Traffic–Aware User Association in Heterogeneous LTE/WiFi Radio Access Networks

Davit Harutyunyan*, Supreeth Herle*, Dimitri Maradin†, George Agapiu‡, and Roberto Riggio*
*FBK CREATE-NET, Trento, Italy; Email: {d.harutyunyan,s.herle,rriggio}@fbk.eu
†CommAgility, Duisburg, Germany; Email: dimitri.marandin@commagility.com
‡OTE, Athens, Greece; Email: gagapiou@oteresearch.gr

***Abstract*—WiFi networks are known to be a cost–efficient traffic offloading solution for mobile networks. The Multi Access Packet Data Network Connectivity is a feature introduced in LTE Release 10 in order to allow users to be simultaneously connected to multiple radio access networks (RAN). Although this feature brings many advantages, such as the possibility to implement QoS–based traffic steering, it poses also many challenges, one of which is distributing traffic among the two radio access technologies. In this paper, we propose a traffic–aware user association algorithm for heterogeneous LTE/WiFi RANs. The proposed algorithm is formulated as an Integer Linear Programming (ILP) problem jointly optimizing user association and resource allocation. A heuristic is also proposed in order to address the scalability issues of the ILP–based algorithm. Numerical simulations are used in order to compare the proposed approaches. Finally, we implemented and tested the heuristic in small–scale testbed using the 5G–EmPOWER platform.**

***Index Terms*—Heterogeneous radio access networks, WiFi, LTE, User Association, ILP, Heuristic**

I. Introduction

Mobile Network Operators (MNOs) have witnessed a tremendous increase in mobile data traffic demand over the last few years [1], [2]. This exponential growth is forecast to reach 49 exabytes per month by 2021 [3]. In order to accommodate this huge traffic demand, MNOs will have to upgrade their infrastructure. Boosting mobile network capacity in terms of both coverage and supported data traffic rate can be achieved in several ways such as exploiting more spectrum, deploying denser radio access networks (RANs) [4], or offloading part of the mobile data traffic to other RANs [5]. Given its low deployment and operation costs, WiFi networks are an efficient traffic offloading solution for mobile networks.

Traffic steering between cellular and WiFi networks is possible through the Access Network Discovery and Selection Function (ANDSF) that has been introduced in LTE Release 8. ANDSF is a 3GPP–defined core network entity that allows User Equipments (UEs) to select the proper radio access technology (if more than one is available). However, Release 8 does not allow a UE to be simultaneously connected to multiple RANs. The Multi Access Packet Data Network (PDN) Connectivity feature introduced in LTE Release 10 addresses this limitation by allowing UEs to attach to multiple RANs (e.g. WiFi and LTE). MNOs can benefit from this feature by offloading the best effort traffic to the WiFi network while keeping demand–attentive traffic on the mobile network.

The contribution of this paper is twofold. First, we formulate a traffic–aware user association problem aiming at optimizing resource utilization in a heterogeneous WiFi/LTE RAN. Integer Linear Programming (ILP) techniques are used in order to derive the optimal solution. Then, a scalable heuristic is proposed in order to tackle the scalability problems of the ILP–based association algorithm. Second, we implement and test the heuristic in real–world conditions over the 5G–EmPOWER platform [6]. The entire implementation of the user association algorithm is released under a permissive APACHE 2.0 License[1].

The rest of this paper is structured as follows. The related work is discussed in Sec. II. The network and resource request models are detailed in Sec. III. The ILP problem formulation and the heuristic are introduced in Sec. IV while Sec. V reports on the numerical simulations. The proof–of–concept implementation and its evaluation are described in Sec VI. Finally, Sec. VII draws the conclusions.

II. Related Work

A sizable body of literature has been published on user association and load balancing in cellular [7], [8], [9], [10], [11] and local area networks [12], [13], [14].

A joint user association and interference mitigation scheme is studied in [7]. The authors show how, by adjusting the proportion of Almost Blank Subframes (ABS), it is possible to increase the data rate of cell–edge users by a factor of three. Similarly, load– and QoS–aware user association is considered in [8]. More specifically, distributed user association and semi–distributed macro–pico load balancing schemes are proposed aiming to jointly determine the ABS ratio that maximizes users association. In [9], a traffic offloading scheme that jointly considers power control and user association is proposed. Assuming that the transmit power of the pico base–stations is fixed and that the available bandwidth of the macro base–stations is divided into coverage and capacity bands, the authors show that significant benefits can be obtained by adjusting only the

Research leading to these results received funding from the European Unions H2020 Research and Innovation Action under Grant Agreement H2020-ICT-644843 (COHERENT Project) and by EIT Digital project (16242-17) ICARO-EU “Seamless Direct Air-to-Ground Communication in Europe.”

978-1-5386-3416-5/18/$31.00 © 2018 IEEE

[1] Available at: http://empower.create-net.org/

transmit power of the capacity bands. In [10], the authors devise a load balancing heuristic combining the admission control and the mobility management. A joint cell association and resource allocation problem is presented in [11]. To reduce the computational complexity of the proposed algorithm, a fractional user association scheme is suggested in which it is assumed that users can be associated with more than one cell.

A transparent load balancing problem is studied in [12] for WLANs. A MILP problem formulation is used in order to find the optimal network–wide scheduling configuration for all the flows in the network, having as un ultimate goal the maximization of the rate assigned to all flows. In contrast to the other works, which assume that users flow rate is known in advance, the authors exploit real–time monitoring information to estimate the desired flow rates. An automatic load balancing algorithm for WiFi networks is proposed in [13]. The algorithm uses different metrics (e.g., resource utilization, connected users, etc.) with the overarching objective of balancing the load across different APs. A joint association and bandwidth allocation problem for WLANs is studied in [14]. The authors propose an algorithm designed to achieve load balancing, and at the same time, fair transmission time allocation to each user.

Recently, more attention has been given to user association and load balancing problems in heterogeneous RANs (e.g., LTE/WiFi) [15], [16], [17], [18], [19], [20], [21]. Reference [15] suggests an energy–efficiency driven load balancing strategy. The energy efficiency is achieved by handing over cell–edge users (with low SINR) from LTE networks to WiFi networks. While the authors of [16] study a load balancing problem in which a load management entity is deployed in the core network, which has a global view of the network and is in charge of offloading users from LTE to WiFi and vice versa based on the UEs access index, which is computed for each UE based on their feedback (e.g., received signal strength from LTE and WiFi networks). To achieve the network–wide load balancing goal, an optimization problem is formulated, aiming at maximizing Jain's fairness index. In [17], the authors formulate an optimal RAT association problem for offloading mobile data as a constrained Markov Decision Process. To select an optimal policy having un ultimate goal of maximizing the expected per–user throughput subject to a constraint on blocking probability of voice users, the authors use value iteration and gradient descent algorithms. This work is extended in [18] considering the same network without the knowledge of the arrival processes. In [19], different resource allocation schemes in multi–homed LTE/WiFi networks are studied. In particular, the authors compare proportional fairness resource allocation strategies applied to LTE and WiFi networks as a whole (where multi–homing is considered) and as a separate entities (without considering multi–homing). The authors conclude that multi–homing has better load balancing capabilities. A QoS–based algorithm is proposed in [20] for vertical handover in LTE/WiFi networks. Handovers are triggered when the QoS score for a certain wireless users is going below a certain threshold. The main advantage of the proposed approach over the existing ones is that the required QoS level is guaranteed without adding new entities in the network and without increasing the signaling load. The study in [21] suggests a network selection strategy for LTE/WiFi networks which accounts also for the WiFi backhaul capacity.

TABLE I: RAN model parameters

Variable	Description
N_n	RAN nodes.
N_{enb}	LTE eNBs in $\in N_n$.
N_{ap}	WiFi APs in $\in N_n$.
$\omega_r(n)$	Available radio resources at node $n \in N_n$.
$loc(n)$	Geographical location of node $n \in N_n$.
$\delta(n)$	Coverage radius of node $n \in N_n$.

TABLE II: Traffic classes and their priorities.

Service ID	Service Type	Share [%]	Priority
1	Audio streaming	7.5	High
2	Video streaming	64	High
3	Web	26	Low
4	File sharing	2.5	Low

Most of the works above study user association for either WiFi or LTE networks. Some other works account for joint WiFi/LTE user association and/or multi–homing, but do not consider different traffic classes nor do they aim at designing an algorithm that takes as input observable network metrics. Conversely, to the best of our knowledge, this is the first attempt at designing, testing, and implementing a traffic–aware user association scheme for heterogeneous WiFi/LTE networks that considers also multi–homing and user observable network metrics.

III. Network Model

A. RAN Model

Let $N_n = (N_{enb}, N_{ap})$ be the set of $n_1 = |N_{enb}|$ eNBs and $n_2 = |N_{ap}|$ APs deployed in the heterogeneous RAN. Each node $n \in N_n$ is associated with a geographic location $loc(n)$, as x, y coordinates. A coverage radius $\delta(n)$ is also associated to each node $n \in N_n$. A single weight $\omega_n^c(n) \in \mathbb{R}$ with $0 \leq \omega_r(n) \leq 1$ is assigned to each node $n \in N$ modeling its available resources. Initially, $\omega_r(n) = 1\ \forall n \in N$. Table I summarizes the RAN model parameters.

B. Resource Request Model

Let N_s be the set of services that can be consumed by wireless clients. Notice how, in this work we consider four types of data traffic services: audio streaming, video streaming, web and file sharing. Each service can have either high or low priority. High priority services should be preferably served by LTE eNBs while low priority services can be served by either LTE eNBs or WiFi APs. As it will be clear in the evaluation section, the mix of services in the actual pool of resource requests is derived considering the global mobile traffic forecast for 2018 found in [3]. Table II summarizes the services considered in this work and their priorities.

TABLE III: Resource request model parameters.

Variable	Description
N_u	Wireless clients.
N_s	Available services.
$\omega_r^s(u)$	Resources requested by client $u \in N_u$ for service $s \in N_s$.
$loc(u)$	Geographical location of client $u \in N_u$.
$\Omega(u)$	Candidate RAN nodes for client $u \in N_u$.

Wireless clients can simultaneously consume multiple services. Moreover, we also assume that wireless clients support multi–homing, i.e., different services of a same wireless client can be served by different RAN nodes. For example, a wireless client can watch a live video stream served by an LTE eNB while a file download is served by a WiFi AP.

Let N_u be the set of wireless clients. Each wireless client $u \in N_u$ is associated with a geographic location $loc(u)$, as x, y coordinates. Multiple weights $\omega_r^s(u) \in \mathbb{N}$ are assigned to each wireless client modeling the amount of resources (in bit/s) requested by the client $u \in N_u$ for the service $s \in N_s$. Table III summarizes the resource request model parameters.

C. Resource Allocation Model

Resource allocation in LTE and WiFi networks is vastly different. LTE uses scheduled access and in particular OFDMA while on the other hand WiFi uses random access based on the CSMA/CA protocol. As a result, a homogeneous way of modeling available and used resources is needed before an optimization problem can be formulated. To this purpose, we introduced the *Equivalent Resource Utilization* ratio $ERU_n(\omega_r^s(u))$ defined as the fraction of radio resource required to support a service request $\omega_r^s(u)$.

In the case of an LTE RAN, the ERU is computed as the fraction of Physical Resource Blocks (PRBs) required to support a given service at a given eNB. The number of PRBs N_{prb} in a subframe required in order to support a given request can be computed as follows:

$$N_{prb} = \frac{\omega_r^s(u)\zeta_{enb}T_{prb}}{N_{sbc}N_{ofdms}N_{modb}N_{ant}}$$

where T_{prb}, N_{sbc}, N_{ofdms}, and N_{ant} are, respectively, PRB duration (1ms), the number of subcarriers (12), the number of OFDM symbols per subcarrier (7), and the number of MIMO streams. Notice how these parameters are unequivocally defined for a given version of the LTE standard. ζ_{enb} is the PRB efficiency considering reference signals, synchronization signals, etc. and can be estimated at 1.25. N_{modb} is the number of modulated bits per symbol. For example, if a 64–QAM modulation is used, then $N_{modb} = 6$.

Finally, the ERU for an LTE RAN can be computed as the ratio between N_{prb} and the total number of PRBs available in a cell. For example, in a 20 MHz cell there are 1000 PRBs in each 10ms–long Radio Frame.

In the case of a WiFi RAN, the ERU is computed as the fraction of the airtime required in order to support a given service at a given AP. Unlike LTE, WiFi relies on random access with exponential back off as a channel access technique. Therefore, we decide to rely on a transactional model in order to estimate the time required to serve a given user. WiFi uses a two–way handshake mechanism where each frame must be acknowledged by the receiver. As a result, for each data packet two frames must be exchanged on the air interface: one for the data itself and one for the WiFi ACK. The time to deliver a frame is thus given by:

$$T_{airtime} = \Big(T_{difs} + T_{data} + T_{sifs} + T_{ack} + 2\sigma\Big)\zeta_{ap}$$

where σ, T_{difs}, T_{sifs}, and T_{ack} are, respectively, the propagation time ($1\mu sec$), the Distributed Interframe Space ($34\mu sec$), the Short Interframe Space ($16\mu sec$), and the time to send an ACK ($24\mu sec$). Notice how these parameters are unequivocally defined for a given version of the IEEE 802.11 standard [22]. This model does not take into account nether the random backoff period nor the lost frames (e.g., due to collisions). As a consequence, we introduce the WiFi efficiency parameter ζ_{ap} in order to account for these impairments. In our simulation, ζ_{ap} has been set to 1.25. More accurate models for estimating this parameter can be found in [23].

Assuming that data packets are encapsulated in maximum length Ethernet frames (1500 bytes) and accounting also for the WiFi header (36 bytes) and for the 6 tail bits that are added for each transmission, the time required to send a frame is:

$$T_{data} = T_{hdr} + \frac{(1500 + 36) * 8 + 6}{N_{sbc}N_{modb}}$$

where T_{hdr}, N_{sbc}, and N_{modb} are, respectively, the synchronization header ($20\mu sec$), the number of subcarriers (48), and the number of encoded bits per subcarrier. For example, assuming a 64–QAM modulation and a $3/4$ coding rate, each subcarrier can encode 4.5 bits.

Finally, the ERU for a WiFi RAN can be computed as $T_{airtime}$ (in seconds) times the number of transactions necessary to support the request $\omega_r^s(u)$ over a unit of time.

Notice how, in both the WiFi and the LTE cases, an $ERU_n(\omega_r^s(u)) > 1$ means that the request $\omega_r^s(u)$ cannot be satisfied by RAN node n. Notice also how, the N_{modb} parameter present in both resource allocation models represents the number of modulated bits for each symbol/subcarrier. This quantity depends on the modulation and coding scheme (MCS) used for the transmission which in time is correlated with the channel quality between wireless terminals and RAN nodes. Several models linking channel quality and MCS can be found in literature [24]. However, since the focus of this paper is on the formulation of the user association problem, the selection of a particular channel model, although important, takes a secondary role. As a result, in the numerical evaluation we will leverage on a simple MCS estimation model which uses as input just the distance between transmitter and receiver.

IV. Traffic–Aware User Association

A. Problem Formulation

In order to find the optimal assignment, we introduce the concept candidate RAN nodes $\Omega(u)$ for the wireless client

$u \in N_u$ defined as the set of RAN nodes that have the wireless client u within their coverage radius:

$$\Omega(u) = \Big\{ n \in N_n | dis(n,u) \le \delta(n) \Big\}$$

Notice that while in the numerical simulations the euclidean distance between wireless clients and RAN nodes is used to identify the candidate RAN nodes for each wireless client, in the proof–of–concept implementation, the candidate RAN nodes are identified based on the signal strength.

We can now provide the optimal ILP formulation for the user association problem. The objective of the ILP problem is to minimize ERU utilization in the network. The chosen objective function is:

$$min\Big(\sum_{n \in N_{enb}} \sum_{u \in N_u} \sum_{s \in N_s} ERU_n(\omega_r^s(u)) \xi_{enb}(u,s) \Phi_n^{u,s} + \\ + \sum_{n \in N_{ap}} \sum_{u \in N_u} \sum_{s \in N_s} ERU_n(\omega_r^s(u)) \Phi_n^{u,s} \Big) \quad (1)$$

The first argument of the objective function aims at minimizing ERU at eNBs. Whereas, the second argument minimizes ERU at APs. $\xi_{enb}(u,s)$ is a coefficient used in order to steer high–priority services toward eNBs. The coefficient $\xi_{enb}(u,s)$ takes values in $(0,1)$. When $\xi_{enb}(u,s) \to 0$ supporting service s becomes progressively cheaper for the client u. Conversely, when $\xi_{enb}(u,s) \to 1$ the full cost of the service s must be sustained. The coefficients $\xi_{enb}(u,s)$ is defined as follows:

$$\xi_{enb}(u,s) = \begin{cases} \alpha_{enb} & \text{if } \frac{dis(u,n')}{dis(u,n'')} \le \beta_{enb} \\ 1 & \text{otherwise} \end{cases} \quad (2)$$

where $n' \in N_{enb}$ and $n'' \in N_{ap}$ are, respectively, the considered candidate eNB and the closest AP (according to the defined distance metric) to the wireless client u, and s is the service class. By tuning β_{enb} and α_{enb} it is possible to steer the algorithm into assigning some service classes to eNBs even when said nodes are not their optimal choice. Essentially, the ξ_{enb} coefficients make placing a certain service at some eNBs *cheaper* from the radio resource utilization perspective.

Wireless clients can attach to RAN nodes as long as the RAN nodes have enough radio resources:

$$\sum_{u \in N_u} \sum_{s \in N_s} ERU_n(\omega_r^s(u)) \Phi_n^{u,s} \le \omega_r(n) \quad \forall n \in N_n \quad (3)$$

where $\Phi_n^{u,s}$ is a binary mapping variable $\in \{0,1\}$ that shows whether service $s \in N_s$ consumed by wireless client $u \in N_u$ is served by the RAN node $n \in N_n$.

The following constraint makes sure that each wireless client $u \in N_u$ is associated with a RAN node that belongs to its list of candidates $\Omega(u)$:

$$\sum_{n \in N \setminus \Omega(u)} \Phi_n^u = 0 \quad \forall u \in N_u \quad (4)$$

where Φ_n^u is a binary mapping variable that shows if the wireless client $u \in N_u$ is associated to the RAN node $n \in N$.

Algorithm 1 Traffic–aware User Association Heuristic.

1: **procedure** *HEU*($N_{enb}, N_{ap}, N_u, N_s$)
2: **for** $u \in N_u$ **do** ▷ List of wireless clients.
3: **for** $s \in N_s$ **do** ▷ List of services.
4: $candidates \leftarrow list()$
5: **if** $\omega_r^s(u) = 0$ **then**
6: **continue**
7: **end if**
8: $cls\_ap\_dis \leftarrow \infty$
9: **for** $n \in N_{ap}$ **do** ▷ List of APs.
10: $cost \leftarrow ERU_n(\omega_r^s(u))$
11: **if** $dis(n,u) \le \delta(n)$ **and** $cost \le res(n)$ **then**
12: $candidates(u,n) = cost$
13: $cls\_ap\_dis \leftarrow$ **min**$(cls\_ap\_dis, dis(n,u))$
14: **end if**
15: **end for**
16: **for** $n \in N_{enb}$ **do** ▷ List of eNBs.
17: $cost \leftarrow ERU_n(\omega_r^s(u))$
18: **if** $dis(n,u) \le \delta(n)$ **and** $cost \le res(n)$ **then**
19: **if** s **is** $high\_pr$ **and** $\frac{dis(n,u)}{cls\_ap\_dis} \le \beta_{enb}$ **then**
20: $cost \leftarrow \alpha_{enb} ERU_n(\omega_r^s(u))$
21: **end if**
22: $candidates(u,n) = cost$
23: **end if**
24: **end for**
25: $mappings(u,s) \leftarrow \arg\min_n [candidates(u,n)]$
26: **end for**
27: **end for**
28: **end procedure**

Finally, the last constraint guarantees that each service $s \in N_s$ consumed by wireless client $u \in N_u$ is served by one and only one RAN node:

$$\sum_{n \in N} \Phi_n^{u,s} = 1 \quad \forall u \in N_u \quad \forall s \in N_s \quad (5)$$

Notice that this formulation does not prevent different services consumed by the same wireless client to be supplied by different RAN nodes.

B. Heuristic

The ILP–based user association algorithm becomes computationally intractable when big networks with tens of thousands of users and hundreds of RAN nodes are considered. In order to tackle this problem, we propose a scalable heuristic.

The pseudo code of the heuristic is reported in Alg. 1. Initially, for each service $s \in N_s$ of each wireless client $u \in N_u$, a list of candidate WiFi APs (lines 9–15) is created, considering the client distance from the APs ($dis(n,u) \le \delta(n)$) and the required resource availability at the APs ($cost \le res(n)$). Also, the distance of the closest AP (i.e., the AP that provides the best channel condition to the client) is stored in the cls_ap_dis variable. The candidate list is then updated, adding also the candidate eNBs (lines 16–24). Notice that after finding valid eNB candidates by checking the condition in line 18, like in the case of the ILP–based algorithm, it is checked whether the ratio of the distances between each candidate eNB and the closest candidate AP is less or equal β_{enb} (line 19). If this condition is true, the cost of the eNBs is reduced by α_{enb} for the service $s \in N_s$, therefore, making the high–priority services cheaper to be supported by eNBs rather than by APs. After having the complete list of candidates encompassing both the candidate APs and eNBs, the service $s \in N_s$ of the wireless client $u \in N_u$ is assigned the candidate node with the minimum cost. It is worthwhile to note that, since we consider a scenario in which wireless clients can employ multi–homing, different services of a same wireless client may be assigned to different RANs (e.g., the high–priority service of a single client may be assigned to an eNB while the low–priority service to an AP), resulting in the client being simultaneously connected to both the eNB and the AP.

V. Evaluation

The goal of this section is to compare the performance of the ILP–based user association algorithm (ILP) with the performance of the proposed heuristic (HEU).

We select $\beta_{enb} = 2$ for all the services while we select $\alpha_{enb} = 0.5$ only for the high–priority services (IDs 1 & 2) and $\alpha_{enb} = 1$ only for the low–priority services (IDs 3 & 4). The rationale behind this choice is to make supporting the high–priority services half as expensive on LTE eNBs when such eNBs are less than twice as distant from the wireless client than an optimal WiFi AP. In this section, we will first describe the simulation environment and the performance metrics used in our study. Then, we will report on the outcomes of the numerical simulations carried out in a discrete event simulator implemented in Matlab®.

A. Simulation Environment

The reference RAN used in this work is composed of 26 eNBs and 50 WiFi 802.11n APs. The eNB distribution is derived from an operational LTE network that provides cellular coverage to 1 million people distributed over an area of $5km^2$. Conversely, APs are randomly deployed within the same area. For simplicity, it is assumed that single sector (omni) cells are used for both eNBs and APs with 2×2 MIMO configuration providing coverage radius of, respectively, 500m and 200m.

Wireless client association requests arrive sequentially in batches, and with each arrival, the algorithms re–associate all the clients in the network. Each batch consists of 5 wireless clients each of them consuming up to 2 services randomly picked among the set of services found in Table III. The actual traffic demand for each service is derived from the global mobile traffic forecast for 2018 [3] (also reported in Table III).

B. Simulation Results

Figure 1a and Fig. 1b plot the distribution of the aggregated traffic served by, respectively, eNBs and APs after 150 association requests. As it can be seen, the median of the traffic served by the eNBs is roughly three times the median of the traffic served by the APs for all the algorithms. This is due to the fact that: (i) the number of APs is twice the number of the eNBs, and (ii) the coverage of the eNBs is larger than the coverage of the APs. Figure 1a shows that both ILP and HEU algorithms distribute the traffic uniformly across the available eNBs. Conversely, in Fig. 1b it can be observed that ILP distributes traffic more uniformly across the APs than the HEU, which is characterized by more outliers and by a bigger delta between the first and the third quartiles. The uniform traffic distribution of the ILP algorithm across both RANs is justified by the fact that it accepts more association requests than the HEU algorithm.

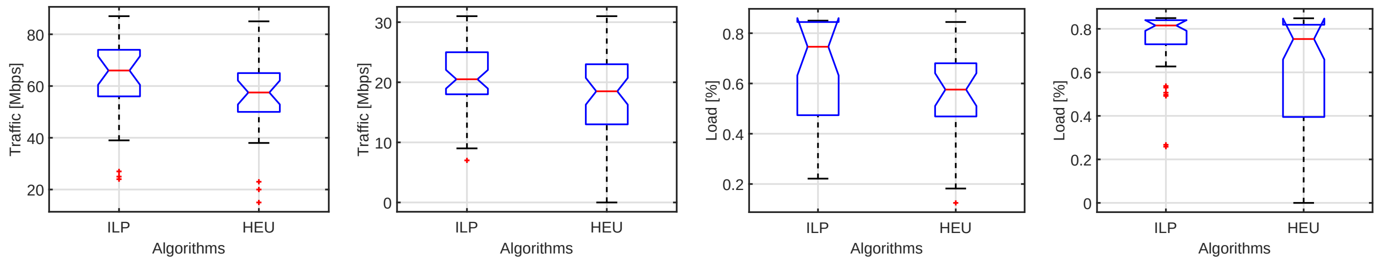
Figure 1c and Fig. 1d plot the distribution of node load (as fraction of available radio resources) at, respectively, eNBs and APs after 150 associations. As expected, the median loads for ILP across both eNBs and APs are higher than the median load for HEU. This is again due to the fact that the ILP–based association algorithm can accept a higher number of requests, therefore, increasing the load at eNBs and APs.

Figure 2 shows the distribution of the 4 possible service classes at both RANs using ILP and HEU algorithms. It can be observed that a higher fraction of services is served by eNBs irrespective of the association algorithm used. This behavior is even more evident in the case of HEU where the fraction of high–priority service classes assigned to eNB reaches approximately the 65%. The rational behind this is that, although there are 2 APs for each eNBs, compared to the coverage of the eNBs, the coverage of the APs is smaller, which in time results in less opportunities for services to be served by APs. Moreover, if the condition (2) is satisfied, the high–priority services demanded by the stations have more chances to be supported by eNBs due to their reduced cost.

The overall acceptance ratio and total number of associated users is plotted in, respectively, Fig. 3a and Fig. 3b. As expected, the acceptance ratio of the ILP algorithm is higher (78%) than the acceptance ratio of the HEU algorithm (68%). However, we can observe that the difference between HEU and ILP is just 10%. The same 10% difference is held also in the number of associated wireless clients (see Fig. 3b). This is because, in our simulation, each association request is composed of a fixed number of wireless clients.

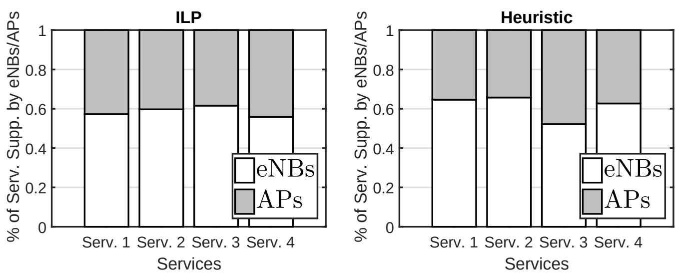
Ideally, an access technology (i.e., an eNB or a WiFi AP) for a client should be (re)selected only if the access technology satisfies the service QoS requirements. In our scenario, audio streaming (*Service 1*) and video (*Service 2*) are considered to be high–priority services (see Table III). Figure 4 shows the traffic share of the high–priority services at eNBs and APs for both algorithms. Although the number of WiFi APs are twice the number of LTE eNBs, we can observe that the traffic share of high–priority services at eNBs is greater that the traffic share at APs for both algorithms. This is because if Eq. 2 is satisfied, the high–priority services become cheaper to be supported by eNBs.

The higher acceptance ratio for the ILP algorithm comes at the expense of an increased execution time. An optimal association can be computed by ILP for 150 group requests (750 clients in total) in 250.34 seconds. Conversely, the heuristic can perform the association in 0.07 seconds.



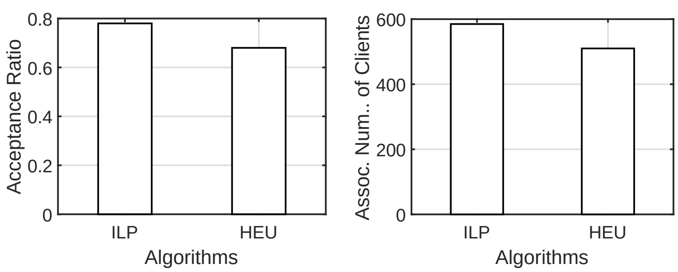
(a) Traffic at eNBs. (b) Traffic at APs. (c) Load at eNBs. (d) Load at APs.

Fig. 1: Traffic and load distributions at eNBs and APs after 150 association requests for all algorithms.



(a) Services supported by eNBs. (b) Services supported by APs.

Fig. 2: Distribution of the 4 possible service classes across both eNBs and APs using the different user association algorithms.



(a) Acceptance ratio. (b) Associated number of UEs.

Fig. 3: Acceptance ratio and total number of served clients.

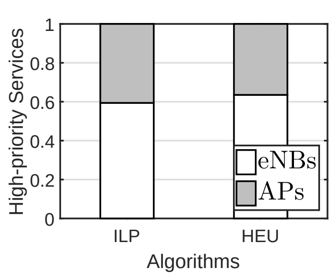


Fig. 4: Share of high–priority services at eNBs and APs.

VI. Proof–of–concept

A. Overview

The proposed user association algorithm has been implemented on the 5G–EmPOWER platform. 5G–EmPOWER is a Multi–access Edge Computing Operating System (MEC–OS) which converges SDN and NFV into a single platform supporting lightweight virtualization and heterogeneous radio access

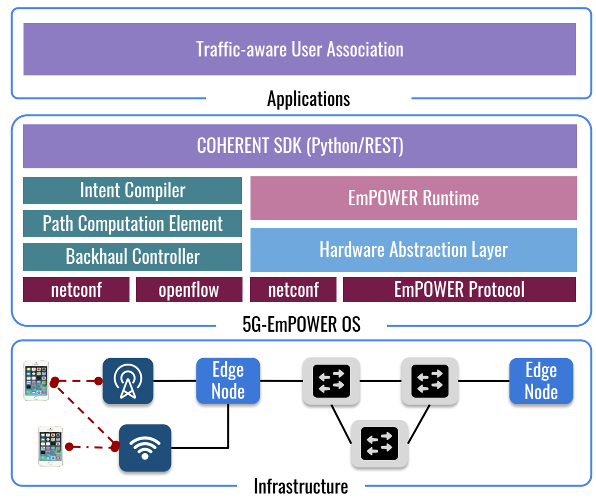
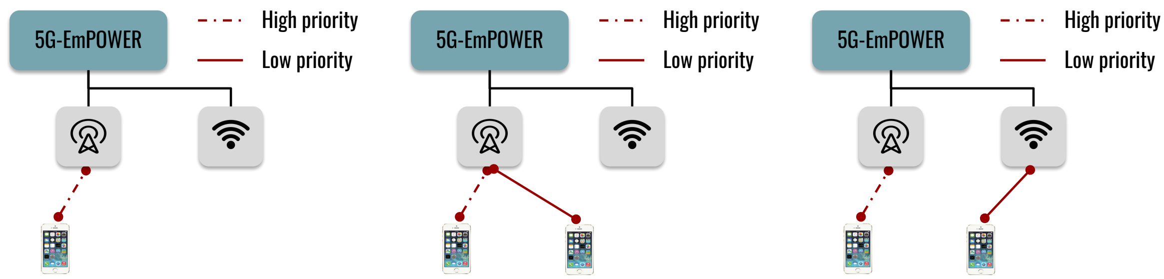


Fig. 5: The 5G–EmPOWER MEC–OS System Architecture.

technologies[2]. A high level view of the the 5G–EmPOWER MEC–OS architecture is sketched in Fig. 5. It is worth noticing that the goal of this section is not to demonstrate the algorithm scalability in a large scale setup, as a matter of fact our deployment consists of one eNB, one AP and two wireless clients. Instead we want to report on a preliminary proof–of–concept implementation of the proposed solution. To the best of the authors' knowledge this is the first real–world open–source SDN platform supporting mobility management applications over heterogeneous Wi–Fi/LTE RANs.

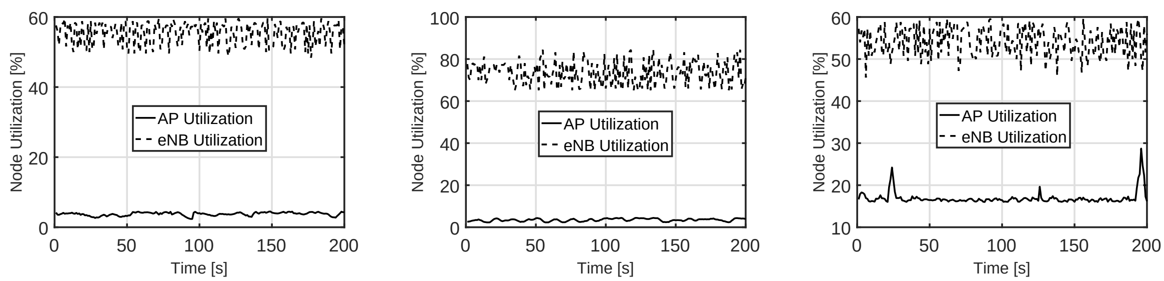
The 5G–EmPOWER MEC–OS consists of a hardware abstraction layer converging several radio access networks control and management protocols into a unified set of abstractions that are then exposed to the application layer. Such abstractions allow the applications layer to implement joint NFV and SDN resource management operations. This includes, for example, joint mobility management and VNF placement/migration schemes as well as radio access and backhaul load balancing. The 5G–EmPOWER MEC–OS currently supports WiFi and LTE radio access nodes. Interaction with SDN–based backhauls is enabled through an Intent–based networking interface. In the rest of this section we will provide a short summary of the *Network Graph* abstraction used to implement the user association algorithm presented in this

[2]Online resources available at: http://empower.create-net.org/



(a) Initial setup. (b) Second client attaches. (c) User re–association.

Fig. 6: Testbed setup. Initially only one client is active with a single high–priority flow served by the LTE eNB. When the second client attaches to the same eNB and starts generating a low–priority flow, the controllers detects a handover opportunity and associates the low–priority user to WiFi AP.



(a) Node utilization (Scenario 1). (b) Node utilization (Scenario 2). (c) Node utilization (Scenario 3).

Fig. 7: The LTE eNB and the WiFi AP cell utilization for the various scenarios.

paper. For a more extensive description we refer the reader to [6].

B. Network Graph

The Network Graph provides network programmers with a full view of the network state. The network graph is exposed as a directed graph $G = (V, E)$ where V is the set of clients and radio access network elements (i.e. WiFi APs and LTE eNBs) and E is the set of edges or links. A weight $\omega(e_{n,m})$ is assigned to each link $e_{n,m} \in E : \omega(e_{n,m}) \in \mathbb{R}$.

The weight assigned to edges can model different aspect of the wireless links. In the current implementation of the 5G–EmPOWER MEC–OS the following types of data–structures can be associated to the edges of the Network Graph:

- *RSSI* (WiFi). The received signal strength indicator as reported by WiFi APs (uplink direction) and wireless clients (downlink direction). Measurements in the downlink direction are taken using the radio resource management features introduced by the 802.11k amendment [22].
- *RSSI/RSRP/RSRQ* (LTE). The carrier received signal strength indicator measures the total power received on reference signals. The RSSI measurement is taken over the full bandwidth while the RSRP is narrow–band. In the RSRQ also the number of PRBs used is considered.
- *Rate Control Statistics* (WiFi). The statistics of the MCS selection algorithm at the AP (downlink). For each supported MCS, the frame delivery ratio and the estimated throughput in the last observation window are reported. Historical, EWMA–filtered values, are also available.
- *Airtime* (WiFi) and *PRB* (LTE) utilization. The fraction of the airtime and of the PRB utilized at, respectively, WiFi APs and LTE eNBs. Notice how, for WiFi APs the airtime utilization is an estimated value while for LTE eNBs it is the actual value.
- *Traffic Matrix* (WiFi/LTE). The number of packets and bytes transmitted/received by each wireless client. The absolute packets/bytes values as well as the bitrate in the last observation window are available to applications. Counters can be defined over an arbitrary portion of the flowspace and are implemented using OpenFlow [25].

The RSSI measurements (in dB) taken at the wireless client side for both WiFi and LTE RAN nodes are used as input for the distance function $dis()$. A wireless client $u \in N_u$ is considered within the coverage radius of a WiFi AP $n \in N_{ap}$ if the RSSI between the WiFi AP and the wireless client $dis(u, n) \leq -70dB$. An RSSI–based metric is used in order to estimate the MCS between WiFi RAN nodes and wireless clients [26]. Whereas, a wireless client $u \in N_u$ is considered within the coverage radius of an LTE eNB $n \in N_{enb}$ if the RSSI between the eNB and the wireless client $dis(u, n) \leq -90dB$. As opposed to WiFi APs, an SINR–based metric is used in order to estimate the MCS between LTE RAN nodes and wireless clients [27]. The packets and bytes counter made available by the network graph trough the *Traffic Matrix* are used to compute the bandwidth requirements of each flow.

C. Evaluation Methodology

The testbed setup consists of one LTE eNB and one Wi–Fi AP. The LTE eNB is based on the Ettus SDR B210 platform [28] and runs the srsLTE software stack [29]. The WiFi is a commercial 802.11n wireless router running a modified version of OpenWRT [30]. Two standard smartphones are used as wireless clients. The 5G–EmPOWER MEC OS as well as the wireless client association heuristic run on a dedicated laptop. The overall network setup is sketched in Fig. 6.

D. Results

Initially, only one wireless client is active and one high–priority video stream with a constant bitrate of 10 Mbps is sent in the downlink direction from the LTE eNB (see Fig. 6a) to the wireless client. As it can be seen in Fig. 7a, in this case the eNB utilization is approximately 50% whereas, the utilization of the AP is negligible (the non–zero utilization is due to the beacon frames that are transmitted periodically by the AP to announce its presence). Then, another wireless client is attached to the LTE eNB and a low–priority traffic stream with an average bitrate of 2 Mbps is sent in downlink direction (see Fig. 6b). As it can be seen in Fig. 7b, the eNB is almost saturated with a utilization of approximately 80%. As for the previous case the AP utilization is still negligible. At this point a handover decision is taken by the heuristic and the second client is moved to WiFi (see Fig. 6c). As it can be seen in Fig. 7c, after the offload is executed by the controller the eNB utilization decreases providing more opportunities for the new clients to be associated with the eNB.

VII. Conclusions

Traffic–aware user association and multi–homing are two promising ways of exploiting the radio resources available in a heterogeneous LTE/WiFi RAN. In this paper we tackled this challenge by presenting a novel formulation of the user association problem for heterogeneous LTE/WiFi RANs. Our formulation builds upon a radio access technology agnostic resource request model and accounts for different traffic classes. Moreover, the problem formulation also supports wireless clients multi–homing allowing different streams from the same wireless client to be served by different RAN nodes. An ILP formulation of the user association problem is compared with a scalable heuristic. Finally, we also reported on a preliminary proof–of–concept implementation of the proposed solution and on its validation over a small scale testbed. As a future work, we want to extend the problem formulation to the wired backhaul and to consider clients mobility using more realistic channel models. We also plan to study the system performances using real traffic traces coming from operational networks. Finally, we intend to increase the testbed size.

References

[1] "MBiT Index," Nokia, Tech. Rep., 2017.

[2] "Ericsson Mobility Report," Ericsson, Tech. Rep., 2017.

[3] "Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016-2021," Cisco, Tech. Rep., 2017.

[4] NOKIA, "Ultra Dense Network (UDN)," 2016. [Online]. Available: http://tools.ext.nokia.com/asset/200295

[5] CISCO, "Architecture for Mobile Data Offload over Wi-Fi Access Networks," 2012. [Online]. Available: https://communities.cisco.com/docs/DOC-35954

[6] R. Riggio, M. K. Marina, J. Schulz-Zander, S. Kuklinski, and T. Rasheed, "Programming abstractions for software-defined wireless networks," *IEEE Transactions on Network and Service Management*, vol. 12, no. 2, pp. 146–162, 2015.

[7] Jin, Wang, Huilin, Jiang, Zhiwen, Pan, Nan, Liu, Xiaohu, You, Tianle and Deng, "Joint user association and ABS proportion optimization for load balancing in HetNet," in *Proc. of IEEE WCSP*, 2015.

[8] J. B. Abderrazak, A. Zemzem, and H. Besbes, "A distributed muting adaptation solution for a qos-aware user association and load balancing in hetnets," in *Proc. of IEEE ICTC*, 2015.

[9] P.-H. Chiang, P.-H. Huang, S.-S. Sun, W. Liao, and W.-T. Chen, "Joint power control and user association for traffic offloading in heterogeneous networks," in *Proc. of IEEE GLOBECOM*, 2014.

[10] Q.-T. Nguyen-Vuong, N. Agoulmine, and Y. Ghamri-Doudane, "Novel approach for load balancing in heterogeneous wireless packet networks," in *Proc. of IEEE NOMS*, 2008.

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

[12] S. De Schepper and J. Famaey, "A transparent load balancing algorithm for heterogeneous local area networks," in *Proc. of IEEE IM*, 2017.

[13] G. Sawma, I. Aib, R. Ben-El-Kezadri, and G. Pujolle, "ALBA: An autonomic load balancing algorithm for IEEE 802.11 wireless networks," in *Proc. of IEEE NOMS*, 2008.

[14] L. Yang, Y. Cui, H. Tang, and S. Xiao, "Demand-aware load balancing in wireless lans using association control," in *Proc. of IEEE GLOBECOM*, 2015.

[15] Zhou, Fanqin and Feng, Lei and Yu, Peng and Li, Wenjing, "Energy-efficiency driven load balancing strategy in LTE-WiFi interworking heterogeneous networks," in *Proc. of IEEE WCNCW*, 2015.

[16] Z. Yang, Q. Yang, F. Fu, and K. S. Kwak, "A novel load balancing scheme in lte and wifi coexisted network for ofdma system," in *Proc. of IEEE WCSP*, 2013.

[17] Roy, Arghyadip and Karandikar, Abhay, "Optimal radio access technology selection policy for LTE-WiFi network," in *Proc. of IEEE WiOpt*, 2015.

[18] Roy, Arghyadip and Chaporkar, Prasanna and Karandikar, Abhay, "An On-Line Radio Access Technology Selection Algorithm in an LTE-WiFi Network," in *Proc. of IEEE WCNC*, 2017.

[19] G. Dandachi, S. E. Elayoubi, T. Chahed, N. Chendeb, and H. Jebalia, "Comparing resource allocation schemes in multi-homed lte/wifi access networks," in *Proc. of IEEE VTC Fall*, 2015.

[20] Bukhari, Jumana and Akkari, Nadine, "QoS based approach for LTE-WiFi handover," in *Proc. of IEEE CSIT*, 2016.

[21] Ting, Alvin and Chieng, David and Kwong, Kae Hsiang and Andonovic, Ivan and Wong, KD, "Dynamic backhaul sensitive network selection scheme in LTE-WiFi wireless HetNet," in *Proc. of IEEE PIMRC*, 2013.

[22] *Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) Specifications*, ANSI/IEEE Std 802.11, LAN/MAN Standards Committee of the IEEE Computer Society Std., 2012.

[23] G. Bianchi and I. Tinnirello, "Remarks on IEEE 802.11 DCF performance analysis," *IEEE communications letters*, vol. 9, no. 8, pp. 765–767, 2005.

[24] J. Du, L. Zhao, J. Feng, J. Xin, and Y. Wang, "Enhanced PSO based energy-efficient resource allocation and CQI based MCS selection in LTE-A heterogeneous system," *China Communications*, vol. 13, no. 11, pp. 197–204, 2016.

[25] ONF, "OpenFlow Switch Specification, Version 1.5.1," https://3vf60mmveq1g8vzn48q2o71a-wpengine.netdna-ssl.com/wp-content/uploads/2014/10/openflow-switch-v1.5.1.pdf.

[26] M. Tauber and S. N. Bhatti, "Low rssi in wlans: Impact on application-level performance," in *Proc. of IEEE ICNC*, 2013.

[27] J. Fan, Q. Yin, G. Y. Li, B. Peng, and X. Zhu, "MCS selection for throughput improvement in downlink LTE systems," in *Proc. of IEEE ICCCN*, 2011.

[28] "Ettus USRP B210," https://www.ettus.com/.

[29] "Software Radio Systems," http://www.softwareradiosystems.com/.

[30] "OpenWRT," https://openwrt.org/.