Mobile Traffic Offloading in Heterogeneous Networks-Based Small Cell Technology

Saied M. Abd El-atty¹,³ · Z. M. Gharsseldien²,³

Received: 24 September 2014 / Accepted: 17 May 2015 / Published online: 4 June 2015
© King Fahd University of Petroleum & Minerals 2015

Abstract In this paper, we propose an efficient heterogeneous network (HetNet) architecture with coexisting small cell technology to provide capacity and coverage expanding in 3G and 4G mobile networks. A hierarchical HetNet layout comprises of three layers macro-, metro- and femtocell. The metrocell is employed as an intermediate layer in the integrated femtocellular/macrocellular network, which operates in a complementary fashion, in order to manage the handover traffic between the edge layers. Consequently, the femtocell serves indoor traffic activity of femto users, while metrocell serves the outdoor traffic activities as well as the overflow traffic from femtocells. The overall HetNet is completed with the macrocell overlay layer, which serves only the macro users and the overflowed traffic from the underlay layer. Then, we develop a realistic teletraffic framework in order to evaluate the performance of the proposed HetNet. We show both analytically and by simulation that the proposed HetNet architecture with coexisting small cells is able to offload traffic from traditional macrocellular network in terms of reducing the blocking probability.

Keywords HetNet · Small cell · QoS · Handover · Teletraffic

1 Introduction

Recently, mobile network operators (MNOs) have developed their infrastructure in order to reduce the deployment and maintenance costs. In addition, deploying traditional cell site in 3G and 4G mobile networks cannot solve the problems of holes “blind spots” coverage [1]. Heterogeneous network (HetNet) is an attractive means of expanding mobile network capacity and enhancing spectral efficiency. Compared to traditional macrocellular networks, a HetNet is more complex in terms of coverage and interference, making for challenging frequency planning, cell selection and handover [2].

A HetNet is typically composed of new small cell, access technologies, architectures, transmission solutions and base stations of varying transmission power [3,4]. As a consequence, MNO should perform their cell planning by using small cell technologies. Femto- and metrocells (small cells) can be used in public or open access areas to provide MNOs a low-cost option to quickly expand network capacity and coverage [5]. Small cells are low-power and short-range access point aiming at improving coverage and capacity. Femtocell is deployed in indoor environment such as indoor residential, small office home office (SOHO) and enterprise [6]. However, metrocell can be deployed in both indoor and outdoor environment such as metropolitan, hot spot and rural to dense urban areas. Meanwhile, metrocell base station (mBS) is easily mounted on walls, lampposts, poles or even the side of a building. Therefore, small cells bring the base station closer to the subscriber, improving air interface quality. Subsequently, small cells can also offload traffic from the
macrocellular networks and improve quality of experience (QoE) of subscribers [7].

Due to the development of small cell technology in the field of mobile communications networks, many works that are interested in the process of handover (HO) between cells are performed. A priority-based dynamic channel reservation scheme and admission control for handover in wireless networks have proposed in [8]. This scheme gives channel reservation for higher priority calls that does not reduce channel utilization significantly. Further, the authors in [9] have presented an efficient radio resource management (RMM) scheme for minimizing traffic in broadband wireless networks, wherein they employed a technique that limits the interference level. On particular relevance to our work are the studies in [10,11]. In [10], the author proposed an efficient resource allocation in hierarchical macrocell and femtocell network-based complete closed access (CCA) method. The numerical results indicated that an appropriate spectrum splitting strategy is significantly able to increase the service availability in the macrocell with femtocell QoS guarantee according to the accommodated indoors traffic in the femtocells. On the other hand, the authors in [11] proposed a call admission control (CAC) policy for traffic modeling in an integrated macrocell/femtocell network. The CAC scheme classifies the HO categories between macrocell and femtocell. The simulation results indicated that the proposed CAC is able to accept huge amount of handover calls as well as new calls in the system. The improvement of the overall forced call termination performance increases the revenue for the mobile operators.

Unlike the above literature, the proposed HetNet consists of a stack of multiple cell layers wherein the upper layer is the macrocell layer, while under this layer, a number of lower small cell layers are formed. The small cells can be femtocells, picocells, relays, metrocells and remote radio heads. However, as all these kinds of small cells are based on similar standards, software and interfaces, we focus on femtocells and metrocells since they can be designed to have distinct functionalities. A femtocell layer serves the indoor traffic activity of femtocell users, while the metrocell serves the outdoor traffic activities as well as the overflow traffic from femtocells. The overall HetNet is completed with the macrocell overlay layer, which serves only the macro users and the overflowed traffic from the underlay layer. Moreover, the metrocell layer is deployed as a complementary layer between the macrocell and femtocell layers and facilitates the handover traffic interaction between the edge layers. Meanwhile, the mobility management in this architecture is critical, and hence, the interaction between successive network layers, due to the HO traffic, is analyzed. Therefore, we develop a realistic teletraffic analytical framework based on queuing theory, and Markov process allows us to evaluate and quantify the performance of the proposed HetNet.

The rest of the paper is organized as follows. The proposed system model is described in Sect. 2. Then, a teletraffic framework for HetNet taking into account the interplays between femto-, metro- and macrocellular layer is developed in Sect. 3. The performance evaluation of the proposed HetNet and numerical results are presented in Sect. 4. Finally, the paper is concluded at Sect. 5.

2 System Model

In this section, we describe the system we have analyzed. The first subsection illuminates the layout of HetNet. The next subsection introduces the spectrum and channel allocation based on splitting strategy. The last subsection provides an illustrative example for describing the HetNet with multilayers architecture.

2.1 Moving Toward HetNet Layout

The challenges of coverage in 3G and 4G such as high-speed packet access (HSPA) and long-term evolution (LTE) multi-standard macrocellular networks are still not solved. Unfortunately, most subscribers are near the cell edge, not near the cell site due to public safety and health care. Therefore, MNOs develop their infrastructures and deploy small cell technologies such as picocell, metroc, and femtocell at the coverage’s holes in 3G and 4G mobile networks. As a result, the MNOs migrate to HetNet layout to improve and densify their existing mobile network infrastructure. HetNets are capable of meeting the expected surging traffic demand from the new services such as Internet applications and cloud-based services including video and other bandwidth-intensive contents due to modern terminal capabilities such as tablets and smart phone [12]. The HetNet makes use of those new technologies to overcome the coverage problems as illustrated in Fig. 1. In sequel, HetNet is able to dynamically leverage the existing macrocellular network topology and increase the proximity between the access network and the subscribers [7]. Therefore, the subscribers who are close to the access point perceived high-quality services and high data rate, while the subscribers further away the access point perceived adequate coverage and service. Therefore, HetNet efficiently enhances performance in all area of the networks through suited coordinated operation between the traditional macrocellular network and the small cell layers [13].

2.2 Spectrum Management and Channel Allocation

The integration of small cells within existing macrocell requires a careful design and optimization to manage interference in HetNets [5]. The MNOs should properly split their limited spectrum between the traditional macrocellu-
lar network and its underlying small cells. The spectrum allocated to HetNet may be used as shared or dedicated spectrum [14]. In shared spectrum, MNO share the same carriers assigned for macrocell to small cells and the interference should be managed. However, in dedicated spectrum, small cells use a dedicated (separate) carrier to avoid interference. We employ splitting spectrum policy in our system model [15]. Hence, the interference among different network layer can be avoided.

As illustrated in Fig. 2, according to the spectrum splitting policy, we consider the following channel allocation. Let $C$ denote the total available number of traffic channels owned by a mobile network operator (MNO) to HetNet. According to spectrum splitting strategy, these channels are partitioned into three channels: $C_M$ channels are allocated to macrocell, $C_m$ channels are allocated to metrocell and $C_f$ channels are assigned to femtocell. Additionally, each (femto or metro or macro) cell reuses the same channels (i.e., reuse frequency factor is unity). A call connection request is rejected as the target (femto or metro or macro) cell has no free channel upon the call arrival. Otherwise, the call request is accepted. Subsequently, in femtocell, new and HO request have the right to occupy any of the $C_f$ channels as long as the channel is idle, while the overflowed requests of femto users (new or HO) can access a higher (metrocell) layer network for possible service. The call blocking is determined by Engset formula due to a finite number of femto users ($m_f$). Similarly, the metro user have the right to occupy any of the $C_m$ channels in metrocell as long as the channel is free and the overflowed requests can access the overlay (macrocell) layer network. In macrocell layer, macro user’s requests and the overflowed requests from underlying layers can access any channel from $C_M$ when it is idle and based on Erlang loss discipline.

2.3 HetNet-based Small Cells

Figure 3 portrays an illustrative example of HetNet architecture-based small cells in a metropolitan. We consider the macrocell base station (MBS) in the center of this area. The femtocell access points (FAPs) are installed inside buildings and serve the indoor traffic originated by home users, while the metrocell base stations (mBSs) are installed outside of the buildings (e.g., mounted on lampposts, on the sides of buildings or even on transport hubs) serving the outdoor users. Femto users can move from the femtocell area to the metrocell area, thus performing a vertical handover, while subsequently another handover maybe required if they move to an area that is covered only by the macrocell. The access to femtocell and metrocell is available to finite number of femtocell subscribers ($m_f$) based on closed group subscribers (CGS) access method [16]. In our approach, the femtocell serves the indoor traffic activity of femto users, while the metrocell serves the outdoor traffic activity as well as the overflow traffic from femtocells. The highest layer in this
heterogeneous network is the macrocell layer, which serves only macro users and the overflowed traffic from the underlying metrocels layer.

3 Teletraffic Framework for HetNet

In this section, we present a teletraffic framework for the proposed HetNet. Let $N_f$ number of femtocells are overlaid by a metrocell and every $N_m$ metrocels are overlaid by a macrocell. We assume a circle shape for femto-, metro- and macrocells. $R_f$, $R_m$ and $R_M$ represent the radii of the femto-, metro- and macrocell, respectively. We consider the overall originating traffic to the HetNet follows Poisson process with arrival rate $\lambda_T$. We study the traffic process in each layer, and at the same time, we take into account traffic interplay between each layer. Further, it considerably reduces interference issues which is of fundamental importance for a layered network where users can handover from a small cell (femto- or metrocell) to the overlaid macrocell and vice versa, or from one small cell to another. To facilitate our analysis, according to the spectrum splitting policy, we assume that overlapping of cells between HetNets layers does not exist and thus inter-cell or cross interference is neglected. Table 1 summarizes the notations used in the model analysis.

3.1 Femtocell Layer

The femto user equipment (FUE) is able to access femto access point (FAP) with completely closed access (CCA) mode. We consider a finite population of femto users ($m_f$) in femtocell. An FAP serves indoor call activities that originated from femto users. Let $q$ represent the fraction of indoor activity of FUEs and $\lambda_u$ denote the traffic arrival rate per FUE. Then, $q\lambda_u$ call requests arrive in an indoors and will access its subscribed FAP. On the other hand, $(1-q)\lambda_u$ call requests arrive in outdoor environment and allow to access metrocell base station (mBS). In addition, according to the splitting spectrum policy, the number of channels assigned to FAP is $C_f$, new and HO requests have the right to occupy any of the $C_f$ channels as long as the channel is idle. Subsequently, there are two categories of traffic processes to femtocell: the originating (new) traffic in femtocell with rate $\lambda_f,n$ and the traffic rate originated from FUEs in metrocels and handed over to the subscribed femtocell with rate $\lambda_f,h$. Therefore, the total incoming traffic rate, $\lambda_f$ to a femtocell is given

$$\lambda_f = \lambda_f,n + \lambda_f,h \quad (1)$$

Due to the finite number of femto users, the new traffic that is originating in the femtocell has a rate

$$\lambda_f,n = m_f q \lambda_u \quad (2)$$

In order to develop the rate $\lambda_f,h$, we should firstly sort the activity of the FUEs in metrocell. Then, we can estimate the HO rate to the femtocell according to the activity of FUEs in metrocell. There are two categories of traffic rates due to the activity of femto users in metrocell:

1. For an accepted call (new or HO) in femtocell, it may not complete its service within femtocell and move out the coverage of FAP; it will request HO to overlay metrocell.
2. The FUEs handover to metrocell, i.e., an accepted FUE in metrocell; it may move out the coverage of metrocell.

Accordingly, we have

$$\lambda_f,h = \left[ (\lambda_f,n + \lambda_f,h)(1 - P_f) P_h, fm + \frac{\ell_f,h}{N_f} \right] \times (1 - P_m) P_h, mf \quad (3)$$

where $P_h, fm$ and $P_h, mf$ are the handover probabilities from femto- to metrocell and from metro- to femtocell, respectively. $P_f$ denotes the blocking probability of (new or HO)
### Table 1 System notations and notifications

<table>
<thead>
<tr>
<th>Notations</th>
<th>Physical meaning</th>
</tr>
</thead>
<tbody>
<tr>
<td>$m_f$</td>
<td>Number of femto users in femtocell</td>
</tr>
<tr>
<td>$N_f$ and $N_m$</td>
<td>Number of femtocells and number of metrocells overlaid in metrocell and macrocell respectively</td>
</tr>
<tr>
<td>$R_f$, $R_m$ and $R_M$</td>
<td>Radius of a femtocell, metrocell and macrocell respectively</td>
</tr>
<tr>
<td>$C$</td>
<td>Spectrum owned by a MNO to HetNet (femto &amp;metro/macrocell)</td>
</tr>
<tr>
<td>$C_f$, $C_m$ and $C_M$</td>
<td>Traffic channels assigned to the femto, metro and macrocell respectively</td>
</tr>
<tr>
<td>$\lambda_T$</td>
<td>Overall originating traffic in HetNet</td>
</tr>
<tr>
<td>$q$</td>
<td>Fraction of indoor activity of femto user equipment (FUE)</td>
</tr>
<tr>
<td>$\lambda_u$</td>
<td>Traffic arrival rate per FUE</td>
</tr>
<tr>
<td>$1/\mu_C$</td>
<td>Mean call duration time</td>
</tr>
<tr>
<td>$1/\eta_f$</td>
<td>Mean residing time of FUE in serving femtocell</td>
</tr>
<tr>
<td>$P_{h,fm}$</td>
<td>Handover probability from femtocell to metrocell</td>
</tr>
<tr>
<td>$P_{h,mf}$</td>
<td>Handover probability from metrocell to femtocell</td>
</tr>
<tr>
<td>$P_{nc}$</td>
<td>Probability that a call is not completed in femtocell</td>
</tr>
<tr>
<td>$1/\eta_m$</td>
<td>Residing time of the FUE in the serving metrocell</td>
</tr>
<tr>
<td>$1/\eta_M$</td>
<td>Residing time of the FUE in the serving macrocell</td>
</tr>
<tr>
<td>$\lambda_{f,n}$</td>
<td>Originating (new) traffic in femtocell</td>
</tr>
<tr>
<td>$\lambda_{f,h}$</td>
<td>Rate of HO traffic to the femtocell due to FUEs originating traffics in the metrocell</td>
</tr>
<tr>
<td>$P_l$ and $P_m$</td>
<td>New blocking probability and blocking probability of FUE in femtocell and metrocell respectively</td>
</tr>
<tr>
<td>$P_M$</td>
<td>Blocking probability in macrocell</td>
</tr>
<tr>
<td>$T_{hf}$</td>
<td>Average channel holding time in femtocell</td>
</tr>
<tr>
<td>$\lambda_{M,n}$</td>
<td>New traffic from the macro user equipment (MUE)</td>
</tr>
<tr>
<td>$\lambda_{M,h}$</td>
<td>Rate of HO traffic from neighboring macrocells due to MUE</td>
</tr>
<tr>
<td>$\ell_{m,n}$</td>
<td>New traffic from metro users (i.e., the originating traffic from outdoor femto users)</td>
</tr>
<tr>
<td>$\ell_{f,overflow}$</td>
<td>Rate of FUEs traffic overflowing from the underlying $N_f$ femtocells</td>
</tr>
<tr>
<td>$\lambda_{m,overflow}$</td>
<td>Rate of FUEs traffic overflowing from the underlying $N_m$ metrocells</td>
</tr>
<tr>
<td>$\ell_{m,h}$</td>
<td>HO traffic from the neighboring metrocells due to the metro users</td>
</tr>
<tr>
<td>$P_{h,mn}$ and $P_{h,MM}$</td>
<td>Handover probability from metrocell to neighboring metrocells and handover probability from macrocell to neighboring macrocells</td>
</tr>
</tbody>
</table>

femto users in femtocell. $\ell_{f,h}$ and $P_m$ represent the total HO rate of femto users to metrocell and the blocking probability in metrocell, respectively. $\ell_{f,h}$ and $P_m$ are developed in the metrocell as shown latter. Also, the division over $N_f$ regarding the traffic distribution among $N_f$ femtocells overlaid in metrocell. Unlike independent statistical for calculating the traffic rate between neighboring cells. Here, as illustrated in Eq. (3), we can see the interaction between the traffic in femtocell and metrocell by determining $\ell_{f,h}$ and $P_m$ and the division over $N_f$ regarding to the traffic distribution among $N_f$ femtocells overlaid in metrocell as illustrated in Eq. (3).

Let the random variable $T_C$ denote the call duration time of femto user, which refers to the time duration of the requested session or service connection. The call duration has an exponential distribution with mean $1/\mu_C$. The time $T_f$ spent by FUE in a femtocell from the onset of a successfully established connection till it crosses the femtocell boundary is called the residing time of FUE in the serving FAP. We assume that $T_f$ is negative exponentially distributed random variable with mean $1/\eta_f$. Therefore, the handover probability from femtocell to metrocell (hand-out), $P_{h,fm}$, is equal to the probability residence time of FUE in femtocell which is smaller than the call duration time ($T_C$). Hence, $P_{h,fm}$ is given by [17],

$$P_{h,fm} = \frac{\eta_f}{[\eta_f + \mu_C]}$$

Due to the feature of the femtocell and metrocell networks, the handover process from metrocell to femtocell is a challenging issue and should be investigated in detail, then the handover probability $P_{h,mf}$ is not straightforward as $P_{h,fm}$.
The handover probability $P_{h,mf}$ is equal to the probability that an accepted FUE in metrocell passes its subscribed femtocell and the session is not completed upon the instant of passing [10]. The ratio of femtocell area to the metrocell area accounts for the probability that the FUE crosses the femtocell. Without loss of generality, let $P_{nc}$ is the probability that a call is not completed when a FUE comes to the border of its subscribed femtocell. Hence, the probability $P_{h,mf}$ can be expressed as in [10]:

$$P_{h,mf} = \frac{\text{Area of femtocell}}{\text{Area of metrocell}} \times P_{nc} = \left(\frac{R_f}{R_m}\right)^2 \times P_{nc} \quad (5)$$

We derive an accurate expression for $P_{nc}$ comparable to the expression of $P_{nc}$ in [10]. $P_{nc}$ is derived as illustrated in “Appendix 1” as follows:

$$P_{nc} = \ln\left(\frac{1 + \mu C/\eta_m}{\mu C/\eta_m}\right) \quad (6)$$

where $T_m$ is a random variable exponentially distributed with mean $1/\eta_m$. $1/\eta_m$ is the average residing time of the FUE in the serving mBS. Then, we can obtain the handover probability from metrocell to femtocell (hand-in), $P_{h,mf}$, as

$$P_{h,mf} = \left(\frac{R_f}{R_m}\right)^2 \left[\ln\left(\frac{1 + \mu C/\eta_m}{\mu C/\eta_m}\right)\right] \quad (7)$$

According to the above assumptions, the average channel holding time of an originated call in femtocell ($T_{hf}$) and the average channel holding time of a HO call in femtocell ($T_{hf}$) are given by

$$T_{hf} = \frac{1}{\eta_f + \mu C} \quad \text{and} \quad T_{hf} = \left(\frac{R_f}{R_m}\right)^2 \left[\ln\left(\frac{1 + \mu C/\eta_m}{\mu C}\right)\right] \quad (8)$$

Let us define

$$\xi_f = \frac{\lambda_{f,n}(1 - P_f)}{\lambda_{f,n}(1 - P_f) + \lambda_{f,h}(1 - P_f)} \quad (9)$$

Therefore, the average channel holding time $T_{hf}$ in femtocell can be obtained as

$$T_{hf} = \frac{1}{\mu_{hf}} = \xi_f T_{hf} + [1 - \xi_f]T_{hf} \quad (10)$$

The femtocell incorporates a finite number of femto users ($m_f$). Therefore, we consider a more realistic approach in order to determine the blocking probability, $P_{b}$, in the femtocell by using Engset formula. Clearly, the arrival rate to FAP will not be constant as in Poisson random arrivals, but it will depend directly on the number of idle femto users (i.e., quasi-random). Therefore, according to a birth–death process and queuing theory, the arrival and departure rate of FAP at state $j$ are illustrated in Fig. 4, and thus, blocking probability for new or HO calls in femtocell, $P_{b}$, is given by

$$P_{b} = \frac{\left(\frac{m_f}{M_f} - 1\right)}{\sum_{i=0}^{C_f} \left(\frac{m_f}{M_f}\right)^i} \quad (11)$$

3.2 Metrocell Layer

The main function of metrocell layer is to complement the service of femto users when they are moving out the coverage area of their FAP. It is an intermediate layer between femtocell and macrocell layer, which operates in a complementary fashion, in order to manage the HO traffic between the edge layers. In the remainder of this paper, the terms femto user and metro user are used interchangeably. We can conclude the categories of traffic processes in metrocell, as follows.

1. The new traffic from metro users (i.e., the originating traffic from outdoor femto users) with rate $\ell_{m,n}$.
2. The traffic rate $\ell_{m,h}$, in metrocell consists of:
   (i) For an accepted call (new or HO) in metrocell, it may not complete its service within metrocell and move out the coverage of mBS, it will request HO to overlay macrocell.
   (ii) The metro user handover to macrocell, i.e., an accepted metro user in macrocell; it may move out the coverage of macrocell.
3. The femto user traffic overflowing from the underlying $N_f$ femtocells with rate $\ell_{f,\text{overflow}}$, and
4. The femto user handover to macrocell, for an accepted metro user in the macrocell, it may move out the coverage of macrocell, $\ell_{f,h}$.

Therefore, the overall traffic rate to metrocell is given by

$$\ell_m = \ell_{m,n} + \ell_{m,h} + \ell_{f,\text{overflow}} + \ell_{f,h} \quad (12)$$
Metro users are able to access all channels in mBS at any time without any differentiation between new and HO calls. The new traffic arrived from metro users have rate

$$\ell_{m,n} = m_l (1-q) \lambda_a \gamma$$  \hspace{1cm} (13)

where $\gamma$ is the fraction of femto users located inside metro’s coverage range, and $(1-\gamma)$ is the fraction of femto users located outside metrocell coverage [17]. Then, the HO rate for metro users can be obtained as follows:

$$\ell_{m,h} = (\ell_{m,n} + \ell_{m,h})(1 - PM) P_{h,m}\frac{A_{m,h}}{N_m} \times (1 - PM) P_{h,Mm}$$  \hspace{1cm} (14)

where $PM$ is the blocking probability in macrocell. $P_{h,mM}$ denotes the handover probability from metrocell to macrocell and is given by [17]

$$P_{h,mM} = \frac{\eta_m}{\eta_m + \mu c}$$  \hspace{1cm} (15)

Like the derivation of $P_{h,mF}$, we can derive the handover probability from macrocell to metrocell, $P_{h,Mm}$, as follows:

$$P_{h,Mm} = \left(\frac{R_m}{R_M}\right)^2 \left[\ln \left(1 + \frac{\mu c/\eta_m}{\mu c/\eta_m}\right)\right]$$  \hspace{1cm} (16)

where $1/\eta_m$ is the average residing time of metro user in the serving MBS.

For an accepted femto user in the femtocell, it may not complete its session or service within the femtocell and move out the coverage of femtocell. In this case, it will request handover to the metrocell to avoid service termination [20]. The overflowing traffic from all femtocells is given by

$$\ell_{f,overflow} = (\lambda_{f,n} + \lambda_{f,h})(1 - P_l) P_{h,fm} \times N_f$$  \hspace{1cm} (17)

In metrocell, the total outgoing rate is equal to the incoming rate to metrocell due to femto users. Hence, the HO traffic rate of femto users in metrocell is given by

$$\ell_{f,h} = (\ell_{f,overflow} + \ell_{f,h})(1 - P_m) P_{h,mm}$$  \hspace{1cm} (18)

where $P_{h,mm}$ denotes the handover probability from femtocell to neighboring metrocell, $P_{h,mm}$ is given by [17].

$$P_{h,mm} = \frac{\eta_{mm}}{\eta_{mm} + \mu c}$$  \hspace{1cm} (19)

where $1/\eta_{mm}$ is the average residing time of the metro users in the handover mBS.

In metrocell, for an accepted session of FUE, it may release the channel due to handover to its subscribed FAP, successful session completion or leaving out of the metrocell. We denote the average channel holding time in metrocell as $T_{hm}$. The average channel holding time is equal to $\min[X, T_C]$ if the FUE does not pass its own femtocell or FUE passes its own femtocell, but is denied to be admitted due to unavailable resources. The average channel holding time is equal to $\min[Y, T_C]$ if FUE passes its own femtocell and is accepted into femtocell. $X$ and $Y$ are negative exponentially distributed random variables. Hence, the channel holding time ($T_{hm}$) can be expressed as follows:

$$T_{hm} = \begin{cases} \min[X, T_C], & \text{with probability } \frac{R_m}{R_M} P_l + \left(1 - \left(\frac{R_m}{R_M}\right)^2 \right) \\ \min[Y, T_C], & \text{with probability } \frac{R_m}{R_M}^2 (1 - P_l) \end{cases}$$  \hspace{1cm} (20)

In sequel, the average value of $T_{hm}$ is $E[T_{hm}]$, calculated as

$$E[T_{hm}] = \frac{1}{\eta_m + \mu c} \left[\left(\frac{R_m}{R_M}\right)^2 P_l + \left(1 - \left(\frac{R_m}{R_M}\right)^2 \right)\right] + \left(\frac{R_m}{R_M}\right)^2 (1 - P_l) \cdot E[Z]$$  \hspace{1cm} (21)

where $E[Z] = E[\min(Y, T_C)]$ is derived in “Appendix 2” as

$$E[Z] = \frac{1}{\mu c} - \frac{1}{\mu c} \left[\ln(1 + \mu c/\eta_m)\right]$$  \hspace{1cm} (22)

The call blocking probability in metrocell can be obtained by Engset formula due to the finite number of femto users; hence, Analogous to $P_l$ at femtocell, the blocking probability $P_m$ in metrocell is given by,

$$P_m = \frac{m_l - 1}{C_m} \left(\frac{\phi_m + \phi_l}{\mu M}\right)^{C_m}$$  \hspace{1cm} (23)

where $\phi_m$ and $\phi_l$ represent the traffic load per user in the metrocell due to metro users and femto users, respectively. $\phi_m$ and $\phi_l$ are given by

$$\phi_m = \left[\frac{\ell_{m,n}}{\mu c + \eta_m} + \frac{\ell_{m,h}}{\mu c + \eta_{mm}}\right]$$

$$\phi_l = (\ell_{f,h} + \ell_{f,overflow}) \cdot E[T_{hm}]$$  \hspace{1cm} (24)

3.3 Macrccell Layer

The macrocell layer is considered the overlay layer of the microcells. Therefore, overflowing traffic from microcells are served by macrocell base station (MBS). In addition, MBS serves the originating traffic from macro users and the
handoff traffic from neighboring macrocells. Hence, there are five categories of traffic processes in macrocell as follows.

1. The new traffic from the macro users with rate $\Lambda_{M,n}$.
2. The HO traffic from neighboring macrocells due to macro users with arrival rate $\Lambda_{M,h}$.
3. The originating traffic from the metrocells due to outdoor activity with rate $\Lambda_{m,n}$.
4. The traffic overflowing from the underlying $N_m$ metrocells with rate $\Lambda_{m,overflow}$ and
5. The HO traffic from the neighboring macrocells due to the metro users with rate $\Lambda_{m,h}$.

Therefore, the overall traffic rate to macrocell is given by

$$\Lambda_M = \Lambda_{M,n} + \Lambda_{M,h} + \Lambda_{m,n} + \Lambda_{m,overflow} + \Lambda_{m,h}$$  \hspace{1cm} (25)$$

Macro users are able to access all channels in the MBS at any time without any differentiation between new and HO requests. As we have previously mentioned, the total traffic rate to the HetNet is $\lambda_T$; hence, the new traffic arrived from macro users have rate

$$\Lambda_{M,n} = \lambda_T - (\Lambda_{m,n} + N_m \ell_{m,n} + N_i \Lambda_f \Lambda_{f,n})$$  \hspace{1cm} (26)$$

The HO rate for macro users can be obtained by equilibrium state as follows [20]:

$$\Lambda_{M,h} = \frac{\Lambda_{M,n}(1 - P_M) P_{h,MM}}{1 - (1 - P_M) P_{h,MM}}$$ \hspace{1cm} (27)$$

where $P_{h,MM}$ denotes the handover probability from macrocell to neighboring macrocell for the originating and HO requests, respectively. $P_{h,MM}$ is given by

$$P_{h,MM} = \frac{\eta_{MM}}{\eta_{MM} + \eta_C}$$ \hspace{1cm} (28)$$

where $1/\eta_{MM}$ is the average residing time of the macro users in the handover MBS.

The total rate of femto users due to outdoor activities in macrocell is given by

$$\Lambda_{m,n} = m_t \lambda_u (1 - q) (1 - \gamma) N_m$$ \hspace{1cm} (29)$$

For an accepted metro user in metrocell, it may not complete its session or service within the metrocell and move out the coverage of metrocell. In this case, it will request handover to the overlaying macrocell to avoid service termination. The overflowing traffic from all metrocells (new or HO) is given by

$$\Lambda_{m,overflow} = (\ell_{m,n} + \ell_{m,h})(1 - P_M) P_{h,mM} \times N_m$$ \hspace{1cm} (30)$$

In macrocell, the total outgoing rate is equal to the incoming rate to macrocell due to metro users. Hence, the HO traffic rate of metro users is given by

$$\Lambda_{m,h} = (\Lambda_{m,n} + \Lambda_{m,h} + \Lambda_{m,overflow})(1 - P_M) P_{h,MM}$$ \hspace{1cm} (31)$$

The average channel holding time, $E[T_{HM}]$, in macrocell can be obtained as discussed in metrocell layer as follows

$$E[T_{HM}] = \frac{1}{\eta_M + \mu_C} \left[ \left( \frac{R_m}{R_M} \right)^2 P_m + \left( 1 - \left( \frac{R_m}{R_M} \right)^2 \right) \right]$$

$$+ \left( \frac{R_m}{R_M} \right)^2 (1 - P_M) \cdot E[W]$$ \hspace{1cm} (32)$$

where $E[W]$ is given by

$$E[W] = \frac{1}{\mu_C} - \frac{1}{\mu_C} \left[ \ln(1 + \mu_C / \eta_M) \right]$$ \hspace{1cm} (33)$$

The blocking probability $P_M$ in macrocell is given using Erlang formula, as

$$P_M = \frac{(\rho_M + \rho_m)^{CM}}{\sum_{j=0}^{CM} (\rho_M + \rho_m)^j / j!}$$ \hspace{1cm} (34)$$

where $\rho_M$ and $\rho_m$ are given by

$$\rho_M = \left[ \frac{\Lambda_{M,n}}{\mu_C + \eta_M} + \frac{\Lambda_{M,h}}{\mu_C + \eta_{MM}} \right]$$

$$\rho_m = (\Lambda_{m,n} + \Lambda_{m,h} + \Lambda_{m,overflow}) \cdot E[T_{HM}]$$ \hspace{1cm} (35)$$

### 4 HetNet Performance Analysis

In this section, we present both numerical and simulation results to evaluate the performance of the proposed HetNet. The radius of macro-, metro- and femtocell is considered 1 km, 200 and 20 m respectively. We consider five metrocells are overlaid in a macrocell, and every metrocell overlays five femtocells to construct the proposed HetNet. The total spectrum owned by a mobile network operator (MNO) for that HetNet is $C = 40$ channels. According to spectrum splitting strategy, $C$ is partitioned among macrocell, metrocell and femtocell. We consider the number of channels assigned to femtocell are only three channels and five channels for metrocell as suggested in [10,18].

The problem emanating from the set of nonlinear equations that are mutually dependent to each other was numerically solved. That was accomplished with the integration of a recursive algorithm and designated convergence criteria, employed to maintain system stability. The input parameters $C_t$, $C_m$, $C_f$, $\mu_C$, $\eta_t$, $q$, $m_t$ etc., are given and the initial conditions for $P_M$, $P_t$ and $P_m$ and the various HO rate
in the system model are set to be 0 initially. Then, the metrics of interest can be obtained. The incoming overall arrival rate, $\lambda_T$, is varied from 0.2 to 0.8 calls/s. However, the traffic rate per femto user per second, $\lambda_u$, is set to 0.002 and the indoor activity factor, $q$, is set to 0.6. The average call duration $1/\mu_C$ is 120 s. The average residence times for originating calls in the serving metro- and macrocell, $1/\eta_M$, and $1/\eta_M$, respectively, as well as the average residence times for HO calls in the handover metro- and macrocell, $1/\eta_{mm}$ and $1/\eta_{MM}$ respectively, are obtained as in [17]. Comparatively, the average residence time of femto users in femtocell is much longer than the macrocell residence time due to the low mobility of FUEs. We assume the speed of femto user in macrocell is $v = 20$ km/h. Some illustrative numerical examples are introduced to demonstrate the performance of our proposed HetNet. Moreover, the numerical results will illustrate the interaction between the performance metrics and critical setting. To summarize, Table 2 includes the default parameters for the proposed HetNet.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_M$</td>
<td>1000 m</td>
</tr>
<tr>
<td>$R_m$</td>
<td>200 m</td>
</tr>
<tr>
<td>$R_f$</td>
<td>20 m</td>
</tr>
<tr>
<td>$C$</td>
<td>40</td>
</tr>
<tr>
<td>$C_M$</td>
<td>32</td>
</tr>
<tr>
<td>$C_m$</td>
<td>5</td>
</tr>
<tr>
<td>$C_f$</td>
<td>3</td>
</tr>
<tr>
<td>$\lambda_T$</td>
<td>0.2,...,0.8</td>
</tr>
<tr>
<td>$\lambda_u$</td>
<td>0.002</td>
</tr>
<tr>
<td>$N_f$</td>
<td>5</td>
</tr>
<tr>
<td>$N_m$</td>
<td>5</td>
</tr>
<tr>
<td>$1/\mu_C$</td>
<td>120 s</td>
</tr>
<tr>
<td>$1/\eta_f$</td>
<td>$9/\mu_C$ s</td>
</tr>
<tr>
<td>$q$</td>
<td>0.6</td>
</tr>
<tr>
<td>$\gamma$</td>
<td>0.5</td>
</tr>
<tr>
<td>$m_f$</td>
<td>3, 4,...,10</td>
</tr>
</tbody>
</table>

4.1 Performance Comparisons

In this scenario, we compare the performance of the proposed HetNet (femto & metro/macro) with the performance of the traditional macrocellular network without deploying femtocell (w/o femto) and the performance of traditional macrocellular network with deploying femtocell (femto/macro). The channel assignment is based on fixed channel assignment discipline. In this scenario, we fix the total number of channels ($C$) that are owned by MNO to that network and according to spectrum splitting strategy, we consider the following different scenarios for HetNet architecture:

Scenario (1) macrocellular network only (w/o femto), $C = C_M = 40$ channels.

Scenario (2) femto/macro, $C = 40$, $C_f = C - C_M$, and we used $C_M = 37$ and thus $C_f = 3$

Scenario (3) femto and metro/macro, $C = 40$, $C_f = C - C_M - C_m$ and we used $C_M = 32$, $C_m = 5$, and hence $C_f = 3$.

Figure 5 shows the blocking probability, $P_M$, in macrocell versus the total arrival rate, and $\lambda_T$ in HetNet when the number of femto users is set to 6. As expected, the blocking probability increases by increasing the arrival traffic rate. The dashed line in the figure represents the blocking probability in traditional macrocellular network without femtocell deployment. Also, we can see that the performance of HetNet outperforms the performance of traditional macrocellular networks without and with femtocell deployment particularly at heavy traffic load. This is due to the fact that metrocell operates in a complementary fashion with femtocell for offloading traffic from hot spot areas in macrocell.

Figure 6 can also verify HetNet offloading capability. This plot shows the relationship between the blocking probabilities $P_f$ and $P_M$ when the number of indoor femto users increases at fixed value of $\lambda_T$ and $q = 0.6$. We can observe $P_f$ smoothly increases by increasing the number of indoor femto users in femtocell. However, $P_M$ significantly decreases when the number of indoor femto users in femtocell increases.

4.2 Effects of Indoor Activity

Figure 7 illustrates the blocking probability, $P_M$, in macrocell versus the total arrival rate, $\lambda_T$, with different value of indoor activity factors. As we can observe in the figure, the blocking probability in macrocell is obviously low at large indoor activity factor. The plot indicates that when the indoor...
activity factor is large, i.e., the traffic load in the femtocell is high, and blocking probability in macrocell is reduced. This is due to the fact that traffic transfer from macrocell to femtocells is major, resulting in decreasing the blocking probability in macrocell. Furthermore, Fig. 8 shows the effect of femto user’s speed into offloading capability. As we can see that when the speed of femto user is high, i.e., the probability of femto user entering the femtocell area is high, the blocking probability ($P_M$) in macrocell will be significantly reduced.

### 4.3 Effects of Number of Metrocells

Figure 9 shows the blocking probability, $P_M$, in macrocell versus the total arrival rate, $\lambda_T$, with different number of overlaid metrocells ($N_m$) in macrocell when the number of femtocells per metrocell is $N_f = 5$. The figure indicates that as the number of deployed metrocells is increased, the blocking probability in macrocell is decreased, i.e., the blocking probability performance in macrocell is improved when the number of overlaid metrocells in macrocell is increased. This is due to the fact that metrocells operate in a complementary fashion with femtocells to accommodate more number of calls, which are diverted from macrocell to small cells. However, as illustrated in Fig. 10, the utilization of spectrum in the HetNet will dramatically decrease, especially from low to medium load, as the number of employed metrocells is large.

Therefore, we can conclude that employing metrocell may have some pros and cons. The main pros of employing metrocell is offloading traffic from macrocellular network as well as reducing HO blocking probability of femtocell’s subscribers when they are roaming in HetNet. However, the cons of metrocell employing is increasing the complexity of interference and deteriorating the spectrum utilization when
it is shared spectrum with macrocell. For that reason, unlike femtocells, metrocells are usually owned and installed by the MNO themselves, who plan, manage and maintain them in the same way as their larger macrocell cousins. Additionally, interference management between HetNet’s base stations can be controlled by enhanced inter-cell interference coordination (eICIC) algorithm as in [19].

4.4 Effects of Number of Femto Users

Figure 11 shows the blocking probability, $P_M$, in macrocell versus total arrival rate, $\lambda_T$, with different number of femto users ($m_f$) in femtocell. We observe that as the number of femto users is increased, the blocking probability in macrocell is decreased. This is due to the capability of femtocell to accommodate more number of calls, which are transferred from macrocell to femtocells.

4.5 Blocking Probability in HetNet Layers

In this scenario, we demonstrate the blocking probability in each layer of the proposed HetNet. Figure 12 illustrates the blocking probabilities in macro-, metro- and femtocell when indoor activity of indoor users increases. As we can see, the blocking probabilities in macro- and metrocells are decreased, while the blocking probability in femtocell is increased by increasing the indoor activity. This is due to the fact that femtocell accommodates the indoor activities, and then, the blocking probability in femtocell increases.

To summarize, the plots indicate that the proposed HetNet architecture is able to offload traffic from macrocellular networks and then lead to much lower blocking probability in macrocellular networks.

5 Conclusions

In this paper, a new architecture of heterogeneous network (HetNet) with coexisting small cell technologies was presented. HetNet architecture consists of three multilayers (femto-, metro- and macrocell) to overcome the blind spots in coverage of 3G and 4G mobile networks. Then, a faithful analytical teletraffic framework was developed to evaluate the performance of the proposed HetNet. We analyzed the teletraffic framework based on queuing theory and Markov model. The numerical results demonstrated that the proposed heterogeneous cooperation framework is a cost-effective alternative to two-layer deployments consisting of macrocells and femtocells, meeting the growing capacity demands and offloading the macrocell traffic loads.
Appendix 1

In this appendix, we derive the probability $P_{nc}$, and let $X$ and $Y$ denote the time duration from an intermediate instant to the border of the metrocell and the time duration from the instant the session starts to the instant the femto user comes to the border of its femtocell. Due to the memoryless property, the random variable $X$ follows the same exponential distribution with mean $1/\eta_m$. However, due to the fact that FUEs and femtocell are randomly distributed within the metrocell, the probability density function (pdf) of $Y$ is given by

$$f_Y(t) = \int_0^{\infty} \frac{1}{x} (\eta_m e^{-\eta_m x}) dx = \eta_m \int_0^{\infty} e^{-t x} dx = \frac{1}{\eta_m} \Gamma(0, \eta_m t),$$

where $\Gamma(a, x)$ is an upper incomplete Gamma function which given by.

$$\Gamma(a, x) = \int_0^{\infty} y^{a-1} e^{-y} dy$$

Taking Laplace transform of (A.2), we have

$$f_\mathcal{L}(s) = \frac{\ln (1 + s/\eta_m)}{s/\eta_m}.$$  \hfill (A.4)

Then, the probability $P_{nc}$ can be expressed as

$$P_{nc} = \Pr \{ Y < T_C \} = f_\mathcal{L}(\mu_C) \quad \text{then}$$

$$P_{nc} = \frac{\ln (1 + \mu_C/\eta_m)}{\mu_C/\eta_m}$$ \hfill (A.5)

Appendix 2

In this appendix, we derive the mean value of $Z$, $E[Z]$. The cumulative distribution function of the random variable $Z$ is given by

$$F_Z(t) = 1 - \left[ 1 - F_Y(t) \right] \left[ 1 - F_{R_c}(t) \right]$$

$$= 1 - e^{-\mu_c t} + F_Y(t) e^{-\mu_c t}$$ \hfill (B.1)

The Laplace transform of the CDF $F_Z(t)$ is given by

$$\mathcal{L}\{ F_Z(s) \} = \frac{\mu_c}{s \cdot (\mu_c + s)} + F_\mathcal{L}(s + \mu_c)$$ \hfill (B.2)

The Laplace transform of the pdf of the random variable $Z$ is given by

$$f_Z(s) = s \cdot \mathcal{L}\{ F_Z(s) \} = \frac{\mu_c}{(\mu_c + s)} + s \cdot F_\mathcal{L}(s + \mu_c)$$ \hfill (B.3)

Then, the mean value of $Z$ is given by

$$E[Z] = - \frac{df_Z(s)}{ds} |_{s=0}$$

$$= \frac{1}{\mu_c} - F_\mathcal{L}(\mu_c) = \frac{1}{\mu_c} - \frac{1}{\mu_c} F_\mathcal{L}(\mu_c)$$

$$= \frac{1}{\mu_c} - \frac{1}{\mu_c} \left[ \ln (1 + \mu_c/\eta_m) \right]$$ \hfill (B.4)

References