le x < +\infty, \quad (34)$$

where x is the distance away from the considered MU. Therefore, F_1 can be rewritten as

$$\begin{aligned}
F_1 &= E_{\tilde{\Phi}_i^{MBS}, H_i^{MBS}} \left[\exp(- I_i^{MBS} d^{\alpha_i}\beta / P_i^{MBS}) \right] \\
&= E_{\tilde{\Phi}_i^{MBS}, H_i^{MBS}} \left[\prod_{D_B \in \tilde{\Phi}_i^{MBS}} \exp(-H_i^{MBS} {D_B}^{-\alpha_i} d^{\alpha_i}\beta) \right].
\end{aligned} \quad (35)$$

To calculate F_1, a property of PPP proved in [29] is exploited as follows,

$$\begin{aligned}
&E_{\Phi, H_i} \left[\prod_{X \in \Phi} \exp(-s H_i g(X)) \right] \\
&= \exp\{ -E_{H_i} \left[\int_0^{+\infty} (1 - e^{-s H_i g(x)}) \lambda(x) dx \right] \},
\end{aligned} \quad (36)$$

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/JIOT.2015.2496161, IEEE Internet of Things Journal

where Φ is a PPP with density function $\lambda(x)$; H_i is generally distributed; s is a nonnegative constant; and $g(x)$ is a nonnegative function of x. The property holds when H_i is independent of PPP Φ. With this property, F_1 can be transformed into

$$\begin{aligned} F_1 &= \exp\{-E_{H_i^{MBS}}\left[\int_d^{+\infty}(1-e^{-\beta d^{\alpha_i} H_i^{MBS} x^{-\alpha_i}})\lambda(x)dx\right]\} \\ &= \exp\{-2\pi\lambda_i^{MBS}\eta(d, H_i^{MBS}, \beta)\}, \\ &where \\ \eta(d, H_i^{MBS}, \beta) &= -\tfrac{1}{2}d^2 + \tfrac{1}{2}d^2 E_{H_i^{MBS}}\{e^{-\beta H_i^{MBS}} + \\ &(\beta H_i^{MBS})^{2/\alpha_i}\left[\Gamma(1-\tfrac{2}{\alpha_i}, 0) - \Gamma(1-\tfrac{2}{\alpha_i}, \beta H_i^{MBS})\right]\}, \\ &and \quad \Gamma(s,t) = \textstyle\int_t^{+\infty} x^{s-1}e^{-x}dx. \end{aligned} \tag{37}$$

And $\lambda(x)$ is given in Eq. (34). The lower limit of the integral in Eq. (37) is d since F_1 is conditioned on that the distance from the closest MBS to the considered MU is no less than d. The calculation of F_2 is similar as that of F_1 and given as,

$$\begin{aligned} &F_2 = \exp\{-2\pi\theta_i^{FBS,usa}\lambda^{FBS}\epsilon(d, H_i^{FBS}, \beta, A)\}, \\ &where \;\; \epsilon(d, H_i^{FBS}, \beta, A) = \\ &\qquad \tfrac{1}{2}d^2\Gamma(1-\tfrac{2}{\alpha_i}, 0)E_{H_i^{FBS}}\left[(A\beta H_i^{FBS})^{2/\alpha_i}\right] \\ &and \quad A = P_i^{FBS}/P_i^{MBS}. \end{aligned} \tag{38}$$

Note that the lower limit of the integral in F_2 is 0 as FBSs can appear arbitrarily close to the considered MU. Similarly, $\theta_i^{FBS,usa}$ denotes the probability that one piece of W_i^{MU} bandwidth in carrier i is occupied by any FS or FNS in one FBS. Therefore, the cdf of $SINR_i^{MU}$ is finally given as

$$\begin{aligned} &\mathrm{P}(SINR_i^{MU} > \beta) = \mathrm{P}(SINR_i^{MU} > \beta) \\ &= \textstyle\int_0^{+\infty} 2\pi\lambda^{MBS} d e^{-\pi\lambda^{MBS}d^2} e^{-n_0 d^{\alpha_i}\beta/P_i^{MBS}} \\ &\qquad \cdot \exp\{-2\pi\lambda_i^{MBS}\eta(d, H_i^{MBS}, \beta)\} \\ &\qquad \cdot \exp\{-2\pi\theta_i^{FBS}\lambda^{FBS,usa}\epsilon(d, H_i^{FBS}, \beta, A)\}d(d), \end{aligned} \tag{39}$$

where $\eta(d, H_i^{MBS}, \beta)$ and $\epsilon(d, H_i^{FBS}, \beta, A)$ are given in Eq. (37) and (38), respectively.

APPENDIX B

Similarly as MUs, the SINR distributions of one FS in carrier i can be expressed as

$$\begin{aligned} &\mathrm{P}(SINR_i^{FS} > \beta) = \mathrm{P}(H > \tfrac{(I_i^{MBS}+I_i^{FBS}+n_0)D_{F_0}{}^{\alpha_i}\beta}{P_i^{FBS}}) \\ &= E_{\Phi_i^{MBS}, \Phi_i^{FBS}\setminus F_0, H_i^{MBS}, H_i^{FBS}, D_{F_0}} \\ &\qquad \left[\exp(-\tfrac{(I_i^{MBS}+I_i^{FBS}+n_0)D_{B_0}^{-\alpha_i}\beta}{P_i^{FBS}})\right], \\ &where \; I_i^{MBS} = \sum_{B\in\Phi_i^{MBS}} P_i^{MBS} H_i^{MBS} D_B{}^{-\alpha_i}, \\ &\qquad I_i^{FBS} = \sum_{F\in\Phi_i^{FBS}\setminus F_0} P_i^{FBS} H_i^{FBS} D_F{}^{-\alpha_i}. \end{aligned} \tag{40}$$

In Eq. (40), F_0 is the associated FBS of the considered FS, and Φ_i^{MBS} (Φ_i^{FBS}) is the subset of MBSs (FBSs) that transmit on the same PRBs with the considered FS in the carrier i. As FSs and FNSs are assigned with different carriers, there is no interference between FSs and FNSs. The distribution of D_{F_0} is different from D_{B_0} in the MU case. Since FSs are only distributed within the coverage of the FBS, i.e., a disk area centred at F_0 with radius R_F, the cdf of D_{F_0} is given as

$$\mathrm{P}(D_{F_0} \le d) = \frac{\pi d^2}{\pi R_F^2}. \tag{41}$$

Then, the pdf of D_{F_0} is

$$f_{D_{F_0}}(d) = \frac{2d}{R_F}, \quad 0 \le d \le R_F. \tag{42}$$

The rest of the calculation of the expectation in Eq. (40) is the same as that of MUs. The results are given directly in Proposition 1.

REFERENCES

[1] R. A. Khan and A. A. Shaikh, "LTE advanced: Necessities and technological challenges for 4th generation mobile network," *International Journal of Engineering and Technology*, vol. 2, no. 8, pp. 1336-1342, Aug. 2012.

[2] J. G. Andrews, "Seven ways that HetNets are a cellular paradigm shift," *IEEE Communications Magazine*, vol. 51, no. 3, pp. 136-144, Mar. 2013.

[3] M. Ismail, A. Abdrabou, and W. Zhuang, "Cooperative decentralized resource allocation in heterogeneous wireless access medium," *IEEE Trans. Wireless Communications*, vol. 12, no. 2, pp. 714-724, Feb. 2013.

[4] N. Zhang, H. Zhou, K. Zheng, N. Cheng, J. W. Mark, and X. Shen, "Cooperative heterogeneous framework for spectrum harvesting in cognitive cellular network," *IEEE Communizations Magazine*, vol. 53, no. 5, pp. 60-67, May 2015.

[5] R. Zhang, M. Wang, L. X. Cai, Z. Zheng, X. Shen, and L. Xie, "LTE-Unlicensed: The future of spectrum aggregation for cellular networks," *IEEE Wireless Communications Magazine*, vol. 22, no. 3, pp. 150-159, June 2015.

[6] D. López-Pérez, X. Chu, A. V. Vasilakos, and H. Claussen, "Power minimization based resource allocation for interference mitigation in OFDMA femtocell networks," *IEEE Journal on Selected Areas in Communications,* vol. 32, no. 2, pp. 333-344, Feb. 2014.

[7] D. T. Ngo, S. Khakurel, and T. Le-Ngoc, "Joint subchannel assignment and power allocation for OFDMA femtocell networks," *IEEE Trans. on Wireless Communications*, vol. 13, no. 1, pp. 342-355, Jan. 2014.

[8] X. Xiang, C. Lin, X. Chen, and X. Shen, "Toward optimal admission control and resource allocation for LTE-A femtocell uplink," *IEEE Trans. on Vehicular Technology*, vol. 64, no. 7, pp. 3247-3261, July 2015.

[9] Z. Shen, A. Papasakellariou, J. Montojo, D. Gerstenberger, and F. Xu, "Overview of 3GPP LTE-advanced carrier aggregation for 4G wireless communications," *IEEE Communications Magazine*, vol. 50, no. 2, pp. 122-130, Feb. 2012.

[10] R. Zhang, Z. Zheng, M. Wang, X. Shen, and L. Xie, "Equivalent capacity in carrier aggregation-based LTE-A systems: A probabilistic analysis", *IEEE Trans. on Wireless Communications*, vol.13, no.11, pp. 6444-6460, Nov. 2014.

[11] R. Zhang, M. Wang, Z. Zheng, X. Shen, and L. Xie, "Cross-layer carrier selection and power control for LTE-A uplink with Carrier Aggregation," *Proc. IEEE Globecom'13*, Atlanta, GA, USA, Dec. 2013.

[12] K. I. Pedersen, F. Frederiksen, C. Rosa, H. Nguyen, L. G. U. Garcia, and Y. Wang, "Carrier aggregation for LTE-advanced: Functionality and performance aspects," *IEEE Communications Magazine*, vol. 49, no. 6, pp. 89-95, June 2011.

[13] D. Cao, S. Zhou, and Z. Niu, "Improving the energy efficiency of two-tier heterogeneous cellular networks through partial spectrum reuse," *IEEE Trans. on Wireless Communications,* vol. 12, no. 8, Aug. 2013.

[14] L. G. U. Garcia, I. Z. Kovcs, K. I. Pedersen, G. W. O. Costa, and P. E. Mogensen, "Autonomous component carrier selection for 4G femtocells - a fresh look at an old problem," *IEEE Journal on Selected Areas in Communications*, vol. 30, no. 3, pp. 525-537, Apr. 2012.

[15] Y. Zhong and W. Zhang, "Multi-channel hybrid access femtocells: A stochastic geometric analysis," *IEEE Trans. on Communications*, vol. 61, no. 7, pp. 3016-3026, July 2013.

[16] R. Zhang, M. Wang, Z. Zheng, X. Shen, and L. L. Xie, "Stochastic geometric performance analysis for carrier aggregation in lte-a systems," *Proc. IEEE ICC'14*, Sydney, Australia, June, 2014.

[17] X. Lin, J. G. Andrews, and A. Ghosh, "Modeling, analysis and design for carrier aggregation in heterogeneous cellular networks," *IEEE Trans. on Communications*, vol. 61, no. 9, pp. 4002-4015, July 2013.

[18] M. Haenggi, J. G. Andrews, F. Baccelli, O. Dousse, and M. Franceschetti, "Stochastic geometry and random graphs for the analysis and design of wireless networks," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 7, pp. 1029-1046, Sept. 2009.

[19] Y. Zhou and W. Zhuang, "Throughput analysis of cooperative communication in wireless ad hoc networks with frequency reuse," *IEEE Trans. on Wireless Communications*, vol. 14, no. 1, pp. 205-218, Jan. 2015.

[20] L. Duan, J. Huang, and B. Shou, "Economics of femtocell service provision," *IEEE Trans. on Mobile Computing,* vol. 12, no. 11, pp. 2261-2273, Nov. 2013.

[21] S. Bu, F. R. Yu, and Y. Qian, "Energy-efficient cognitive heterogeneous networks powered by the smart grid," *Proc. IEEE INFOCOM'13*, Turin, Italy, Apr. 2013.

[22] R. E. Miles, "On the homogeneous planar Poisson point process," *Mathematical Biosciences*, vol. 6, pp. 85-127, 1970.
[23] D. Wu and R. Negi, "Effective capacity: A wireless link model for support of quality of service," *IEEE Trans. on Wireless Communicaitons*, vol. 2, no. 4, pp. 630-643, July 2003.
[24] F. Aurenhammer, "Voronoi diagrams - a survey of a fundamental geometric data structure," *ACM Computing Surveys (CSUR)*, vol. 23, no. 3, pp. 345-405, Sept. 1991.
[25] 3GPP TR 36.808 v10.0.0, *Evolved universal terrestrial radio access (E-UTRA); carrier aggregation; base station (BS) radio transmission and reception (release 10)*, Tech. Spec. Group Radio Access Network, June 2012.
[26] B. Błaszczyszyn, M. K. Karray, and H.-P. Keeler, "Using poisson processes to model lattice cellular networks," *Proc. IEEE INFOCOM'13*, Turin, Italy, Apr. 2013.
[27] J.-S. Ferenc and Z. Néda, "On the size distribution of Poisson Voronoi cells," *Physica A: Statistical Mechanics and its Applications*, vol. 385, no. 2, pp. 518-526, Aug. 2007.
[28] E. L. Lawler and D. E. Wood, "Branch-and-bound methods: A survey," *INFORMS Operations Research*, vol. 14, no. 4, pp. 699-719, 1966.
[29] F. Baccelli, B. Błaszczyszyn, and P. Mühlethaler, "Stochastic analysis of spatial and opportunistic aloha," *IEEE Journal on Selected Areas in Communications*, vol. 27, no. 7, pp. 1105-1119, Sept. 2009.