Trio-Connectivity for Efficient Uplink Performance in Future Mobile HetNets

Ali Mahbas, Huiling Zhu, *Member, IEEE*, and Jiangzhou Wang, *Fellow, IEEE*
School of Engineering and Digital Arts,
University of Kent, Canterbury, CT2 7NT, United Kingdom
Email: ali.mahbas@hotmail.com, [H.Zhu, J.Z.Wang]@kent.ac.uk

***Abstract*—The technical challenges, e.g. the mobility management and the offloading process, hinder the conventional cellular systems to meet the huge data traffic requirements of the next generation mobile communications. The traditional system (e.g dual-connectivity (DC)) has been proposed to improve the mobility management, however, it will inherit the big trade-off in the offloading process between the energy consumption for the small cell (SC) discovery (SCD) process and the efficiency of utilizing the system resources (e.g. frequency and signaling). In this paper, we present a framework to model the potential offloading opportunities as well as the offloading loss when a typical user equipment (UE) performs the inter-frequency (IRF) scan periodically. The proposed framework also studies the impact of the SCD on the energy efficiency. To improve the system performance and reduce the power consumption at the UEs, a new scheme, trio-connectivity (TC), is proposed in this paper to tackle the aforementioned challenges. The TC includes three planes: control-plane (C-plane), user-plane (U-plane) and indication-plane (I-plane). The I-plane works as an indicator to help the UE to identify and discover the SCs in the system prior to offloading. The role of the I-plane is to keep the SCD on one frequency channel regardless of the number of frequency channels in the system. In the proposed offloading mechanism, some of the energy consumption is transferred from the UE to the network. By using the proposed framework, UE energy efficiency and system energy efficiency as well as the total energy consumption are derived as performance metrics to compare between the TC and the DC. The results show that the TC can outperform the DC in dense cellular systems.**

***Index Terms*—Energy efficiency, heterogeneous network (HetNet), small cell discovery, offloading.**

I. Introduction

In order to cope with the future ultra-high data rate demand in wireless communications, spectral efficiency needs to be improved significantly from legacy systems [1], [2]; and the heterogeneous network (HetNet) has been considered as one of the essential techniques for this purpose [3]. The offloading process, in which the users and their traffic are offloaded from the overloaded cells (e.g. macro cells (MCs)) to the small cells (SCs), plays an important role to utilize the potential benefits of HetNets. In the future HetNet, the SCs use a frequency channel other than the frequency channel used by the MCs (inter-frequency HetNet (IRF HetNet)), therefore the offloading process requires a SC discovery process (SCD), in which the user equipment (UE) performs IRF scan periodically to discover the surrounding SCs before being offloaded from the serving MC to the potential serving SC [4], [5].

However, there is a compromise in the offloading process between the energy consumption required for performing IRF scan and the offloading loss [6]–[8]. Due to limited battery capacities at the UEs, it is expected that the issue in the offloading process will be more serious when the SCs are deployed on multiple frequency channels and the UEs need to perform the IRF scan on each of these channels frequently in order to exploit the potential offloading opportunities and to free the resources at the overloaded MCs.

A. Cell Discovery

The cell discovery is a procedure done by UEs to identify the nearby cells and measure their signal quality. There are two types of cell discovery: i) intra-frequency (IAF) and ii) IRF. In the current systems, UEs perform IAF scans on a predefined time-frequency basis on pilot or reference signals (RS) of the cells that are deployed on the same frequency of the serving cells. These scans are sent as reports to the serving cells. They are important not only for inter-cell mobility, but also for estimating the channel between the UEs and the serving cells. The IRF scans take place when UEs are triggered by their serving cells to discover the cells deployed on a different frequency. In the current mobile systems, periodic scan has been introduced to allow UEs to detect the SCs. UEs perform IRF scans and send them to their serving

MCs either periodically or in an event based manner to identify the nearby SCs [5], [9]. The offloading is initiated when a number of offloading requirements are met, e.g. a UE receives signal of enough quality and meets the access policy of the discovered SCs.

B. Related Work

The offloading in the co-channel HetNets has been studied extensively [10], [11], where UEs are offloaded from large cells (e.g. MCs) to SCs. In [10], the coverage rate and the signal to interference plus noise ratio (SINR) distribution were studied in a multi-tier HetNet under a flexible cell association. The impact of the association on the offloading was captured and it was also shown that an optimum fraction of traffic, which depends on the ratio of the frequency resources at each cell to the UE's requirements, should be offloaded to achieve the maximum coverage rate. In [11], a multi-tier HetNet was studied by using stochastic geometry where the base stations (BSs) of each tier form independent Poisson point process (PPP). [11] provided a framework with a flexible cell association to model and derive the downlink (DL) SINR distribution and the average ergodic rate. It was shown that the cell association (bias) has a significant impact on the load balancing since the nodes with low transmit power (e.g. SCs) have very small footage due to the transmit power difference. Although increasing the bias to the SCs increases the outage probability as some UEs are associated to the cells that do not provide the highest received power, it improves the data rate in the lightly-loaded HetNets. Although it is widely believed that the future HetNets will include SCs deployed on high frequency either than the MCs frequency (it is also known as IRF deployment) due to the limited low frequency bands, most of the studies such as [10] and [11] studied the effects of different system parameters on the offloading such as node density and bias in the co-channel deployment (SCs and MCs use the same frequency channel) where the impact of the SCD process was not taken into consideration due to the type of deployment in which no IRF scan was required. These studies also assumed that the UEs are in a stationery state and the cell association was derived based on this assumption.

Some work has studied the offloading and the SCD in the IRF HetNet [6], [8], [12]–[14]. In [6], stochastic geometry was used to investigate the impact of IRF scan periodicity (t_r) on the average energy efficiency. The offloading loss was approximated as a function of t_r via a polynomial curve fitting for a fixed SC density and the UE's speed. This approximation may not be accurate for more diverse system parameters, for example, it did not consider the impact of other system parameters such as the SC transmit power, the minimum time for successful offloading, the distribution of the SCs around the UE's path and the overlaps of SCs on the path as will be shown later in this paper. Furthermore it was assumed that the fraction of a reference UE's path covered by the SCs is a product of the second tier association (which was derived in [11] and based on a UE being in stationery state) and the total of transition time. In [12], an IRF scan scheme was proposed based on the UE mobility. In this scheme, the IRF scan periodicity is vaired according to the speed of the UE in order to reduce the power consumption. The UE mobility based scheme can help to reduce the power consumpation for UEs with high mobility. However, the UEs with low mobility consume significant amount of power as they are forced to perform the IRF scan frequently. In [13], [14], radio fingerprint schemes were proposed to overcome the challenges in the periodic IRF scan mechanism. In the radio fingerprint, received signal strength (RSS) are used to estimate the UE's location with respect to the SCs. The MCs instruct the UEs to carry out IAF scan to check if the RSS matches entries in a fingerprint database stored at the MCs or at the UEs. The accuracy of UE location estimations and the large amount of signaling required for storing and updating the fingerprint database are the main challenges of this scheme.

In [8], the impact of the periodic IFR scan on the system performance was modelled by taking into consideration the effect of some system parameters such as the transmit power of the SCs, the SC density and the speed of UEs. The proposed framework in [8] did not take into account the SCs overlaps on the reference UE's path and the minimum time for successful offloading when modelling the offloading loss and the fraction of the reference UE's path covered by the SCs. Since the SCs will be deployed densely, the overlaps will have a significant impact on the accuracy of the offloading opportunities and the offloading loss analysis. All [6], [8], [12]–[14] assumed that the SCs were deployed on one frequency channel. It is expected that the role of the IRF scan will be critical in the future HetNets when the number of UEs increases significantly and these UEs have to perform IRF scans on multiple frequency bands (different frequency channels from different bands may be deployed in the future HetNets).

C. Contribution

It is shown in our paper that the number of system parameters next to the UE speed need to be considered in

order to boost the system performance. The contributions of this paper are summarized as follows:

- By using the stochastic geometry tool, a mobility framework to obtain the SC coverage on a reference UE's path is proposed by taking into consideration some overlaps taking place among the SCs in the network. By using the framework, the potential offloading opportunity in the system is modeled by taking into consideration different system parameters such as the SC density, the SC transmit power, the overlaps among SCs and the speed of UEs.
- In order to obtain the energy consumption for SCD process and evaluate the system performance, the offloading opportunities missed by a UE which is also known as offloading loss, is studied in this paper. It is defined as the offloading missed by the UE due to long periodicity of IRF scan (small number of scans per unit time). The offloading loss is derived when the UE moving with a constant speed and performing the IRF scan periodically with taking into account the periodicity of IRF scan (t_r), the potential offloading opportunities in the system and the minimum time required for successful offloading.
- By considering the offloading loss and the potential offloading opportunities, the UE energy efficiency and the system energy efficiency as well as the energy consumption in the SCD process are given as performance metrics to evaluate different system parameters and study their impacts on the system performance, where the locations of the SC base stations (SBS) and MC base stations (MBS) are modeled as PPPs.
- A novel trio-connectivity system (TC) is proposed to overcome the current problem in the offloading process. The proposed system which includes control-plane (C-plane), user-plane (U-plane) and indication-plane (I-plane) takes into consideration the need to add more frequency channels from different frequency bands to the future cellular system. The I-plane is a function of broadcasting the system information of the SCs on the low frequency in parallel with the high frequency to enable UEs discovering the surrounding SCs without performing IRF scans periodically. The TC can help to exploit the system resources (e.g. frequency, signaling load and power) efficiently by not only overcoming some of the mobility management challenges, but also minimizing the energy consumption and the signaling load in the offloading process.
- The energy consumption in the proposed SCD mechanism is also studied and a comparison with the conventional mechanism (e.g. periodic scan) is presented by using the above performance metrics.

The rest of this paper is structured as follows. Section II explains the system model and the mobility model. The total potential offloading time is investigated in Section III . In Section IV the total offloading loss is derived. In Section V the proposed TC is introduced. The total energy consumption for both uplink (UL) traffic and SCD in both traditional system and the proposed system in addition to UL and system energy efficiency are presented in Section VI. Results are shown in Section VII. Finally, conclusions are drawn in Section VIII.

II. System Model

Consider a two-tier HetNet as shown in Figure 1, where the first-tier includes MBSs and uses low frequency to provide a wide coverage. The second tier includes SBSs and uses high frequency to provide the U-plane (heavy traffic). It is assumed that UEs, MBSs and SBSs are spatially distributed as independent PPPs, Φ_u, Φ_m and Φ_s with density λ_u, λ_m and λ_s respectively [15]. Since a two-tier HetNet is considered in this paper, any UE can be served (associated) by either the first tier (the MCs) or the second tier (the SCs). This depends on the location of the UE and efficiency of the adopted SCD process in the system. For instance, some UEs located in SCs' coverage keep associated to MCs as they fail to discover the SCs promptly. Next, the UL SINR at UE of interest ($\mathcal{U}_0$) is obtained:

$$SINR_k = \frac{P_k(x_k) g_0 L_f(x_k)}{\sigma^2 + \sum_{i \in \Phi_k} \bar{P}_i(\hat{z}_i) h_i \bar{L}_{f,i}(z_i)} \qquad (1)$$

where $SINR_k$ is the received SINR at the serving BS of $\mathcal{U}_0$, k is the tier indicator and takes a value of either s when $\mathcal{U}_0$ is served by a SC or m when $\mathcal{U}_0$ is served by MC, $P_k(x_k) = p_u x_k^{\alpha_k \epsilon}$ is $\mathcal{U}_0$ transmit power when associated to the kth tier, p_u is the baseline UE transmit power, x_k represents the distance between $\mathcal{U}_0$ and its serving BS, ϵ is the UL power control factor and takes a value between 0 and 1, α_k is the path-loss exponent of the kth tier, g_0 is the channel gain between $\mathcal{U}_0$ and the serving BS (this is due to the multi-path propagation of the signal in the environment) and it is assumed that g_0 has a Rayleigh distribution, $L_f(x_k) = \mathcal{L}_f x_k^{-\alpha_k}$ is the desired link path-loss, $\mathcal{L}_f = (\frac{\omega}{f 4\pi})^2$ is the path-loss at 1 meter with the frequency f and obtained by using the Frii's transmission equation, ω is the velocity of light. σ^2 is the additive noise power, $\bar{P}_i(\hat{z}_i) = \hat{z}_i^{\alpha_k \epsilon} p_u$ is the transmit power of the ith interfering UE, $\hat{z}_i$ is the distances from the ith interfering UE to its serving

BS, $\bar{L}_f(z) = \mathcal{L}_f z_i^{-\alpha_k}$ is the interfering link path-loss, z_i is the distance between the ith interfering UE and the serving BS of $\mathcal{U}_0$, and h_i is the channel gain between the ith interfering UE and the serving BS of $\mathcal{U}_0$. Note that shadowing is not considered in this paper.

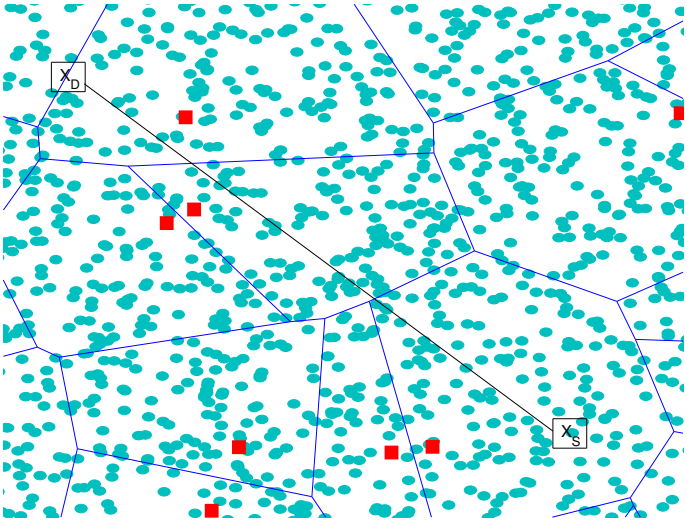

Fig. 1: System Model. Red squares represent the MBSs (first tier), blue dots represent the SCs coverage, X_S and X_D represents the starting point and destination point, and the black line represents the reference UE's path between X_S and X_D.

We make the following assumptions to clarify the system model:

- **Association Assumption**: Offloading UEs from the first tier to the second tier is highly desired and prioritized in order to increase the system capacity and exploit the abundant resources at the SCs efficiently. It is expected that the resources at the SCs will not be exploited efficiently in the IRF HetNets if the cell association is based on the best received power due to the transmit power difference between the first and the second tiers as well as the path-loss difference between the low frequency at the first tier and the high frequency at the second tier [16]. Since there is no interference between the first and the second tiers, it is assumed that the second tier association is based on a RS received power threshold (ρ_{min}). This means that the UE can associate to the second tier when it receives the RSs from any SC with power greater or equal to (ρ_{min}). Any UE will be considered in the coverage of the ith SC when:

$$\rho_i > \max(\rho_{min}, \max_{j \in \Phi_s} \rho_j) \tag{2}$$

where ρ_i and ρ_j represent the average value of the RSs received power from the ith SC and the jth SC respectively. Note, the average value of the RSs received power is considered for cell association to avoid the ping pong effect [17]. However, the instantaneous value of the received power is considered when calculating the SINR and average data rate.

- **Open-Access Assumption**: It is assumed that all SCs operate on open-access mode.

We consider the random waypoint (RWP) proposed in [18], where the $\mathcal{U}_0$'s movement can be modeled as $\{X_{Sl}, X_{Dl}, V_l, S_l\}$, X_{Sl} is the lth starting point, X_{Dl} is the lth destination point, V_l is the speed of $\mathcal{U}_0$ during its movement from X_{Sl} to X_{Dl} and S_l the pause time that $\mathcal{U}_0$ spends at X_{Dl}. Note that X_{Sl} and X_{Dl} are chosen randomly on the system area ($\mathscr{A}$) where the destination of the lth movement will be the starting point of the next movement ($X_{S(l+1)} \equiv X_{Dl}$).

III. Total Potential Offloading Time

The total potential offloading time can be defined as the total time that $\mathcal{U}_0$ resides in the second tier's coverage during one movement. For brevity, we consider the first movement where the starting point and the destination point are donated by X_S and X_D respectively. The total distance that $\mathcal{U}_0$ travels during the movement is obtained. Assuming that $\mathscr{A}$ is a circle with radius $\mathscr{R}$, the expectation of the distance between X_S and X_D is [19]:

$$D_{X_S - X_D} = \mathbb{E}[\| X_D - X_S \|] = \frac{125\mathscr{R}}{45\pi} \tag{3}$$

where $\| \ . \ \|$ indicates the Euclidean distance. It has been widely accepted that the cells form regular shapes when studying the time that UEs spend in the cells (e.g. sojourn time) [20], [21]. Since each tier uses a different frequency channel, the footprint of each SC is assumed to be a circle. When a SC is crossed by $\mathcal{U}_0$, the coverage of this SC on $\mathcal{U}_0$'s path is given by:

$$C = 2r \sin\left(\cos^{-1}\left(\frac{\eta}{r}\right)\right), \qquad \eta \leq r \tag{4}$$

where r is the radius of the SC's footprint, η is the length of the vertical line between the SBS and $\mathcal{U}_0$'s path, and it is assumed to be randomly distributed in the range $[0, r]$ as shown in Figure 2. The expectation of the second tier coverage on $\mathcal{U}_0$'s path is obtained as follows.

Lemma 1 The expectation of the second tier's coverage (including the overlap coverage) on $\mathcal{U}_0$'s path can be expressed as:

$$\mathbb{E}[C_T] = \frac{125 r^2 \lambda_s \mathscr{R}}{45} \tag{5}$$

Proof: See Appendix A.

The above result also includes some of the overlap areas on $\mathcal{U}_0$'s path. The overlap coverage can be neglected for low SC density. But it is anticipated that the SC density in the future cellular networks is very high and the overlap coverage needs to be considered in order to find the potential offloading time accurately. Therefore, the SCs crossed by $\mathcal{U}_0$ needs to be studied carefully. Next, we investigate the total overlap coverage among SCs on $\mathcal{U}_0$'s path in order to obtain the total potential offloading time.

Theorem 1 When the speed of $\mathcal{U}_0$ is constant, the total expected time that $\mathcal{U}_0$ spends in the second tier's coverage or the total potential offloading time can be expressed as:

$$\mathbb{E}[\mathcal{T}_s] = \frac{\mathbb{E}[C_T] - \mathbb{E}[\bar{\bar{C}}_T]}{v} + S\mathbb{P}_s \tag{6}$$

where $\mathbb{E}[\bar{\bar{C}}_T]$ is the expectation of the overlap coverage on $\mathcal{U}_0$'s path, $S\mathbb{P}_s$ is the expectation of pause time $\mathcal{U}_0$ spent in the second tier coverage, S is the total pause time of the movement and $\mathbb{P}_s$ is the probability of X_D located in any SC.

Proof: The total potential time includes the time that $\mathcal{U}_0$ travels in the SC coverage as well as the time that $\mathcal{U}_0$ resides in the coverage of a SC if the destination point X_D is located in that SC. Since $\mathcal{U}_0$ is associated to the nearest SBS as long as the the average value of RSs received power is greater than ρ_{min} (Association and Open-Access Assumptions), the probability of X_D located in any SC can be obtained:

$$\begin{aligned}\mathbb{P}_s &= \mathbb{P}\big[\rho_0 > \rho_{min}\big] \\ &= \mathbb{P}\big[x_s < \big(\frac{\mathcal{L}_2 p_{s2}^{rs}}{\rho_{min}}\big)^{\frac{1}{\alpha_2}}\big] \\ &= 1 - \exp\Big(-\pi\lambda_s\big(\frac{\mathcal{L}_2 p_{s2}^{rs}}{\rho_{min}}\big)^{\frac{2}{\alpha_2}}\Big)\end{aligned} \tag{7}$$

where ρ_0 is the RSs received power from the nearest SC, x_s represents the distance between $\mathcal{U}_0$ and the nearest SBS, p_{s2}^{rs} is the transmit power of the RSs from the high frequency at the SCs, $\exp\big(-\pi\lambda_2\big(\frac{PL_2 p_{s2}^{rs}}{\rho_{min}}\big)^{\frac{2}{\alpha_2}}\big)$ is the probability of no SBS within the area $\pi\big(\frac{PL_2 p_{s2}^{rs}}{\rho_{min}}\big)^{\frac{2}{\alpha_2}}$ which can also be interpreted as the probability of $\mathcal{U}_0$ associated to MCs (null probability [15]). The total times that $\mathcal{U}_0$ stays in the coverage of the SCs during pause time can be expressed as:

$$S_s = S\mathbb{P}_s \tag{8}$$

The total distance that $\mathcal{U}_0$ travels in the coverage of SCs between X_S and X_D is expressed as:

$$\mathbb{E}[C_s] = \mathbb{E}[C_T] - \mathbb{E}[\bar{\bar{C}}_T] \tag{9}$$

where $\mathbb{E}[\bar{\bar{C}}_T] = \mathbb{E}[N_{\bar{\bar{C}}}]\mathbb{E}[\bar{\bar{C}}]$ is the total overlap coverage on the path, $\mathbb{E}[N_{\bar{\bar{C}}}]$ is the expected number of SC overlaps on the path and $\mathbb{E}[\bar{\bar{C}}]$ is the expected value of one overlap coverage between any two consecutive SCs on the path. $\mathbb{E}[\bar{\bar{C}}]$ is derived in [22] as:

$$\mathbb{E}[\bar{\bar{C}}] = \int_0^{4r}\int_0^{\bar{c}} \frac{\bar{c}c(\bar{c}-c)\,dc}{32r^3\sqrt{8r^2 - c^2 - (\bar{c}-c)^2}} d\bar{c} \tag{10}$$

$\mathbb{E}[N_{\bar{\bar{C}}}]$ can be obtained as:

$$\mathbb{E}[N_{\bar{\bar{C}}}] = \mathbb{E}[N_{\bar{\bar{C}},max}]\mathbb{P}_{\bar{\bar{C}}} \tag{11}$$

where $\mathbb{P}_{\bar{\bar{C}}}$ is the probability of an overlap occurring on the path and $\mathbb{E}[N_{\bar{\bar{C}},max}]$ is the maximum number of overlap taking place on the path and can be assumed to take a value equal from 0 to $\mathbb{E}[N_s]-1$ when $\mathbb{E}[N_s]$ is the expected number of SC crossed by $\mathcal{U}_0$ during its movement to X_D. Since the SCs form a PPP in the system, $\mathbb{E}[N_s]$ is obtained as:

$$\mathbb{E}[N_s] = 2r\lambda_s D_{X_S - X_D} \tag{12}$$

where $2rD_{X_S-X_D}$ represents the area surrounding the path from X_S to X_D. $\mathbb{P}_{\bar{\bar{C}}}$ is obtained as

$$\mathbb{P}_{\bar{\bar{C}}} = 1 - e^{-\frac{2\lambda_s}{16r^2}\int_0^{4r}\int_0^{\bar{c}} \frac{\bar{c}c(\bar{c}-c)\,dc}{\sqrt{8r^2 - c^2 - (\bar{c}-c)^2}} d\bar{c}} \tag{13}$$

The result in Eq. (13) was proven in [22]. The total potential offloading time is expressed as

$$\mathbb{E}[\mathcal{T}_s] = \frac{\mathbb{E}[C_s]}{v} + S_s \tag{14}$$

The desired result in Eq. (6) is reached after using Eq. (8) and Eq. (9) in Eq. (14). ■

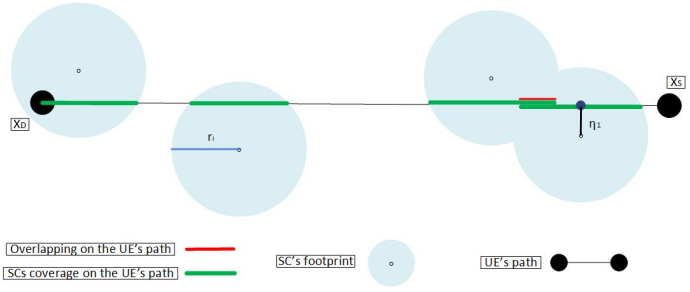

Fig. 2: The $\mathcal{U}_0$'s path between X_S and X_D.

IV. Total Offloading Loss

The total potential offloading time is given in Section III. However, $\mathcal{U}_0$ will not be able to exploit all the potential offloading time due to time loss during the SCD process. The time loss, which is also known as total offloading loss, can be defined as the potential offloading opportunities to SCs that UEs fail to exploit when the IRF scan periodicity is long (a small number of IRF scans per unit time). As a result, $\mathcal{U}_0$ fails to detect the crossed SCs during the movement from X_S to X_D. The IRF scan periodicity (t_r) plays an important role in the SCD and the total offloading loss.

$\mathcal{U}_0$ performs IRF scan every t_r, it stops the SCD process when finding a SC and resumes scanning when leaving the SCs' coverage. Although UEs sometimes need to repeat the scan for more accurate scans, it is assumed that $\mathcal{U}_0$ discovers the SCs from the first IRF scan in their coverage. Assuming that t_{min} is the minimum time needed to consider the offloading beneficial (the offloading procedure costs significant amount of signaling at both UE and network), or it can also be defined as the minimum time needed to complete the offloading procedure successfully. During its movement, $\mathcal{U}_0$ moves with speed V and performs scanning every t_r. The offloading is considered useful if $\mathcal{U}_0$ stays associated to the SC for t_{min}. According to $\mathcal{U}_0$'s speed and the value of t_r as well as the locations of SBSs, the crossed SCs during the movement from X_S to X_D can be classified into three sets of SCs. The set 1 (Ξ_1) includes the SCs that $\mathcal{U}_0$ travels time greater than $t_{rm} = t_r + t_{min}$ in each one. Note that the number of SCs in Ξ_1 increases when $\mathcal{U}_0$'s speed decreases. In set 2 (Ξ_2), $\mathcal{U}_0$ spends time less than t_{rm} and greater than t_{min} in each SC. While, $\mathcal{U}_0$ travels time less than t_{min} in each SC of set 3 (Ξ_3).

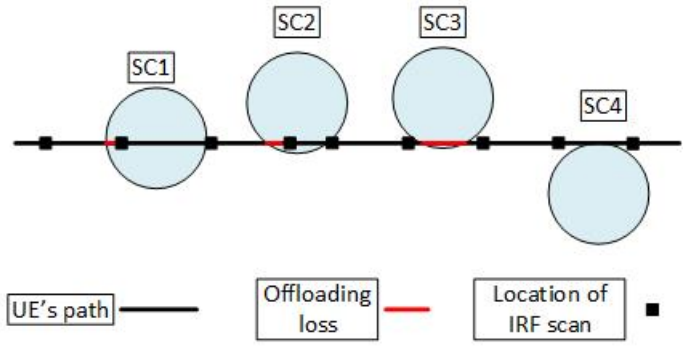

Fig. 3: The Offloading Loss.

According to the SC classification, SC1 in Figure 3 belongs to Ξ_1. Since $\mathcal{U}_0$ performs IRF scan every t_r, $\mathcal{U}_0$ will not miss SC1 entirely as it travels in SC1's coverage time greater than the gap between two consecutive IRF scans. However, it may miss the offloading opportunity partially. Therefore the time passes before $\mathcal{U}_0$ detects SC1 (t_1) can be assumed to be uniformly distributed in the range $[0, t_r]$. When $\mathcal{U}_0$ moves with a constant speed $V = v$, the expectation of the offloading loss from each SC of Ξ_1 (T_{Ξ_1}) can be obtained as:

$$\begin{aligned}\mathbb{E}[T_{\Xi_1}] &= \mathbb{P}_{\Xi_1} \int_0^{t_r} t_1 f_{t_1}(t) dt \\ &= \left(1 - \int_0^{C_{rm}} f_C(c) dc\right) \int_0^{t_r} t_1 f_{t_1}(t) dt\end{aligned} \tag{15}$$

where $f_{t_1}(t)$ is the PDF of t_1, $\mathbb{P}_{\Xi_1}$ is the probability of any crossed SC by $\mathcal{U}_0$ belonging to Ξ_1, which can also be interpreted as the probability of crossed SC having a coverage of $C_{rm} = vt_{rm}$ on $\mathcal{U}_0$'s path, and $f_C(c)$ was obtained in proof of Lemma 1 APPENDIX A (Eq (A1)). Differently, $\mathcal{U}_0$ either detects or misses each SC of Ξ_2 before leaving its coverage. This is because $\mathcal{U}_0$ either performs the IRF scan in some of the SC's coverage or travels in some of the SC during the gap between two consecutive IRF scans. For instance, both SC2 and SC3 belong to Ξ_2 as shown in Figure 3, where SC2 is missed partially and SC3 is missed completely.
Assuming that $\mathcal{U}_0$ travels time of t_2 in each SC of Ξ_2, where t_2 is a random variable and takes a value between t_{min} and t_{rm}. The offloading loss from each SC of Ξ_2 (T_{Ξ_2}) can be t_2 or less (when the T_{Ξ_2} is equal to t_2, it means that $\mathcal{U}_0$ misses the SC entirely). The expected value of (T_{Ξ_2}) can be expressed as:

$$\begin{aligned}\mathbb{E}[T_{\Xi_2}] &\overset{(a)}{=} \mathbb{P}_{\Xi_2} \mathbb{E}_{t_2}\left(\left(1 - \frac{t_2}{t_{rm}}\right) t_2 + \frac{t_2}{t_{rm}} \int_0^{t_2} t f_t(t) dt\right) \\ &\overset{(b)}{=} \frac{1 - \left(\mathbb{P}_{\Xi_1} - \int_0^{C_{min}} f_C(c) dc\right)}{V} \int_{C_{min}}^{C_{rm}} \left(c - \frac{c^2}{2vt_{rm}}\right) f_C(c) dc\end{aligned} \tag{16}$$

where $\int_0^{C_{min}} f_C(c) dc$ represents the probability of any SC covering a distance of C_{min} from $\mathcal{U}_0$'s path, $\mathbb{P}_{\Xi_2}$ is the probability of any SC belonging to Ξ_2 (or the probability of any SC having a coverage greater than C_{min} and less than C_{rm}), $C_{min} = vt_{min}$, the first term and the second term in (a) represent that $\mathcal{U}_0$ misses any SC of Ξ_2 completely with probability $1 - \frac{t_2}{t_{rm}}$ and partly from 0 to t_2 with probability $\frac{t_2}{t_{rm}}$ respectively, and (b) follows from $t_2 = \frac{C_2}{v}$. The expectation of total offloading loss $\mathcal{U}_0$ experiencing is obtained in the following Theorem.

Theorem 2 The expectation of the total offloading loss during one movement is expressed as:

$$\mathbb{E}[\mathcal{T}_{miss}] = \Big(\mathbb{E}[N_{sc}] - \mathbb{E}[N_{\bar{L}}]\Big)\Big(\mathbb{E}[T_{\Xi_1}] + \mathbb{E}[T_{\Xi_2}]\Big) \tag{17}$$

Proof: Since there are some overlaps on $\mathcal{U}_0$'s path, no offloading loss takes place when $\mathcal{U}_0$ moves between two overlapped SCs. The expectation of the $\mathcal{T}_{miss}$ is obtained in Eq. (17). ■

Our analysis shows that the total offloading loss depends on a number of system parameters such as t_r, the SC density, ρ_{min}, the SC transmit power, t_{min} and the UE's speed. Therefore, choosing the value of t_r carefully with taking into consideration the UE's speed is important to minimize the total offloading loss. It is shown that minimizing the value of t_r helps to reduce the total offloading, however, the value of t_r can not be chosen without studying its impact on the energy consumption regarding the SCD. Therefore, the impact of t_r on the energy consumption is studied in the next section.

V. Trio-Connectivity

The dual-connectivity (DC) has been proposed to overcome some of the current architecture's technical challenges such as load signalling and mobility management [23], [24]. The DC includes the C-plane and U-plane, where the C-plane is always provided by the MCs and the U-plane is provided by either the MCs or the SCs. Since DC is based on the IRF deployment, it is expected that DC inherits the offloading issue. The trend of offloading more UEs from the MCs to the SCs in the future HetNets will not only make the energy consumption more serious but also cause significant signalling load. Therefore, an effective offloading process will be of vital importance to ease the aforementioned issues and to exploit the system resources efficiently.

To meet the future requirements and overcome the technical challenges, a new approach needs to be proposed by taking into consideration all the limitations and issues that the current systems suffer from. It is anticipated that the SCD will have a great impact on the performance of the future cellular systems due to limited power at the UEs, multiple high frequency channels and limited resources at the MCs. Therefore we propose a novel TC to solve the issues regarding the mobility management, and the offloading process, since these challenges will have a great impact on the quality of service (QoS) of UEs and the utilization of the system resources. The proposed TC includes three planes, C-plane, U-plane and I-plane. C-plane is maintained by the first tier to provide the required control information for maintaining wireless links with the UEs. The U-plane is maintained by either the first tier or the second tier to deliver the UEs' traffic. Splitting the C-pane and U-plane can provide smoother mobility management [24]–[26]. The I-plane is provided by the second tier by reusing fraction of the low frequency at every SC to advertise the SCs' information (e.g. CID). The purpose is to keep the offloading process (discovering and identifying the SCs) on one frequency (the low frequency) regardless of the number of frequency channels in the network. In other words, the whole SCD process will be based on the IAF scan instead of the IRF scan. The I-plane will act as an indicator for UEs to detect and discover the nearby high frequency SCs in order to offload the UEs traffic to the discovered SCs as shown in Figure 4, which can save power at the UEs as well as the signaling load at both UEs and network. However, this mechanism will cause the network some power consumption for maintaining the I-plane. Although there are substantial differences between the current cellular systems and the future cellular systems, and the technical specifications and the details of the offloading procedure and mobility management have not been characterized yet. Therefore, the proposed TC is modeled and designed by considering the current cellular systems.

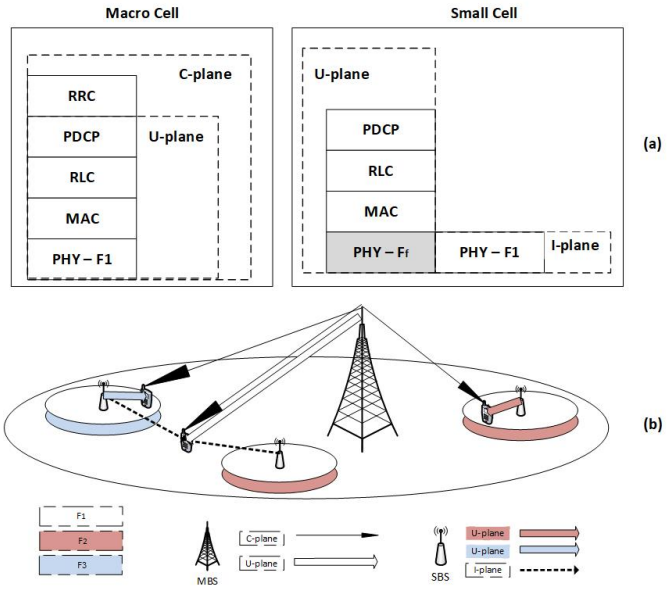

Fig. 4: (a) The proposed protocol stack (b) The proposed TC.

A. Three-Plane

In the proposed TC, the C-plane, which includes the control information, and some of the services that do not require large frequency resources are provided by

the MCs. For instance, the voice service is one of the services needing limited frequency resources, but it requires high reliability. Providing this service by the first tier will minimize the interruption in service continuation and also minimize the signalling, in addition to guarantee more reliable communication. The control information includes all information needed for setting up and maintaining communication links between the UEs and their serving cells. Therefore, the C-plane will not only control the U-plane when provided by the first tier, but also control the U-plane when provided by the second tier [24]. This implies that in the TC, the protocol stack at the MCs will match the conventional protocol stack in the current cellular systems and include radio resources control (RRC), packet data convergence protocol (PDCP), radio link control (RLC), medium access control (MAC) and physical (PHY - F1) as shown in Figure 4a. The RRC includes a number of functions such as the broadcast of system information, RRC connection control, measurements configuration and reporting, support of self-configuration and optimization and others [5]. The first tier in the TC system will also be responsible for configuring, and managing measurements and reports regarding the cell search procedure for both IAF and IRF cells.

The U-plane will be provided by the high frequency at the second tier or by the first tier according to the service requested by UEs and availability of the second tier. The U-plane will be provided by the first tier when the UEs are not located in the second tier's coverage or when light services (e.g. voice service) are requested by the UEs. Since the control information is provided by C-plane, the second tier's role will be restricted to deliver the UEs traffic supported by the C-plane. The protocol stack at the second tier will include all the protocols except the RRC. However, the U-plane at the second tier still needs some functions of the RRC (e.g. broadcast of system information). The second tier will need to include RSs and the synchronization signals for the channel estimation between UEs and the serving SCs, and also for the cell search.

Fraction of the low frequency will be reused at each SC, which will be functioning as an indicator to enable UEs to discover the surrounding SCs and also to estimate the proximity to the SBSs. The system information of each SC will be broadcasted not only on the high frequency (U-plane), but also on the low frequency (I-plane). Moreover, the I-plane will not be used to serve the UEs in the SCs to protect the UEs served by the first tier and also to minimize the interference with the C-plane. Note that the I-plane requires limited frequency resources, therefore, it is assumed that limited frequency resources are allocated for this purpose to minimize the inter-plane interference (the interference between the I-plane and the C-plane). The I-plane only broadcasts the cell information (e.g. CID), RSs and synchronization signals. Decoding the synchronization signals allows UEs to obtain the CID [27]. RSs exist at the physical layer to deliver a reference point for the DL power. Since the locations of RSs in the channel are based on the CID, the synchronization signals enable UEs to discover the SCs in the system and provide the serving cells with the required measurements of the surroundings [27], [28]. Furthermore, the protocol stack at the I-plane will be limited to the physical layer (PHY - F1) and broadcast the system information as shown in Figure 4a.

B. Discovery Mechanism

The low frequency is used at each SC next to one of the high frequencies as shown in Figure 4b. It is assumed that both frequencies at each SC are overlaid and cover the same area. Covering the same area by both spectrum bands requires compensation for the path-loss difference between the low frequency and the high frequency. The long-term average value of the received power from any frequency of the ith SC at a reference point (e.g. cell edge) is expressed as $\rho_i = p_{s,f}^{rs} \bar{r}_0^{-\alpha_s} \mathcal{L}_f$, where $p_{s,f}^{rs}$ is the transmit power of RSs from the SCs on the fth frequency channel (F_f), $\bar{r}_0$ is the distance between the reference point and the SBS and $f \in [2, 3, \cdots, \mathcal{F}+1]$. The RSs transmit power of F_1 to receive the same received power from F_f at the same reference point, can be expressed as:

$$p_{s,1}^{rs} = p_{s,f}^{rs} \left(\frac{F_1}{F_f} \right)^2 \tag{18}$$

The proposed SCs discovery mechanism in the TC can be summarized as:

- The serving MCs manage both the IAF and the IRF scans. When a typical UE is provided with the C-plane and the U-plane by the same serving MC, the typical UE is triggered to perform IAF scans only. The typical UE performs IAF scans and sends them back to the serving MC. These scans will not only be used to estimate the channel between the typical UE and its serving MC, but also used to discover the nearby SCs since all SCs broadcast their information on both low and high frequencies.
- Offloading requirements are the criterion to trigger the IRF scans at the typical UE as a final step before offloading the typical UE from the first tier to the second tier. These requirements could include signal level and quality, the current load at the discovered

SCs (full or not) and the access restriction. The serving MC can use the IAF scans provided by the typical UE to identify the discovered SCs, and check their current load status and whether the access to these SCs restricted or not.

- When the requirements are not met, the U-plane will remain maintained by the serving MC without performing the IRF scan. It is expected that keeping the cell search in the system on one frequency will help to save the power at the UEs and also minimize the signaling overhead at both the UEs and the network.
- When the offloading requirements are met, the typical UE is triggered by the serving MC to perform the IRF scan to initiate the offloading to the discovered SC.

According to the IRF scan sent to the serving MCs from the UEs, the offloading decision to the SC will be made. Figure 5 shows the signaling diagram of the proposed SCD mechanism.

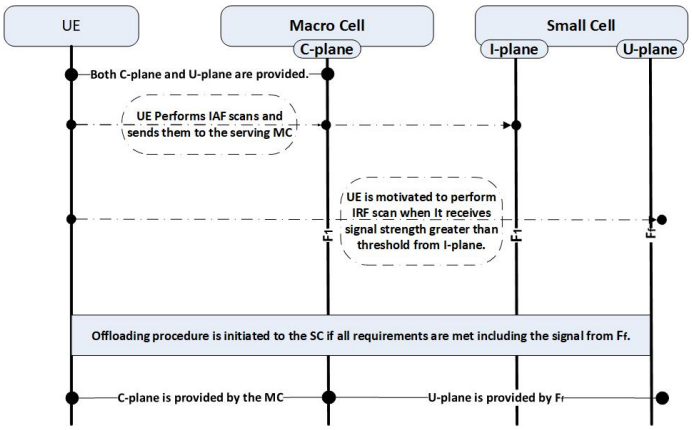

Fig. 5: The signaling diagram of the proposed mechanism

VI. Energy Efficiency

In this section, the performance of the SCD process in the current system and the proposed SCD are studied. Energy efficiency is considered as performance metric to demonstrate the impact of system parameters. In this paper, the energy efficiency is defined as the total energy required to achieve a specific total average data rate, where the total energy includes the energy for the SCD process in addition to the energy for maintaining the UL. Note, the total energy consumption for the SCD process takes place at the UEs in the traditional system, while the total energy consumption for the SCD process in the proposed system (Trio-Connectivity TC) takes place at both the UE and the network. Therefore, in order to have a fair comparison between the traditional systems and the proposed system in this paper, two energy efficiency scenarios, UE energy efficiency and system energy efficiency, are considered.

Before we find the energy efficiency for each scenario, the power consumption and achievable data rate are found in Lemma 2 and Lemma 3.

Lemma 2 The total power consumption when associated to the kth tier is given by:

$$\mathbb{E}[P_k] = p_u \left(\frac{\Gamma\left(\frac{\alpha_k \epsilon + 2}{2}\right)}{(\pi \lambda_k)^{\frac{\alpha_k \epsilon}{2}}} \right) \tag{19}$$

Proof: Under assumption that the distance between $\mathcal{U}_0$ and the serving BS of the kth tier has a PDF of $f_{x_k}(x) = 2\pi\lambda_k e^{-2\pi\lambda_k x_k^2}$, the UL power consumption of $\mathcal{U}_0$ when associated to the kth tier is expressed as:

$$\mathbb{E}[P_k] = \int_0^\infty p_u x_k^{\alpha_k \epsilon} f_{x_k}(x) dx \tag{20}$$

The results in Eq. (19) is reached by using [29, 3.326.2]. ■

Lemma 3 Assuming $\mathcal{U}_0$ is at x_k distance from the serving cell of the kth tier, the achievable data rate for UL becomes:

$$\mathcal{R}_k = \int_0^\infty \int_0^\infty \mathscr{L}_{I_k}(.) d\tau \, f_{x_k}(x) dx_k \tag{21}$$

where $\mathscr{L}_{I_k}(.) = \exp\Big(- 2\pi\lambda_k \int_{x_k}^\infty \big(1 - \int_0^\infty \frac{2\pi\lambda_k \hat{z}_i e^{-\pi\lambda_k \hat{z}_i^2} d\hat{z}}{1 + s(\hat{z}_i^{\alpha_k})^\epsilon p_u z^{-\alpha_k}}\big) z \, dz \Big)$ and $f_{x_k}(x) = 2\pi\lambda_k x e^{-\pi\lambda_k x^2}$ is the PDF of the random variable x.

Proof: See Appendix B. ■

A. UE Energy Efficiency

The UE energy efficiency is defined as the total energy required to achieve a specific total UL average data rate. Due to limitation of power supply at the UEs, this metric is considered essential to study the impact of different system parameters on the energy efficiency at the UEs. It is assumed that the handover process between any two cells from different tiers is seamless and the short periods of zero rate are negligible. Therefore, the UE energy efficiency in the traditional system can be expressed as:

$$\begin{aligned} \Pi &= \frac{\mathcal{R}_{\mathcal{U}_0}}{E_{\mathcal{U}_0}} \\ &= \frac{\Upsilon_m B_m \mathcal{R}_m + \Upsilon_s B_s \mathcal{R}_s}{\mathbb{E}[\mathcal{T}_T]\big(\Upsilon_m \mathbb{E}[P_m] + \Upsilon_s \mathbb{E}[P_s]\big) + \mathbb{E}[E_{IRF}]} \end{aligned} \tag{22}$$

where $\mathcal{R}_{\mathcal{U}_0}$ is the total UL achievable data, $E_{\mathcal{U}_0}$ is the total energy consumption, B_m and B_s are the allocated bandwidths of first tier and second tier to each UE in traditional system respectively, $\mathbb{E}[\mathcal{T}_T]\Upsilon_m \mathbb{E}[P_m]$ and

$\mathbb{E}[\mathcal{T}_T]\Upsilon_s\mathbb{E}[P_s]$ are UL energy consumption when associated to the first tier and second tier respectively, and $\mathbb{E}[E_{IRF}]$ is the energy consumption to perform IRF scans in the SCD process. Both $\mathbb{E}[P_m]$ and $\mathbb{E}[P_s]$ are given by Eq. (19) and B_m and B_s are given in the next Lemma.

Lemma 4 The average allocated bandwidths to each UE associated to the first tier B_m and the second tier B_s are found as:

$$B_m = \frac{\lambda_m W_m}{\lambda_u \Upsilon_m} \tag{23}$$

$$B_s = \frac{\lambda_s W_s}{\lambda_u \Upsilon_s} \tag{24}$$

where W_s and W_m are the SC and MC bandwidths respectively.

Proof: The number of UEs associated to the first tier can be expressed by $\mathscr{A}\Upsilon_m\lambda_u$. The expected number of UEs associated to each MC is obtained by dividing the expected number of UEs associated to the first tier by the number of MCs in the system $\frac{\mathscr{A}\Upsilon_m\lambda_u}{\lambda_m\mathscr{A}}$ where $\mathscr{A}\lambda_m$ represents the expected number of MCs in the system. Note that UE being associated to one of the BSs does not necessarily mean that frequency resources are allocated to her incessantly due to limited bandwidth and a large number of UEs associated to each BS. For analysis simplicity, it is assumed that the average bandwidth assigned to each UE is obtained by dividing the available bandwidth by the expected number of UEs associated to one BS:

$$B_m^{2P} = \frac{\lambda_m W_m}{\lambda_u \Upsilon_m} \tag{25}$$

The average bandwidth assigned to each UE associated to each SC can be found similarly. The results in Eq. (23) and Eq. (25) are reached. ■

Previously, the impact of t_r on the total offloading loss was modeled. In the conventional system (periodic scan based systems, e.g. DC), most of the energy consumption for the SC discovery takes place at the UEs. The UEs perform the IRF scan frequently, stop scanning when being offloaded to the second tier and resume scanning when leaving the second tier coverage. The energy consumption of the discovery process at the UEs depends on the time that UEs spend associating to the first tier and the value of t_r. $\mathbb{E}[E_{IRF}]$ can be expressed as:

$$\mathbb{E}\big[E_{IRF}\big] = E_{scan}\mathcal{F}\frac{\mathbb{E}[\mathcal{T}_T]\Upsilon_m}{t_r} \tag{26}$$

where $\mathcal{F}$ is the number of high frequency channels deployed in the system, E_{scan} represents the energy needed for one IRF scan, and $\frac{\mathbb{E}[\mathcal{T}_T]\Upsilon_m}{t_r}$ represents the expected number of IRF scans that $\mathcal{U}_0$ performs on each frequency channel during its movement. Although minimizing t_r helps to exploit more potential offloading opportunities, it may maximize the energy consumption at the UEs and in the whole system significantly. The result in Eq. (26) also shows that the energy consumption is linearly proportional to the number of frequency channels and the UE density in the system for the same SC density and SC's footprint. Since the number of UEs in the system will increase and also there will be a need to deploy more frequency channels in the system, the energy consumption will degrade the energy efficiency in the whole system.

In the TC, the UEs will only perform the IRF scan after they have detected and discovered the surrounding SCs on the low frequency and when the offloading requirements are met. Therefore the total number of IRF scans will be minimized and also the energy consumption at the UEs will be reduced significantly. But the SCs in the TC will have to consume more energy to broadcast their information on two frequency channels. The UL energy efficiency in the proposed system is expressed as:

$$\begin{aligned}\bar{\Pi} &= \frac{\bar{\mathcal{R}}_{\mathcal{U}_0}}{\bar{E}_{\mathcal{U}_0}} \\ &= \frac{\bar{\Upsilon}_m\bar{B}_m\mathcal{R}_m + \bar{\Upsilon}_s\bar{B}_s\mathcal{R}_s}{\mathbb{E}[\mathcal{T}_T]\big(\bar{\Upsilon}_m\mathbb{E}[P_m] + \bar{\Upsilon}_s\mathbb{E}[P_s]\big) + \mathbb{E}[\bar{E}_{IRF}]}\end{aligned} \tag{27}$$

where $\bar{\mathcal{R}}_{\mathcal{U}_0}$ is the average total data rate achieved by $\mathcal{U}_0$ in the TC, $\bar{E}_{\mathcal{U}_0}$ is the total energy consumption at $\mathcal{U}_0$ in the TC, $\bar{\Upsilon}_m$ and $\bar{\Upsilon}_s$ represent the fractions of total time $\mathcal{U}_0$ associated to the first tier and second tier in the TC respectively. Note that both $\bar{\Upsilon}_m$ and $\bar{\Upsilon}_s$ are function of the IAF scan (t_a) instead of IRF scans (t_r) as the proposed SCD discovery is not based on frequent IRF scan as shown earlier. $\bar{B}_m = \nu B_m$ and $\bar{B}_s = B_s$ are the bandwidths allocated to $\mathcal{U}_0$ when associated to the first tier and second tier respectively, ν is the fraction of W_m used at the MCs and $1-\nu$ represents the fraction of W_m used for I-plane at the SCs where ν takes a value between 0 and 1, $\bar{E}_{IRF}$ is the energy consumption for performing IRF scans in the proposed system during the movement. Under the same assumption made in the offloading loss section, each UE discovers the SCs from the first IRF scans in their coverage, the expected value of $\bar{E}_{IRF}$ can be expressed as:

$$\mathbb{E}[\bar{E}_{IRF}] = \big(\mathbb{E}[N_s] - \mathbb{E}[N_{\bar{C}}]\big)E_{scan} \tag{28}$$

Unlike the energy consumption at the UE in the traditional system (Eq. 26), the energy consumption at the UE in the proposed system does not increase when the number of frequency channels deployed in the system increases. This eases one of the main challenges

in the traditional SCD and boosts the UL energy efficiency when deploying more frequency channels in the system.

B. System Energy Efficiency

Unlike the periodic mechanism where most of the energy consumption takes place at the UEs, most of the energy consumption in the proposed mechanism will take place at the network. Transferring the energy consumption from the UEs with limited power to the network with an unlimited power supply is considered highly desirable. However, for fair comparison, the total energy consumption in the system (at both UEs and network) for both mechanisms will also be considered in the system energy efficiency. The system energy efficiency is defined as the total energy required to achieve a specific average data rate in the whole system. Therefore, the system energy efficiency is expressed as:

$$\Pi^S = \frac{\sum_{i\in\Phi_u} \mathcal{R}_{\mathcal{U}_i}}{\sum_{i\in\Phi_u} E_{\mathcal{U}_i}} \tag{29}$$

In the traditional system, all the energy is consumed from the UE's battery. However, this is different from the proposed system when some of the energy consumption will be transferred to the network. Before we find the system energy efficiency in TC, the total energy consumption needs to be investigated. Some of the energy is spent at the network to help UEs discovering the surrounding SCs as explained earlier. The total system energy consumption is expressed as

$$E_{3p} = \sum_{i\in\Phi_u} \bar{E}_{\mathcal{U}_i} + \sum_{j\in\Phi_s} E_{IP,j} \tag{30}$$

where $E_{IP,j}$ is the total energy that the jth SC consumes to maintain the I-plane in TC. As explained in the previous section, each SC broadcasts its information on the low frequency simultaneously to form the I-plane. As a result, some energy consumption takes place at each SC and this energy consumption depends on the RS transmit power. The expectation of E_{IP} during the time $\mathcal{T}_T$ can be expressed as:

$$\mathbb{E}[E_{IP}] = p^{rs}_{s,1}\mathbb{E}[\mathcal{T}_T] \tag{31}$$

where $p^{rs}_{s,1}$ is RS transmit power from the SC on the low frequency. To ensure that low frequency at the SC (I-plane) covers the same area covered by the high frequency, the path-loss difference needs to be considered when obtaining the RS transmit power on the low frequency. Therefore, $p^{rs}_{s,1}$ is obtained in Eq. (18). The first term on the left hand side of Eq. (30) represents the energy consumption at the UEs while the second term represents the energy consumption at the network.

It can be observed that unlike in the DC, the energy consumption in the proposed mechanism is independent of the IRF scan periodicity (t_r), since it is based on the IAF scan and the UEs do not need to perform the IRF scan periodically. It is also observed that increasing the density of SCs will reduce the energy consumption in the periodic scan mechanism and increase the energy consumption in the proposed mechanism (at the network). However, the density of UEs will have a great impact on the energy consumption in the periodic scan mechanism as the number of UEs will be very large in the future cellular systems. Therefore, the system energy efficiency can be expressed as

$$\bar{\Pi}^S = \frac{\bar{\Upsilon}_m\bar{B}_m\mathcal{R}_m + \bar{\Upsilon}_s\bar{B}_s\mathcal{R}_s}{\mathbb{E}[E_{3p}]} \tag{32}$$

where $\mathbb{E}[E_{3p}] = \mathbb{E}[\mathcal{T}_T]\big(\bar{\Upsilon}_m\mathbb{E}[P_m] + \bar{\Upsilon}_s\mathbb{E}[P_s] + \frac{\lambda_s}{\lambda_u}p^{rs}_{s,1}\big) + \big(\mathbb{E}[N_{sc}] - \mathbb{E}[N_{\bar{C}}]\big)E_{scan}$. Unlike the energy consumption of the SCD process in the DC system, the energy consumption of the SCD process in the TC system is dependent on the SC density. This is because some of the energy consumption in the SCD process is brought to the network and each SC broadcasts its information on the low frequency in addition to the high frequency.

VII. Results

In this section, we present some representative results to show the impact of some parameters such as t_r and the SC density on the cellular system performance, and also to show a comparison between the conventional SCD and the proposed mechanism in terms of energy efficiency and energy consumption for the SCD process in a IRF HetNet. In the current mobile systems, the IAF scan (t_a) is performed every 200 ms for channel estimation and inter-cell handover [4], [9]. Since the SCD mechanism is based on the IAF scan, the UEs in the TC will also use these scans for discovering SCs by detecting the I-plane at each SC. In the conventional system, the UEs have to perform IRF scans in addition to IAF scans to discover the surrounding SCs. Therefore, different values of t_r will be considered for achieving different system performance. The different system parameters can be found in Table I, unless given otherwise.

The analysis in this paper is validated through simulations. Figure 6 shows offloading opportunities and offloading loss for different values of SC density and t_r. It can be seen from the figure that offloading opportunities increase when the density of SCs increases in the system. The offloading loss decreases when the value

TABLE I: System Parameters

Parameter	Symbol	Value
Minimum received power	ρ_{min}	-90 dBm
MBS density	λ_1	0.2 MBS/Km^2
SBS density	λ_2	30 SBS/Km^2
Path-loss exponent	α_k	4
RSs transmit power from F_2	$p^{rs}_{s,2}$	33 dBm
RSs transmit power from F_1	$p^{rs}_{s,1}$	19 dBm Eq. (18)
UE baseline transmit power	p_c	20 dBm
UE speed	v	5 Km/h
Number of high frequency channels	$\mathcal{F}$	1
Minimum time for beneficial offloading	t_{min}	0
Low frequency	F_1	2 Ghz
High frequency	F_2	10 Ghz
Power control factor	ϵ	0.5
Energy cost for one IRF scan	E_{scan}	2.25 mJ [9]
UE density	λ_u	200 UEs/Km^2
Fraction of bandwidth at MCs	ν	0.5
MCs bandwidth	W_m	5 MHz
SCs bandwidth	W_s	20 MHz

of t_r decreases, for instance, the UEs will miss small fraction of the offloading opportunities in the system and the offloading loss is minimized when the value of t_r is low (e.g $t_r = 1$ sec), while they miss more offloading opportunities and the offloading loss is maximized when the value of t_r is high (e.g $t_r = 40$ sec). This is because the gap between any two consecutive IRF scans increases when adopting high value of t_r and the likelihood to miss more offloading opportunities increases.

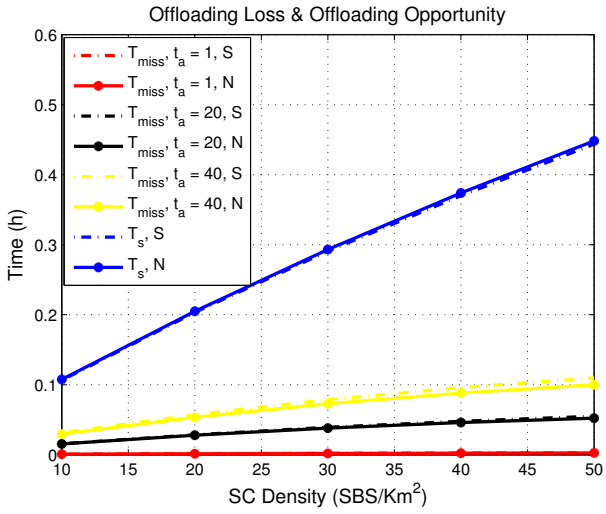

Fig. 6: Offloading opportunity and offloading loss for different values of λ_s and t_r.

Although more potential offloading opportunities are added to the system when adding more SCs in the system, the UEs may miss a greater fraction of these opportunities or consume more energy for performing IRF scans if the value of t_r is not set correctly. It can be seen from Figure 6 that lower values of t_r help to exploit more offloading opportunities, however, choosing small values of t_r means performing more IRF scans per unit time and as a result more energy consumption takes place at the UE as shown in Figure 7. The UEs consumes more energy when t_r set to 1 sec and they save significant amount of energy when t_r is set to 40 sec. It can also be seen from Figure 7 that energy consumption is minimized when the density of SCs increases for the same value of t_r, this is because a smaller number of IRF scans per unit time takes place due to more SCs are discovered and the UEs stay associated longer time to these SCs.

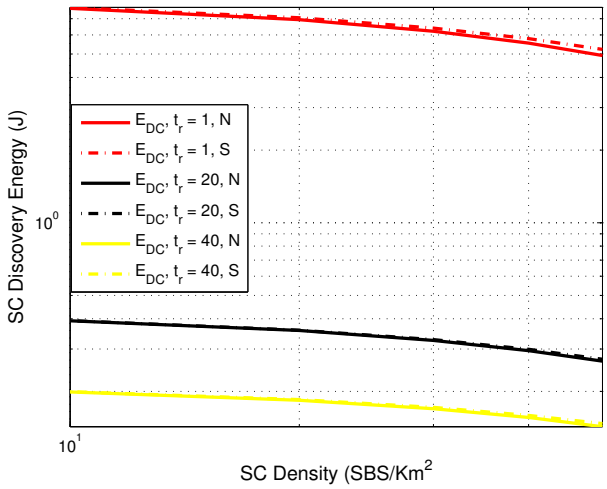

Fig. 7: Energy consumption in the SCD of the traditional system for different values of λ_s and t_r.

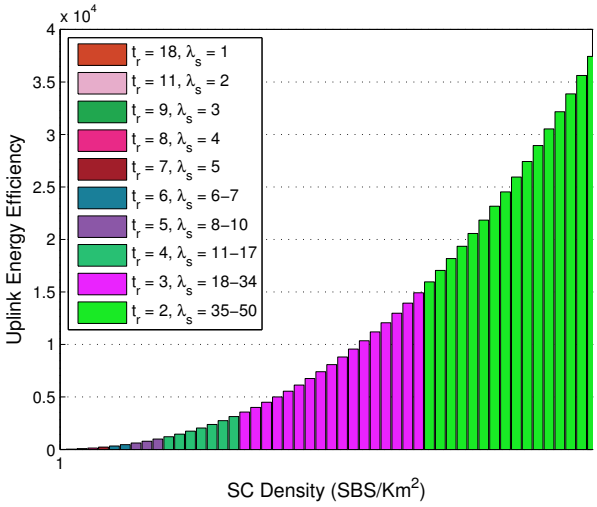

Fig. 8: Optimum value of t_r for different SC densities.

It is also seen from both Figure 6 and Figure 7 that there is trade-off between the offloading loss and exploiting the offloading opportunities in cellular systems. Therefore, the energy efficiency is considered in this paper as a performance metric to evaluate the performance of the system parameters e.g t_r. Fig 8 shows the energy

efficiency in the traditional system where the UL energy efficiency is improved when the density of SCs increases, not only more offloading opportunities available in the system with higher SC density, but also lower energy consumption takes place at the UE. The UEs spend less power energy consumption for performing IRF scans when the SC density increases for the same value of t_r as shown in Figure 7. Since the UEs spend more time associating to the SCs due to increment in the SC density, energy consumption to maintain the UL is minimized as the UEs are associated to the SCs with shorter distances in comparison to the MCs that are placed at longer distances. It is also seen that there is only one value of t_r to achieve the best system performance for different SC densities. For instance, when the SC density is in the range of $50 - 35SBS/Km^2$, the best performance (the highest energy efficiency) is achieved when t_r is set to 2 sec. While the best energy efficiency for lower SC density is obtained when t_r is higher. For example, the optimum value of t_r is 18 sec when the density of SCs is $1SBS/Km^2$.

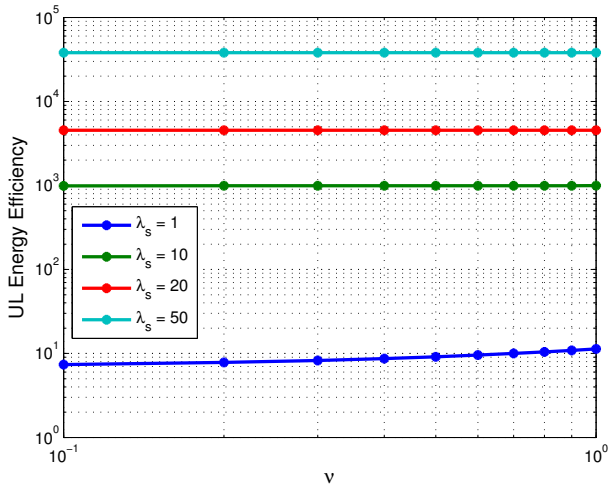

Fig. 9: TC energy efficiency for different values of ν and λ_s.

Before we investigate the TC performance in comparison to the traditional system performance, the impact of parameter ν on the TC performance since fraction of the low frequency band will be used at each SC to form the I-plane. Figure 9 shows the energy efficiency in the proposed system (3P) for different values of ν. It can be seen that the impact of ν is very limited for different SC densities. For instance, when the SC density is set to $50SBS/Km^2$, the energy efficiency is not affected, this is because of the limited frequency resources at the MCs being used to serve a large number of UEs. While the UEs can spend longer time associated to the SCs where the frequency resources are shared by smaller number of UEs. Therefore, using part of the low frequency band to form the I-plane will not affect the system performance in the medium and high dense networks. However, the impact of ν increases when the density of SC is very low. For instance, when the SC density is $1SBS/Km^2$, the energy efficiency slightly increases with higher values of ν until it reaches the highest value at $\nu = 1$. This is reasonable due to the UEs spend longer time associated to the MCs and using the whole bandwidth at the MCs will provide slightly better energy efficiency when the offloading opportunities are very few in the system (low SC density).

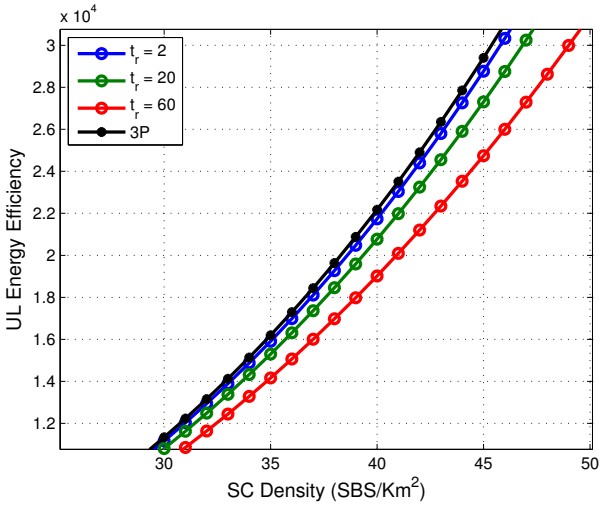

Fig. 10: Energy efficiency comparison between the proposed system (3P) and the traditional system (2P) with different values of t_r.

Figure 10 shows a comparison between the traditional system (2P) and the proposed system (3P) in terms of energy efficiency. It can be seen in this figure that 3P can achieve better energy efficiency for medium and high SC densities. Such performance improvement is due to saving some of the energy unnecessarily consumed during SC discovery process and being associated to the first tier for longer period. One of the main purposes to propose this system is that it is necessary to minimize the energy consumption and boost the energy efficiency at the UE due to the limitation of the UE's battery capacity. Therefore, Figure 11 illustrates the difference between the traditional system and the proposed system in terms of energy consumption at the UE. Although the energy consumption in the traditional system decreases when the SC density increases because it spends more time associated to the SCs, the energy consumption in the 3P is less without trade-off in investing the offloading opportunities. It can be seen from Figure 11 also shows that the energy consumption for the SC discover process is doubled when the number of high frequency chan-

nels becomes 2. This is because the UEs will have to perform IRF scans on two different frequency channels in order to discover all the SCs in the system prior to offloading to the discovered SCs. Note that increasing the number of frequency channel will not maximize the energy consumption in the 3P, this is because the SC discovery process occurs on one frequency channels (low frequency) regardless the number of frequency channels in the system.

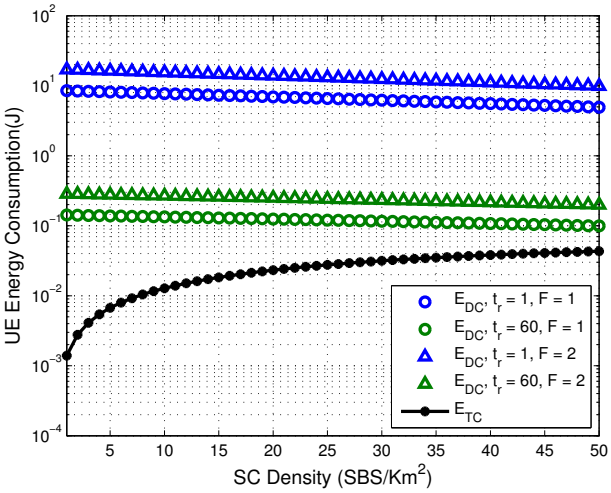

Fig. 11: Energy consumption at the UE.

As explained earlier that in the proposed system, some of the energy consumption will be transferred from the UE with limited battery capacity to the network with unlimited power supply. Although it is very desirable to minimize the energy consumption at the UE, the system energy efficiency is also addressed. Figure 12 shows a comparison between the two systems in terms of system energy efficiency. The system performance in the proposed system is slightly better than one in the traditional system. The difference in the system performance between the two systems increases when the density of UEs in the system increases. When the number of UEs increases, more energy consumption is caused in the SC discovery process in 2P, however, the increase of the energy consumption in the 3P is limited when the density of SCs is fixed. It is also seen that the system performance is degraded in both systems when the density of UEs increases. The reason is that there is an increment in the energy consumption due to more UEs deployed in the system while the system frequency resources are fixed.

VIII. Conclusion

In this paper, the three-plane system (Trio-connectivity), including C-plane, U-plane and I-plane,

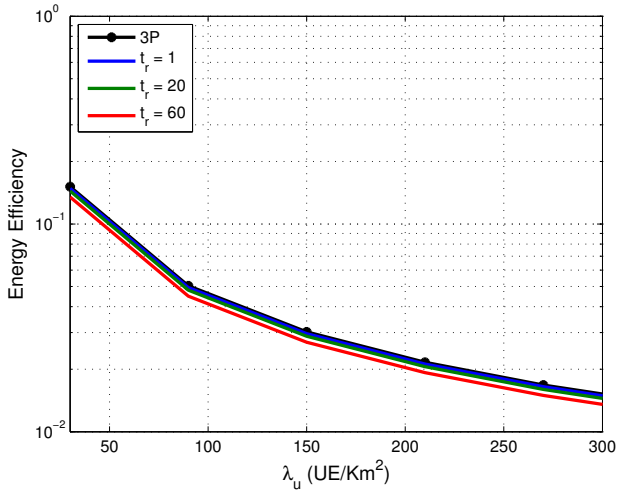

Fig. 12: System energy efficiency comparison between the proposed system (3P) and the traditional system (2P) for different values of t_r and λ_u.

has been proposed to enhance both the mobility management and offloading process for the future mobile HetNet. A system design (i.g each plane's role and functions, and signaling diagram of the proposed offloading mechanism) was also presented in this paper. The stochastic geometry tool was used to model the total offloading opportunity, the total offloading loss, average achievable rate and energy consumption for the SCD process in a two-tier HetNet. The TC and the traditional system (e.g DC) were compared in terms of offloading loss, UL energy efficiency and system energy efficiency and energy consumption. The simulation results showed that the TC saves not only power at the UE but also power in the offloading process as a whole in the high dense HetNet with no trade-off in exploiting the offloading opportunities in the system. Deploying more frequency channels in the system can increase the energy consumption and degrade the energy efficiency in the DC. But the energy consumption in the proposed mechanism is not affected when more frequency channels are deployed, as the offloading process in the TC (i.e discovering and identifying the SCs) is kept on one frequency regardless of the number of frequency channels.

Appendix A
Proof of Lemma 1

The expected coverage of the ith SC crossed by $\mathcal{U}_0$ is $\mathbb{E}[C_i] = \int_0^\infty c f_{C_i}(c) dc$, where $f_{C_i}(c)$ is the probability density function (PDF) of the ith SC coverage and is

found by using the transforming density function:

$$\begin{aligned} f_{C_i}(c) &= f_{\eta_i}(\eta(C_i))\,|\frac{d\eta}{dc}| \\ &\overset{(a)}{=} \frac{1}{r_i}\,\frac{d}{dc}\big(r_i\cos(\sin^{-1}(\frac{c}{2r_i}))\big) \\ &= \frac{c}{4r_i^2\sqrt{1-\frac{c^2}{4r_i^2}}} \end{aligned} \tag{A1}$$

where $f_{\eta_i}(\eta) = 1/r_i$ is the PDF of the distance between the ith SC and the $\mathcal{U}_0$'s path, $f_{\eta_i}(\eta(C_i)) = 1/r_i$, and (a) follows from $\eta_i = r_i\cos(\sin^{-1}(\frac{c}{2r_i}))$ by using Eq. (4). Since the coverage of the SCs crossed by $\mathcal{U}_0$ are uncorrelated and have the same distribution, the total expected value of the second tier coverage on the $\mathcal{U}_0$'s path is the summation of the individual expected coverages of SCs (linearity of expectation). When $r_i = r$, $\forall i$, the expected total coverage on the path can be expressed as:

$$\mathbb{E}[C_T] = \mathbb{E}[N_{sc}]\int_0^{r_i} c f_{C_i}(c)dc \tag{A2}$$

where N_{sc} is the number of SCs crossed by $\mathcal{U}_0$. Consider d_i as the nearest distance between $\mathcal{U}_0$'s path and the ith SC with radius r_i. The number of SCs crossed by $\mathcal{U}_0$ during its movement can be expressed as:

$$N_{sc} = \sum_{i\in\Phi_2}\mathbf{1}(d_i \le r_i)$$

where $\mathbf{1}(.)$ is the indicator function. When $\mathscr{A}_{X_D-X_S} = 2rD_{X_D-X_S}$ denotes the area surrounding the $\mathcal{U}_0$'s path, any SC will be crossed by $\mathcal{U}_0$ if its SBS is located in this area. Since SBSs are distributed as PPP, the expected number of SCs crossed by $\mathcal{U}_0$ is obtained as:

$$\mathbb{E}[N_{sc}] = \lambda_2\mathscr{A}_{X_D-X_S} \tag{A3}$$

The expectation in Eq. (5) is reached after substituting Eq. (A3) in Eq. (A2).

Appendix B
Proof of Lemma 3

The average ergodic rates of $\mathcal{U}_0$ associated to the kth tier can be expressed as:

$$\mathcal{R}_k = \int_0^\infty \mathbb{E}_{SINR}\big[\ln(1+SINR_k(x))\big]\,f_{x_k}(x)dx \tag{B1}$$

where $\mathbb{E}_{SINR_k}\big[\ln(1+SINR_k)\big]$ is obtained as:

$$\begin{aligned} &= \int_0^\infty \mathbb{P}\big[\ln(1+SINR_k(x_k)) > \tau\big]\,d\tau \\ &= \int_0^\infty \mathbb{P}\Big[\ln\big(g_0 > \frac{x_k^{\alpha_k(1-\epsilon)}(\sigma^2+I_k)(\mathrm{e}^\tau-1)}{\mathcal{L}_f p_u}\big)\Big]d\tau \\ &= \int_0^\infty e^{\frac{-x_k^{\alpha_k(1-\epsilon)}\sigma^2(\mathrm{e}^\tau-1)}{p_u\mathcal{L}_f}}\,\mathscr{L}_{I_k}\big(\frac{x_k^{\alpha_k(1-\epsilon)}(\mathrm{e}^\tau-1)}{p_u\mathcal{L}_f}\big)d\tau \end{aligned} \tag{B2}$$

where τ is the threshold of the achievable data rate, $I_k = \sum_{i\in\Phi_k}\bar{P}_i(\hat{z}_i)h_i\bar{L}_{f,i}(z_i)$ is the interference from other UEs, $\mathscr{L}_k(.)$ is the Laplace transform of the cumulative interference from UEs in UL. $\mathscr{L}_{I_k}(s) = \mathbb{E}_{I_k}\left[\mathrm{e}^{-sI_k}\right]$ can be obtained as:

$$\begin{aligned} \mathscr{L}_{I_k}(s) &= \mathbb{E}_{I_k}\left[\mathrm{e}^{-sI_k}\right] \\ &\overset{(a)}{=} e^{-2\pi\lambda_k\int_{x_k}^\infty\left(1-\mathbb{E}_{\hat{z}}[\frac{1}{1+s(\hat{z_i}^{\alpha_k})^\epsilon p_u z^{-\alpha_k}}]\right)z\,dz} \\ &\overset{(b)}{=} e^{-2\pi\lambda_k\int_{x_k}^\infty\left(1-\int_0^\infty\frac{2\pi\lambda_k\hat{z_i}e^{-\pi\lambda_k\hat{z_i}^2}d\hat{z}}{1+s(\hat{z_i}^{\alpha_k})^\epsilon p_u z^{-\alpha_k}}\right)z\,dz} \end{aligned} \tag{B3}$$

where (a) is obtained similar to [30]. In the interference-limited network ($\sigma^2 = 0$), Eq. (21) is obtained when substituting Eq. (B3) in Eq. (B1)

References

[1] H. Zhu and J. Wang, "Chunk-Based Resource Allocation in OFDMA Systems - Part I: Chunk Allocation," *IEEE Transactions on Communications*, vol. 57, pp. 2734 – 2744, Sep. 2009.

[2] H. Zhu and J. Wang, "Chunk-Based Resource Allocation in OFDMA Systems - Part II: Joint Chunk, Power and Bit Allocation," *IEEE Transactions on Communications*, vol. 60, pp. 499 – 509, Feb. 2012.

[3] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, "What Will 5G Be?," *IEEE Journal on Selected Areas in Communications*, vol. 20, pp. 1056–1082, Jun 2014.

[4] 3GPP TS 36.133 ver 12.7.0 Release 12, "Requirements for Support of Radio Resource Management," 06 2015.

[5] 3GPP TS 36.331 ver 13.1.0 Release 13, "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA)," 3 2016.

[6] O. Onireti, A. Imran, M. A. Imran, and R. Tafazolli, "Energy efficient inter-frequency small cell discovery in heterogeneous networks," *IEEE Transactions on Vehicular Technology*, vol. 65, pp. 7122–7135, Sept 2016.

[7] A. Prasad, O. Tirkkonen, P. Lunden, O. N. Yilmaz, L. Dalsgaard, and C. Wijting, "Energy-Efficient Inter-Frequency Small Cell Discovery Techniques for LTE-Advanced Heterogeneous Network Deployments," *IEEE Communications Magazine*, vol. 51, pp. 72–81, May 2013.

[8] A. Mahbas, H. Zhu, and J. Wang, "The Optimum Rate of Inter-Frequency Scan in Inter-Frequency HetNets," *to appear in IEEE International Conference on Communications (ICC'17)*, May 2017.

[9] 3GPP TR 36.839 V11.1.0 Release 11, "Mobility Enhancements in Heterogeneous Networks," Sep 2012.

[10] S. Singh, H. S. Dhillon, and J. G. Andrews, "Offloading in Heterogeneous Networks: Modeling, Analysis, and Design Insights," *IEEE Transactions on Wireless Communications*, vol. 12, pp. 2484–2497, May 2013.

[11] H.-S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous Cellular Networks with Flexible Cell Association: A Comprehensive Downlink SINR Analysis," *IEEE Transactions on Wireless Communications*, vol. 11, pp. 3484–3495, Oct. 2012.

[12] A. Prasad, P. Lundn, O. Tirkkonen, and C. Wijting, "Energy-Efficient Flexible Inter-Frequency Scanning Mechanism for Enhanced Small Cell Discovery," *IEEE 77th Vehicular Technology Conference: VTC2013-Spring*, pp. 1–5, 2013.

[13] R. Mondal, J. Turkka, T. Ristaniemi, and T. Henttonen, "Positioning in Heterogeneous Small Cell Networks using MDT RF Fingerprints," *2013 First International Black Sea Conference on Communications and Networking (BlackSeaCom)*, pp. 127 – 131, 2013.

[14] A. Prasad, P. Lundn, O. Tirkkonen, and C. Wijting, "Energy Efficient Small-Cell Discovery Using Received Signal Strength Based Radio Maps," *IEEE 77th Vehicular Technology Conference: VTC2013-Spring*, pp. 1–5, 2013.

[15] D. Stoyan, W. S. Kendall, and J. Mecke, *Stochastic Geometry and Its Applications*. John Wiley and Sons Ltd., 1995.

[16] A. Mahbas, H. Zhu, and J. Wang, "Unsynchronized Small Cells with a Dynamic TDD System in a Two-Tier HetNet," *IEEE 83rd Vehicular Technology Conference: VTC2016-Spring*, pp. 1–5, 2016.

[17] H.-S. Jo, Y. J. Sang, P. Xia, and J. G. Andrews, "Heterogeneous Cellular Networks with Flexible Cell Association: A Comprehensive Downlink SINR Analysis," *IEEE Transactions on Wireless Comm.*, vol. 11, pp. 3484–3495, Oct. 2012.

[18] D. B. Johnson and D. A. Maltz, "Dynamic Source Routing in Ad Hoc Wireless Networks," *Mobile Computing*, pp. 153 – 181, 1996.

[19] C. Bettstetter, H. Hartenstein, and X. Perez-Costa, "Stochastic Properties of the Random Waypoint Mobility Model," *Wireless Networks*, vol. 10, p. 555567, Sep. 2004.

[20] S. Shin, U. Lee, F. Dressler, and H. Yoon, "Analysis of Cell Sojourn Time in Heterogeneous Networks With Small Cells," *IEEE Communication Letters*, vol. 20, pp. 788 – 791, Apr. 2016.

[21] X. Lin, R. K. Ganti, P. J. Fleming, , and J. G. Andrews, "Towards Understanding the Fundamentals of Mobility in Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 12, pp. 1686–1698, Apr. 2013.

[22] A. J. Mahbas, H. Zhu, and J. Wang, "Impact of small cells overlapping on mobility management," *IEEE Transactions on Wireless Communications*, vol. 18, pp. 1054–1068, Feb 2019.

[23] H. Ishii, Y. Kishiyama, and H. Takahashi, "A Novel Architecture for LTE-B : C-plane/U-plane Split and Phantom Cell Concept," *2012 IEEE Globecom Workshops (GC Wkshps)*, pp. 624 – 630, 2012.

[24] A. Mohamed, O. Onireti, M. A. Imran, A. Imran, and R. Tafazolli, "Control-data separation architecture for cellular radio access networks: A survey and outlook," *IEEE Communications Surveys Tutorials*, vol. 18, pp. 446–465, Firstquarter 2016.

[25] H. Ishii, Y. Kishiyama, and H. Takahashi, "A novel architecture for lte-b," in *Globecom Workshops, 2012 IEEE*, pp. 624–630, 2012.

[26] H. Song, X. Fang, and L. Yan, "Handover Scheme for 5G C/U Plane Split Heterogeneous Network in High-Speed Railway," *IEEE Transactions on Vehicular Technology*, vol. 63, pp. 4633 – 4646, Nov. 2014.

[27] 3GPP TS 36.211 ver 13.1.0 Release 13, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical channels and modulation," 04 2016.

[28] 3GPP TS 36.213 ver 13.0.0 Release 13, "Evolved Universal Terrestrial Radio Access (E-UTRA); Physical layer procedures," 05 2016.

[29] I. Gradshteyn and I. Ryzhik, *Table of Integrals, Series, and Products*. San Diego, USA: ELSEVIER, 2007.

[30] T. D. Novlan, H. S. Dhillon, and J. G. Andrews, "Analytical Modeling of Uplink Cellular Networks," *IEEE Transactions on Wireless Communications*, vol. 12, pp. 2669–2679, Jun. 2013.