A Novel Load Balancing Scheme for Cellular-WLAN Heterogeneous Systems with Cell Breathing Technique.

Shobanraj Navaratnarajah*, Mehrdad Dianati† *Senior Member, IEEE,*
and Muhammad Ali Imran‡ *Senior Member, IEEE*
*Institute for Communication Systems, University of Surrey, Guildford, Surrey, United Kingdom.
†Warwick Manufacturing Group, University of Warwick, Coventry, United Kingdom.
‡School of Engineering, University of Glasgow, Glasgow, United Kingdom.
Email: s.navaratnarajah@surrey.ac.uk, m.dianati@surrey.ac.uk, m.imran@surrey.ac.uk

***Abstract*—This paper proposes a novel load balancing scheme, for an operator-deployed cellular-WLAN heterogeneous network (HetNet), where the user association is controlled by employing cell breathing technique for the WLAN network. This scheme, eliminates the complex coordination and additional signalling overheads between the users and the network, by allowing the users to simply associate with the available WLAN networks similar to the traditional WLAN-first association, without making complex association decisions. Thus, this scheme can be easily implemented in an existing operator-deployed cellular-WLAN HetNet. The performance of the proposed scheme is evaluated in terms of load distribution between cellular and WLAN networks, user fairness, and system throughput, which demonstrates the superiority of the proposed scheme in load distribution and user fairness while optimising the system throughput. In addition, a cellular-WLAN interworking architecture and signalling procedures are proposed for implementing the proposed load balancing schemes in an operator-deployed cellular-WLAN HetNet.**

***Index Terms*—Heterogeneous Networks, Multi-RAT, Cellular, WLAN, Load Balancing, Cell Breathing.**

I. Introduction

The explosive growth of wireless data-traffic causes huge challenges for the wireless network operators to significantly increase the network capacity. One of the most promising solutions for this challenge is the *heterogeneous network* (HetNet) architecture. A HetNet may consist of different sizes of cells with different radio access technologies (RATs), with overlapping coverage that complement each other [1]. In this regard, off-loading mobile data traffic to Wireless Local Area Network (WLAN) becomes increasingly popular among cellular network operators. This is due to the fact that the utilisation of unlicensed spectrum in WiFi brings additional bandwidth resources, instead of sharing the much scared and costly cellular frequency spectrum for small cells. Thus, cellular network operators have already started using WiFi to meet the capacity demands in their cellular networks [2]. However, there are many challenges related to the operation of such integrated cellular-WLAN HetNet. One of the major challenges is optimally balancing the load between cellular and WLAN networks [3]–[5].

Traditionally, in cellular-WLAN HetNet, the user equipment (UE) always tries to connect to WiFi, if there is a WiFi coverage available, regardless of the status of the WiFi network, which is known as *WLAN-first* network selection [6]. This kind of simple scheme can be beneficial to legacy cellular networks (e.g., 2G and 3G networks), which have relatively low system capacity compared to WiFi. However, capacity of the latest cellular network (e.g., LTE) is relatively higher. Moreover, latest cellular networks are relatively more efficient in terms of the spectral-efficiency (SE). Therefore, off-loading all data traffic to WiFi in the dual coverage area may not always be advantageous; especially, when the user density is much higher under the WiFi coverage.

In order to overcome the shortcoming of mainstream WLAN-first scheme, some load balancing schemes have been proposed in literature. Notably, a load balancing scheme based on fuzzy logic algorithm is proposed in [3] for a 3G and WLAN HetNet, which focuses on the utilisation fairness. In [4], the authors propose a load balancing algorithm to improve network utilisation and call blocking probability. A policy based resource management framework is presented in [5] for cellular-WLAN integrated networks, which improves the network utilisation by dynamically balancing the offered traffic load via admission control and vertical handoff.

However, most of the existing schemes are user-centric. Although, such distributed user-centric schemes can be advantageous for uncoordinated cellular-WLAN HetNet, they impose additional complexity to the limited hardware resources on user terminals. In addition, such schemes add more signalling overhead to the network and the scared wireless link, in order to enable distributed decision making process. However, in a coordinated (i.e., an operator-deployed) cellular-WLAN HetNet, it is beneficial to balance the network load in a centralised network-centric scheme. This is due to the fact that the up-to-date network and user related information can be efficiently gathered within the network without overloading the scared wireless link. In addition, the network-centric approach alleviates complex decision making process on the UE, which has much less hardware resources compared to the network equipment (NE).

Recently, in [7] the authors investigate the application of Cell Range Expansion (CRE) in WiFi network to optimise the load balance in a cellular-WiFi HetNet. The CRE technique is primarily developed to balance the load in single-RAT HetNet (e.g., LTE macro-femto HetNet), where small-cells in a single RAT HetNet have relatively very small coverage compared to the macro-cells; and thus, attract very few users. In such circumstances, the load can be balanced by applying some biasing to the received power from small-cells (i.e, the received signal from the small cell is multiplied by a bias-factor before comparing it for cell association), which is referred to as CRE [8], [9].

However, applying CRE in multi-RAT HetNet such as cellular-WLAN HetNet is questionable, due to the fact that this technique can only work for a system where the user terminal compares the received signal strength (RSS) from all potential point of attachments (POAs) in order to make the association decision. However, just comparing RSS to make association decision in multi-RAT HetNet is not practical due to the fact that in reality, different RATs employ different multiple access schemes at MAC layer, even though they may employ the same physical layer (PHY) technique (e.g., OFDM-PHY in LTE and WiFi). The difference between cellular and WLAN in MAC layer techniques has a considerable impact on the users' perceived throughput as well as on the system performance of each RAT. In addition, the study in [7] does not consider hot-spot deployment of WiFi APs. Moreover, it assumes that the WiFi access points (APs) operate same as femto-cells except they utilise different carrier frequency. However, since WiFi is a random access technology, these assumptions are questionable in practical cellular-WLAN HetNet.

Considering the above limitations in the existing solutions, in this paper, we propose a novel load balancing scheme, where the user association is controlled by employing cell breathing technique to the WLAN networks. In this scheme, the coverage of the WiFi AP is optimally adjusted according to the objective of the optimisation. Thus, the users can associate to the available strongest WiFi AP similar to the traditional WLAN-first scheme, without making any complex network selection decisions.

The rest of the paper is organised as follows. Section II describes the considered system model and user association schemes in detail. The optimisation problem is formulated in Section III. A suboptimal heuristic algorithm with less computational complexity is proposed in Section IV. In Section V, the performances of the considered association schemes are compared and analysed. Then in Section VI, a practical implementation scenario of the proposed scheme is presented. Finally, Section VII concludes the paper.

II. System Model

In this section, first, we discuss the considered cellular-WLAN HetNet and the main assumptions in detail; then the considered user association method and the user throughput estimation model are provided. Fig. 1 depicts the considered system model. The main assumptions are as follows:

1) Network architecture: A cellular-WLAN HetNet that comprises of Orthogonal Frequency Division Multiplexing (OFDM) based cellular macro-cells overlaid with the coverage of multiple WiFi APs per macro-cell. The macro base stations

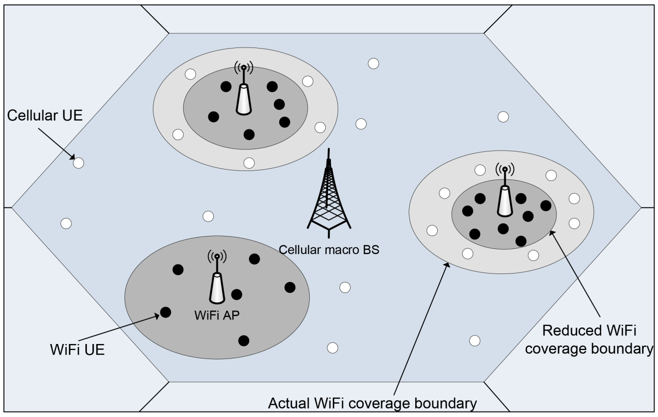

Fig. 1: Considered system model

(BSs) are deployed in a hexagonal grid with a fixed inter site distance (ISD) of D. The WiFi APs are randomly deployed with the average density of λ_w APs per macro-cell. Typically, in an operator-deployed cellular-WLAN HetNet, the WiFi APs are deployed in areas where the user density is high (i.e., hot-spot zones), in order to ease traffic burden on the cellular network. Hence, the users in the coverage area are divided into two categories: one is the *hot-spot users* who have both cellular and WiFi coverage; and the other one is *non-hot-spot users* who have only cellular coverage.

2) User distribution: The users are assumed to be randomly deployed with the total density of λ_u per macro-cell, with the ratio of hot-spot users to non-hot-spot users being r. This scenario coincides with the evaluation scenario 2a of 3GPP specification TR 36.872 [10].

3) Traffic model: According to Ericsson Mobility Report [11], by the end of 2016 mobile data traffic accounted for more than 95% of total mobile traffic (voice + data). The data traffic grew around 50% year-on-year from 2011 to 2016, and it is expected to grow around the same rate in the future. At the same time, the growth of voice traffic was almost flat during the same period, and expected to be same in the future. In addition, most of the mobile data traffic is best-effort traffic such as video, audio, social networking, web browsing, software download, and file sharing. Only a small fraction of data traffic is real-time (e.g., VoIP). Thus, in our work we only consider best-effort data traffic, which is some times referred to as *saturated* or *full-buffer* traffic. We consider a downlink scenario, where each user downloads data from its serving cell (i.e., BS or AP) at a maximum achievable data-rate.

4) Resource allocation in cellular system: Typically, in cellular network, the resources are allocated by a centralised network entity, such as BS. Thus, it is assumed that the system bandwidth in cellular network is allocated by the BS according to the underlying resource allocation scheme (i.e., according to the outcome of optimisation process).

5) Resource allocation in WiFi system: In contrast to the cellular network, in WiFi, the resources are randomly accessed by the UEs. Thus, it is assumed that WiFi operates under the default random access scheme, called Distributed Coordination Function (DCF).

A. User association

In this subsection, first, we introduce the notations that are used for the following analysis. To this end, we denote $\mathcal{C} = \{1, ..., M\}$ and $\mathcal{A} = \{1, ..., N\}$ be the set of cellular macro-cells and WiFi APs in the considered coverage area, respectively. Let $\mathcal{U} = \{1, ..., K\}$ denote the set of users within the considered coverage area. In our system model, there is a total number of $(M + N)$ POAs (i.e., BS or AP), which refers to both macro-cells and the WiFi access points. We use index i to identify each of these POAs. To this end, $i = 1, ..., M$ are used for macro-cells, while $i = M+1, ..., M+N$ are used for WiFi access points. We assume that a user is associated to a single POA at a time (also referred to as unique association). A binary variable $x_{i,k}$ indicates the user association such that $x_{i,k} = 1$ when the user k is associated with POA i, otherwise $x_{i,k} = 0$.

In this paper, the user association is considered to be similar to the traditional association scheme for cellular-WLAN HetNet, called WLAN-first. In WLAN-first scheme, the users always select WiFi whenever the WiFi coverage is available [6]. However, the association region of a WiFi AP can be controlled by cell breathing, where the power of beacon signal of the AP is reduced, thus limiting the coverage of the AP without reducing the transmit power of the traffic signal [12]. This will allow the AP to control the number of users associated with it, without suffering degraded signal quality. Fig. 1 shows an example of congested WiFi AP limiting the number of associated users with cell breathing.

The following analysis describes the above user association for the considered system model. Let $P^r_{i,k}$ represent the received power of user k from POA i which is given as

$$P^r_{i,k} = P^t_i g_{i,k}, \tag{1}$$

where P^t_i is the transmit power of POA i and $g_{i,k}$ is the channel power gain from the POA i to user k, which is given by

$$g_{i,k} = L_0 \left(\frac{d_{i,k}}{d_0} \right)^{-\alpha_i}, \tag{2}$$

where L_0 is the path loss at the reference distance d_0, α_i is the pathloss exponent, and $d_{i,k}$ is the distance from POA i to user k. The reference path loss L_0 is typically considered as a free space path loss at unit distance; hence,

$$L_0 = \left(\frac{\omega}{4\pi} \right)^2, \tag{3}$$

where ω is the wavelength of the signal. The above channel model is a simple yet sufficient tool to characterise the multi-RAT system for trade-off analysis without the complexity of small-scale power fluctuations caused by shadowing and fast fading, thus leading to simpler analysis [13]. However, the short term fading effects can be taken into account by introducing an appropriate fading margin or an additional random variable directly into the channel power gain function (2).

From the received power obtained by (1), the POA of user k can be identified as follows:

$$POA_k = \begin{cases} \arg \max\limits_{i: i \in \mathcal{A}} y_i P^r_{i,k}, & \text{if } \max(y_i P^r_{i,k}) \geq P^r_{th} \ \forall i \in \mathcal{A}, \\ \arg \max\limits_{i: i \in \mathcal{C}} P^r_{i,k}, & \text{otherwise}, \end{cases} \tag{4}$$

where y_i (≤ 1) is the power reduction factor that is used to reduce the beacon power of the AP i. P^r_{th} represents the minimum required received power threshold (i.e., receiver sensitivity) to successfully establish the connection with the WiFi AP.

B. User throughput

In this subsection, we model the user achieved throughput. First, lets find the value of the association indicator $x_{i,k}$ $\forall i, k$, which is given as

$$x_{i,k} = \begin{cases} 1, & \text{if } POA_k = i, \\ 0, & \text{otherwise}. \end{cases} \tag{5}$$

Since the resource allocation amongst the attached users in cellular and WiFi is different (i.e., centralised scheduled resource allocation in macro-cell, while the resources are randomly accessed in WiFi), the user achieved throughput model for cellular and WiFi will be different. Therefore, first, we model the user throughput of a user associated with a cellular BS. Hence, for the analysis that follows, let $\gamma_{j,k}$ represent the perceived Signal to Interference plus Noise Ratio (SINR) of the user k associated with the macro-cell $j \in \mathcal{C}$, which is given by

$$\gamma_{j,k} = \frac{P^r_{j,k}}{I_{j,k} + \sigma^2_{j,k}}, \tag{6}$$

where $I_{j,k}$ is the total interference power and $\sigma^2_{j,k}$ is the noise power. Considering frequency re-use of one for the macro-cells, the interference power will be

$$I_{j,k} = \sum_{l \in \mathcal{C} \setminus j} P^r_{l,k}. \tag{7}$$

Note that here the interference is only considered from the neighbouring macro-cells, not within the macro-cell, since we assume orthogonal resource allocation amongst the associated users within the macro-cell. In addition, there is no interference between cellular and WiFi, since typically different RATs utilise orthogonal carrier frequency. Let $b_{j,k}$ be the portion of the bandwidth allocated to the user k from the system bandwidth Bj of macro-cell j. Thus, the noise power will be

$$\sigma^2_{j,k} = b_{j,k} B_j N_0, \tag{8}$$

where N_0 is the noise power spectral density. Thus, the throughput of user k associated with the macro-cell j will be

$$R_{j,k} = b_{j,k} B_j \log_2(1 + \frac{\gamma_{j,k}}{\eta^c_{phy}}) \eta^c_{mac}, \tag{9}$$

where η^c_{phy} is the system efficiency of cellular network, which is determined by the efficiency of the practical modulation and coding scheme used in the PHY layer, and η^c_{mac} reflects the system efficiency at MAC layer due to the signaling overheads.

In contrast to the cellular network, the radio channel is randomly accessed in WiFi network. This random access method is called DCF. A detailed theoretical analysis of DCF mechanism is given in [14]. However, in [14], only single physical layer (PHY) transmission rate is considered. The performance analysis of multi-rate WiFi is carried out in [15] and [16]. In these works, it has been shown that under the DCF mechanism, the throughput of all users attached to the same WiFi AP will be equal and lower than the PHY rate of the lowest rate user. This is referred to as *Performance anomaly*. The reason for this performance anomaly is that the underlying CSMA/CA mechanism of WiFi, which guarantees equal long term channel access probability for all users [15]. Hence, the WiFi is classified as the throughput fair access network in [17]. Although the users attached to the same AP achieve same throughput, different user combinations (i.e., number of users attached to the AP and their channel conditions) result in distinct throughput values [17]. Hence, the users associated with different WiFi AP may have different throughput. Thus, in this paper, it is considered that within an AP coverage, the channel is randomly accessed with equal opportunity, and there is no interference amongst the users attached to the same AP. However, similar to the cellular system, there will be interference from the neighbouring APs, since all APs utilise the same frequency spectrum. Hence, similar to the cellular system, the perceived SINR of a user k associated with the WiFi AP $\hat{j} \in \mathcal{A}$, can be obtained by

$$\gamma_{\hat{j},k} = \frac{P^r_{\hat{j},k}}{I_{\hat{j},k} + \sigma^2_{\hat{j}}}, \tag{10}$$

where $\sigma^2_{\hat{j}}$ is the noise power and $I_{\hat{j},k}$ is the total interference power, which is given by

$$I_{\hat{j},k} = \sum_{l \in \mathcal{A} \setminus \hat{j}} P^r_{l,k}. \tag{11}$$

However, in contrast to the cellular system, the whole bandwidth is dedicated to a single user at a given instant of time. Thus, the instantaneous noise power will be

$$\sigma^2_{\hat{j}} = B_{\hat{j}} N_0, \tag{12}$$

where $B_{\hat{j}}$ is the system bandwidth of WiFi AP $\hat{j}$. Hence, the instantaneous achievable rate of the user k associated with the WiFi AP $\hat{j}$ during the reception of the data frame is

$$\beta_{\hat{j},k} = B_{\hat{j}} \log_2(1 + \frac{\gamma_{\hat{j},k}}{\eta^a_{phy}}) \eta^a_{mac}, \tag{13}$$

where η^a_{phy} is the PHY layer efficiency of WiFi system, which is determined by the efficiency of the practical modulation and coding scheme, and η^a_{mac} reflects the system efficiency at MAC layer of WiFi network due to the DCF access mechanism. Let L be the frame length in bits; hence, the time duration of the user k to receive a frame from AP $\hat{j}$ is

$$\tau_{\hat{j},k} = \frac{L}{\beta_{\hat{j},k}}. \tag{14}$$

Thus, the total throughput of WiFi AP $\hat{j}$ will be

$$\begin{aligned} R_{\hat{j}} &= \frac{\sum\limits_{\hat{k} \in \mathcal{U}} x_{\hat{j},\hat{k}} L}{\sum\limits_{\hat{k} \in \mathcal{U}} x_{\hat{j},\hat{k}} \tau_{\hat{j},\hat{k}}} \\ &= \frac{\sum\limits_{\hat{k} \in \mathcal{U}} x_{\hat{j},\hat{k}}}{\sum\limits_{\hat{k} \in \mathcal{U}} \frac{x_{\hat{j},\hat{k}}}{\beta_{\hat{j},\hat{k}}}}. \end{aligned} \tag{15}$$

Hence, the throughput of a user k associated with the AP $\hat{j}$ is

$$R_{\hat{j},k} = \frac{1}{\sum\limits_{\hat{k} \in \mathcal{U}} \frac{x_{\hat{j},\hat{k}}}{\beta_{\hat{j},\hat{k}}}}. \tag{16}$$

Therefore, the general expression of throughput of user k will be

$$R_k = \begin{cases} x_{i,k} b_{i,k} B_i \log_2(1 + \frac{\gamma_{i,k}}{\eta^c_{phy}}) \eta^c_{mac}, & \text{if } i \in \mathcal{C}, \\ \dfrac{x_{i,k} B_i \eta^a_{mac}}{\sum\limits_{k \in \mathcal{U}} \dfrac{x_{i,k}}{\log_2(1 + \frac{\gamma_{i,k}}{\eta^a_{phy}})}}, & \text{if } i \in \mathcal{A}. \end{cases} \tag{17}$$

III. Problem Formulation

Considering a utility function perspective, we assume that user k obtains utility $U_k(R_k)$ when the user throughput is R_k, where $U_k(.)$ is a utility function. Thus, we formulate the general optimisation problem that maximises the aggregated utility function, which involves jointly finding the optimal power reduction factor y_i for WiFi AP $i \in \mathcal{A}$ and bandwidth allocation factor $b_{i,k}$ for each user k associated with cellular BS $i \in \mathcal{C}$ as follows:

$$\begin{aligned} \max_{y,b} \quad & \sum_{k \in \mathcal{U}} U_k(R_k), \\ \text{s.t.} \quad & \sum_{i \in \mathcal{A} \cup \mathcal{C}} x_{i,k} = 1, \quad \forall\, k \in \mathcal{U}, \\ & \sum_{k \in \mathcal{U}} b_{i,k} \leq 1, \quad \forall\, i \in \mathcal{C}, \\ & 0 < y_i \leq 1, \quad \forall\, i \in \mathcal{A}. \end{aligned} \tag{18}$$

Using a linear utility function for the above optimisation problem results in a trivial solution, where each AP serves only one UE which has the strongest channel condition, while the BS allocates all the resources to one user with the best channel condition. Although, the linear utility can produce optimal system throughput, it will lead to very unfair resource allocation, as a result most of the users will have zero throughput. Thus, it is not a desirable solution. However, instead of using linear utility function, using a logarithmic utility function naturally achieve load balancing and some fairness among the user throughput (i.e., proportional-fair). Hence, the utility function can be updated as follows:

$$U_k(R_k) = \log(R_k). \tag{19}$$

In addition, it has been proven that the logarithmic utility function reduces the complexity of the optimisation, since the optimal bandwidth allocation for logarithmic utility is equal allocation [8]. Hence, the user throughput with equal bandwidth allocation will be

$$c_k = \begin{cases} \dfrac{x_{i,k} B_i \log_2(1+\frac{\gamma_{i,k}}{\eta^c_{phy}})\eta^c_{mac}}{\displaystyle\sum_{k\in U} x_{i,k}}, & \text{if } i \in \mathcal{C}, \\[2ex] \dfrac{x_{i,k} B_i \eta^a_{mac}}{\displaystyle\sum_{k\in\mathcal{U}} \frac{x_{i,k}}{\log_2(1+\frac{\gamma_{i,k}}{\eta^a_{phy}})}}, & \text{if } i \in \mathcal{A}. \end{cases} \tag{20}$$

As a result, the optimisation problem is reduced to only finding the optimal power reduction factor y_i for WiFi AP $i \in \mathcal{A}$. Thus, the new optimisation problem is

$$\begin{aligned} \max_{y} \quad & \sum_{k\in\mathcal{U}} U_k(R_k) = \sum_{k\in\mathcal{U}} \log(R_k), \\ \text{s.t.} \quad & \sum_{i\in\mathcal{A}\cup\mathcal{C}} x_{i,k} = 1, && \forall\, k \in \mathcal{U} \\ & 0 < y_i \leq 1, && \forall\, i \in \mathcal{A}. \end{aligned} \tag{21}$$

IV. Sub-optimal Heuristic Approach

The problem in the above section is combinatorial due to the binary variable $x_{i,k}$ (i.e., unique association). Hence, the complexity of brute force algorithm to find the optimal solution y is $\mathcal{O}((y_{max}/\delta)^N)$ where y_{max} is the maximum power reduction factor and δ is the step size of power reduction per iteration. In order to reduce the complexity of the above problem, we propose a heuristic algorithm where the power reduction is carried out to all APs at each iteration, based on a weighted approximation. Typically, the optimal power reduction factor is influenced by two main factors: one is the AP location relative to BS location; and the second one is the current load of the AP relative to the BS load (i.e., ratio of users attached to AP and BS). Hence, we define two weighting factors w^d_i and w^l_i that represent the weights associated to the AP location relative to BS location and the current load of the AP relative to the BS load, respectively, such that:

$$w^d_i = c_1 d_{i,j} + c_2, \tag{22}$$

$$w^l_i(t) = c_3 \frac{\displaystyle\sum_{k\in\mathcal{U}} x_{i,k}(t-1)}{\displaystyle\sum_{k\in\mathcal{U}} x_{j,k}(t-1)}, \tag{23}$$

where $j \in \mathcal{C}$ is the corresponding macro BS of AP $i \in \mathcal{A}$, and $d_{i,j}$ is the distance of AP i from BS j, and c_1, c_2 and c_3 are some constants. Note, that the parameter w^d_i is independent from the current load, hence it will be constant (assuming that the locations of APs and BS are constant). However, w^l_i can change depending on the current load, which in turn can vary depending on the applied power reduction to the AP. Hence, w^l_i will change for every iteration t. Thus, the power reduction factor will be incremented for all APs $i \in \mathcal{A}$ for each iteration t as follows:

$$y_i(t) = y_i(t-1) + \{\psi^d w^d_i + \psi^l w^l_i(t-1)\}\delta, \tag{24}$$

where ψ^d and ψ^l are some weighting constants that are used to adjust the importance of w^d_i and w^l_i according to their influence over the optimal power reduction factor; such that $\psi^d + \psi^l = 1$. The proposed heuristic algorithm is given in Algorithm 1.

Algorithm 1 Proposed Heuristic Algorithm

Input: Channel State Information (CSI) of all UEs
Initialise the iteration ID $t = 0$
Initialise the power reduction factor $y_i(t) = 0, \ \forall i \in \mathcal{A}$
Set power reduction factor incremental step size δ
Compute $w^d_i, \ \forall i \in \mathcal{A}$ by (22)
Set a termination flag $TermFlag = FALSE$
Obtain $x_{i,k} \ \forall k \in \mathcal{U}$ from (4) and (5)
Compute utility $U(t) = \sum_{k\in\mathcal{U}} \log(R_k)$ by (20)
do
  {
  $t = t + 1$
  Compute $w^l_i(t), \ \forall i \in \mathcal{A}$ by (23)
  Update $y_i(t), \ \forall i \in \mathcal{A}$ by (24)
  Obtain $x_{i,k} \ \forall k \in \mathcal{U}$ from (4) and (5)
  Compute utility $U(t) = \sum_{k\in\mathcal{U}} \log(R_k)$ by (20)
  if $\max\{y_i(t)\} > y_{max}$ **then**
    $TermFlag = TRUE$
  end if
  }
while $(TermFlag)$;
$y_i(t^*) = \arg\max_t U(t)$
Output $y_i(t^*), \ \forall i \in \mathcal{A}$

V. Performance Evaluation

In this section, performance of the proposed scheme is evaluated, and the results are analysed and compared with three benchmark schemes, namely; WLAN-first, Max-RX, and CRE. In addition, the complexity of the proposed heuristic algorithm is analysed and compared to the optimal algorithm. The performances are evaluated through multiple random simulations and the results are averaged over the random iterations. Three performance metrics are analysed, namely; system throughput, user fairness, and load distribution between the cellular and WLAN, which reflect the system performance, fairness among the users, and load balancing in the network, respectively. Three different scenarios are considered for this evaluation. In the first scenario, the user density λ_u varies from 50 to 250 users per BS with a fixed AP density of 3

APs per BS (i.e., $\lambda_w = 3$) and a fixed hot-spot user ratio of 10 (i.e., $r = 10$). In the second scenario, the user density and AP density are fixed such that $\lambda_u = 100$ and $\lambda_w = 3$, but the hot-spot user ratio r varies from 2.5-25. Similarly, in the third scenario, the user density and the hot-spot user ratio are fixed such that $\lambda_u = 100$ and $r = 10$, but the WiFi AP density λ_w varies from 1-5.

For the system efficiency parameters of cellular network η^c_{mac} and η^c_{phy}, the values of 0.75 and 1.25 dB are used (same as the bandwidth and SINR efficiencies of AWGN channel as given in [18]), respectively. For the WiFi network, the value of η^a_{phy} is considered to be equal to η^c_{phy}, since WiFi also has OFDM PHY. However, the value of η^a_{mac} is considered to be much lower than η^b_{mac} due the high overheads of CSMA/CA and DCF mechanisms of WiFi [15]. In [19], it has been observed that the achieved throughput of an 802.11g WiFi AP is nearly 50% lower compared to the given transmission rate. Therefore, the value of η^a_{mac} is considered as 0.5. Moreover, for the proposed heuristic algorithm, the values for the constants and the weighting factors are set as follows: $c_1 = 0.02, c_2 = 10, c_3 = 0.3, \psi^d = 0.2$ and $\psi^l = 0.8$. The rest of the system parameters are listed in Table I.

For the benchmark scheme of WLAN-first, it is assumed that the UE will always be associated to WiFi if there is a WiFi coverage. In the Max-RX scheme, the UE will be associated to the POA (i.e., either cellular or WiFi) that has the strongest received signal. Similarly, in CRE scheme the UE will be associated to the POA that has the strongest biased received signal. In [7], it has been shown that the optimal biasing values for out-of-band (e.g., cellular-WiFi) off-loading is about 20-25dB for a small cell density five times that of a macrocell. Thus, for this evaluations, we apply 20dB biasing towards WiFi signals for CRE scheme.

TABLE I: System Parameter Settings

Parameter	**Notation**	**Value**
Macro BS ISD	D	1000 m
AP density per BS	λ_w	Variable (1-5)
User density per BS	λ_u	Variable (50-250)
Hot-spot user ratio	r	Variable (2.5-25)
Cellular carrier wave length	ω^c	0.150m (2GHz)
WiFi carrier wave length	ω^a	0.125m (2.4GHz)
Path loss exponent of cellular	$\alpha_i(\forall i \in C)$	3.5
Path loss exponent of WiFi	$\alpha_i(\forall i \in A)$	4
Cellular carrier bandwidth	$B_j(\forall j \in C)$	10 MHz
WiFi carrier bandwidth	$B_{\hat{j}}(\forall \hat{j} \in A)$	10 MHz
Noise power spectral density	N_0	-174 dBm/Hz
BS transmit power	$P^t_j(\forall j \in C)$	46 dBm
AP transmit power	$P^t_{\hat{j}}(\forall \hat{j} \in A)$	23 dBm
WiFi receiver sensitivity	P^r_{th}	-100 dBm

A. System throughput

Fig. 2-4 show the system performance in terms of system throughput against varying user density λ_u, hot-spot user ratio r, and AP density λ_w, respectively. In these figures, the considered network association schemes are denoted as follows: the optimal system throughput scheme in (18) is denoted as “Opt-SysTP”; the optimal log utility scheme in (21) is denoted as “Opt-Util”; the proposed heuristic scheme in Algorithm 1 is denoted as “Heu-Alg”, respectively.

In Fig. 2, we can observe that the system throughput is almost fixed for the Opt-Util, Heu-Alg, CRE and WLAN-First schemes, regardless of the number of users in the system. This is due to the fact that the considered traffic model is full-buffer. However, in the Opt-SysTP and Max-RX schemes, the system throughput increases as the number of users increases at the beginning and then it tends to saturates to a fixed value. The increment at the beginning is seen as a result of increased spacial diversity of users that provides better channel condition between some users and the BS/AP. Moreover, this figure clearly demonstrates the inefficiency of mainstream WLAN-first scheme in terms of system performance. This is due to the fact that in WLAN-first scheme, all users under WiFi coverage will be served by the WiFi AP. In addition, WiFi AP will be highly congested due to high user density in the hot-spot area. This leads to unbalanced load distribution between BS and APs, where the WiFi APs are highly loaded while the cellular BS is not fully utilised. Similarly, performance of the CRE scheme also very low, and it is almost equal to the WLAN-first. This is due to the fact that in CRE, UEs are steered toward WiFi, which leads to more UEs attached to the WiFi than cellular in the considered scenario. The performance of Opt-SysTP is superior as expected, and the Max-RX also performs very well in terms of system throughput. However, they significantly lack in terms of user fairness, which will be discussed in detail in the following subsection. The performance of the Opt-Util and the Heu-Alg fall between the two extreme of Opt-SysTP and WLAN-First. This is due to the fact that both of these schemes try to maximise the system throughput while attaining certain level of fairness among the users. In addition, we can observe from all these figures that the performance of the proposed heuristic scheme is almost identical to the optimal log utility scheme, which is achieved with much less complexity as explained later in this section.

In Fig. 3, we can observe that the system throughput of WLAN-first scheme does not change against the increased hot-spot user ratio. In the other schemes, there is a slight improvement at the beginning and then it tends to saturate. In addition, we can observe that the performance of Opt-Util and the Heu-Alg is equal to the performance of WLAN-first at low hot-spot user ratio. This is due to the fact that at low hot-spot user ratio, the number of users under the WIFi coverage will be very low due to relatively small coverage of WiFi AP compared to the cellular BS. Hence, serving all the users within the coverage of WiFi AP will be the best option for the considered utility optimisation. At the same time, extending the coverage of WiFi degrade the performance even below WLAN-First as seen in the case of CRE, at the beginning. In contrast to the Fig. 2 and 3, in Fig. 4, we can see that the system performance in terms of system throughput continues to increase as the number of WiFi AP per BS increases. This is due to the fact that the increased number of APs introduces additional resources into the system, hence the overall system performance increases in all schemes. In summary, the relative performances of all the schemes in the considered three scenarios are similar (i.e., WLAN-First scheme demonstrates the worst performance, while the performance of CRE is closer to WLAN-First, and the Opt-SysTP demonstrates the best performance, while the performance of Max-RX is closer to Opt-SysTP, and the performance of the Opt-Util and Heu-Alg fall between those cases).

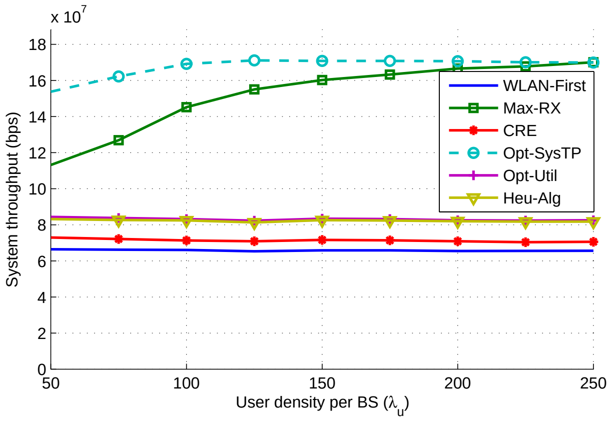

Fig. 2: ***System throughput vs. Number of users***

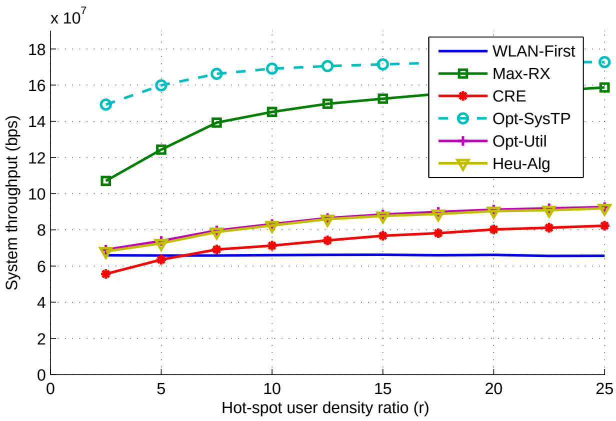

Fig. 3: ***System throughput vs. Hot-spot user density ratio***

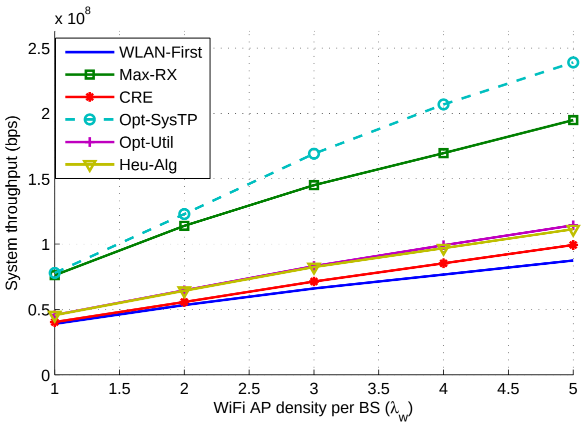

Fig. 4: ***System throughput vs. Number of WiFi AP per BS***

B. Fairness of user throughput

The user fairness is evaluated in terms of GINI coefficient of user throughput. Fig. 5-7 show the performance in terms of GINI coefficients against varying user density λ_u, hot-spot user ratio r, and AP density λ_w, respectively. From Fig. 5-7, we can notice that WLAN-First scheme demonstrates the best performance in terms of fairness. Note that the lower the GINI coefficient the fairer the system. The reason behind this observation is that the users served by the WiFi APs will inherently have equal throughput due to the underlying random access scheme. Since most of the users will be served by WiFi APs in WLAN-First, most of the users will be served with equal throughput. As a result, the WLAN-First demonstrates better fairness. However, it leads to lower system throughput as explained earlier. On the other hand, in the Opt-SysTP and Max-RX schemes, only few UEs will be served by the WiFi APs. Hence, those users will have much higher throughput than the users served by the BS. As a result, these schemes demonstrates the worst performance in terms of fairness. The performances of Opt-Util and Heu-Alg are lower than WLAN-First, however they are much closer to WLAN-First. Although, the performance of CRE is much better than Max-RX and lower than Opt-Util and Heu-Alg. This is due to the fact that in the proposed Opt-Util and Heu-Alg, the coverage of each AP optimally adjusted based on the current load and their location relative to the BS, whereas in CRE, a fixed bias is applied to all APs regardless of their load and location. Thus, the proposed scheme performs better.

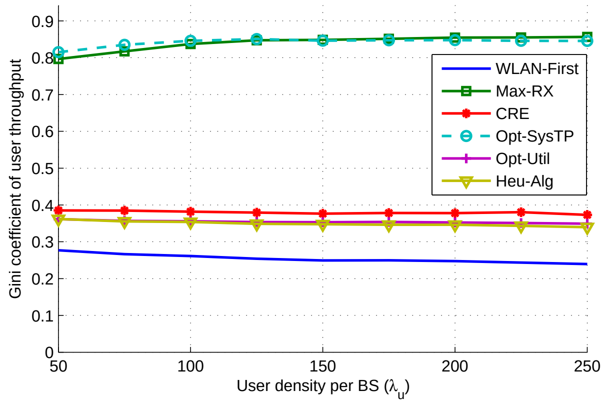

Fig. 5: ***GINI coefficient of user rate vs. Number of users***

C. Load distribution between cellular and WiFi

We evaluate the load distribution between the cellular and WiFi in terms of the percentage of the users served by WiFi APs in the system. Fig. 8-10 show the performance in terms of percentage of the users served by WiFi APs against varying user density λ_u, hot-spot user ratio r, and AP density λ_w, respectively. From Fig. 8, we can observe that the percentage of WiFi users does not change with increased number of users for all schemes as expected. As we can see in this figure, about 65% and 61% of the users are served by the WiFi in WLAN-First and CRE schemes, respectively. On the other hand, only

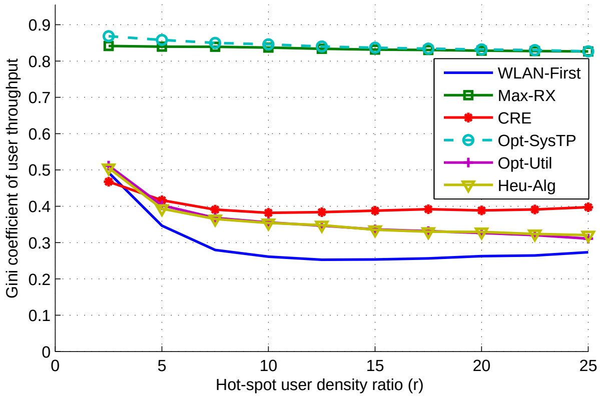

Fig. 6: ***GINI coefficient of user rate vs. Hot-spot user density ratio***

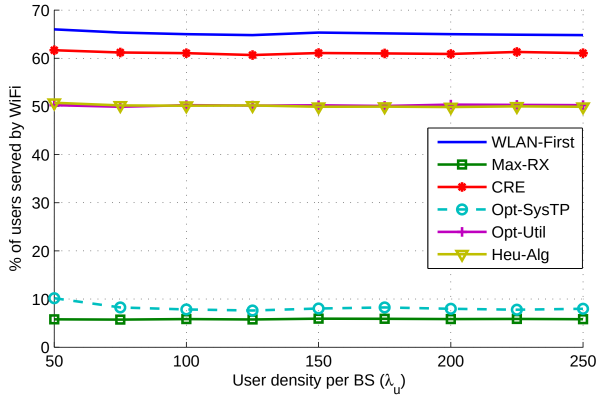

Fig. 8: ***Percentage of users served by WiFi vs. Number of users***

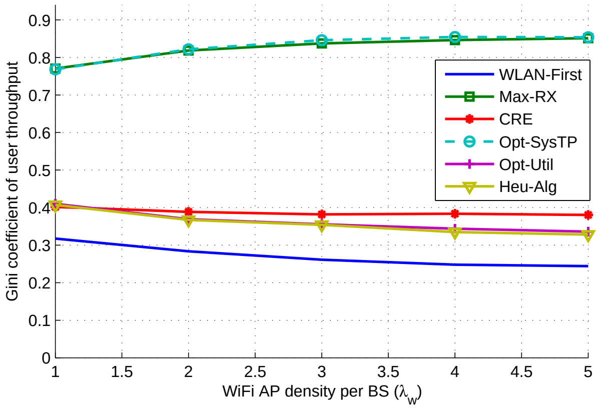

Fig. 7: ***GINI coefficient of user rate vs. Number of WiFi AP per BS***

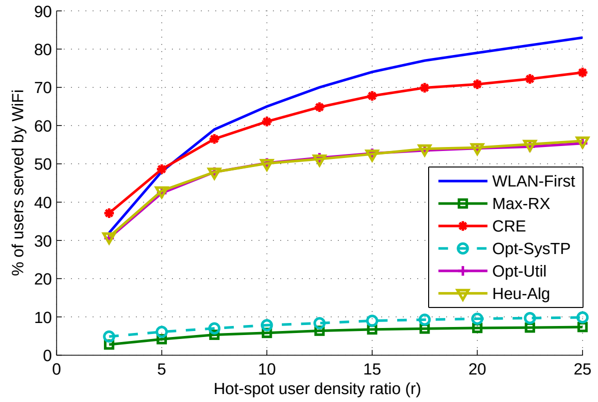

Fig. 9: ***Percentage of users served by WiFi vs. Hot-spot user density ratio***

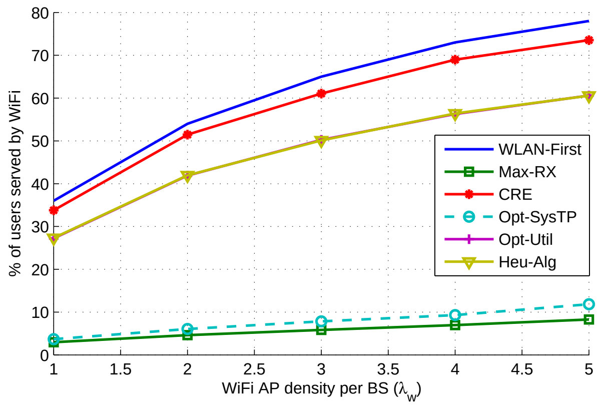

Fig. 10: ***Percentage of users served by WiFi vs. Number of WiFi AP per BS***

less than 10% is served by WiFi in Opt-SysTP and Max-RX. In Opt-Util and Heu-Alg schemes the percentage of WiFi users is about 50%. Considering that in this scenario, there are 3 WiFi APs per cellular BS and the BS has high transmission power with higher system efficiency than the WiFi APs, this load distribution can be considered as a fair load distribution.

In Fig. 9, we can observe that the percentage of WiFi users increases with increased hot-spot user ratio in all schemes. This is due to the fact that there will be more users under the WiFi coverage area with increased hot-spot user ratio. Similarly, in Fig. 10 also the percentage of WiFi users increases with the increased WiFi APs in all schemes. However, this is due to the increased WiFi resources in the system. In both these scenarios, Opt-Util and Heu-Alg schemes demonstrate better load balancing between cellular and WiFi compared to the other schemes.

D. Analysis of computational complexity

In this subsection, we analyse the computational complexity of the Heu-Alg scheme and compare it with Opt-Util scheme, since both of the schemes demonstrate desired system performances. Fig. 11 shows the computational complexity of both schemes in terms of number of iteration required to find the optimal solution with increased number of APs per BS. The complexity of the Opt-Util exponentially increases with respect to the number of APs per BS, where as in the Heu-Alg scheme it decreases with increased APs. Notably, the complexity is higher for the Heu-Alg when the AP density is less than 2 per BS. However, with higher AP density, the Heu-Alg scheme requires much less iterations compared to the Opt-Util scheme. This is due to the fact that in the Heu-Alg scheme, the power reduction factor is incremented based on the current load distribution and the location of AP with respect to the BS. Hence, if there are only few APs per BS, the power reduction factor will be incremented with smaller value due to much less variation of the load distribution in each iteration. However, with higher number of APs per BS the load variation for each iteration will be considerably higher. Hence the power reduction factor will be incremented at higher rate, until it reaches to the maximum value for at least one AP. Thus, the number of iteration required decreases with the increased AP density for the proposed Heu-Alg scheme.

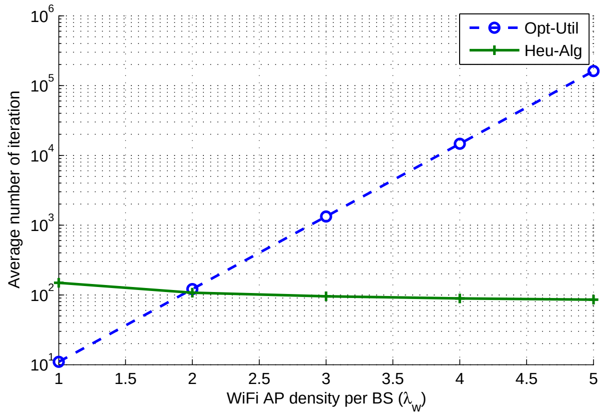

Fig. 11: Percentage of users served by WiFi vs. Number of WiFi AP per BS

VI. Practical Implementation Scenario

In this section, we propose an interworking architecture for implementing the proposed load balancing schemes. To this end, this section first describes the existing cellular-WLAN interworking solutions, followed by the proposed cellular-WLAN (particularly, LTE-WiFi). Then the detailed description of signalling procedures for the proposed interworking architecture is given.

A. Existing Cellular-WLAN Interworking Solutions

As mentioned earlier, in order to cope with the dramatic growth of cellular data traffic, mobile operators seek for cost-effective and easily deployable solutions to increase the capacity of their networks. One of these solutions is to use WiFi networks to off-load the cellular network. However, there are several challenges in integrating WiFi network to the cellular network, due to the architectural and technical differences of both RATs. To this end, the latest standardization work by 3GPP pays lots of attention to the integration of WiFi to LTE, including trusted access to cellular services for WiFi-only devices, seamless WiFi-LTE handover, and ANDSF. Especially, in 3GPP Release 12 and 13 solutions for tighter coupling between LTE and WiFi at RAN level have been investigated. The are numerous advantages of such integration, ranging from transfer of simple network assistance information to fully centralised radio resource management [20]–[22].

Although, the current 3GPP integration options can provide better integration of multi-RAT network, they are not as flexible to enable efficient multi-RAT connectivity at the RAN level of an already deployed cellular-WLAN HetNet as one would expect. For example, by convention, all user data in LTE is represented as IP packets, and all IP packets are hauled with a fixed QoS level through their respective EPS-bearers, which act like virtual circuits. As a result, the LTE network internally operates as a circuit-switched system, while externally appearing to be packet-switched [13]. Although, this provides the flexibility that the cellular system requires, it is very different from how IP works. As a result, no external IP traffic from the users is actually allowed inside the LTE network, which has significant implications on LTE-WiFi integration [13].

Moreover, most of the existing solutions for integrating cellular network with WLAN network, such as 3GPP I-WLAN, ANDSF, and HotSpot2.0 mainly focus on the integration of the core network [2]. Although, this kind of integration is beneficial for integrating multiple networks that are owned by different network operators, the lack of integration in the RAN leads to a suboptimal system performance gain from off-loading process of the cellular traffic to the WLAN. This is due to the fact that the current solutions in the standard mainly provide long term network centric information to the users in order to facilitate intelligent network selection by the users. However, the dynamic network status such as current network load and channel conditions are not shared with the users to make such intelligent network selection decision. For example, ANDSF provides a useful framework for distributing flexible operator-defined network selection information and policies. However, the dynamic network status information which could be used to improve the network selection decisions by the UE is not captured in the current iteration of ANDSF [2].

To this end, in [13] the authors have introduced an network entity called AAGW, which mirrors 3GPP functionalities into a WLAN RAT and vice-versa. As a result, the WiFi AP appears as an eNodeB to the LTE core network, which enables tighter integration of WiFi to LTE network. However, to realise the proposed solution, there is a need for change in the network architecture due to the introduction of AAGW. In addition, UE will need a special driver installed, and the UE needs to know the IP address of the nearest AAGW through an external mechanism such as ANDSF. Thus, in the following subsection, we propose an interworking architecture to implement the proposed off-loading scheme, which does not require any changes to the UE, and only requires a software update to the network entity such as eNodeB and AP. Therefore, the proposed solution can be easily implemented in an already deployed cellular-WLAN HetNet.

B. Proposed Cellular-WLAN Interworking Architecture

In this subsection, a network architecture is proposed, which comply with the current standard approach for integrating WLAN network with the 3GPP network according to the current 3GPP standard. Notably, a logical signalling link is introduced in the interworking architecture between the

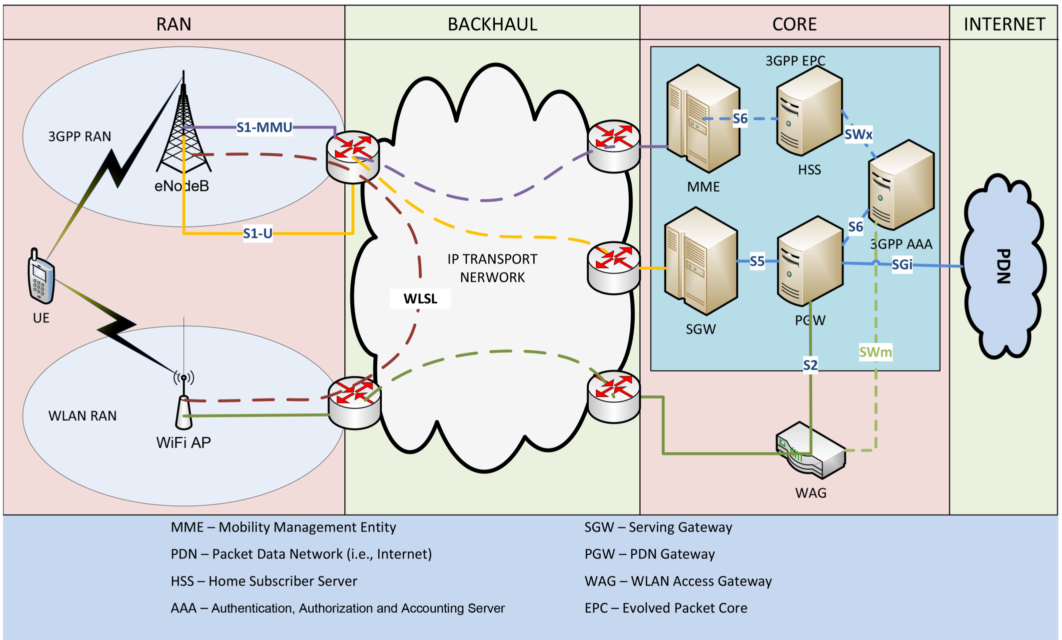

Fig. 12: ***Cellular-WLAN Interworking Architecture***

eNodeB and the WiFi APs under the converge of that eNodeB. This link is called WLAN Logical Signalling Link (WLSL). Fig. 12 depicts the proposed architecture with WLSL. Since LTE is an all IP network, and the considered HetNet is an operator deployed one, it is possible to create a signalling link between the WiFi APs and the eNodeB with logical IP or VPN, through existing backhaul transport network. Thus, the proposed WLSL does not require any additional direct physical communication links between the eNodeB and WiFi AP; hence, it can be implemented cost effectively in the existing HetNet.

The main advantage of this WLSL is that it enables close co-ordination at the RAN side of the network, without introducing any external entity or complex network upgrades. Moreover, this link enables the eNodeB to act as a central controller to perform optimal traffic steering between the eNodeB and the WiFi APs. For example, the eNodeB can be updated dynamically with the information regarding the current load of the APs and the channel condition of the attached UEs of the APs. With these information, the eNodeB can perform optimal traffic steering by optimising the coverage area of each AP under the coverage of that eNodeB by updating the power reduction factor of each AP for cell breathing. However, to make use of the proposed WLSL, the AP and eNodeB require some software changes. In addition, there will be some difficulties in implementing the WLSL, when the LTE, WLAN and backhaul networks are owned and operated by different providers. Moreover, for the future deployments of integrated cellular-WLAN HetNet, this interworking architecture may not be necessary to implement the proposed load balancing scheme, since the future cellular networks are expected to have tighter coordination with WLAN including the RAN side. In such networks, the proposed load balancing scheme can be easily incorporated as a part of network operation. Nevertheless, the proposed architecture provides a cost-effective and immediate solution for the existing (i.e., already deployed and operational) cellular-WLAN HetNet, to implement the proposed cell breathing scheme.

C. Signalling Procedure for the Proposed Interworking Architecture

Fig. 13 shows the signaling procedure for two use cases; one is the service arrival at the WiFi coverage area; and the other one is the handover of active secession from eNodeB to WiFi AP. In the case of new session arrival at WLAN coverage area, first, the UE listen to the WiFi beacons to find the availability of WLAN networks. If there is a WiFi coverage, the UE sends connection request to the WiFi AP. Since this is a new service request, the UE has to be authenticated for the purpose of security and billing. Therefore, the WiFi AP sends an authentication request to the core network (i.e., EPC). This authentication procedure can follow the standard approach such as the one defined in 3GPP I-WNAL standard. Once the UE is authenticated, the EPC sends the acknowledgement to the WiFi AP. Then the AP responds to the connection request to the UE. Once the connection between the UE and the AP is established, the user traffic is routed to the EPC via the WAG according to the standard procedure. At this point, the WiFi AP updates its current status to the eNodeB via the WLSL. This triggers the optimisation process in the eNodeB, which optimises the coverage of the APs under its coverage. Then the eNodeB sends the updated AP coverage adjustment (e.g., power reduction factor) to each AP. Then the APs adjust their coverage accordingly. It is worth noting that the coverage of WiFi AP is adjusted by only changing the beacon power. Hence, if the coverage of a WiFi AP become small, the existing active sessions will not be effected by this

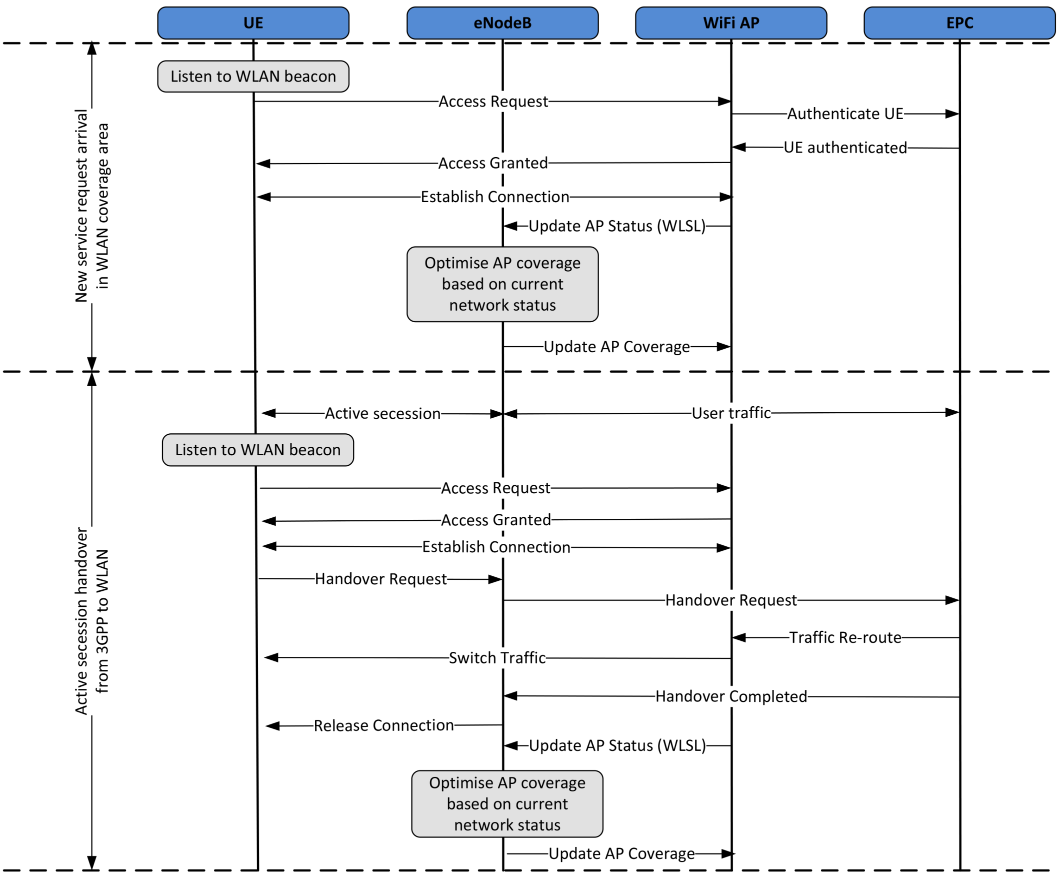

Fig. 13: Signalling Procedure for new Service Arrival at WLAN, and Handover from Cellular to WLAN

action. However, any new session arrivals or handover requests will be effected.

In the case of active secession handover, when the UE enters to the WLAN coverage area, it scans for the WiFi beacons to find the available WiFi APs. Once the UE identifies the availability of WiFi coverage, it sends the connection request to the WiFi AP. In this case, the AP does not need to authenticate the UE since it is already have authenticated with EPC via eNodeB. Therefore, the WiFi AP immediately responds to the connection request of the UE and establishes the connection with the UE. Note that at this point, the UE have active connections with both eNodeB and WiFi AP. This kind of handover is called “soft handover”. Then the UE requests the eNodeB to handover the active session to the new connection via WiFi. Since this is a inter-RAT handover, the eNodeB cannot simply switch the connection. Hence, the eNodeB forwards the request to EPC. The EPC re-route the traffic through the new connection, and notify the UE, and completes the handover. The traffic re-routing of this handover can be done according to the existing protocols such as “mobile IP”. Once the handover is completed, similar to the previous case, the AP updated its status to the eNodeB via WLSL, and the eNodeB carries out the optimisation process and update the coverage of the APs.

VII. Conclusion

In this paper, we proposed a novel load balancing scheme for an operator-deployed cellular-WLAN HetNet. In this scheme, the user association is controlled by employing cell breathing technique for the WLAN network. Hence, the users can simply associate with available strong WiFi AP without making complex association decision, since the computational complexity is limited to the network only. This makes this scheme easily implementable in existing cellular-WLAN HetNet. The objective of the proposed scheme is to optimise the system throughput while ensuring certain level of fairness amongst the users and balance the load between cellular and WLAN. We proposed a suboptimal heuristic algorithm in order to reduce the complexity of the optimisation process. The performance of the proposed scheme is evaluated in terms of system throughput, user fairness and load distribution between cellular and WLAN networks, and compared with three benchmark schemes such as WLAN-First, Max-RX and CRE. The simulation results show that the proposed scheme demonstrates better load distribution and user fairness while optimising the system throughput. For example, in the proposed scheme about 50% of the users are served by the WiFi APs compared to 65% and 61% of the users are served by the WiFi in WLAN-First and CRE schemes, respectively. On the other hand, just less than 10% is served by WiFi in Max-RX. In addition, the proposed heuristic algorithm achieves similar performance with reduced complexity compared to the optimal scheme. For example, with 5 APs in the system, the proposed heuristic algorithm only requires less than 100 iterations, while the exhaustive search requires more than 10^5 iterations. Moreover, an interworking architecture is proposed with detailed signalling procedures in order to facilitate the implementation of the proposed scheme on an existing operator-deployed cellular-WLAN HetNet.

ACKNOWLEDGMENT

The authors would like to acknowledge the support of the University of Surrey 5GIC (http://www.surrey.ac.uk/5gic) members for this work.

REFERENCES

[1] S. Landström, A. Furuskãr, K. Johansson, L. Falconetti, and F. Kronestedt, "Heterogeneous networks–increasing cellular capacity," *Ericsson Review*, p. 4, Feb 2011.

[2] "Integration of cellular and Wi-Fi networks," 4G Americas, www.4gamericas.org, Tech. Rep., Sep 2013.

[3] S. Jin, W. Xuanli, and S. Xuejun, "Load balancing algorithm with multi-service in heterogeneous wireless networks," in *Communications and Networking in China (CHINACOM), 2011 6th International ICST Conference on*, Aug 2011, pp. 703–707.

[4] J. Ha, J. Kim, J. up Kim, and S.-H. Kim, "Dynamic load balancing architecture in heterogeneous wireless network environment," in *Communications and Information Technology, 2009. ISCIT 2009. 9th International Symposium on*, Sep 2009, pp. 248–253.

[5] W. Song, W. Zhuang, and Y. Cheng, "Load balancing for cellular/WLAN integrated networks," *Network, IEEE*, vol. 21, no. 1, pp. 27–33, Jan 2007.

[6] W. Song, H. Jiang, and W. Zhuang, "Performance analysis of the WLAN-First scheme in Cellular/WLAN interworking," *Wireless Communications, IEEE Transactions on*, vol. 6, no. 5, pp. 1932–1952, May 2007.

[7] J. Andrews, S. Singh, Q. Ye, X. Lin, and H. Dhillon, "An overview of load balancing in HetNets: old myths and open problems," *Wireless Communications, IEEE*, vol. 21, no. 2, pp. 18–25, April 2014.

[8] Q. Ye, B. Rong, Y. Chen, M. Al-Shalash, C. Caramanis, and J. Andrews, "User association for load balancing in heterogeneous cellular networks," *Wireless Communications, IEEE Transactions on*, vol. 12, no. 6, pp. 2706–2716, June 2013.

[9] A. Damnjanovic, J. Montojo, Y. Wei, T. Ji, T. Luo, M. Vajapeyam, T. Yoo, O. Song, and D. Malladi, "A survey on 3gpp heterogeneous networks," *Wireless Communications, IEEE*, vol. 18, no. 3, pp. 10–21, June 2011.

[10] *Small cell enhancements for E-UTRA and E-UTRAN*, 3GPP Std. 3GPP TS 36.872, 2013.

[11] Ericsson, "Ericsson mobility report: On the pulse of the networked society," Ericsson, Tech. Rep., November 2016. [Online]. Available: https://www.ericsson.com/res/docs/2016/ericsson-mobility-report-2016.pdf

[12] Y. Bejerano and S. jae Han, "Cell breathing techniques for load balancing in wireless lans," *Mobile Computing, IEEE Transactions on*, vol. 8, no. 6, pp. 735–749, June 2009.

[13] O. Galinina, A. Pyattaev, S. Andreev, M. Dohler, and Y. Koucheryavy, "5g multi-rat lte-wifi ultra-dense small cells: Performance dynamics, architecture, and trends," *IEEE Journal on Selected Areas in Communications*, vol. 33, no. 6, pp. 1224–1240, June 2015.

[14] G. Bianchi, "Performance analysis of the ieee 802.11 distributed coordination function," *Selected Areas in Communications, IEEE Journal on*, vol. 18, no. 3, pp. 535–547, March 2000.

[15] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, "Performance anomaly of 802.11b," in *INFOCOM 2003. Twenty-Second Annual Joint Conference of the IEEE Computer and Communications. IEEE Societies*, vol. 2, March 2003, pp. 836–843 vol.2.

[16] D.-Y. Yang, T.-J. Lee, K. Jang, J.-B. Chang, and S. Choi, "Performance enhancement of multirate IEEE 802.11 WLANs with geographically scattered stations," *Mobile Computing, IEEE Transactions on*, vol. 5, no. 7, pp. 906–919, July 2006.

[17] E. Aryafar, A. Keshavarz-Haddad, M. Wang, and M. Chiang, "RAT selection games in HetNets," in *INFOCOM, 2013 Proceedings IEEE*, April 2013, pp. 998–1006.

[18] P. Mogensen, W. Na, I. Kovacs, F. Frederiksen, A. Pokhariyal, K. Pedersen, T. Kolding, K. Hugl, and M. Kuusela, "LTE capacity compared to the shannon bound," in *Vehicular Technology Conference, 2007. VTC2007-Spring. IEEE 65th*, April 2007, pp. 1234–1238.

[19] Cisco, "White paper: Capacity, coverage, and deployment considerations for IEEE 802.11g," Cisco Systems, Inc, Tech. Rep., 2005.

[20] Ericsson, "LTE Release 13: Expanding the networked society," Ericsson, Tech. Rep., 2015. [Online]. Available: http://www.ericsson.com/res/docs/whitepapers/150417-wp-lte-release-13.pdf

[21] S. Andreev, M. Gerasimenko, O. Galinina, Y. Koucheryavy, N. Himayat, S. P. Yeh, and S. Talwar, "Intelligent access network selection in converged multi-radio heterogeneous networks," *IEEE Wireless Communications*, vol. 21, no. 6, pp. 86–96, December 2014.

[22] *Access to the 3GPP Evolved Packet Core (EPC) via non-3GPP access networks (Release 12)*, 3rd Generation Partnership Project (3GPP) Std. 3GPP TS 24.302, June 2014.