Performance Analysis of Two-tier Wireless Networks with Dynamic Traffic, Backhaul Constraints and Terminal Mobility

Elena Bernal-Mor, Vicent Pla, Jorge Martinez-Bauset, and Luis Guijarro

Abstract—**During the last years, mobile cellular networks have witnessed an enormous growth in the carried data-traffic volume. The current networks' features are not enough to cope with this traffic trend and the concept of small cells has emerged as a feasible solution to increase the network capacity. However, the deployment of small cells introduces several technical challenges such as the cross-tier interference between the macrocell and the small cells, or the use of the subscriber land-line to send the backhaul data. In this paper, an analytical model is proposed to study the impact that the user traffic dynamics, the mobility of macrocell users, the scheme chosen to associate macrocell users to the small cells and the changing available capacity of the small cells backhaul have on the system performance. To make the solution of the model computationally feasible, we exploit the time-scale decomposition approach. In most practical scenarios, the arrival and departure rates of traffic flows are much larger than the rate of events associated with the mobility of macrocell users. Then, flows perceive that macrocell users are still. This model is applied to identify the scheme to associate macrocell users to the small cells which maximizes the performance perceived by the small cell users.**

I. INTRODUCTION

IN the last years, mobile cellular networks have experienced a major growth and progress due to a change in the way today's society creates, shares and consumes information. The increasing popularity of some applications has lead to a spectacular growth in the carried data-traffic volume and it will continue to grow as mobile systems are expected to support a larger variety of multimedia services. Moreover, according to recent surveys [1], the traffic which is expected to produce the bulk of the network load will mainly occur indoor. Unfortunately, the current networks' features are not enough to cope with this development paradigm. In this context, the novel concept of small cells [2], [3], [4], [5] has emerged as a solution to increase both network capacity and indoor coverage.

Small cells provide service to small geographical areas and require low-cost and low-power base stations called Small cell Access Points (SAPs). These SAPs are installed by the users for better indoor voice and data reception and benefit

both users and operators. Users improve their quality of service (QoS), while operators can manage the growth of traffic without the need to deploy new network infrastructure. Moreover, the SAPs send the backhaul data to the cellular operator network over the user land-lines [2], thus allowing operators to release resources of their own backhauls for other users connected to the macrocell.

From the perspective of the small cell connectivity priority, two types of users are defined: i) *Small cell Users* (SUs) which are registered in the small cell and can always connect to the small cell, i.e., they are the rightful users; ii) *Macrocell Users* (MUs) which are not registered in the small cell. Depending on the Registration Policy (RP) at the small cell, they might be allowed to connect to the small cell or not. In addition, a user land-line is shared by the *Land-line Users* (LUs), which are non-cellular users that generate traffic carried by the same land-line connection that the SAP uses as a backhaul.

The deployment of small cells introduces several technical challenges [2], [6], [7]. One of the most crucial performance-limiting factors is the cross-tier interference between the macrocell and the small cell [8], [9], [10]. This problem has been widely addressed in the literature and many approaches have been proposed, which, for example, involve the use of power control [11], [12], [13]; or advanced spectrum management techniques [14], [15]. The level of interference perceived by SUs is influenced by the number of MUs allowed to connect to the SAP. Hence, besides using power control or spectrum management, the interference can also be managed by allowing strong MUs interferers to become associated to the SAP [16], [17].

The RP is a key mechanism to regulate the access of users to the small cells. This mechanism provides different levels of priority to SUs and MUs, allowing the small cell to control which users can have access to it. Three basic RPs have been defined [16]: i) *closed access*, only the subscribed users, i.e. the SUs, have access to the small cell (MUs cannot connect to the small cell); ii) *open access*, all users can make use of any small cell (in that sense, there are no differences between SUs and MUs); and iii) *hybrid access*, a limited amount of the small cell resources are available to MUs, while SUs have no restrictions. The hybrid access mode is proposed as a trade-off between open and closed access modes [18], [17], [19]. Depending on the system resources, users traffic profile, number of MUs and their mobility pattern, number of SUs, degree of interference and small cell backhaul capacity constraints, the RP in the hybrid access mode has to be

carefully configured in order to guarantee a certain degree of QoS to the SUs.

To the best of our knowledge, the small cell architecture has been mostly studied considering a static traffic setting in which there is a constant number of infinitely long flows under transmission. In contrast, in this paper we consider a dynamic scenario where terminals generate finite flows that randomly arrive to the system and eventually terminate. As a result, the number of flows and the load of the system vary dynamically over time. In [20], we introduced a preliminary version of this approach. However, there we restricted to the time scale of flow-level dynamics, and the position of terminals was considered static. The new model proposed here considers the time scales beyond flow-level dynamics. Moreover, in this extended approach it is also contemplated that the position of MUs inside the macrocell can vary over time.

The major contribution of the new analytical model is that it considers in an integrated way: i) the traffic profile of the users, ii) the number of MUs and their mobility pattern, iii) the number of SUs, iv) the interference of MUs over SUs, v) different MUs registration policies in the SAPs, and vi) the traffic activity of the LUs. To make the analytical model computationally tractable, we exploit its time scale decomposition [21]. In most practical scenarios, the arrival and departure rates of traffic flows are much larger than the rate of events associated with the mobility of MUs. Then, flows perceive that MUs are still. This technique considerably reduces the computational cost, which allows us to evaluate the performance of different RP and to identify the policy that maximizes the performance perceived by SUs. Moreover, in this work we also validate the analytical model by simulation.

This paper is organized as follows. In Section II, we describe the interference model proposed. In Section III, we present the system model used to evaluate the performance of the small cell. Then, in Section IV we define the time scale decomposition approach and apply it to our system model. In Section V, we discuss the numerical results and validate the analytical model by simulation. Finally, Section VI concludes the paper.

II. INTERFERENCE MODEL

We analyze the performance of a small cell which is inside a macrocell. We consider the uplink direction, but our study is independent of whether the uplink or downlink is considered. It is well-known that the path-loss is a dominant factor in the reduction of the received signal power. Hence, the interference produced by an MU connected to the BS upon the SAP will depend on the distances from this MU to the SAP and the BS. According to these distances, i.e., according to the amount of interference caused to the SAP by the MUs connected to the BS, we consider different interference regions within the macrocell. We define K regions denoted by A_i, $i = 1, 2, \cdots, K$. MUs connected to the macrocell base station (BS) cause the strongest interference when located in A_1, and the weakest interference when located in A_K. The interference caused to the SAP by each active MU in A_i contributes to the reduction of the achievable bitrate in the

small cell by a fraction γ_i $(1 > \gamma_1 > \cdots > \gamma_K \geq 0)$ of its maximum value (i.e., the value with no interference). The total achievable bitrate reduction is then obtained by adding the contributions of all MUs in each interference region. To illustrate this point, consider the following simple example: let $K = 2$, $\gamma_1 = 0.08$, $\gamma_2 = 0.03$, and let R_s be achievable bitrate in the small cell. If there were 2 MUs in A_1 and 5 MUs in A_2 (all of the them connected to the BS and continuously transmitting), then the bitrate that a single SAP user can achieve is $(1 - 2 \cdot 0.08 - 5 \cdot 0.03)R_s = 0.69R_s$. To keep this example simple, constraints to the bitrate of SAP users other than interference have not been considered. In Section III-D, the rest of relevant constraints as well as the fact that the MUs connected to the BS are not transmitting all the time are taken into account to obtain Eq. (20).

Without loss of generality, we consider that the BS and the SAP are in the abscissa axis symmetrically situated respect to the origin. If the distance between the BS and the SAP is D, the BS is at $(-\frac{D}{2}, 0)$ and the SAP at $(\frac{D}{2}, 0)$. Let d be the distance between the BS and an MU at (x, y) connected to the BS, and d' the distance between this MU and the SAP; see Fig 1. These distances can be written as:

$$d = \sqrt{\left(x + \frac{D}{2}\right)^2 + y^2}; \qquad d' = \sqrt{\left(x - \frac{D}{2}\right)^2 + y^2}. \quad (1)$$

The transmission power of an MU connected to the BS is proportional to d^ζ, where ζ is the path-loss exponent. Similarly, the interference that this MU (which is connected to the BS) produces at the SAP is inversely proportional to d'^ζ. As a result, the Signal to Interference Ratio (SIR) at the SAP, which determines the uplink achievable bitrate for the users connected to the SAP, is proportional to d'^ζ and inversely proportional to d^ζ, that is

$$SIR \propto \left(\frac{d'}{d}\right)^\zeta. \quad (2)$$

Then, the thresholds $\delta_1 < \delta_2 < \cdots < \delta_{K-1}$ divide the macrocell into interference regions as:

$$A_i = \left\{(x, y) : \delta_{i-1} \leq \frac{d'}{d} \leq \delta_i\right\}, \quad (3)$$

with $\delta_0 = 0$ and $\delta_K = \infty$.

The border between A_i and A_{i+1} is given by the points where $d'/d = \delta_i$, i.e. by the curve given by the points where the quotient of the distances d and d' is constant. Then, we are searching the points (x, y) where:

$$\sqrt{(x - D/2)^2 + y^2} / \sqrt{(x + D/2)^2 + y^2} = \delta_i. \quad (4)$$

Further manipulation of this expression results in

$$\left(x + \frac{\frac{D}{2}(\delta_i^2 + 1)}{\delta_i^2 - 1}\right)^2 + y^2 = \frac{4(\frac{D}{2})^2 \delta_i^2}{(\delta_i^2 - 1)^2}, \quad (5)$$

which is a circumference with center:

$$\left(-\frac{\frac{D}{2}(\delta_i^2 + 1)}{\delta_i^2 - 1}, 0\right), \quad (6)$$
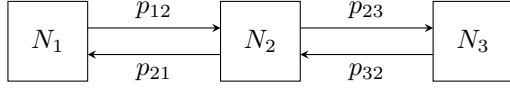
Fig. 1. Different interference regions, $K = 3$.

Fig. 2. State transitions of the CTMC which models the backhaul.

and radius

$$\frac{2\frac{D}{2}\delta_i}{|\delta_i^2 - 1|}.\tag{7}$$

Notice that if $\delta_i = 1$, the curve is the perpendicular bisector of the segment which connects the BS and the SAP points, i.e., it is the ordinate axis. In Figure 1, an example with $K = 3$ regions with $D = 70$, $\delta_1 = 3/2$ and $\delta_2 = 9/4$ is shown.

We consider that MUs can be associated either to the BS or to the SAP, and that any user (terminal) can be either idle or active, i.e. it can be idle or uploading a data flow. In this work, we evaluate two schemes to associate MUs to the SAP. For each scheme we determine the performance obtained by the users connected to the SAP. As it will be apparent later, the proposed wireless network system model is independent of the interference model used, as far as it is based on the number of mobiles in each macrocell region. This has been a common assumption in the literature [22].

III. SYSTEM MODEL

We begin this section by describing the backhaul traffic model. Then, we present the RP at the small cell, the mobility model and the system model, which considers the interference, the mobility of MUs and the backhaul load. A summary of all the parameters introduced in this paper appears in Table I in Section V.

A. Backhaul traffic model

The traffic going through the SAP competes with the LUs' traffic for the wireline bandwidth. We assume that priority is given to LUs' traffic since it is the primary reason for the Internet access service subscription. Nevertheless, it is assumed that there is a minimum bitrate that the SAPs would be able to get at anytime in order to prevent starvation and also to guarantee that the SUs can get access to the voice service at anytime. This is required as the SUs may not have land-line telephone subscription nor a sufficiently good coverage from the BS.

The load generated by the LUs is modeled by a finite-state Continuous Time Markov Chain (CTMC) with J states. The transition rate from state j to state j' is denoted by $\beta_{j,j'}$, $j, j' = 1, 2, \cdots, J$ (see Figure 2). When the CTMC is in state j the available bitrate in the backhaul is C_j, where $C_1 > C_2 > \cdots > C_J > 0$. The minimum bitrate reserved to the SAP in order to avoid starvation is given by C_J. Our model does not impose any limitations on the pattern of possible transitions between the states of the CTMC. For the sake of illustration we consider a scheme in which the traffic generated by LUs resembles TCP behavior: increases in the utilized bandwidth occur one step at a time, but it can drop several steps at once,

$$\beta_{j,j'} \begin{cases} \geq 0 & \text{if } j' > j \text{ or } j' = j - 1, \\ = 0 & \text{if } j' < j - 1. \end{cases}\tag{8}$$

Note that the proposed wireless network system model defined later in Section III-D is independent of the model used to characterize the backhaul traffic, and any Markovian backhaul traffic model might be used. The fitting of the parameters of the backhaul traffic model from real traffic data is out of the scope of our study [23], [24].

B. MU association scheme

Each user (MU or SU) can be either idle or active. When active, it is uploading an elastic flow, which corresponds to the transfer of a digital document. If the uploading occurs through the SAP (BS) we say that the user is associated to the SAP (BS). Let H be the number of SUs permanently registered, or associated, to the SAP. Let M be the total number of MUs in the cell and M' the number of MUs associated to the SAP. Let m_i be the total number of MUs in A_i, and $m'_i \leq m_i$ be the number of MUs in A_i which are associated to the SAP. Then, $M = \sum_i m_i$ and $M' = \sum_i m'_i$. Note that m_i may vary over time due to the mobility of MUs.

We consider three interference regions ($K = 3$). The MUs in A_1 and A_2 are considered to be close enough to the SAP to become candidates to be handed over from the BS to the small cell, while those in A_3 are not ($m'_3 = 0$). We also consider that MUs in A_3 do not produce interference to the SAP ($\gamma_3 = 0$). Obviously, if there are still users in A_1, users in A_2 cannot be associated to the SAP (if $m'_1 < m_1$ then $m'_2 = 0$). At a given time, the number of users associated to the BS, n_m, and to the SAP, n_s, are given by

$$\begin{aligned} n_m &= m_1 + m_2 + m_3 - m'_1 - m'_2, \\ n_s &= H + m'_1 + m'_2. \end{aligned}\tag{9}$$

In this work, we study two different schemes to associate MUs to the SAP, that we denote by A and B. In Scheme A,

Fig. 3. Equivalent closed queuing network for the mobility model

the number of MUs associated to the SAP in A_1 and in A_2 are given by:

$$m_1' = \min(\nu, m_1); \quad m_2' = \min(\nu - m_1', m_2), \quad (10)$$

where ν is a fixed parameter that defines the maximum number of MUs that can be associated to the SAP. If $\nu = 0$ the RP is *closed access*, and *hybrid access* otherwise ($\nu > 0$); in practice, however, a sufficiently high value of ν (so that $m_1 + m_2 < \nu$ can be guaranteed in all practical situations) would amount to an *open access* RP. If $\nu \leq m_1$, the ν MUs associated to the SAP are in A_1. Otherwise, all the MUs in A_1 are associated to the SAP and the rest of MUs up to ν are in A_2. Note that the number of MUs associated to the SAP will be less than ν only when $m_1 + m_2$ is lower than ν. In this case, since only the MUs in A_1 and A_2 can be associated to the SAP, there will be exactly $m_1 + m_2$ MUs associated to the SAP.

In Scheme B, let $\kappa \in [0, 2]$ be a parameter that determines m_1' and m_2' as follows,

$$m_1' = \min\left(\lceil \kappa \cdot m_1 \rceil, m_1\right),$$
$$m_2' = \left[\left\lceil (\kappa - 1) \cdot m_2 \right\rceil\right]^+, \quad (11)$$

where the notation $[x]^+ \triangleq \max(0, x)$ is used. Then, no MUs are allowed to connect to the SAP for $\kappa = 0$. The number of associated MUs to the SAP is half of the MUs in A_1 for $\kappa = 0.5$, all MUs in A_1 for $\kappa = 1$, all MUs in A_1 and half of the MUs in A_2 for $\kappa = 1.5$ and all MUs in A_1 and A_2 for $\kappa = 2$. Note that the RP is: *closed access* when $\kappa = 0$, *hybrid access* when $0 < \kappa < 2$, and *open access* when $\kappa = 2$.

C. Mobility model

We consider that the residence time of an MU in a certain region A_i before entering another one is generally distributed. We denote by λ_i the rate of transitions from region i to region $i - 1$ ($i = 2, 3$), and by μ_i the rate of transitions from region i to region $i + 1$ ($i = 1, 2$), as represented in Fig 1. We model the mobility of MUs by an equivalent closed queuing network with $K = 3$ nodes and a constant number of M MUs roaming inside the BS coverage area. We denote by N_1, N_2 and N_3 the three nodes of the closed network, as shown in Fig. 3. Each node N_i is of the infinite-server type.

Let $\vec{m} = (m_1, m_2, m_3)$ be the vector whose ith component gives the number of users at N_i and p_{ij} be the routing probabilities from N_i to N_j. The routing probabilities can be obtained from the region transition rates as

$$p_{12} = p_{32} = 1; \quad p_{23} = \frac{\mu_2}{\mu_2 + \lambda_2}; \quad p_{21} = \frac{\lambda_2}{\mu_2 + \lambda_2}. \quad (12)$$

Let τ_i be the mean residence time in N_i. Then,

$$\tau_1 = \frac{1}{\mu_1}; \quad \tau_2 = \frac{1}{\mu_2 + \lambda_2}; \quad \tau_3 = \frac{1}{\lambda_3}. \quad (13)$$

Let Λ_i denote the total arrival rate to node N_i. Then, Λ_i are the solution to the flow balance conditions:

$$\Lambda_i = \sum_{j=1}^{3} \Lambda_j p_{ji}, \quad i = 1, \cdots, 3. \quad (14)$$

It is known that these equations are not independent and hence they cannot be solved to uniquely find the Λ_i. However, the Λ_i's can be obtained up to a multiplicative constant. Let Λ_j^* be a particular solution of Eq. (14). We assume that $\Lambda_1^* = \lambda_2$ and then, we solve Eq. (14) to obtain Λ_2^* and Λ_3^* as:

$$\Lambda_2^* = \mu_2 + \lambda_2; \quad \text{and} \quad \Lambda_3^* = \mu_2. \quad (15)$$

We have a three-node closed network of quasi-reversible queues, and the steady-state distribution of the number of customers in each node, $p(\vec{m})$, is given by [25]:

$$p(\vec{m}) = P(\mathbf{0}) \frac{(\Lambda_1^* \tau_1)^{m_1}}{m_1!} \frac{(\Lambda_2^* \tau_1)^{m_2}}{m_2!} \frac{(\Lambda_3^* \tau_3)^{m_3}}{m_3!}, \quad (16)$$

where $P(\mathbf{0})$ is obtained by normalization. This distribution is known to be insensitive to the residence time distributions at each node, i.e. it only depends on them through their means.

D. System model

Each user (MU or SU) can be either idle or uploading an elastic flow. Both, the periods of time that users are idle between two consecutive uploads and the size (in bits) of the flows generated by users form sequences of independent and identically distributed random variables. The lifetime of a flow will depend on its size and on the amount of available resources, which vary with time. For the sake of mathematical tractability, we assume that the duration of an idle period is exponentially distributed with mean $1/\alpha$. Likewise, the size of a flow is also exponential with mean σ.

We model the proposed system using a multidimensional process. The states of this process are represented by the state vector

$$\boldsymbol{s} = (m_1, m_2, m_3, x, j, z_1, z_2, z_3), \quad (17)$$

where m_i denotes the total number of MUs in region A_i and $\sum_i m_i = M$; $x \leq n_s$ denotes the number of active users associated to the SAP; j denotes the state of the backhaul, with the corresponding available bitrate C_j, $j = 1, 2, \cdots, J$; and $z_i \leq m_i - m_i'$ denotes the number of active MUs in A_i which are associated to the BS. For convenience, we will simplify the notation and write $\boldsymbol{s} = (\vec{m}, x, j, \vec{z})$.

The maximum bitrate in the macrocell (small cell) that could be achieved by employing all the macrocell (small cell) resources and assuming a sufficiently high Signal to Interference plus Noise Ratio (SINR) is R_m (R_s). Due to the impatience of users [26] the minimum bitrate of users is restricted to be r_m, and due to hardware limitations of terminals the maximum bitrate of users is r_M. While the maximum bitrate r_M is imposed by hardware limitations and is a hard bound, the minimum r_m is a design target and might not be guaranteed at all times, as discussed later.

The throughput obtained by each active MU associated to the BS at state s is given by:

$$\phi_m(s) = \min\left\{r_M, \frac{R_m}{z_1 + z_2 + z_3}\right\}. \tag{18}$$

Note that each active MU receives an equal share of the available BS radio resources.

The throughput obtained by each active SAP user is constrained by: i) hardware limitations; ii) the interference caused by the MUs connected to the BS upon the SAP; and iii) the available backhaul resources. In order to determine the impact of the interference caused by the MUs on the achievable bitrate at the SAP, we use the interference model described in Section II. Since each active MU connected to the BS obtains an equal bandwidth share, the average fraction of time that a flow is using the BS resources in state s is given by:

$$\eta_t(s) = \frac{\phi_m(s)}{R_m} = \min\left\{\frac{r_M}{R_m}, \frac{1}{z_1 + z_2 + z_3}\right\}. \tag{19}$$

Then, the throughput at the SAP is reduced by the fraction $\eta_t(s) \cdot (\gamma_1 z_1 + \gamma_2 z_2)$ due to the interference produced by active MUs. Taking the three constrictions into account, the throughput obtained by each active SAP user at state s is given by:

$$\phi_s(s) = \min\left\{r_M, \frac{R_s}{x}\left[1 - \eta_t(s) \cdot (\gamma_1 z_1 + \gamma_2 z_2)\right]^+, \frac{C_j}{x}\right\}. \tag{20}$$

E. Admission Control Scheme

We consider admission control at the flow level, both at the SAP and BS. A flow from an SU or an MU associated to the SAP is accepted to the small cell if, after acceptance, all ongoing flows in the small cell obtain a bitrate equal or larger than r_m. When a flow is not accepted, it is lost. Let $a_s(s)$ be the probability that a flow is accepted by the SAP and $a_{m,i}(s)$ the probability that a flow from a terminal in interference region A_i is accepted by the BS, both at state s. Also, let e_k be an 8-dimensional vector with a 1 on the k-th position and 0's elsewhere. Then, these probabilities can be expressed as:

$$a_s(s) = \begin{cases} 1 & \text{if} \quad \phi_s(s + e_4) \geq r_m \\ 0 & \text{otherwise} \end{cases} \tag{21}$$

$$a_{m,i}(s) = \begin{cases} 1 & \text{if} \quad \phi_m(s + e_{5+i}) \geq r_m \\ 0 & \text{otherwise} \end{cases} \tag{22}$$

Note that a flow is not forced to terminate if, after accepted, its throughput falls below r_m and it is served with an unacceptably low QoS, due, for example, to a drop in the backhaul available capacity.

Observe also that the cardinality of the state space of this CTMC can become very large, and its solution might become computationally unfeasible. However, if the time scale at which terminals move from one interference region to another is much slower that the traffic flow dynamics, time scale decomposition can be used to greatly reduce the computational complexity associated to the CTMC solution [21].

IV. SOLVING THE SYSTEM MODEL THROUGH TIME SCALE DECOMPOSITION

In most practical scenarios, the arrival and departure rates of flows to and from the system are much larger than the rate of events associated with the mobility of MUs, i.e., changes from one interference region to another. Then, flows perceive that MUs are still. This allows us to decompose the model defined by (17) in the previous section into two independent models. We refer to them as the fast time scale subsystem (FTSS) and the slow time scale subsystem (STSS). Then, the stationary probability distribution for the complete system, π, is given by

$$\pi(\vec{m}, x, j, \vec{z}) = p(\vec{m}) \cdot p(s_f|(\vec{m})), \quad s_f = (x, j, \vec{z}), \tag{23}$$

and the set of feasible states is given by

$$\mathcal{S} := \{s: \quad \vec{m}, \vec{z} \in \mathbb{N}^3; \quad x, j \in \mathbb{N} \quad x \leq X; \quad j \leq J$$
$$z_i \leq Z_i; \quad \sum_i m_i = M\}, \tag{24}$$

where $X = n_s$, $Z_1 = m_1 - m_1'$, $Z_2 = m_2 - m_2'$ and $Z_3 = m_3$.

The stationary probability distribution for the FTSS $p(s_f|(\vec{m}))$ must be found for each combination of MUs in the interference regions. That is, a different CTMC for the FTSS must be solved for each state of the STSS ($\vec{m}$). For the STSS, the movement of users over different interference regions does not depend on the flow dynamics. Therefore, its stationary probability distribution is given by (16).

For the FTSS, let $q_{(f|\vec{m})}(s_f, s_f')$ be the transition rate from state s_f to state s_f' conditioned on $\vec{m}$. For simplicity, we denote these transition rates by q_f. Then,

$$q_f = \begin{cases} a_s(s_f)(n_s - x)\alpha & s_f' = s_f + e_4, \\ x\phi_s(s_f)/\sigma & s_f' = s_f - e_4, \\ \beta_{j,j'} & s_f' = s_f + (j' - j)e_5, \\ a_{m,1}(s_f)(m_1 - m_1' - z_1)\alpha & s_f' = s_f + e_6, \\ z_1\phi_m(s_f)/\sigma & s_f' = s_f - e_6, \\ a_{m,2}(s_f)(m_2 - m_2' - z_2)\alpha & s_f' = s_f + e_7, \\ z_2\phi_m(s_f)/\sigma & s_f' = s_f - e_7, \\ a_{m,3}(s_f)(m_3 - z_3)\alpha & s_f' = s_f + e_8, \\ z_3\phi_m(s_f)/\sigma & s_f' = s_f - e_8, \\ 0 & \text{otherwise.} \end{cases}$$

Recall that $\beta_{j,j'}$ are the transition rates due to variations in the backhaul bandwidth occupied by LUs, as defined in Section III-A.

The average bitrate obtained by users associated to the SAP (SUs or MUs) is thus given by:

$$E[\phi_s] = \frac{\sum_{s \in \mathcal{S}} x\phi_s(s)\pi(s)}{\sum_{s \in \mathcal{S}} x\pi(s)}. \tag{25}$$

Let us denote by P_b the blocking probability in the small cell, which is the probability that the system is in a state in which the SAP will not accept the flows generated by the users associated to it. Then,

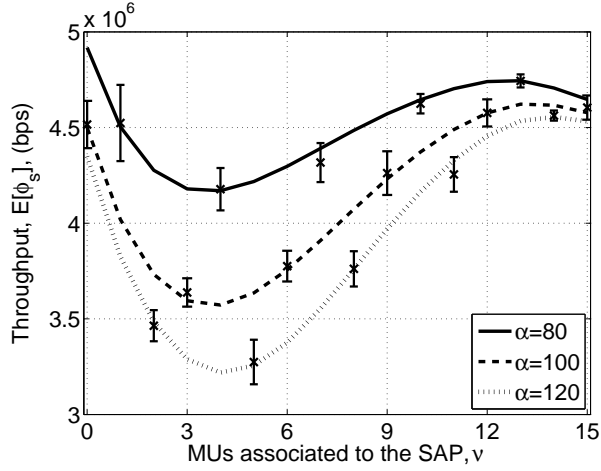$$P_b = \sum_{s \in \mathcal{S}, x < n_s} (1 - a_s(s))\pi(s). \tag{26}$$

Fig. 4. Association Scheme A. Average throughput of users associated to the SAP, $E[\phi_s]$, as a function of ν for different values of α.

Let $\mathcal{L}$ be the set of states where the active users connected to the SAP are served with a bitrate lower than the minimum,

$$\mathcal{L} := \{ \boldsymbol{s} \in \mathcal{S}; \quad \phi_s(\boldsymbol{s}) < r_m \}. \tag{27}$$

The probability that active users associated to the SAP are served with a bitrate lower than the minimum r_m and therefore with low QoS is obtained as

$$P_{QoS} = \sum_{\boldsymbol{s} \in \mathcal{L}} \boldsymbol{\pi}(\boldsymbol{s}). \tag{28}$$

V. NUMERICAL EVALUATION

To show the flexibility and features of the proposed model, we evaluate the performance perceived by the users associated to the SAP in different scenarios. First, in section V-A, the impact of the two schemes proposed to associate MUs to the SAP is evaluated. The optimal association approach will then be used in Section V-C to evaluate the impact of the backhaul traffic. Finally, in Section V-D, we evaluate the impact that the mobility of MUs has on the performance perceived by SUs.

A list of all model parameters and their values, unless otherwise specified, is given in Table I.

In Table II, the transition rate matrices $[\beta_{jj'}]$ are set to obtain different load conditions. In scenario 1 the average LU's traffic is light, in scenario 2 it is medium, and in scenario 3 it is heavy. The backhaul average available bitrates for the three scenarios are: 89.8, 53.1 and 14.4 Mbps, respectively. Numerical results are obtained for scenario 1, unless otherwise stated.

A. Impact of the scheme used by the SAP to associate MUs

We consider the association schemes defined in Section III-B. For Scheme A, Fig. 4 shows the evolution of the average throughput of users associated to the SAP ($E[\phi_s]$) with the maximum number of MUs that can be associated to the SAP (ν). Each curve corresponds to a different duration of the idle period $1/\alpha$. Note that the load offered by users increases with α. When α is smaller, i.e. longer idle periods, the capacity of the system is shared among less active users

Parameter	Symbol	Value
Interference		
Distance BS–SAP	D	$70\,\text{m}$
Distance BS–MU	d	-
Distance SAP–MU	d'	-
Number of interference regions	K	3
Interference region i	A_i	-
Region interference thresholds	δ_1, δ_2	$3/2, 9/4$
Interference penalty	γ_1, γ_2	$0.8, 0.4$
Backhaul		
Number of states of the CTMC	J	3
State of the backhaul	j	$1, 2, 3$
Transition rate $j \to j'$	$\beta_{j,j'}$	See Table II
Backhaul bitrate state 1	C_1	$100\,\text{Mbps}$
Backhaul bitrate state 2	C_2	$50\,\text{Mbps}$
Backhaul bitrate state 3	C_3	$6\,\text{Mbps}$
Association Scheme		
Association parameter Scheme A	ν	$0, 1, \cdots, 15$
Association parameter Scheme B	κ	$[0, 2]$
Number of SAP MUs	M'	Depends on ν or κ
Number of BS MUs	n_m	Depends on ν or κ
Total number of SAP users	n_s	Depends on ν or κ
Number of MUs in A_i	m_i	Varies over time
Number of SAP MUs in A_i	m'_i	Varies over time $m'_3 = 0$
Number of active SAP users	x	Varies over time
Number of active BS MUs in A_i	z_i	Varies over time
Mobility		
Transition rate $i \to i-1$	λ	$8 \cdot 10^{-4} s^{-1}$
Transition rate $i \to i+1$	μ	$10^{-4} s^{-1}$
Routing probability	p_{ij}	See Eq. (12)
System		
Total number of SUs	H	6
Total number of MUs	M	15
Maximum Macrocell bitrate	R_m	$500\,\text{Mbps}$
Maximum Small cell bitrate	R_s	$100\,\text{Mbps}$
Maximum user bitrate	r_M	$100\,\text{Mbps}$
Minimum user bitrate	r_m	$1\,\text{Mbps}$
Mean flow size	σ	$800\,\text{kbits}$
Mean duration of an idle period	$1/\alpha$	$0.01\,\text{s}$
State vector	$\boldsymbol{s}$	-
Performance parameters		
Fraction of time an MU uses the BS resources (in state $\boldsymbol{s}$)	$\eta_t(\boldsymbol{s})$	-
Throughput of an MU connected to the BS (in $\boldsymbol{s}$)	$\phi_m(\boldsymbol{s})$	-
Throughput of a SAP user (in $\boldsymbol{s}$)	$\phi_s(\boldsymbol{s})$	-
Mean throughput of SAP users	$E[\phi_s]$	-
Small cell blocking probability	P_b	-
Small cell low QoS probability	P_{QoS}	-

Scenario 1 light			Scenario 2 medium			Scenario 3 heavy		
-6	5	1	-51	50	1	-41	1	40
40	-50	10	10	-15	5	5	-55	50
0	50	-50	0	40	-40	0	10	-10

on average, therefore the mean throughput perceived by a user associated to the SAP ($E[\phi_s]$) is higher.

For a given α, the throughput first decreases, reaches a minimum and then increases until it reaches a maximum, beyond which it decreases again. As expressed by (20), ϕ_s depends on two somehow opposite effects. It decreases as the number of users associated to the SAP increases, i.e. as

the denominator x of (20) increases. On the other hand, ϕ_s increases as more MUs from regions A_1 and A_2 get associated to the SAP, i.e. the available capacity of the SAP increases when the number of interferers decreases. We refer to the first one as the *bandwidth sharing effect*, and to the second as the *interference reduction effect*.

Three parts can be observed for each curve in Fig. 4. In the first one, the bandwidth sharing effect dominates and $E[\phi_s]$ decreases as more MUs get associated. In the second one, the interference reduction effect dominates and $E[\phi_s]$ increases as more MUs get associated upto a maximum. In the third one, $E[\phi_s]$ decreases again as more MUs get associated. The effect in the third part is due to the fact that when the number of MUs associated to the SAP is large, further MUs associations most likely corresponds to MUs in A_2, which contribution to the SAP interference is lower that those in A_1. Observe also that when α is low, i.e less loaded systems, the maximum for $E[\phi_s]$ is achieved when MUs are not allowed to connect to the SAP, i.e. closed access RP. However, when α is high, the maximum for $E[\phi_s]$ is achieved when the number of MUs associated to the SAP is close to the maximum number of MUs in the system. In this case the RP follows a hybrid access. Finally, note that as α decreases, ν must be decreased to achieve the maximum throughput. This is because for less loaded scenarios $\eta_t(s)$, and therefore the interference, is lower. Then, less MUs are needed to be associated to the SAP in order to balance the advantage of less interference and the penalty of the bandwidth sharing effect.

For association Scheme B, Fig. 5 shows the average throughput of users associated to the SAP ($E[\phi_s]$) as a function of κ. Recall that in Scheme B the maximum number of MUs that can be associated to the SAP is a non-decreasing function of κ. As before, each curve corresponds to a different α. For a given α, we see again the same behavior observed with Scheme A. When α is small, the highest throughput is achieved when no MUs are associated to the SAP, i.e. closed access. However, when α is high, the highest throughput is achieved when $\kappa = 1$, i.e. when all the MUs in A_1 are associated to the SAP and no MUs in A_2 are associated to the SAP, which corresponds to a hybrid access. Again, this can be explained in the same way as in Scheme A. For high α, the interference reduction effect due to the association of MUs in A_2 does not compensate the additional bandwidth sharing effect.

From now on, we only consider Scheme B, as it outperforms Scheme A by achieving a larger $E[\phi_s]$.

B. Validation of the analytical model

In order to validate the assumptions made in the analytical model concerning the movement of users and the traffic generated by them, extensive simulations have been performed deploying more realistic mobility and traffic profiles. A discrete event based simulation program implemented in C has been developed to mimic the system dynamics. That is, by generating flows that arrive and terminate, random idle times at flow terminations and random movement patterns for users. To model the residence time in the three regions more
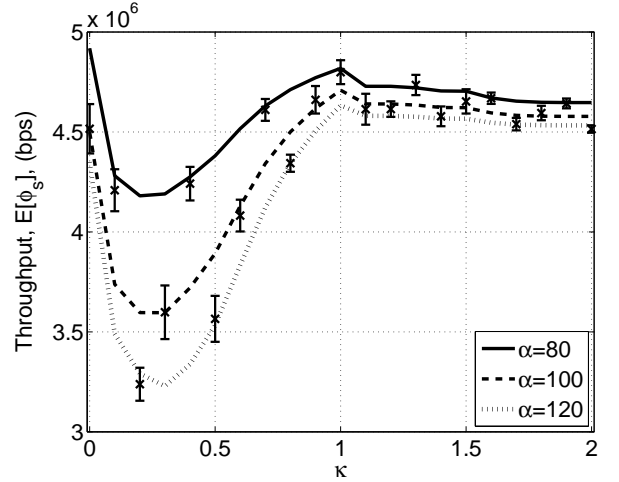

Fig. 5. Association Scheme B. Average throughput of users associated to the SAP, $E[\phi_s]$, as a function of κ for different values of α.

realistically, the lognormal distribution is chosen [27]. The duration of an idle period and the size of a flow are also considered to be lognormally distributed [28], [29]. For the lognormal distributions of the residence time, flow idle period and flow size, their means coincide with the means of the corresponding exponential distributions used in the analytical model. However, their coefficients of variation are set to 2.

In Figures 4 and 5, the results of the analytical model are represented using lines, while the results of the simulations are represented using crosses. For each simulation point, 10 simulations are executed and 90% confidence intervals are shown. Observe the excellent agreement between the analytical and simulation results. In addition, the results seem to suggest that the stationary distribution of the CTMC that models the system dynamics might be insensitive to the flow size and idle time distributions. The formal proof of this fact is outside the scope of the paper.

C. Impact of backhaul dynamics

Figure 6 shows $E[\phi_s]$ as a function of κ, for the three different backhaul load scenarios shown in Table II. As expected, the higher the LUs traffic, the lower the throughput achieved by the SUs. In Scenario 1, where the LUs traffic is light, the highest throughput is achieved when all MUs in A_1 are associated to the SAP. In Scenario 2, where the LUs traffic is medium, associating more MUs to the SAP reduces the throughput. Clearly, associating more MUs increases the available wireless capacity of SAP users, but the capacity bottleneck is now the backhaul. Then, adding more MUs only strengthens the bandwidth sharing effect. For Scenario 3, where LUs traffic is heavy, the average throughput also decreases as κ increases, falling even bellow the minimum $r_m = 1$ Mbps. This figure clearly shows the importance of taking into consideration the backhaul state when defining the MUs association scheme.

The evolution of P_b, the blocking probability of the flows generated by users associated to the SAP, as a function of κ is shown in Fig. 7. Each curve corresponds to one of the
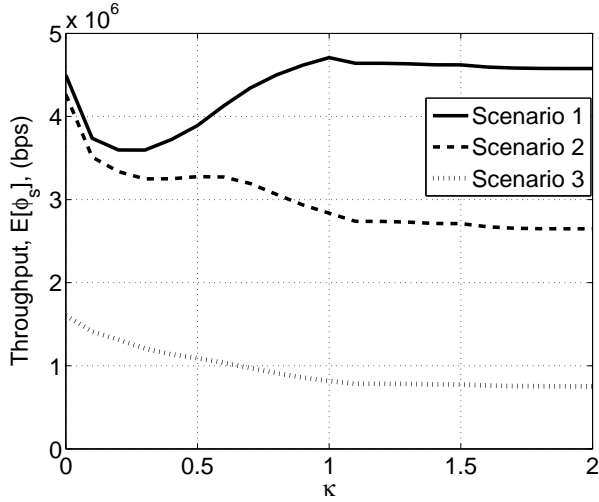
Fig. 6. Throughput achieved by users associated to the SAP as a function of κ for different backhaul scenarios.
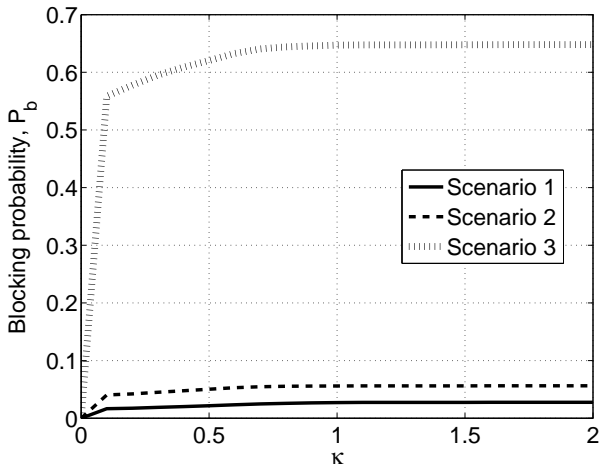

Fig. 8. Throughput achieved by users associated to the SAP as a function of the ratio λ/μ for different values of κ.

Fig. 7. Blocking probability of the users associated to the SAP as a function of κ for different backhaul scenarios.

Fig. 9. Low QoS probability of the users associated to the SAP as a function of the ratio λ/μ for different values of κ.

three different backhaul load scenarios shown in Table II. The blocking probability is zero when no MUs are allowed to connect to the SAP. As the minimum available bitrate at the backhaul is $C_3 = 6$ Mbps, this guarantees that all $H = 6$ SUs will get the minimum rate $r_m = 1$ Mbps when no MUs are associated. As expected, in Scenario 3, where the LUs traffic is high, the blocking probability increases rapidly as MUs are associated to the SAP. Also, note in Fig. 6, particularly in Scenario 3, that $\mathrm{E}[\phi_s]$ would be much lower if no flow admission control would be exercised.

D. Impact of MUs movement

Let $\lambda = \lambda_2 = \lambda_3$ and $\mu = \mu_1 = \mu_2$. In Fig. 8, $\mathrm{E}[\phi_s]$ is shown as a function of the ratio λ/μ. Since $\lambda/\mu = \lambda_2/\mu_2 = p_{21}/p_{23}$, this ratio can be interpreted as the number of times that an MU in A_2 moves to A_1 (i.e., closer to the SAP) per each time an MU in A_2 moves to A_3 (i.e., closer to the BS). Therefore, for low values of λ/μ, on average, more MUs will be located in A_3 (closer to the BS) than in A_1 (closer to the
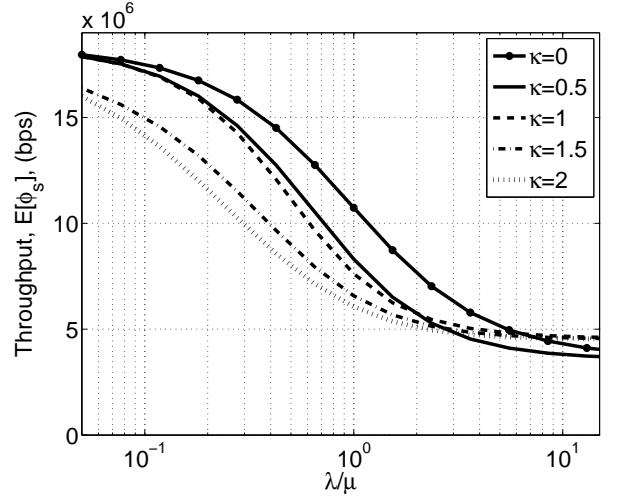
SAP), whereas for high values of λ/μ the opposite situation will be observed.

As observed in Fig. 8, the lower the ratio λ/μ, the higher the throughput, as less interference is produced by MUs at the SAP. Clearly, in this scenario, associating more MUs only strengthens the bandwidth sharing effect which decreases the throughput. The opposite is also true, the higher the ratio λ/μ, the higher the number of interferers close to the SAP, which decreases the available capacity. However, now associating more MUs strengthens the interference reduction effect which slightly increases the throughput.

In Fig. 9, P_{QoS}, the probability that users associated to the SAP are served with low QoS, is shown again as a function of the ratio λ/μ. For any curve with $\kappa > 0$, the lower the ratio λ/μ, the lower probability of being served with low QoS, as less interference is suffered by the SAP. For a low λ/μ ratio, the higher the value of κ, the higher the probability of being served with low QoS. This occurs because when interferers are far away, associating more MUs only reinforces the bandwidth sharing effect with not much gain in interference reduction.

For $\kappa = 0$, $P_{QoS} = 0$, since the minimum backhaul bandwidth guarantees that all the SUs are served at their minimum bitrate if no MUs are associated to the SAP. Conversely, if $\kappa > 0$ and the value of λ/μ is sufficiently high, the number of MUs associated to the SAP will also be high. In this case, when the available bitrate in the backhaul is at its lowest value, $C_3 = 6$ Mbps, the probability that the number of active users connected to the SAP (either SUs or MUs) is above $C_3/r_m = 6$ would be approximately one, i.e., $P(x > 6) \approx 1$. This explains why in Fig. 9, all curves except the one for $\kappa = 0$ saturate to a common value as λ/μ grows; and that this common value is approximately the probability that the available bitrate in the backhaul is $C_3 = 6$ Mbps.

VI. CONCLUSIONS

We proposed a traffic model for two-tier wireless networks composed of macrocells and femtocells. We evaluated the impact that the user traffic dynamics, the mobility of MUs, the scheme chosen to associate macrocell users to the small cells and the capacity constraint of the small cell backhaul have on the system performance. We modeled the interference produced by the MUs to the SUs by considering different geographical regions of interference. We also modeled the mobility of MUs between interference regions and the random load generated by the land-line users using finite-state CTMCs. To make the solution of the model computationally feasible, we exploited the time-scale decomposition approach.

We validated the analytical model by comparing the analytical results with results obtained by simulations. The results show that the association Scheme B achieves higher average throughput than Scheme A. We have also found that the Registration Policy that maximizes the average throughput achieved by the users connected to the SAP mainly depends on the backhaul state. When the backhaul is highly loaded, a closed access (i.e., $\kappa = 0$) should be implemented in order to maximize the throughput and minimize the flow blocking probability and the probability that SUs are served with low QoS. When the backhaul is not highly loaded, the best Registration Policy depends on the load offered by users and the movement pattern of the MUs, i.e. it depends on the interference produced by the MUs. When MUs tend to be far away from the SAP, the closed access mode ($\kappa = 0$) is better. When MUs tend to be close to the SAP and the load offered by users is high, a hybrid mode is better, where MUs close to the SAP are associated to it (i.e., $\kappa = 1$).

REFERENCES

[1] G. Mansfield, "Femtocells in the US market-business drivers and consumer proposition," FemtoCells Europe, ATT, U.K., Tech. Rep., Jun. 2008.
[2] V.Chandrasekhar and J.G. Andrews, "Femtocell networks: A survey," *IEEE Communications Magazine*, vol. 46, no. 9, pp. 59-67, Sep. 2008.
[3] H. Claussen, L. Ho. and L. Samuel, "An overview of the femtocell concept," *Bell Labs Technical Journal*, vol. 13, no. 1, pp. 221-245, 2008.
[4] I. F. Akyildiz, D. M. Gutierrez-Estevez, R. Balakrishnan, and E. Chavarria-Reyes, "LTE-Advanced and the evolution to beyond 4G (B4G) systems," *Physical Communications (Elsevier) Journal*, vol. 10, pp. 31 – 60, 2014.
[5] J. Andrews, H. Claussen, M. Dohler, S. Rangan, and M. Reed, "Femtocells: past, present, and future," *IEEE J. Sel. Areas Commun.*, vol. 30, no. 3, pp. 497–508, 2012.

[6] D. Lopez-Perez, 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, 2009.
[7] X. Y. Yu and H. B. Zhu, "An efficient method for loss performance modeling of hierarchical heterogeneous wireless networks," *International Journal of Communication Systems*, 2013.
[8] Y.-S. Liang, W.-H. Chung, G.-K. Ni, I.-Y. Chen, H. Zhang, and S.-Y. Kuo, "Resource allocation with interference avoidance in OFDMA femtocell networks," *IEEE Transactions on Vehicular Technology*, vol. 61, pp. 2243–2255, 2012.
[9] T. Zahir, K. Arshad, A. Nakata, and K. Moessner, "Interference management in femtocells," *IEEE Communications Surveys and Tutorials*, vol. 15, no. 1, pp. 293–311, 2013
[10] P. Palanisamy and S. Nimala, "Downlink interference management in femtocell networks- a comprehensive study and survey," *International Conference on Information Communication and Embedded Systems (ICICES)*, Chennai, 2013, pp. 747–754.
[11] 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, 2009.
[12] D.-C. Oh, H.-C. Lee, and Y.-H. Lee, "Power control and beamforming for femtocells in the presence of channel uncertainty," *IEEE Transactions on Vehicular Technology*, vol. 60, pp. 2545–2554, 2011.
[13] M. Iturralde, T. Yahiya, and A. Wei, a amd Beylot, "Interference mitigation by dynamic self-power control in femtocell scenarios in LTE networks," *IEEE GLOBECOM 2012*, Anaheim, CA, 2012, pp. 4810–4815.
[14] N. Saquib, E. Hossain, L. B. Le, and D. I. Kim, "Interference management in OFDMA femtocell networks: Issues and approaches," *IEEE Wireless Communications*, vol. 57, no. 10, pp. 3059–3068, 2009.
[15] V. Chandrasekhar and J. Andrews, "Spectrum allocation in tiered cellular networks," *IEEE Transactions on Communications*, vol. 57, no. 10, pp. 3059–3068, 2009.
[16] G. de la Roche, A. Valcarce, D. Lopez-Perez, and J. Zhang, "Access control mechanisms for femtocells," *IEEE Communications Magazine*, vol. 48, no. 1, pp. 33–39, 2010.
[17] J.-S. Lin and K.-T. Feng, "Femtocell access strategies in heterogeneous networks using a game theoretical framework," *IEEE Transactions on Wireless Communications*, vol. 13, pp. 1208–1221, 2014.
[18] R. Balakrishnan and B. Canberk, "Traffic-aware QoS provisioning and admission control in OFDMA hybrid small cells," *IEEE Transactions on Vehicular Technology*, vol. 63, pp. 802–810, 2014.
[19] X. Ge, T. Han, Y. Zhang, G. Mao, C.-X. Wang, J. Zhang, B. Yang, and S. Pan, "Spectrum and energy efficiency evaluation of two-tier femtocell networks with partially open channels," *IEEE Transactions on Vehicular Technology*, vol. 63, pp. 1306–1319, 2014
[20] E. Bernal-Mor, V. Pla, J. Martinez-Bauset, and D. Pacheco-Paramo, "A model of resource management in small cells with dynamic traffic and backhaul constraints," in Proceedings of the 19th European Wireless Conference (EW), Guildford, UK, pp. 1–6, 2013.
[21] S. Liu and J. Virtamo, "Performance analysis of wireless data systems with a finite population of mobile users," *Proceedings of the 19th International Teletraffic Congress (ITC19)*, 2005.
[22] G. Fodor and M. Telek, "Bounding the blocking probabilities in multirate CDMA networks supporting elastic services," *IEEE/ACM Transactions on Networking*, vol. 15, no. 4, pp. 944–956, 2007.
[23] Z. Avramova, D. D. Vleeschauwer, K. Laevens, S. Wittevrongel, and H. Bruneel, "Modelling H.264/AVC VBR video traffic: comparison of a Markov and a self-similar source model," *Telecommunication Systems*, vol. 39, no. 2, pp. 91–102, 2008.
[24] G. Casale, E. Zhang, and E. Smirni, "Kpc-toolbox: Simple yet effective trace fitting using markovian arrival processes," *5th International Conference on Quantitative Evaluation of Systems*, (QEST'08), Saint Malo, France, 2008, pp. 83–92.
[25] R. Nelson, *Probability, Stochastic Processes and Queueing Theory*. Springer-Verlag, 1995.
[26] E. Bernal-Mor, V. Pla, and J. Martinez-Bauset, "Handover performance for elastic flows in mobile cellular networks," *IEEE Communications Letters*, vol. 16, no. 10, pp. 1632–1635, 2012.
[27] E. Zola and F. Barcelo-Arroyo, "Impact of mobility models on the cell residence time in WLAN networks," *Proceedigs of the IEEE SARNOFF'09*, Mar. 2009, pp. 1–5.
[28] A. B. Downey, "Lognormal and pareto distributions in the internet," *Computer Communications*, vol. 28, no. 7, pp. 790–801, May 2005.
[29] G.-F. Zhao, Q. Shan, S. Xiao, and C. Xu, "Modeling web browsing on mobile internet," *IEEE Communications Letters*, vol. 15, no. 10, pp. 1081–1083, Oct. 2011.