Load-aware Handover Decision Algorithm in Next-generation HetNets

Konstantinos Alexandris, Nikolaos Sapountzis, Navid Nikaein, Thrasyvoulos Spyropoulos
Mobile Communications Department
EURECOM, Biot, France 06410
Email: firstname.lastname@eurecom.fr

Abstract—In this work we propose a novel handover (HO) algorithm, that considers system performance from both user and network perspective, in the context of heterogeneous networks (HetNets), i.e., networks composed of BSs with asymmetrical transmission power. In such an environment, conventional HO algorithms that consider only the user perspective, e.g., received signal strength (RSS)-based, might offer suboptimal performance, since they mainly push users to cells with high transmission powers. Thus, new algorithms that take into account also the network perspective, e.g., cell load, are needed. In this work, a load-aware algorithm is proposed considering the service delay that a user experiences from the network. In addition, an implementable framework based on Software Defined Networking (SDN) architecture is sketched to support the algorithm. The proposed algorithm is compared with the traditional one we implementable framework based on Software Defined Networking (SDN) architecture is sketched to support the algorithm. The proposed algorithm is compared with the traditional one we meet in long-term evolution (LTE) systems and a distance-based one. Extracted cell assignment probability and user service delay performance results show that the load-aware approach outperforms both of them.

I. INTRODUCTION

Handover (HO) decision algorithms in heterogeneous networks (HetNets) are emerging with the extreme densification and offloading of cellular systems as the key technologies to get 1000x data rate [1]. Many works have been done in HO decision algorithms considering the overall system performance from the user and network perspective, in the context of long-term evolution (LTE)/LTE–Advanced (LTE–A) networks [2]. Specifically, from the network perspective, a HO algorithm that adapts the hysteresis and the time-to-trigger (TTT) based on certain network key-metrics (e.g., HO failure or ping-pong ratios) is studied in [3], along with the self-organising networks (SONs). Advanced self-organizing map (SOM) is examined to suppress handovers in regions that coincide with many unnecessary HOs [4]; also the HO offset tuning is performed via network load-balancing for assuring handover to possible target cells and no return to their serving cell [5]. The main focus so far has been to enhance the system performance from the user-perspective. The most common used are based on: (i) received signal strength (RSS), (ii) user speed and (iii) interference-management [6], [7].

However, denser deployments in HetNets experience high spatio-temporal load variations and thus require more advanced HO algorithms that consider both perspectives, jointly. In addition, power based algorithms (e.g., RSS-based) proposed in the literature cannot be applied, due to asymmetrical transmission powers among macrocells and picocells (also called small cells (SCs) in the literature). Especially, in such environments, a user equipment (UE) usually remains connected to a macrocell that offers high transmission power (and is potentially overloaded due to its high number of concurrent users), while is placed close to one or more underloaded SCs. According to the latter, authors in [8] propose to scale down macrocell RSS with an appropriate factor to force connection to a SC. This factor is selected optimally based on the maximization of the SC assignment probability. Therefore, it was left as future work to examine policies that consider user throughput, delay or other user service requirements. To this end, we revisit the problem of HO, in the context of future HetNets deployments. Our contributions can be summarized as follows:

1) We construct a framework considering both user (i.e., RSS) and network (i.e., cell load) perspectives. Specifically, the service delay of a random flow is analytically derived, as a key QoS criterion for the HO decision algorithm under the assumption of uneven transmission power regimes. This approach overcomes the shortcomings created by only considering RSS criteria in HO decision for HetNets (Section II and III).

2) We propose a HO decision algorithm (Section III) and sketch an implementable framework based on Software Defined Networking (SDN) architecture, as suggested to be a key enabler for the realization of 5G networks [9] (Section IV).

3) We investigate the related trade-offs involved, in different load variation scenarios. Further, the proposed algorithm is compared with the (i) one used in traditional LTE systems and (ii) a distance-based algorithm proposed in [8] and significantly outperforms them (Section V).

II. SYSTEM MODEL

A. Signal model

This work considers a network consisting of a macrocell and a group of picocells located at given distances from it, where \( j \in \{1, \ldots, P\} \) and \( P \) is the total number of the picocells. The cells, i.e., macrocell/picocells, are denoted by \( m, p_j \), respectively. In addition, the picocells are assumed circular on average, regardless of the fading effects presence. Each base station (BS) is considered to be located in the center of each circular cell \( i \in \{m, p_j\} \) of a given radius \( R_i \) corresponding to the \( (x_i, y_i) \) coordinates. Let \( r_{m,i}[k] \) and \( r_{p_j}[k] \) denote the reference signal received power (RSRP) from the macrocell BS and the picocell BS at time \( k \) in dBm scale, respectively:

\[
r_{i}[k] = P_{\text{dBm}} + P_{\text{dB}}(d_i^k) + \psi_i[k], \quad i \in \{m, p_j\},
\]

where \( k \) corresponds to the discretization of the continuous time \( t \) sampling at \( \tau_s \) intervals, where \( \tau_s \) stands for the measurement sampling period.
where $P_{i,k}^{dB}$ denotes the BS transmission power, $P_{i,k}^{dB}(d_{i,k})$ denotes the path loss in dB, $d_{i,k}$ represents the distance from the BS to the user that is greater than a reference distance $d_r$ and $\psi_{i,k}^{dB}[k]$ stands for the shadowing fading in dB for each cell $i$. The pathloss $P_{i,k}^{dB}(d_{i,k}) \triangleq 10 \log_{10}(K) + 10 \eta \log_{10}(d_{i,k}/d_r)$, where $K$ is the path loss measured at reference distance $d_r$ and $\eta$ is the pathloss exponent. The $K$ parameter is given by $K = [c_0/(4\pi f_d d_r)]^2$, where $c_0$ is the universal speed of light in vacuum and $f_d$ is the carrier frequency. The shadow fading $\psi_{i,k}^{dB}[k] \sim N(0, \sigma_{dB,i}^2/\xi^2)$ and is assumed to be independent across $i$ and $k$, where $\xi \triangleq 10/\ln(10)$. Averaging is performed by an exponential moving average (EMA) filter, i.e., low-pass filter, for smoothing any RSRP abrupt variations. High frequency fluctuations (i.e., Rayleigh fading) are filtered out and can be neglected. Consequently, the output filtered signal is:

$$r_{i,k}^{dB}[k] \triangleq (1 - \alpha)r_{i,k}^{dB}[k-1] + \alpha \psi_{i,k}^{dB}[k],$$

where $\alpha \triangleq 2^{-\eta/2}$ and $q \in \mathbb{N}$.

### B. Mobility and Network Users

The users are distinguished in two different categories depending on their mobility status: a) **Static users (SU):** Users that are not moving on average and they don’t intend to handover. The Static users are divided into two subcategories based on their on-going traffic: i) **Active users (AU):** Users that have already associated with a BS and generate new flows. ii) **Disconnected users (DU):** Users that are considered to be switched-off. This implies no association with a specific BS. b) **Mobile users (MU):** Users that are moving and intend to handover. The users mobility can be described with any mobility model.

The SUs are getting active with probability $p_r$, in each cell $i$ at each time $k$. On that account, the number of the AUs in each cell $i$ at time $k$, $N_{AU,i}^k$, is distributed according to the binomial distribution with parameters $N_{SU,i}$ and $p_r$, where $N_{SU,i}$ stands for the total number of the static users in each cell $i$. The exact value of $N_{AU,i}^k$ is considered to be known to the BS at each time $k$. The MUs are always active and $N_{MU}$ stands for their total number.

### C. Traffic Model

New flows (all considered as best-effort) by active users are assumed to arrive according to a Poisson process with the total arrival rate in each cell $i$ at time $k$ [10]:

$$\lambda_{i,k} = \lambda (N_{MU,i}^k + N_{AU,i}^k),$$

where $\lambda$ denotes the flow arrival rate per user. The flow size follows a general distribution with mean $Y$. We assume the Processor Sharing (PS) scheduling discipline and adopt the stationary M/G/1/PS system [10]. $N_{MU,i}^k \triangleq \mathbb{1}(u \not\in i) + N_{MU,i}^0$, where $\mathbb{1}(\cdot)$ stands for the indicator function and becomes one when a mobile user $u$ that intends to handover is not associated with the cell $i$ (i.e., $u \not\in i$) at time $k$: $N_{MU,i}^0$ denotes the total number of mobile users that are associated with the cell $i$ at time $k$ (i.e., a subset of $N_{MU}$). $N_{MU,i}^k$ is also acquainted by each cell $i$ BS. If the mobile user $u$ is already associated with the cell $i$, then $1(\cdot)$ is zero and $u$ is already included in $N_{MU,i}^k$. Thus, after computing $N_{MU,i}^k$ and plug-inning it to Eq. (3), we get an estimation of the total arrival rate after the HO procedure. This is also used for the prediction of the service delay, as it will be discussed later in Section III.

### III. Handover Decision Algorithm

A well-known criterion, commonly used in conventional HO decision algorithms for mobile communication systems (e.g., 3GPP LTE), is based on RSRPs comparison method in which hysteresis and threshold are included [11]. Specifically, the criterion for the handover scenario is expressed as:

$$\bar{r}_{p,j}^{dB}[k] > \bar{r}_{m}^{dB}[k] + \Delta \land \bar{r}_{m}^{dB}[k] < r_{m,th}^{dB},$$

where $\Delta$ denotes the hysteresis and $r_{m,th}$ is the minimum RSRP macrocell threshold value.

There are scenarios where a macrocell is overloaded and it becomes crucial for the user to connect to a picocell that is probably underloaded. Due to significant differences in transmission power among a macrocell and picocells, there is a huge gap between their received powers at the UEs (i.e., $\bar{r}_{p,j}^{dB}[k] \gg \bar{r}_{m}^{dB}[k]$). Hence, the above criterion does not often hold and the conventional HO algorithm cannot be triggered to maintain the required QoS (e.g., in terms of rate, delay and throughput). Thus, load-aware algorithms development are crucial for uneven received powers regimes found in HetNet scenarios.

The proposed load-aware (LA) policy is described in Algorithm 1. The algorithm is applied only to the mobile users by taking into account also the static active ones. The disconnected users are not taken into account, since they are assumed to be switched-off. $\bar{r}_{i}[k]$ stands for the RSRP from each cell $i$ at time $k$ in linear scale. The picocell that the user intends to handover is the one with the maximum received power, denoted as $p_j$. After the selection of the $p_j$, picocell, the condition $\bar{r}_{p,j}[k] > r_{p,th}$, with $r_{p,th}$ standing for the picocell RSRP threshold, ensures that the user stays within the picocell coverage area and the received signal is retained adequately strong. Thus, the algorithm is running only inside the picocell. Subsequently, we define $\bar{r}_{j}[k] \triangleq \bar{r}_{p,j}[k] + \left[1 - f\left(\bar{D}_k, \bar{D}_{p,j}\right)\right] \bar{r}_{m}[k]$. $\bar{D}_k$ denotes

```
Algorithm 1 Handover algorithm

\textbf{Input:} $\bar{D}_k^m$: predicted average delay of the macrocell $m$
$\bar{D}_k^p_j$: predicted average delay of the picocell $p_j$
$r_{p,th}$: picocell RSRP threshold

\textbf{Output:} user cell association

$\hat{j} = \arg\max_j \bar{r}_{p,j},$
if $\bar{r}_{p,j}[k] > r_{p,th}$ then
\text{if} $\bar{r}_{j}[k] > \bar{r}_{m}[k]$ then
\text{connect to picocell}
\text{else}
\text{connect to macrocell}
end if
end if
```

2 The $x \sim N(\mu, \sigma^2)$ denotes that random variable $x$ is Gaussian with mean $\mu$ and variance $\sigma^2$. 
the prediction of the average (service) delay of a new flow of a cell \(i\) at time \(k\) that the user experiences when staying in the picocell. An analytical expression for \(\tilde{D}_i^k\) is given later in Eq. (7). It is noted that the prediction of the average delay \(\tilde{D}_i^k\) is performed at time \(k\), but it represents an estimation for the entire time period that the user is staying within the picocell (i.e., user’s delay assumed to be \(\tilde{D}_i^k\) during this time period). As it is mentioned later, the computations for \(\tilde{D}_i^k\) remain consistent with this assumption. \(f(\cdot)\) stands for a continuous function and is constructed properly to force the user to connect in the picocell \(p_j\), if \(D_{p_j}^k < D_m\), despite the fact that \(\bar{r}_m[k] \gg \bar{r}_{p_j}[k]\). This approach meets the user needs that cannot be supported by the congested macrocell and its association with a picocell is critical in terms of QoS. The condition that must hold to connect to the picocell is given analytically by the following inequality that applies the above policy:

\[
\bar{r}_j[k] > \bar{r}_m[k] \iff \bar{r}_{p_j}[k] > f\left(\tilde{D}_i^k, \tilde{D}_{p_j}^k\right) \bar{r}_m[k], \quad (5)
\]

otherwise the user is connected to the macrocell. Finally, the function \(f\left(\tilde{D}_{p_j}^k, \tilde{D}_m^k\right) \equiv \exp\left[-cD_{p_j}^k/D_m^k\right], \quad c \geq 1\) with a codomain in \([0,1]\) provides such a property as intended. Specifically, the exponential nature of the function provides quick convergence to zero when \(D_{p_j}^k < D_m^k\). This manages to scale down the received powers asymmetry in order to hold Eq. (5) and associate the user with the underloaded picocell. Finally, the constant \(c\) stands for a tuning parameter. Next, we describe the computations needed for the prediction of \(\tilde{D}_i^k\).

The \(\tilde{C}(d_{i,i}) \triangleq \left[\tilde{C}_L(d_{i,i}) + \tilde{C}_U(d_{i,i})\right]/2\) is defined as the arithmetic mean of the \(i^{th}\) static active user distance, \(d_{i,i}\), for each cell \(i\) (see Fig. 1). A detailed description of their analytical expression is given in Appendix A (see Eq. (9), (10)). The latter computations are consistent with the prediction of \(D_i^k\) for the entire user’s sojourn within the picocell, since \(d_{i,i}\) remains invariant for the static users across time \(k\). For the mobile users, there is no prior knowledge of their future positions, thus their corresponding lower and upper bounds of the averaged capacity as a function of the \(i^{th}\) static active user distance, \(d_{i,i}\), for each cell \(i\) (see Fig. 1). A detailed description of their analytical expression is given in Appendix A (see Eq. (9), (10)). The latter computations are consistent with the prediction of \(D_i^k\) for the entire user’s sojourn within the picocell, since \(d_{i,i}\) remains invariant for the static users across time \(k\). For the mobile users, there is no prior knowledge of their future positions, thus their corresponding lower and upper bounds of the averaged capacity for each cell \(i\), \(\tilde{C}_L^i, \tilde{C}_U^i\), are averaged over all the possible distances, assuming that the user passes uniformly through all the points within the picocell after multiple visits during its staying within it. The uniform distribution of the points as well as the corresponding capacity analysis are enclosed in Appendix B. Thus, the aforementioned computations

\[
\tilde{C}(d_{i,i}) \triangleq \left[\tilde{C}_L(d_{i,i}) + \tilde{C}_U(d_{i,i})\right]/2.
\]

are consistent also with the prediction of \(\tilde{D}_i^k\) for the total user’s sojourn within the picocell. Finally, their respective arithmetic mean is given by \(\bar{C}_i = \left(\bar{C}_L^i + \bar{C}_U^i\right)/2\).

According to the above computations, the expected average rate \(\bar{R}_i^k\) at time \(k\) for each cell \(i\) is determined by [12]:

\[
\bar{R}_i^k = \frac{\bar{N}_{\text{MU},i} \bar{C}_i + \sum_{\ell=1}^{N_{\text{AU},i}} \bar{C}(d_{\ell,i})}{\bar{N}_{\text{MU},i} + N_{\text{AU},i}^k}. \quad (6)
\]

Consequently, based on the system assumptions (i.e., stationary \(M/G/1/PS\)), the prediction of \(D_i^k\) is given by [10]:

\[
D_i^k = \frac{1}{\mu_i^k - \lambda_i^k}, \quad (7)
\]

where \(\mu_i^k = \bar{R}_i^k/\bar{Y}\) stands for the service rate at time \(k\) for each cell \(i\).
The MUs mobility model is based on a two-dimensional random walk (2D RW) limited in a finite space (i.e., circular macrocell) with step $s$ or $s/\delta$ depending on where the user is moving inside (i.e., macrocell-picocell) and $\delta > 0$ denotes a degrading step factor. This variable step simulates the scenario that the users are staying inside the picocell area for a sufficient amount of time compared to the macrocell one (e.g., in airports, malls, etc). As next, Fig. 3–4 present a scenario with three picocells placed in different distances from the macrocell ($D_1 = 250$ m, $D_2 = 350$ m, $D_3 = 450$ m) and one mobile user (i.e., $N_{MU} = 1$), as shown in Fig. 2.

Fig. 3 demonstrates the mobile user cell assignment probabilities, $P_r(u \in m)$, $P_{r}(u \in p)$ for the macrocell and the picocell, respectively. These mobile user cell assignment probabilities are presented as a function of the flow arrival rate $\lambda$ for different number of static users $N_{SU,m}$ = 200, 400, 600 and $N_{SU,p_j} = 10$. Their computation is performed via Monte-Carlo simulations within the picocells $p_j$. Especially, $P_r(u \in p)$ is also averaged out for all picocells $p_j$. Specifically, with the increment of $\lambda$, the total load in the macrocell augments, since the number of the macrocell static users is greater than the picocell ones, causing macrocell overload. Consequently, the difference in the corresponding delays increases (i.e., macrocell delay is greater than that of the picocells). Thus, the LA algorithm associates the user with the underloaded cell (i.e., picocell) or equivalently $P_r(u \in p)$ increases with the augmentation of $\lambda$. Also, as $N_{SU,m}$ increases the macrocell becomes overloaded with smaller values of $\lambda$, so $P_r(u \in p)$ tends to 1 in different $\lambda = 0.04, 0.06, 0.12$ for $N_{SU,m} = 200, 400, 600$, respectively. What is more, we include the conventional (CONV) algorithm, described in Eq. (4) and compare it with the proposed LA one. In sharp contrast with the LA algorithm, the CONV algorithm always suggests the user to be associated with the macrocell. In case of congestion, this is suboptimal, since the user is placed close to one or more underloaded cells. $P_r(u \in m)$, is also plotted, showing the complementary behavior of the algorithm.

Fig. 4 shows the predicted delays, $D_m, D_p$ for the macrocell and the picocell, as a function of the flow arrival rate $\lambda$. Specifically, these are the $D^k_m, D^k_p$ that are also averaged out across time $k$. Their computation is performed via Monte-Carlo simulations for different number of static users $N_{MU}$.

The proposed algorithm as well as the performed analysis hold for the general type of cells (i.e., either overlapping or non-overlapping cells) with any radius $R_i$. Without loss of generality, we relax this assumption considering non-overlapping cells and the same radius for the picocells. The consideration of one mobile user is sufficient enough for studying the behavior of the algorithm without affecting its validity for the general case.

**Table 1. Simulation setup parameters**

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
<th>Parameter</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_m$</td>
<td>1000 m</td>
<td>$f_c$</td>
<td>700 MHz</td>
<td>$P_{rec}$</td>
<td>0.8</td>
</tr>
<tr>
<td>$R_{p_j}$</td>
<td>200 m</td>
<td>$B$</td>
<td>10 MHz</td>
<td>$P_{pp_j}$</td>
<td>0.8</td>
</tr>
<tr>
<td>$P_{pm}$</td>
<td>43 dBm</td>
<td>$d_{ic}$</td>
<td>1 m</td>
<td>$\Delta$</td>
<td>3 dB</td>
</tr>
<tr>
<td>$P_{pm}$</td>
<td>21 dBm</td>
<td>$\eta$</td>
<td>1</td>
<td>$Y$</td>
<td>1</td>
</tr>
<tr>
<td>$N_{SU,m}$</td>
<td>-104.5 dBm</td>
<td>$\eta$</td>
<td>2</td>
<td>$s$</td>
<td>50</td>
</tr>
<tr>
<td>$N_{SU,p}$</td>
<td>-50 dBm</td>
<td>$\sigma_{db,m}$</td>
<td>8 dB</td>
<td>$\delta$</td>
<td>10</td>
</tr>
<tr>
<td>$R_{pm}$</td>
<td>-55 dBm</td>
<td>$\sigma_{db,p_j}$</td>
<td>6 dB</td>
<td>$\epsilon$</td>
<td>1</td>
</tr>
</tbody>
</table>

Fig. 3. Cell assignment probability (CONV vs LA) for different number of static macrocell users $N_{SU,m}$ and number of static picocell users $N_{SU,p_j} = 10$.

Fig. 4. Average delay in the macrocell/picocell for different number of static macrocell users $N_{SU,m}$ and number of static picocell users $N_{SU,p_j} = 10$.

$N_{SU,m} = 200, 400, 600$ and $N_{SU,p_j} = 10$ within the picocells $p_j$. Especially, $D_p$ is also averaged out for all picocells $p_j$. It is noticed that the average delay, $D_m$, increases sharper for larger values of $N_{SU,m}$ and smaller values of $\lambda$, according to Eq. (7). $N_{SU,m}$ and $\lambda$ affect $D_m$ increment and consequently enlarge the difference between $D_m, D_p$. Hence, when their ratio becomes large ($D_m \gg D_p$, $f(\cdot)$ tends to zero and vanishes the gap between the received powers. Thus, the impact of $N_{SU,m}$, $\lambda$ on $D_m, D_p$ ratio exceeds the impact of received powers asymmetry. This validates the LA algorithm functionality that associates the user with the underloaded picocells in the case of an overloaded macrocell ($P_r(u \in p)$ goes to one), regardless the gap in received powers. On the other hand, HO is not triggered if there is not much difference in $D_m, D_p$, i.e., the user QoS is satisfied ($P_r(u \in p)$ goes to zero). Consequently, RSS-based HO that maximizes the instantaneous rate of a user (i.e., the best modulation and coding scheme (MCS) is used), reflects user QoS only when the BS is lightly loaded. However, user performance, in terms of per flow delay, may be severely affected if the BS offering the best RSS is congested.

In Fig. 5, the picocell assignment probability, $P_r(u \in p)$, is depicted as a function of the user’s distance from the macrocell for fixed number of $N_{SU,m} = 200$ and $N_{SU,p_j} = 10$ in the picocell. Our algorithm is compared with the CONV and a distance-based (DIST) one provided in [8], for different $\lambda$. The latter is applied in hierarchical macrocell/SCs network that is also our case. The DIST approach associates the user
with its closest SC regardless the asymmetry in BSs powers. This is achieved using a scale down factor $\alpha$ based on the maximization of the SC assignment probability. Further, some of the system parameters are changed to adapt the simulated scenario in [8]. For the DIST algorithm, $P_r(\lambda < p) = 1$ after a specific user’s distance from the picocell despite the fact that the macrocell is overloaded or not. However, with the LA approach, $P_r(\lambda < p)$ increases, depending on the macrocell load for different $\lambda$, no matter the user’s proximity to the picocell. Thus, if the user is close to a picocell, its association with it may be unnecessary, since its QoS requirements may be already met. Finally, the CONV algorithm does not trigger the HO, due to BSs power asymmetry.

The overall delay that the user experiences (i.e., $D_{\text{m}}$ from the macrocell and $D_{\text{p}}$ from the picocell depending on where it is associated) is presented in Table II. The setup is considered the same as in Fig. 5. Using the CONV algorithm, it is noted that the overall delay, $D_{\text{CONV}}$, is identical to the delay of the macrocell (see Fig. 4 for $N_{\text{STU,m}} = 200$ and $N_{\text{STU,p}} = 10$), since the user is always connected to the macrocell. The DIST algorithm retains the user connected to the macrocell up to a specific distance threshold and after that it attaches it to the picocell (see Fig. 5) with $P_r(\lambda < p)$ close to 1. Thus, the augmentation of the macrocell delay, when it becomes overloaded, compared to the picocell ones, increases the overall delay, $D_{\text{DIST}}$. The LA algorithm keeps a conservative association with the underloaded picocell earlier than the DIST one, since the latter connects the user to the cell with the lower delay (i.e., picocell) earlier. Consequently, unnecessary handovers to the picocell and related signaling overhead are avoided with the LA approach. On the other hand, if the load is high in the macrocell, the LA algorithm associates the user with the underloaded picocell earlier than the DIST algorithm that is load-unaware giving significant performance to the overall delay, $D_{\text{LA}}$. The respective ratios of the CONV/DIST algorithms over the LA algorithm (i.e., $D_{\text{CONV}}/D_{\text{DIST}}$) overall delays are demonstrated to show the corresponding gains in Table II. The same trend holds for higher values of $N_{\text{STU,m}}$, but with different values of $\lambda$, as explained in Fig. 4. Finally, significant gains are provided in high load scenarios (i.e., $\approx 4$ and $\approx 8$ compared to the CONV and DIST approach).

### Table II: Delay Gain for $N_{\text{STU,m}} = 200$, $N_{\text{STU,p}} = 10$ and Different $\lambda$

<table>
<thead>
<tr>
<th>$\lambda$</th>
<th>0.05</th>
<th>0.08</th>
<th>0.1</th>
<th>0.11</th>
<th>0.12</th>
</tr>
</thead>
<tbody>
<tr>
<td>$D_{\text{LA}}$ (sec)</td>
<td>0.0701</td>
<td>0.0599</td>
<td>0.1261</td>
<td>0.1125</td>
<td>0.0621</td>
</tr>
<tr>
<td>$D_{\text{DIST}}$ (sec)</td>
<td>0.0624</td>
<td>0.0807</td>
<td>0.1109</td>
<td>0.1479</td>
<td>0.2808</td>
</tr>
<tr>
<td>$D_{\text{CONV}}$ (sec)</td>
<td>0.0743</td>
<td>0.1160</td>
<td>0.1864</td>
<td>0.2708</td>
<td>0.5496</td>
</tr>
<tr>
<td>$D_{\text{CONV}}/D_{\text{LA}}$</td>
<td>1.0596</td>
<td>1.1701</td>
<td>1.4790</td>
<td>2.4071</td>
<td>8.8547</td>
</tr>
</tbody>
</table>

6 Specifically, one macrocell and one picocell are considered in distance $D_1 = 200$ m, $a^2 = 0.96$ as the optimal $\alpha$ for the specific $D_1$, $\Delta = 0$, $R_p = 150$ m and $R_{\text{p,th}} = -51$ dBm. In addition, the mobility model is changed, since the authors in [8] assume that one mobile user moving across a line from the macrocell to the picocell coverage area.

### VI. Conclusion & Future Work

This work focuses on user-centric HO decision algorithms for HetNets. It was shown that the proposed load-aware algorithm can significantly enhance the system performance by considering both user and network perspective in order to improve user QoS. To this end, it was compared with a conventional and a distance-based LTE-handover algorithm and we showed that it can outperform both of them according to the extracted cell assignment probability and user service delay performance results. As future work, we plan to investigate more complex scenarios with more mobile users and study the impact of our load-aware algorithm on user-distribution and further on load-balancing. Finally, although we treated all flows as best-effort, we plan to consider also dedicated flows based on different QoS metrics (e.g., blocking probabilities).

### Acknowledgments

The research leading to these results has received funding from the European Union’s Seventh Framework Programme under grant agreement no 612050 (FLEX Project).

### References


APPENDIX A
CAPACITY ANALYSIS—I

The average capacity, assuming shadowing and Rayleigh flat fading, is expressed as:

\[ \bar{C}(d_i) = E_{\psi_i|h_i^2} \left[ B \log_2 \left( 1 + \rho_i K (d_i/d_r)^{-\eta} \psi_i |h_i|^2 \right) \right] \tag{8} \]

where \( \rho_i \triangleq P_{S,i}/N_0 \) denotes the transmit signal-to-noise ratio (SNR)\(^9\), \( P_{S,i} \) is the transmission power in linear scale, \( d_i \) represents the distance from the BS cell \( i \) to the user that is greater than a reference distance \( d_r \), \( B \) stands for the available bandwidth, \( N_0 \) represents the aggregated noise power over \( B \) in linear scale\(^10\), \( \psi_i \sim \text{Log-}\mathcal{N} (0, \sigma^2_{\text{dBm}}/\zeta^2) \)\(^11\) represents the shadowing fading in linear scale and \( h \sim \mathcal{CN} (0, 1) \)\(^12\) defines the Rayleigh flat fading channel coefficient. To the best of our knowledge, an analytical expression for the averaged capacity does not exist. Therefore, its lower and upper bound, \( \bar{C}_b(d_i) \), \( b \in \{ L, U \} \), can be easily computed, as follows.

A. Lower bound

Exploiting the concavity of the logarithmic function, an analytical lower bound is given by:

\[ C(d_i) \leq \bar{C}(d_i) = E_{\psi_i|h_i^2} \left[ B \log_2 \left( 1 + \rho_i K (d_i/d_r)^{-\eta} \psi_i |h_i|^2 \right) \right] \]

\[ \Rightarrow \]

\[ \bar{C}(d_i) \geq \bar{C}_L(d_i) \geq B \log_2 \left( \rho_i K (d_i/d_r)^{-\eta} \right) - \gamma / \ln(2) \tag{9} \]

where \( \gamma \) denotes the Euler-Mascheroni constant.

B. Upper bound

Using Jensen’s inequality for concave functions, the upper bound is computed analytically as:

\[ C(d_i) = E_{\psi_i|h_i^2} \left[ B \log_2 \left( 1 + \rho_i K (d_i/d_r)^{-\eta} \psi_i |h_i|^2 \right) \right] \]

\[ \Rightarrow \]

\[ \bar{C}(d_i) = E_{\psi_i|h_i^2} \left[ B \log_2 \left( 1 + \rho_i K (d_i/d_r)^{-\eta} \psi_i |h_i|^2 \right) \right] \]

\[ \Rightarrow \]

\[ \bar{C}_U(d_i) \leq B \log_2 \left( 1 + \rho_i K (d_i/d_r)^{-\eta} \mu_{\psi_i} \right) \]

where \( \mu_{\psi_i} = E[\psi_i] = \exp \left[ \sigma^2_{\text{dBm}}/2 \zeta^2 \right] \).

APPENDIX B
CAPACITY ANALYSIS—II

The picocell points \( x, y \) are distributed uniformly over a circle according to the following distribution:

\[ f_{X,Y}(x,y) = \begin{cases} \frac{1}{\pi (R_{p,i}^2 - d_r^2)} & , \ d_r \leq r \leq R_{p,i} \\ 0 & , \ o/w \end{cases} \]

\[ \Rightarrow \]

\[ \bar{C}_b(d_i) = \int_{XY} \bar{C}_b(d_i(x,y)) f_{X,Y}(x,y) \, dx \, dy , \ b \in \{ L, U \} \tag{12} \]

A. Lower bound

The average lower bound of the average capacity by applying cartesian to polar coordinates transformation in Eq. (12) is computed as:

\[ \bar{C}_L^i = \frac{1}{\pi (R_{p,i}^2 - d_r^2)} \int_0^{2\pi} \int_{d_r}^{R_{p,i}} r \bar{C}_L \left( \bar{d}_i(r, \theta) \right) \, dr \, d\theta = \]

\[ = B \log_2 (\rho_i) + \]

\[ \frac{1}{\pi (R_{p,i}^2 - d_r^2)} \int_0^{2\pi} \int_{d_r}^{R_{p,i}} r \log_2 \left( K \left( \bar{d}_i(r, \theta) / d_r \right)^{-\eta} \right) \, dr \, d\theta \]

\[ - \gamma / \ln(2) \tag{13} \]

B. Upper bound

Applying the same methodology used for the lower bound, the corresponding average upper bound is given by:

\[ \bar{C}_U^i = \frac{1}{\pi (R_{p,i}^2 - d_r^2)} \int_0^{2\pi} \int_{d_r}^{R_{p,i}} r \bar{C}_U \left( \bar{d}_i(r, \theta) \right) \, dr \, d\theta = \]

\[ = \frac{\pi (R_{p,i}^2 - d_r^2)}{B} \times \]

\[ \int_0^{2\pi} \int_{d_r}^{R_{p,i}} r \log_2 \left( 1 + \rho_i K \left( \bar{d}_i(r, \theta) / d_r^{-\eta} \mu_{\psi_i} \right) \right) \, dr \, d\theta \tag{14} \]

The function \( \bar{d}_i(r, \theta) \) is given by:

\[ \bar{d}_i(r, \theta) = \sqrt{\frac{r \cos(\theta) - x^*}{x}} + \sqrt{\frac{r \sin(\theta) - y^*}{y}} \tag{15} \]

where \( x^* = x_m - x_{p,i} \) and \( y^* = y_m - y_{p,i} \).

This function represents the user distance from the macrocell and the picocell BS in polar coordinates, when the user is located within the picocell. Finally, the \( J_1 \) and \( J_2 \) integrals are numerically computed.