Radio Access Technology Selection in Heterogeneous Networks

Kinda Khawam[a], Samer Lahoud[b], Marc Ibrahim[c], Mohamad Yassin[b,c], Steven Martin[d], Melhem El Helou[c], Farah Moety[e]

[a]*kinda.khawam@uvsq.fr, University of Versailles.*
[b]*IRISA, University of Rennes 1.*
[c]*ESIB, Saint-Joseph University of Beirut.*
[d]*LRI Laboratory, Paris-Sud University, France*
[e]*Orange Labs, Issy-Les-Moulineaux.*

Abstract

The migration of wireless networking towards the 5G era is distinguished by the proliferation of various Radio Access Technologies (RAT). As no existing technology can be surrogated by another one, the coexistence of today wireless networks is the best solution at hand when dealing with the incessantly growing user demand for bandwidth. Hence, in this heterogeneous environment, users will be able to utilize services through diverse RATs. RAT selection is crucial and must be designed astutely to avoid resource wastage. In this paper, we consider the downlink of a heterogeneous network with two broadband RATs: a primary RAT such as LTE, and a secondary RAT such as WiFi. We start by formulating a centralized approach for the RAT selection as an optimization problem. Then, two distributed approaches are proposed for adequate RAT selection: first, we put forward distributed heuristic algorithms based on the peak rate perceived by users from available RATs. Second, we devise a distributed RAT selection scheme portrayed as a non-cooperative game with a learning-based algorithm to reach the Nash Equilibriums of the RAT selection game. Extensive simulation results show that the proposed distributed algorithms give efficient results compared to the centralized optimal approach. The analysis of the simulation results enables to define pertinent use cases that delimit the scope of the proposed optimal centralized and distributed approaches.

Keywords: RAT Selection; Centralized and Distributed Resource Allocation; Heterogeneous Networks; Non-linear Optimization; Non-cooperative Game Theory.

1. Introduction

Every year, the demand in mobile broadband communications increases spectacularly. Practical solutions need to be proposed to face this imminent thousand-fold traffic augmentation. To address this challenge, ubiquitous radio access will be offered by forthcoming 5G heterogeneous network deployments. On the one hand, the advent of mmWave technology and carrier aggregation mechanisms is inevitable to support higher capacity [1]. On the other hand, improved spectral efficiency and novel heterogeneous network deployments with astute resource sharing are vital to meet the predicted traffic demands for the next decade.

For increased efficiency, heterogeneous networks will be self-organized. Operators will profit from the abundance of diverse Radio Access Technologies (RATs) in the same operating area and devise advanced Radio Resource Management (RRM) schemes to take advantage of the available system resources. Hence, RATs will need to be integrated, with any combination of 3G, WiFi [8], WiMAX [7] and LTE [2]. As highlighted by the seminal paper [3], heterogeneous networks are undeniably presented as a major cornerstone of the upcoming 5G networks. The authors particularly focus on the challenge of integrating different RATs. Moreover, the Horizon 2020 [4] European framework programme for research and innovation identifies that future networks will need to become significantly more heterogeneous and use multiple RATs. This challenge is tackled particularly by the METIS project that lays down the foundations of 5G networks [5]. In such a heterogeneous network, when a new or a handover session arrives, a decision must be astutely made as to which technology it should be associated with. This is known as *RAT Selection*. In such a context, a mobile user will be able to connect concurrently to different RATs by enabling device support for carrier aggregation.

The straightforward approach in apprehending the RAT selection issue is to formulate the problem as centralized optimization task whose objective is to maximize throughput or equivalently minimize delay. In order to derive the expression of the delay, we use an analytical model whose key feature lies in accounting for the effect of interference as well as for the physical layer and channel characteristics in an easy and straightforward manner. On the one hand, the model takes into consideration frequency planning and scheduling aspects; and on the other hand, it provides tractable formulas of the end user mean delay.

While optimization models give an insight into the upper bounds of achievable RAT selection gains, the implementation of these centralized optimal mechanisms are cost prohibitive in real systems. Indeed, RRM mechanisms studied in the state-of-the-art build upon markedly lower complexity distributed schemes. Consequently, the present work is threefold: the first part addresses the RAT selection issue as a centralized optimization problem. The second part proposes simple but cost effective and fully distributed heuristic algorithms. The third part resorts to non-cooperative game theory to put forward a distributed algorithm based on replication dynamics where each mobile user selfishly strives to improve its own performances.

Results are validated through extensive simulations in the practical setting of a geographical area covered by a global LTE network acting as the primary RAT overlapping with several local WiFi hotspots acting as the secondary RAT. This typically corresponds to a WiFi offloading scenario [6]. We begin by examining static scenarios chosen randomly then assess the algorithms performances in a dynamic setting.

The paper is organized as follows. Related work is presented in Section 2. The system model is described in Section 3 with adopted models for network structure, traffic, perceived rates, and cost function. The optimal centralized RAT selection scheme is formulated in Section 4 as a non linear optimization problem. The heuristic distributed approach is explained in Section 6. Furthermore, the RAT selection policy performances are assessed in a realistic dynamic setting. The game theoretic distributed approach is explained in Section 7 and a realistic distributed algorithm based on replicator dynamics is explained in 7.1 to reach the Nash equilibriums of the RAT selection game. Finally, in Section 8, simulations were conducted in a dynamic setting to compare all formulated approaches. We conclude in Section 9.

2. Related Work

The need to fully profit from the large number of currently available RATs is the main driver behind the growing relevance of heterogeneity for future 5G networks. The subject is not only a hot topic for the scientific community but also for the related standardization bodies that are duly specifying procedures to support the interoperability between heterogeneous networks. In fact, the IEEE 802.21 group has defined ([10, 11]) a framework to enable seamless handovers between RATs.

In the state-of-the-art, two approaches are proposed to tackle the RAT selection problem. First, the *centralized approach* where the network performs resource allocation in a way to satisfy all mobile users. Second, the *distributed approach* where mobile users strive to improve their performances on their own.

2.1. Centralized approach

The centralized approach is studied in ([13]-[19]). In [13], [14], and [15], a Semi-Markov Decision Process (SMDP) is proposed to find the optimal RAT selection that maximizes a long-term reward function. Beside load conditions and user spatial distribution, the authors in [15] have also considered different service classes. Recently, the paper in [16] proposes an optimal Joint Call Admission Control approach for initial RAT selection for heterogeneous wireless networks composed of two co-located networks supporting two different service classes. The framework of SMDP is used to formulate the problem as a joint call admission control and RAT selection problem that maximizes the system capacity while selecting the RAT that consumes the least amount of energy. A Markov model for users' network selection in a heterogeneous network is introduced in [19] that accounts for users' mobility, the quality of service, and the price charged by the operator for a given RAT.

Authors in [17] devise a utility-based resource management framework with multiple client classes distinguished by means of a risk-averse utility function. The optimization problem was not solved; instead simple heuristic algorithms were planned to approximate the optimal solution in a narrow set environment. In [18], various wireless networking scenarios embracing diverse technologies and operators are modeled as an optimization problem, using a utility function, able to modulate the weight given to multiple merit parameters (connectivity, preferred operator, handover, link quality), which reflect the requirements of both the network and end-users.

2.2. Distributed approach

The distributed approach is tackled in ([20]-[21]). In [20], a measurement-based network selection technique that estimates QoS information by bootstrap approximation is proposed. Unnecessary handovers between RATs are filtered using Bayesian estimation and cumulative sum monitoring. In [21], a distributed RAT selection scheme based on utility function and integer linear programming is proposed, taking into account bandwidth, packet loss, delay and energy information. Unfortunately, it is unrealistic to establish base requirements for every service and to assume that the total network bandwidth is enough to satisfy all traffic.

In ([22]-[26]), game theory is used for the distributed RAT selection problem. A comprehensive review of the solutions and challenges of game theory-based network selection is provided in [27]. Players (*i.e.,* the individual users) have actually no precise idea on the global network state and will try to reach a mutually agreeable solution, or equivalently, a set of strategies they will unlikely want to change. In [22], the selection between WiFi and 3G is formulated as a symmetric non-cooperative game that reduces to a threshold policy. The work in [23] load balances traffic between WiMAX and WiFi using replicator dynamics. In [24], vertical hand-offs are initiated in order to maximize resources consumption while striving to meet QoS requirement of end-users. Vertical hand-off with coalition game can reduce the service time for hand-off calls. The conceived algorithm in [25] enables users to single out the optimal RAT based on Bayesian Nash-equilibrium point. The reached equilibrium reduces the hand-off delay while maximizing the offered QoS at the lowest price. In [26], non-cooperative game is again used to single out the best RAT. The paper shows that RAT selection games with a single traffic class converge to Nash equilibriums, while an improvement path can be repeated infinitely with a mixture of classes. The Pareto-efficiency of the Nash equilibriums of the proposed games is analyzed and conditions to attain it are derived.

2.3. Our contributions

This work starts by introducing and formulating the Radio Access Technology (RAT) Selection issue in heterogeneous wireless networks. In this formulation, we use a novel analytical radio model whose key feature lies in accounting for the effect of interference as well as for the physical layer and channel characteristics in an easy and straightforward manner. Our aim is to study the RAT selection problem from various perspectives that cover the state-of-the-art approaches:

- The first part in this work addresses the RAT selection as a global centralized non-linear optimization problem. After analyzing the properties of the problem, we examine different algorithmic solutions based on an exhaustive search and on a Mixed Integer Linear Program (MILP) re-formulation. We finally introduce a new approach named enhanced search algorithm that enables to drastically reduce the computation complexity. This approach takes into account the aforementioned analytical interference model in order to discretize the coverage area into zones.

- We propose and study heuristic approaches that are based on simple decisions made by the users that necessitate no signaling information.

- We propose and study a game-theoretic approach that mimics the behavior of selfish users. A fully distributed learning-based algorithm is adopted to reach the Nash equilibriums of the RAT selection game.

The complexity analysis, the discussions, and the numerical results provided in this work confirm that no approach prevails and a thorough study consists in handling all facets.

3. Network Model on the Downlink

We consider the downlink of a heterogeneous network with two broadband RATs: a primary RAT such as LTE, and a secondary RAT serviced by WiFi. An LTE cell range is in the order of a few kilometers while the WiFi cell range spans from a few tens to a few hundreds of meters only [29]. Hence, typically, an LTE cell will be covered with several WiFi antennas as in Figure 1. We will use the term BS (Base Station) to designate the serving antenna in any RAT: a WiFi antenna of an LTE BS.

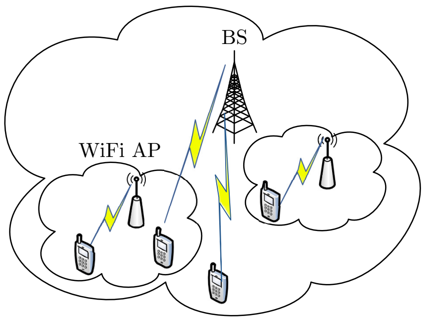

Figure 1: Heterogeneous wireless access

We assume that mobile users compare the received power levels from all WiFi BSs and LTE BSs. Then, they are associated to the WiFi BS and to the LTE BS that provide the highest signal strength. Users benefiting from the coverage of both LTE and WiFi are coined *hybrid* and engage in the selection process. The other mobile users profit only from the primary RAT coverage.

In what follows, we will present our model in terms of traffic and data rate. Accordingly, we define a suitable cost function.

3.1. Network Structure

The network is set to be a two-dimensional disc of radius R_{Net}. The index x is used throughout the paper to designate a given RAT, $x = P$ indexing the primary RAT (LTE) and $x = S$ referring to the secondary RAT (WiFi). We consider a central cell in the network as the reference cell. Our reference cell comprises one LTE antenna and multiple WiFi antennas. Let B_0^x denote the considered BS of RAT x in our reference cell. Let then $\{B_i^x, i = 1, .., N^x\}$ be the set of the N^x interfering BSs deployed for RAT x (co-channel BSs with B_0^x).

Let n_x be the total number of users serviced by B_0^x and n_S the number of WiFi users ($n_S \leq n_x$). These n_S users are engaged in the RAT selection and coined *hybrid* users; let $H = \{1, ..., n_S\}$ be the set of hybrid users.

For LTE, the BSs produce a hexagonal lattice since we assume a constant distance d_P between any two nearby BSs. This could be seen as a grid with honeycomb shape where each BS is the center of a hexagon of side $\mathcal{R}_P = d_P/\sqrt{3}$. For WiFi, our model assumes a uniform distribution of BSs of coverage radius $\mathcal{R}_S$ (that depends on the power detection threshold of mobile users).

The main notations used in the present paper are reported in Table 1. These notations will be used gradually in the following sections.

3.2. Radio Model

In this section, our goal is to derive the probability distribution of the Signal to Interference plus Noise Ratio (SINR) of hybrid users in the reference cell. The SINR depends on the user location in the cell and on its radio conditions (described based on path loss and Rayleigh fading models). We denote by $r(k)$ the distance from user k to B_0^x, while $r_i(k)$ represents the distance from interfering B_i^x to that same user k.

The power received by user k depends on the BS emitted power and radio channel attenuation, and varies with time due to fading. In RAT x, let P^x be the power emitted by B_0^x. The received power is then:

$$Pr^x(k) = P^x \cdot \gamma^x(k) \cdot X_k, \tag{1}$$

where the random variables X_k are i.i.d. and follow an exponential distribution of parameter λ as we consider fast fading [32]. The path loss for user k, $\gamma^x(k)$, depends on the distance $r(k)$ from B_0^x and is given by:

$$\gamma^x(k) = A^x/r(k)^\beta \tag{2}$$

where β is the path loss exponent and A^x a constant characterizing the radio propagation in B_0^x. Finally, the SINR of user k in RAT x is given by:

$$\mathrm{SINR}^x(k) = \frac{P^x \cdot \gamma^x(k) \cdot X_k}{\sigma^2 + \sum_{i=1}^{N^x} P^x \gamma_i^x(k) X_i} \tag{3}$$

Notation	Definition
x	The index x used to designate the RAT
B_0^x	The considered BS of RAT x
n_x	The number of users serviced by B_0^x
N_x	The number of interfering BSs (B_i^x)
$H = \{1, ..., n_S\}$	The set of hybrid users engaged in the RAT selection.
π_j	The transmit power at level j
$\mathcal{R}_x$	The coverage radius of B_0^x
P^x	The amount of power emitted by B_0^x
$r(k)$	The distance between user k and its serving BS B_0^x
$\gamma^x(k)$	The path loss for user k in B_0^x
$\gamma_i^x(k)$	The path loss between interference BS B_i^x and user k
σ^2	The background noise
X_k	The fast fading variables (i.i.d. following an exponential distribution of parameter λ)
β	The path loss exponent
A^x	A constant characterizing the radio propagation in B_0^x
$\text{SINR}^x(k)$	The Signal to Interference and Noise Ratio of user k in RAT x
$\Delta_{x,j}$	The j^{th} peak rate realized in RAT x
M_x	The total number of peak rates realized in RAT x
$\chi_{k,x}$	The peak rate of user k in RAT x
$R_{k,x}$	The mean rate of user k in RAT x
$T_{k,x}$	The bit transfer time of user k in RAT x
C_k	The cost function of user k
θ_k	The probability that user k is associated with the secondary RAT (or equivalently the percentage of its traffic serviced by B_0^S)
C_{tot}	The total cost of hybrid users
Z_m^x	Zone m in RAT x where users realize peak rate $\chi_{x,m}$
$p_{k,x}$	The ratio of hybrid users in RAT x that perceive peak rate $\chi_{k,x}$

Table 1: Notation Summary

where σ^2 is the background noise. Further, $\gamma_i^x(k)$ is the path loss between B_i^x and user k given by $\gamma_i^x(k) = \frac{A^x}{r_i(k)^\beta}$.

Hereafter, we compute the probability $S^x(k, \delta)$ that the SINR of a hybrid user k is larger than a threshold δ as in $S^x(k, \delta) = \mathbb{P}(\text{SINR}^x(k) > \delta)$. Based on the above notations and according to a previous result obtained in [30], we have what follows:

$$S^x(k,\delta) = \exp\left(\frac{-\delta\lambda\sigma^2}{\gamma^x(k)P^x}\right)\prod_{i=1}^{N^x}\frac{1}{1+\frac{\delta\gamma_i^x(k)}{\gamma^x(k)}} \approx \exp\left(\frac{-\delta\lambda\sigma^2}{\gamma^x(k)P^x}\right)\cdot\exp\left(-\delta\sum_{i=1}^{N^x}\frac{\gamma_i^x(k)}{\gamma^x(k)}\right) \tag{4}$$

As we can see in equation (4), the computation of the total amount of interference involves a discrete sum over all BSs in RAT x that we denote by I^x:

$$I^x(k) = \sum_{i=1}^{N^x}\frac{\gamma_i^x(k)}{\gamma^x(k)} \tag{5}$$

Note that $I^x(k)$ can only be evaluated numerically. Therefore, we have recourse to the so-called fluid model proposed in [30] and [31] to propose a tractable formula given by what follows:

$$I^x(k) = \frac{2\pi\kappa_x\varrho_x}{R_{Net}^{\beta-2}}\left(\frac{(1-\tau_{k,x}{}^{2-\beta})}{2-\beta} + \frac{(\tau_{k,x}{}^{-\beta}-1)(\frac{r(k)}{R_{Net}})^2}{4/\beta} + \frac{(\tau_{k,x}{}^{-\beta-2}-1)(\frac{r(k)}{R_{Net}})^4}{8/\beta}\right) \tag{6}$$

where $\tau_{k,x} = \frac{\mathcal{R}_x}{R_{Net}}$, ϱ_x is the density of interfering BSs in RAT x and κ_x is a given constant such as: for WiFi, $\kappa_S = 1$; whereas for LTE, κ_P depends on the path loss factor β and has a constant value for a given β; these values are obtained numerically and given in [30]. Finally, $\varrho_P = \frac{2}{3\sqrt{3}RF_P(\mathcal{R}_P)^2}$ and $\varrho_S = \frac{1}{RF_S\pi(\mathcal{R}_S)^2}$ where RF_x is the frequency reuse factor in RAT x. For LTE, we consider full frequency reuse (*i.e.*, $RF_P = 1$) [28]. For WiFi, a proper deployment typically uses three non-overlapping independent channels (*i.e.*, $RF_S = 3$) [42].

3.3. Rate Model

3.3.1. Peak Rate

Modulation and coding constraints result in a discrete set of M_x peak rates in RAT x. Let $\Delta_{x,j}$ be the j^{th} peak rate realized by a mobile user if its SINR lies within the interval $[\delta_j^x, \delta_{j+1}^x[$, $j = 1,\ldots,M_x$ (by convention $\delta_{M_x+1} = \infty$). The peak rate of a user k, denoted by $\chi_{k,x}$, is given by what follows:

$$\chi_{k,x} = \begin{cases} 0 & \text{if } \mathrm{SINR}^x(k) < \delta_1^x \ , \\ \Delta_{x,1} & \text{if } \delta_1^x \leq \mathrm{SINR}^x(k) < \delta_2^x \ , \\ ... & \\ \Delta_{x,M_x} & \text{if } \delta_{M_x}^x \leq \mathrm{SINR}^x(k) < \delta_{M_x+1}^x = \infty. \end{cases}$$

The SINR can take all possible values due to fading, which means that the peak rate varies for a given user in time. Hence, the probability that the SINR of user k lies within the interval $[\delta_j^x, \delta_{j+1}^x[$ is given by:

$$\mathbb{P}(\delta_{j-1} < \mathrm{SINR}^x(k) < \delta_j) = S^x(k, \delta_j) - S^x(k, \delta_{j+1}) \tag{7}$$

The standardized numerical values for peak rates are displayed hereafter:

- In the primary RAT, we have $\chi_{k,P} \in \{11.2, 22.4, 25.2, 33.6, 50.4, 67.2, 75.6, 100.8\}$ Mbits/s for LTE [2].
- In the secondary RAT, we have $\chi_{k,S} \in \{6, 9, 12, 18, 24, 36, 48, 54\}$ Mbits/s for WiFi [8].

Hence, in the present heterogeneous network, at any time instant, each hybrid user perceives two peak rates $\chi_{P,k}$ and $\chi_{S,k}$ provided by the two co-localized RATs.

3.3.2. Scheduling

The aforementioned peak rates are realized by mobile users if they are alone in the cell. In a practical setting, the actual data rates obtained are contingent upon the scheduling scheme adopted in each RAT. The scheduling scheme must inevitably provide fairness to serviced users. The fairness issue arises whenever a limited amount of resources is to be shared by many users. In a wireless environment, due to random channel variations, we must distinguish between temporal fairness, which means that all users get the same amount of time slots, and utilitarian fairness, where all users get the same share of the overall system capacity, which leads to equal rates. While these two types of fairness are equivalent in a wire-line environment, they can be substantially different in a wireless environment. Therefore, we consider in this paper two relevant fair scheduling policies: *Fair Time Sharing* which ensures temporal fairness and *Fair Rate Sharing* which insures utilitarian fairness.

Fair time sharing. All active users are given the same chance to access resources. Hence, the data rate of user k when assigned to RAT x in the reference cell is the peak rate of user k divided by the number of users associated with the same RAT x:

$$R_{k,x} = \frac{\chi_{k,x}}{\sum_{i=1}^{n_x} \mathbb{1}_{\{\text{user i joins RAT x}\}}} = \frac{\chi_{k,x}}{1 + \sum_{i \neq k}^{n_x} \mathbb{1}_{\{\text{user i joins RAT x}\}}}, \quad (8)$$

where,

$$\mathbb{1}_{\{\text{user } i \text{ joins RAT x}\}} = \begin{cases} 1 \text{ if user } i \text{ joins RAT x} \\ 0 \text{ otherwise.} \end{cases} \quad (9)$$

Fair rate sharing. The data rates of all active users are made equal. Hence, the data rate of user k when assigned to RAT x in the reference cell is:

$$R_{k,x} = 1/\sum_{i=1}^{n_x} \frac{\mathbb{1}_{\{\text{user } i \text{ joins RAT x}\}}}{\chi_{i,x}}. \quad (10)$$

In the primary RAT, users will get assigned different sets of subcarriers over different time slots (called resource blocks), hence the multiple access division is done in both frequency and time. We consider that all resource blocks are allocated to a given user at a time and that fair time sharing scheme is applied to service users in turns. In the secondary RAT, we consider the fair rate scheme because it is the only possible resource sharing model that stems from the CSMA (Carrier Sense Multiple Access) protocol adopted in WiFi. The seminal work in [36] provides proof for the latter. In fact, the uplink traffic is neglected in this paper which leads to a fair access scheme on the downlink channel in WiFi. However, when a low rate user captures the channel, it will use it for a long time, which penalizes high rate users and reduces the fair access policy to a case of fair rate policy. In this context, we assume no collisions, a constant MAC frame size and neglecting the 802.11 waiting times (*i.e.*, DIFS, SIFS, ...) in comparison with transmission times.

3.4. Traffic Model

We denote by θ_k the probability that hybrid user k is associated with the secondary RAT or equivalently the percentage of its traffic serviced by the WiFi BS. Hence, $1-\theta_k$ is the probability that hybrid user k is associated with the primary RAT. We only consider elastic traffic for which TCP protocol is typically used at the transport layer. This type of application adapts to available resources and is delay tolerant. Moreover, a worst case scenario is chosen where each user has persistent traffic.

3.5. Cost Function

The objective of the present traffic allocation is to set the amount of traffic that every user should convey through each RAT so that all users are satisfied. Satisfaction for a user is defined here as the minimization of the cost it perceives subsequent to a given RAT selection. The cost function adopted is an image of service time. As we consider elastic traffic, the user satisfaction increases with the perceived throughput [37] or equivalently decreases with the service time. We denote by $T_{k,x}$ the mean amount of time necessary to send a data unit through RAT x, given by $T_{k,x} = \mathbb{E}[1/R_{k,x}]$.

Thus, for the primary RAT in our reference cell (serviced by B_0^P) where we consider fair time sharing, we have the following mean data transfer time:

$$T_{k,P} = \frac{\sum_{i=1,i\neq k}^{n_S}(1-\theta_i)+\overline{n}+1}{\chi_{k,P}} \tag{11}$$

where $\theta_i = \mathbb{E}[\mathbb{1}_{\{\text{user i joins RAT S}\}}]$, and $\overline{n} = n_P - n_S$ is the number of mobile users that are only covered by B_0^P.

For the secondary RAT (serviced by B_0^S) where we consider fair rate sharing, we have the following mean data transfer time:

$$T_{k,S} = \frac{1}{\chi_{k,S}} + \sum_{i=1, i\neq k}^{n_S} \frac{\theta_i}{\chi_{i,S}} \tag{12}$$

Therefore, the cost function of a hybrid user k, defined as its expected time necessary to send a unit of data in this heterogeneous environment, is given by the following:

$$C_k = T_{k,P} \times (1 - \theta_k) + T_{k,S} \times \theta_k \tag{13}$$

4. The Optimal RAT Selection Problem

In the present section, we formulate a centralized approach for the RAT selection as a global non-linear optimization and analyze its salient properties. Based on these properties, we examine different algorithmic solutions based on an exhaustive search and on a Mixed Integer Linear Program (MILP) reformulation. We finally introduce a novel enhanced search algorithm that enables to drastically reduce the computation complexity. We assume the existence of a central entity responsible of routing the downlink traffic of each user (for instance Base Band Units in a Cloud-RAN architecture [38]). The central entity may connect the two RATs to a common core network or directly to the Internet.

4.1. Problem Formulation

The network assigns the traffic of each hybrid mobile user among RATs in order to minimize the total network cost. In the following, the network cost is computed for the reference cell, comprising primary and secondary RATs. This cost denoted by C_{tot} is defined as the sum of the individual costs of hybrid users and given by:

$$C_{tot}(\theta_k, k \in H) = \sum_{k=1}^{n_S} C_k(\theta_k) \tag{14}$$

where H denotes the set of hybrid users.

This optimization problem $(\mathcal{P})$ is introduced hereafter:

$$(\mathcal{P}) : \text{Minimize } C_{tot}(\theta_k, k \in H) =$$
$$\sum_{k=1}^{n_S} \left(\frac{\sum_{i=1, i \neq k}^{n_S} (1-\theta_i) + \overline{n} + 1}{\chi_{k,P}} \times (1-\theta_k) + \left(\frac{1}{\chi_{k,S}} + \sum_{i=1, i \neq k}^{n_S} \frac{\theta_i}{\chi_{i,S}}\right) \times \theta_k \right) \tag{15}$$
$$\text{Subject to: } 0 \leq \theta_k \leq 1, \ \forall k \in H \tag{16}$$

$(\mathcal{P})$ is a non-linear optimization problem that consists in minimizing the objective function $C_{tot}(\theta_k, k \in H)$ over the convex polytope $\{ 0 \leq \theta_k \leq 1, \ k = 1, .., n_S\}$. The vertices of this polytope correspond to the points where $\theta_k = 0$ or $\theta_k = 1$ for each $k = 1, .., n_S$. Theorem 1 demonstrates that the optimal solution for Problem $(\mathcal{P})$ is always reached on the vertices of the polytope.

Theorem 1. *An optimal solution to Problem $(\mathcal{P})$ is reached on the vertices of the convex polytope.*

Proof. Suppose that an optimal solution to Problem $(\mathcal{P})$ is reached for $\hat{\theta} = (\hat{\theta}_k, k \in H)$. Let us proceed by contradiction and suppose that there exists an integer j with $1 \leq j \leq n_S$ for which $\hat{\theta}_j \notin \{0, 1\}$. The function $C_{tot}(\theta_j) := C_{tot}(\theta_j, \theta_k = \hat{\theta}_k, k \in H - \{j\})$ is obtained by taking $\theta_k = \hat{\theta}_k, k \in H - \{j\}$ as a linear function of variable θ_j. Trivially, $C_{tot}(\theta_j)$ attains an optimal value for $\theta_j = 0$ or $\theta_j = 1$. Therefore, $\hat{\theta}_j \in \{0, 1\}$ and our proof is completed by contradiction. $\square$

Theorem 1 implies that each hybrid user k is either assigned to the primary RAT or to the secondary RAT. Thus, no load sharing between RATs is required which avoids expensive repetitive transfers between RATs. In fact, the different technologies may have different delays, packet sizes or coding systems. Hence, reassembling messages sent via two RATs may be hazardous.

4.2. Algorithmic Solutions

4.2.1. Exhaustive Search Algorithm

The optimal solution for Problem $(\mathcal{P})$ can be directly computed by using an exhaustive search algorithm. However, the time complexity of an exhaustive search algorithm is exponential in the number of hybrid users n_S. Indeed, Theorem 1 implies that the optimal solution to Problem $(\mathcal{P})$ is obtained on the vertices of the polytope. Therefore, the objective function C_{tot} should be evaluated on each of the vertices, corresponding to the points where $\theta_k = 0$ or $\theta_k = 1$ for each $i = 1, .., n_S$. The cardinality of the set of vertices is given by 2^{n_S}, hence the time complexity for computing the minimum value of the function over this set is in $\mathcal{O}(2^{n_S} \log(2^{n_S}))$, using for instance a quicksort algorithm.

4.2.2. MILP Reformulation

Problem $(\mathcal{P})$ is non-linear due to the quadratic terms $\theta_k \times \theta_i$ in the objective function C_{tot}. Yet, following theorem 1, these variables can be considered binary without affecting the optimality of the problem. Hence, the product expressions of binary variables $\theta_k \times \theta_i$ can be reformulated into linear expressions and the optimization problem becomes a Mixed Integer Linear Program (MILP) as explained hereafter.

In order to linearize our objective function, we replace the non-linear terms by new variables and additional inequality constraints, which ensure that new variables behave according to the one they replace. Let us start by replacing each occurrence of $\theta_k \times \theta_i$ in the objective function by a new variable z_{ki}. Thus, the reformulated objective function $C_{tot}(\theta_k, z_{ki}, k \in H, i \in H - \{k\})$ is *linear* in the variables θ_k and z_{ki}. Then, we add the following *linear* inequalities to the set of constraints:

$$z_{ki} - \theta_k \leq 0, \quad \forall k \in H, \forall i \in H - \{k\}, \tag{17}$$

$$z_{ki} - \theta_i \leq 0, \quad \forall k \in H, \forall i \in H - \{k\}, \tag{18}$$

$$\theta_k + \theta_i - z_{ki} \quad \leq 1, \quad \forall k \in H, \forall i \in H - \{k\}. \tag{19}$$

Inequalities (17) and (18) ensure that z_{ki} is equal to zero when either θ_k or θ_i is equal to zero, while inequalities (19) force z_{ki} to be equal to one if both θ_k and θ_i are equal to one. Further, the bound constraints on the variables z_{ki} are given by:

$$0 \leq z_{ki} \leq 1, \forall k \in H, \forall i \in H - \{k\}. \tag{20}$$

Finally, the problem $(\mathcal{P})$ is reformulated into a MILP denoted $(\mathcal{P}')$ and given by:

$$\begin{aligned} (\mathcal{P}'):\quad & \text{Minimize} \quad && C_{tot}(\theta_k, z_{ki}, k \in H, i \in H - \{k\}) \\ & \text{Subject to:} \quad && (17), (18), (19), (20), \\ & && \theta_k = \{0, 1\}, \forall k \in H. \end{aligned} \tag{21}$$

$(\mathcal{P}')$ is typically solved using a branch-and-bound approach based on linear-programming. The idea of this approach is to solve Linear Program (LP) relaxations of the MILP and to look for an integer solution by branching and bounding on the decision variables provided by the LP relaxations. Thus, in a branch-and-bound approach the number of integer variables determines the size of the search tree and influences the execution time of the algorithm. For large instances of the problem, obtaining an optimal solution of problem $(\mathcal{P}')$ remains a hurdle. Thus, in the following section, we introduce a novel approach that drastically reduces the computation time of a solution to the RAT selection problem.

5. Cell Decomposition

In this section, we propose to discretize the coverage area into *zones* characterized by similar radio conditions, in particular similar peak rates, as illustrated in Figure 2. The reason behind this decomposition is to make the computation tractable. In fact, the interference and the path-loss factor are different for every user depending on its position in the cell. In practical network dimensioning, it is

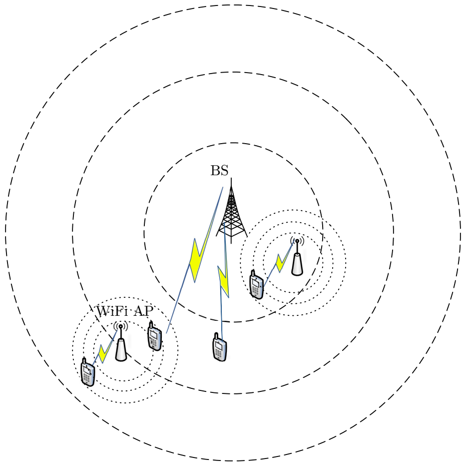

Figure 2: Geographical zones for primary and secondary RAT

not possible to use, for each user, exact values for these parameters but average values among users [33]. Hence, for each RAT x, the cell will be logically divided into M_x concentric discs of radii r_m^x, $m = 1, ..., M_x$ (recall that we have M_x peak rates in RAT x), and the area between two adjacent circles of radii r_{m-1}^x and r_m^x is called zone $\{Z_m^x, m = 1, .., M_x\}$.

5.1. Average Parameters per Zone

Accordingly, we compute for any user in zone Z_m^x, the SINR within that zone by replacing, in (3), $\gamma^x(k)$ by its sample average $\hat{\gamma}^x(Z_m^x)$:

$$\hat{\gamma}^x(Z_m^x) = \frac{A^x}{\pi(r_k^x)^2 - \pi(r_{k-1}^x)^2} \int_{r_{k-1}^x}^{r_k^x} \frac{2\pi r}{r^\beta} dr \tag{22}$$

Our target is to define the above mentioned zones in such a way that users in RAT x, located in zone Z_m^x, are most likely to get peak rate $\chi_{x,m}$ (with a high probability e.g. $\mathbb{P}_{th} = 0.9$). Let us perform the computation necessary for localizing users belonging to the first zone, where $\mathbb{P}(\text{SINR}^x(k) \geq \delta_1^x) \geq \mathbb{P}_{th}$ (r_0^x being zero). According to (4) and (6), we get:

$$F(r(k)) \leq \frac{1}{\delta_1^x} \ln(\frac{1}{\mathbb{P}_{th}}) \tag{23}$$

where $F(r(k))$ is a positive increasing function in $r(k)$ such as:

$$F(r(k)) = \frac{\lambda\sigma^2(\beta+2)(r(k))^\beta}{2A^x \cdot P_x^T} + \varrho_{k,x} \times (\frac{(1-\tau_{k,x}{}^{2-\beta})}{\beta-2} + \frac{(\tau_{k,x}{}^{-\beta}-1)(\frac{r(k)}{R_{Net}})^2}{4/\beta} + \frac{(\tau_{k,x}{}^{-\beta-2}-1)(\frac{r(k)}{R_{Net}})^4}{8/\beta})$$

Thus, inequality (23) yields $r_1^x = F^{-1}(\frac{1}{\delta_1^x} ln(\frac{1}{\mathbb{P}_{th}}))$. Users situated at a distance not exceeding r_1^x from their serving B_0^x perceive the lowest peak rate $\chi_{x,1}$ with high probability $\mathbb{P}_{th}$. We conclude that zone Z_1^x is the disc of radius r_1^x. Similarly, we compute the radii delimiting the remaining zones.

5.2. Enhanced Search Algorithm

As shown in the previous section, a zone in RAT x is defined as the geographical area where users perceive a similar peak rate $\chi_{k,x}$. Suitably, the number of zones is low: each RAT x offering M_x peak rates leads to a total of $M_P \times M_S$ possible combinations of primary and secondary peak rates. In practice, as the coverage of a WiFi BS is relatively small compared to the primary RAT coverage [29], a hybrid user perceives a maximum of two primary peak rates. Thus, the number of zones reduces to $2 \times M_S$ which gives 16 active zones with WiFi as the secondary RAT. This is typically the case when the coverage of a WiFi hotspot is smaller than the difference between the radii of two successive LTE zones.

Owing to the zone discretization, we introduce an enhanced search algorithm that computes the optimal solution of the RAT selection problem with reduced complexity. Contrary to the exhaustive search algorithm that associates a decision variable with each user, the enhanced search associates one variable with the set of users in each zone designating the ratio of users associated with the primary RAT. Then, the enhanced search evaluates the cost function C_{tot} for each possible ratio. The output of the algorithm is the optimal ratio of users in each zone associated with each RAT.

Compared to the complexity of the exhaustive search in $\mathcal{O}(2^{n_S}\log(2^{n_S}))$, the enhanced search complexity is in $\mathcal{O}(\prod_{i=1}^{z}(N_i+1))$, where N_z is the number of users in zone z. Thus, the latter remains tractable for a reduced number of zones, whereas the former has a combinatorial complexity explosion for a large number of users. For example, if the secondary RAT has a total of 25 users equally distributed among 5 zones, the exhaustive search algorithm needs to compare 2^{25} values of cost function C_{tot}, whereas the enhanced search compares only $(5+1)^5$ values.

By virtue of the complexity reduction, we adopt in the rest of this document the enhanced search algorithm as the solution method for the centralized optimal RAT selection problem.

6. Heuristic Distributed Approach

Although optimal, the centralized approach presented in the previous sections can be costly and resource consuming (even for the enhanced search algorithm). From a system design perspective, distributed mechanisms using pre-configured and simple resource allocation *rules* are quite appealing. Therefore, in section 6.1, we propose two lightweight heuristic distributed algorithms that approximate the optimal solution. In Section 6.2, extensive simulations are conducted to compare the two algorithms in static and dynamic scenarios.

6.1. Definition of Heuristic Algorithms

We propose two distributed heuristic algorithms for RAT selection called respectively *Peak Rate Based Algorithm* denoted by **R** and *Probabilistic Peak Rate Based Algorithm* denoted by **PR**. These heuristics were introduced in our previous work in [41]. In the present work, the performances are thoroughly assessed in static as well as dynamic scenarios. Recall that for the optimal solution, each user selects a single RAT. Therefore, our proposed heuristics are built in order to mimic such behavior: each user exploits available signaling information and chooses the best RAT to connect to. Results show that the proposed algorithms give efficient results (in terms of total network cost) compared to the optimal centralized algorithm depending on the spatial users distribution. The two algorithms are described hereafter. The Peak Rate Based Algorithm **R**, presented in Algorithm 1, in which each hybrid user, compares the peak rates provided by the primary and the secondary RAT. Then, the hybrid user chooses to connect to the RAT providing the highest peak rate.

Algorithm 1 Peak Rate Based Algorithm

Require: Hybrid user k measures peak rates $\chi_{k,P}$ and $\chi_{k,S}$
1: **if** $\chi_{k,P} > \chi_{k,S}$ **then**
2: User k connects to primary RAT
3: **else**
4: User k connects to secondary RAT
5: **end if**

Compared to the first heuristic algorithm, the Probabilistic Peak Rate Based Algorithm **PR**, presented in Algorithm 2, adds uncertainty to the choice made by the hybrid user. The probability to connect to a given technology is proportional to the peak rate perceived in it. Particularly, we draw a random uniform variable $x \in [0, \chi_{k,P} + \chi_{k,S}]$ as noted in step 1 of the algorithm. Thus, the probability that x is lower than $\chi_{k,P}$ equals $\frac{\chi_{k,P}}{\chi_{k,P}+\chi_{k,S}}$, whereas the probability that x is greater than $\chi_{k,P}$ equals $\frac{\chi_{k,S}}{\chi_{k,P}+\chi_{k,S}}$. The rationale behind this choice is to enable by simple comparisons (x greater or lower than $\chi_{k,P}$) to perform a random RAT selection decision proportional to the ratio of the peak rates.

In practice, in LTE, mobile users measure the channel quality based on pilots, *i.e.,* Cell-Specific Reference Signals (CRS) that are spread across the whole band independently of the individual users allocation. The peak rate can be easily inferred from evaluated channel quality. In WiFi, the mobile users rely on the beacon frame (sent at least every 100 ms) to evaluate their performances (through

Algorithm 2 Probabilistic Peak Rate Based Algorithm

Require: Hybrid user k measures peak rates $\chi_{k,P}$ and $\chi_{k,S}$
1: User k draws uniformly a random variable $x \in [0, \chi_{k,P} + \chi_{k,S}]$
2: **if** $x < \chi_{k,P}$ **then**
3: User k connects to primary RAT
4: **else**
5: User k connects to secondary RAT
6: **end if**

their received signal power level) and hence their peak rate.

6.2. Comparison of Distributed Heuristic Algorithms

In this section, we compare the distributed heuristic algorithms. We consider 15 hybrid users in the reference cell. We begin in 6.2.1 with a preliminary study that deals with specific scenarios. We carry on with a thorough comparison in 6.2.2 where general scenarios are analyzed. To study more thoroughly the problem at hand, a dynamic setting is studied in 6.2.3 with users arriving to an arbitrary zone and leaving the system after being serviced. In the following, we resort to what we call probability vectors, a practical means to generate and thereby study a given scenario in depth: we denote by $p_{k,x}$ the ratio of hybrid users in RAT x that perceive peak rate $\chi_{k,x}$. The probability vector is then given by $[p_{1,x}\, p_{2,x} \ldots p_{M_x,x}]$, where M_x is the number of peak rates in RAT x. Note that the value of the peak rates in the probability vector take already into account the interference according to the introduced radio model. Moreover, the scheduling enables to compute the perceived rates of each user.

6.2.1. Specific Scenarios

We present below the probability vectors of three different scenarios [1], along with the results gathered after running 30 simulations with 15 hybrid users. We

[1] On each boxplot, the central mark is the median, the edges of the box are the 25^{th} and 75^{th} percentiles, the whiskers extend to the most extreme data points not considered outliers, and outliers are plotted individually.

denote by $\frac{R}{O}$ the network cost ratio of the peak rate algorithm against that of the enhanced search optimal algorithm. The network cost for each algorithm (*e.g.*, **R** or **O**) is given by equation (14)). Similarly, $\frac{PR}{O}$ is the ratio corresponding to the probabilistic peak rate algorithm. Finally, $\frac{PR}{R}$ is used to compare the heuristic algorithms.

- Scenario 1: Primary RAT probabilities are [0 0 0 0 0 0 0.3 0.7] and secondary RAT probabilities are [0.7 0.25 0.05 0 0 0 0 0]. This choice implies that about 70% of the hybrid users perceive a peak rate of 6 Mbits/s from the secondary RAT, while 70% perceive around 100 Mbits/s from the primary RAT (cf. the peak rates given in Section 3.3). In this case, the primary technology is dominant since most users are close to its tower, but distributed at the frontier of the WiFi coverage, far away from its antenna. Results in Figure

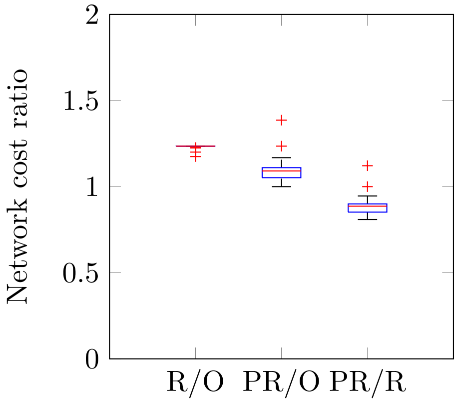

Figure 3: Specific scenario 1 for comparison of distributed heuristics

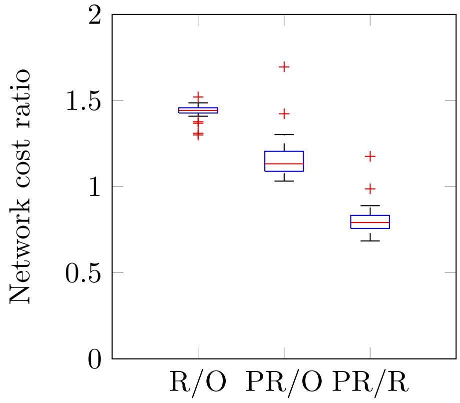

Figure 4: Specific scenario 2 for comparison of distributed heuristics

3 show expectedly that the optimal solution is the best as both ratios $\frac{PR}{O}$ and $\frac{R}{O}$ are greater than 1. In fact, this is always true since the optimal solution minimizes the cost function. In addition, since $\frac{PR}{R}$ is smaller than 1, we deduce that, in this case, the probabilistic peak rate algorithm is more advantageous than the peak rate algorithm; a result that could have

been predicted since with the **R** algorithm, all users are going to choose the primary RAT making the throughput perceived by each user less than $100/15 = 6.66$ Mbits/s (fair time model).

However, in this scenario, the **PR** algorithm will force some hybrid users to connect to the secondary RAT, thus reaching a state of balance between these two technologies and raising users data rates thus reducing global cost.

- Scenario 2: Primary RAT vector probabilities is [0.7 0.3 0 0 0 0 0 0] and secondary RAT vector probabilities is [0 0 0 0 0 0.1 0.3 0.6]. In this case, 70% of hybrid users get 11.2 Mbits/s from the primary RAT while 60% get 54 Mbits/s from the secondary RAT. The secondary technology is dominant since most users are close to its BS, but distributed at the frontier of the LTE cell, far away from its antenna.

  Here, the secondary RAT offers higher rates and users tend to favor it over the primary RAT especially with **R**. However, **PR** will act in the same way as in scenario 1, forcing some hybrid users to choose the primary RAT and reducing the global cost of the network. The results obtained in Figure 4 validate this logic.

- Scenario 3: Primary RAT probabilities [0 0.5 0.5 0 0 0 0 0] and secondary RAT probabilities are all equal to 1/8. Users in the secondary RAT are equally distributed on the entire coverage area. With at least 50% of hybrid users getting greater rates from the primary RAT, we can say that both technologies are approximately equivalent and that a state of equilibrium is achieved. In this case, Figure 5 shows that **R** is superior to **PR**.

In brief, we give the relevant conclusions that stem from analyzing the specific scenarios:

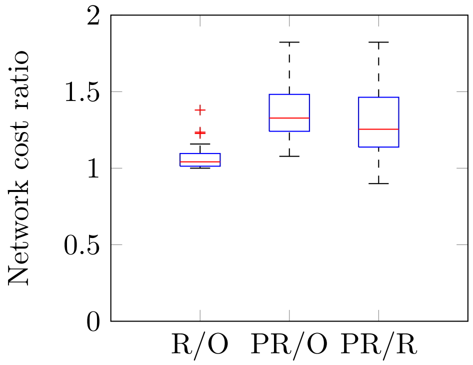

Figure 5: Specific scenario 3 for comparison of distributed heuristics

- When one RAT surrogates the other in terms of peak rates, hybrid users will largely choose the dominant RAT according to the **R** algorithm which will deteriorate overall performances. In this case, the **PR** algorithm is precious as it balances the load among both technologies.
- Whereas when the two RATs are equivalent in terms of peak rates, the **R** algorithm is appropriate as there is no risk that hybrid users overcrowd any preferred RAT. In this case, the randomization brought by **PR** algorithm tends to temper the network performances.

6.2.2. General Scenarios

In the previous section, we studied punctual specific scenarios that gave us insights into the performances of the **R** and **PR** algorithms respectively. What happens then when a random scenario is at hand? And how to express such a random scenario? To answer both questions, we need to model each scenario taking into account the main characteristics (*e.g.*, the peak rates perceived by users) in order to single out the most suitable algorithm.

In order to create such representation or function, we need first to pinpoint what defines a given scenario in terms of the technologies present in the covered geographic area and the distribution of the users relatively to each of these technologies. In other words, these parameters are the rate ratios $\frac{\chi_{i,S}}{\chi_{j,P}}$ for $i = 1, .., M_S$ and $j = 1, .., M_P$ perceived by hybrid users from both technologies and the probability vectors $[p_{1,x}\, p_{2,x} \dots p_{M_x,x}]$ introduced earlier for each RAT x. Thus, we introduce a function deemed Q that captures the aforementioned parameters. Q is the expected value of the rate ratios: it is computed by multiplying the joint probability $p_{i,S} \times p_{j,P}$ by the corresponding rate ratios as given in the following:

$$Q = \sum_{i=1}^{M_S} \sum_{j=1}^{M_P} p_{i,S} \times p_{j,P} \times \frac{\chi_{i,S}}{\chi_{j,P}} \tag{24}$$

In the following simulation results, our goal is to predict which heuristic algorithm is the most suitable for a given random scenario. We compute the average $\overline{\frac{PR}{R}}$ over 50 values for 100 users. We plot $\overline{\frac{PR}{R}}$ as a function of Q for a thousand scenarios in order to observe the pattern they generate.

We consider two different distributions of users:

- A *Uniform distribution* where the coverage area of a WiFi BS is homogeneously occupied by users. For this reason, all components of the secondary RAT probability vector are the same.

- A *Cluster distribution* where users are close to each other. Since hybrid users have a tendency to gather in groups when accessing web pages, downloading media content or attending an e-learning course. Thus, all components of the secondary RAT probability vector, except one, are equal to zero.

For each of the aforementioned user distributions, we plot $\overline{\frac{PR}{R}}$ as a function of Q and analyze the results obtained for 100 hybrid users.

Note that, in the uniform scenario, hybrid users uniformly span the different available peak rates in the secondary RAT while they still perceive one or two possible peak rates in the primary RAT. Extreme Q values correspond to the

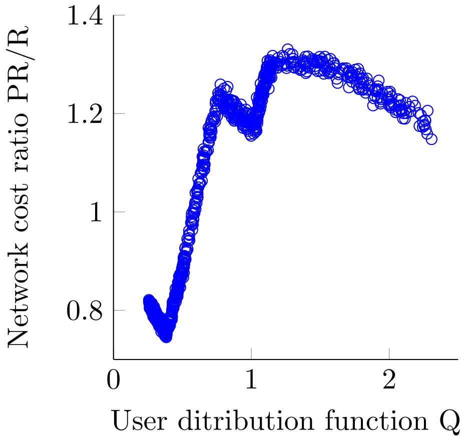

Figure 6: General scenario: uniform distribution

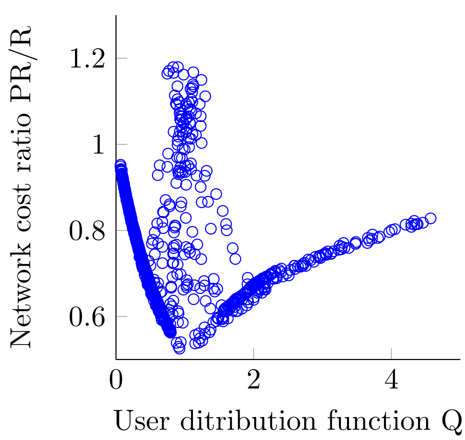

Figure 7: General scenario: cluster distribution

following: very high (low) Q values match the case where the majority of users perceive much higher (lower) peak rates in the secondary RAT in comparison with the primary RAT. In Figure 6, for low Q values, we notice that **PR** is less costly than **R** because the majority of users perceive a higher rate in the primary RAT and will all join - according to **R** - the primary RAT causing congestion and overall bad performances. Whereas, the **PR** load balances the users between both RATs. When the primary peak rates start decreasing in the primary RAT in comparison with those in the secondary RAT, Q increases leading to a gradual improvement in the performances of **R** until it over-tops **PR**. This is due to the fact that the perceived primary peak rates take an intermediary value in comparison with the set of secondary peak rates. After reaching a maximum value for Q around 1, $\overline{\frac{PR}{R}}$ starts decreasing as the number of users having better peak rates in the secondary RAT starts increasing. Therefore, the deterministic algorithm **R** is less attractive as it leads to congestion in the secondary RAT.

In the cluster scenario in Figure 7, hybrid users perceive only one of the available peak rates in the secondary RAT. We note that extreme Q values in the cluster scenario correspond to the following: very high (low) Q values match the case where the perceived peak rate in the secondary RAT is much higher (lower) than those perceived in the primary RAT. In general, **PR** is more efficient than **R** except for Q values around 1 for which the deterministic **R** performs better. In fact, for such cases, both RATs are equivalent in terms of peak rates and hence the **PR** algorithm disturbs the natural load balancing attained through **R**. In other cases including scenarios having extreme Q values, perceived peak rates in one RAT are greater than the other and therefore **PR** performs better than **R** that singles out the more advantageous RAT leading to congestion.

6.2.3. Dynamic Scenarios

To study in depth the problem at hand, the system is henceforth dynamic with users arriving to an arbitrary zone and leaving the system after being serviced. Now that the system is dynamic, we need to adapt the proposed existing algorithms. For the optimal solution, we propose to apply the enhanced search algorithm in 5.2, each time a user enters or leaves the secondary RAT. This adaptation may induce changes in the RAT selection process for active users which are commonly known as vertical handovers (HO). Whereas for the distributed heuristics, we choose to implement the RAT selection decisions only for each incoming session. Although, this will impact the perceived performances of ongoing sessions, active users will not revise their former decisions. Hence, there are no vertical HOs for the proposed distributed heuristics. Considered as a costly operation that is not directly charged by the operator, it is important to compute the number of HOs occurring at each transition for the optimal solution. In particular, a large number of HOs can hinder the benefits of an efficient algorithm.

We consider a uniform distribution of users to compare the optimal solution denoted **O** against the **PR** and **R** algorithms. The arrival of users follows a

Poisson distribution and each hybrid user leaves the system after downloading a file whose size follows an exponential distribution of mean 5 MBytes. Each simulation accounts for 100,000 events (arrivals or departures). In the following, we introduce the cumulative cost as an evaluation criterion. The cumulative cost at time t is defined as:

$$C_{cum}(t) = C_{tot}(t) \times I_{t,t+1} + C_{cum}(t-1),$$

where $I_{t,t+1}$ is the inter-event time. The expression of $C_{cum}(t)$ takes into account the time spent in a given network state and multiply it by the corresponding total network cost.

In Figures 8 and 9, we display the cumulative cost and the mean number of HO for the enhanced search algorithm as a function of arrival rate λ.

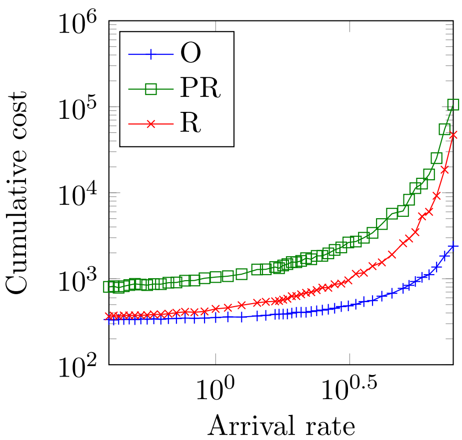

Figure 8: Cumulative cost in dynamic scenarios

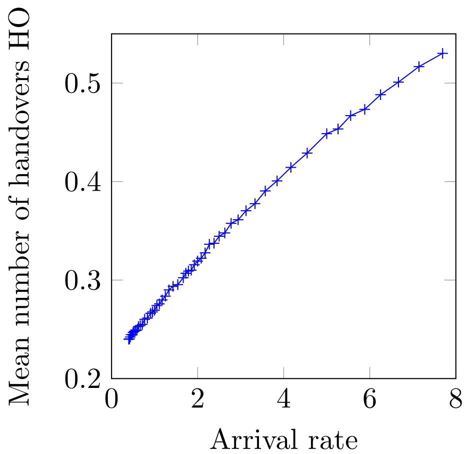

Figure 9: Number of handovers HO for the enhanced search algorithm in dynamic scenarios

We see that the optimal algorithm with the enhanced search generates the best performances when compared with the heuristics specifically at high load. At moderate to low load, the discrepancy between the heuristics and optimal algorithm is fairly low. Furthermore, we notice that the maximum rate of HO is approximately 1 HO every 2 events which is very costly from the operator's point of view since a HO is an uncharged service that is hard to accomplish. This leads us to the undeniable conclusion that there is no perfect solution but only a compromise between the optimal solution and its heavy HO requirements, and the suboptimal distributed heuristic algorithms and their very low complexity.

7. Game Theoretic Distributed Approach

To fully assess the relevance of our distributed heuristics in section 6, we will compare their performances against another distributed RAT selection scheme based on non-cooperative game theory that we proposed in [35]. Game theory is well adapted to model the interactions between players competing selfishly for a common resource. Initially, the game consists for each end-user in allocating the traffic among the primary and secondary RATs in a way to minimize selfishly its own cost. However, we showed in [35] that after convergence, each user is connected to a single RAT which avoids costly traffic splitting between RATs.

For each state of the system, defined by the number of hybrid users n_S, we define a multi-player non-cooperative game G between the n_S hybrid mobile users present in a geographic area covered simultaneously by the primary RAT and secondary RAT. In this model, there is a sequence of one-stage games, each corresponding to a given state of the system, defined by the number of hybrid users. In [35], the game was static but we have recourse here to a realistic dynamic setting: whenever a new hybrid mobile is admitted in the system, the game is played again with an additional player. Mobile users are assumed to make their decisions without knowing the decisions of each other. The formulation of this non-cooperative game $G = \langle N, S, C \rangle$ can be described as follows:

- A finite set of players: $N = (1, ..., n_S)$.
- The space of pure strategies $\mathcal{S}$ formed by the Cartesian product of each set

of pure strategies $\mathcal{S} = S_1 \times ... \times S_{n_S}$. An action of a hybrid user k is the proportion of traffic θ_k sent via the secondary RAT. Hence $\mathcal{S}_k = [0, 1]$. The strategy chosen by player k is then θ_k while θ_{-k} denotes the set of actions taken by all other players. A *strategy profile* $\theta = (\theta_1, ..., \theta_k, ..., \theta_{n_S})$ specifies the strategies of all players.

- A set of cost functions $C = (C_1(\theta), C_2(\theta), ..., C_{n_S}(\theta))$ that quantify players' costs for a given strategy profile θ where $C_k(\theta)$ for $k \in N$ is given by equation (13).

Since for every hybrid user k, C_k is convex w.r.t. θ_k and continuous w.r.t. $\theta_i, i \neq k$, a Nash equilibrium exists [39]. Proving the existence of the Nash Equilibrium (NE) for a non-cooperative game is paramount as we need an equilibrium point to which selfish players are willing to adhere. However, it is far from sufficient, as we need to compute those particular equilibrium points. Moreover, we need to find a realistic distributed algorithm to help non-cooperative end-users learn autonomously those NEs. All those goals were attained in [35] where two search algorithms were proposed to characterize NEs of the RAT selection game.

However, in a real environment, the search algorithms cannot be practically applied. On the one hand, they are time consuming and not tractable for real size scenarios. On the other hand, they necessitate that every user knows the strategy of all other users present in the cell. The latter property requires expensive signaling and hinders the benefits of a distributed resource management policy. As a consequence, we fall back to replicator dynamics algorithm proposed in [40] to learn NEs. With a replicator dynamics based algorithm, each hybrid user needs only to be aware of its own cost and strategy at each time iteration.

7.1. Learning NE: Replicator Dynamics

For clarity, we remind here of the replicator dynamics based algorithm mechanism as devised in [35]. Players have pure strategies and the devised algorithm will give the mixed strategies corresponding to a probability distribution over pure strategies.

In the present context, this algorithm can be applied by considering the discrete version of our problem: hybrid users have a finite set of strategies, each strategy amounts to selecting only one RAT (the special case where $\theta_k = \{0, 1\}$). Therefore, the set of pure strategies of any hybrid user k in the finite game, termed s_k, is $s_k = \{S, P\}$ corresponding to choosing secondary or primary RAT respectively. The hybrid user gets thereby the following cost:

$$c_k = \begin{cases} T_{k,P} & \text{if} \quad s_k = P \\ T_{k,S} & \text{if} \quad s_k = S \end{cases} \tag{25}$$

Thus, a mixed strategy matches our probability distribution $(\theta_k, 1 - \theta_k)$ on s_k.

Definition 2. *The game mechanic works as follows: at $t = 0$, we begin with $\theta(0) = (\theta_1(0), ..., \theta_{n_S}(0))$ any random vector of probabilities. At each iteration $t > 0$:*

1. *Each hybrid user k selects the WiFi BS with probability $\theta_k(t)$ which leads to outcome $c_k(t)$.*
2. *Each hybrid user k updates $\theta_k(t)$ as follows:*

$$\theta_k(t+1) = \begin{cases} \theta_k(t) + b(1 - \frac{c_k(t)}{c_{max}})(1 - \theta_k(t)) \\ \textit{if} \quad s_k(t) = S, \\ \theta_k(t) - b(1 - \frac{c_i(t)}{c_{max}})\theta_k(t) \\ \textit{if} \quad s_k(t) = P, \end{cases} \tag{26}$$

where $0 < b < 1$ is a parameter that controls the convergence rate [12] and c_{max} is the maximum cost perceived by any user in a given RAT.

Proposition 3. *For any initial condition where $\forall i \in N$, $\theta_i \neq 0$ or $\theta_i \neq 1$, the considered learning algorithm converges to a pure Nash equilibrium.*

Thus, according to proposition 3 (proof of convergence is given in [35]), a single RAT is selected similar to the proposed algorithm in Section 4 and the devised heuristics in Section 6. Consequently, costly recurrent shifts between RATs are eschewed.

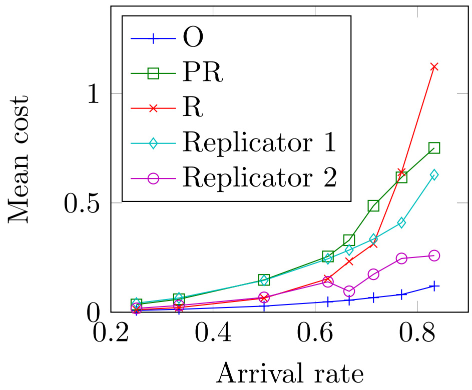

Figure 10: Cost as a function of arrival rate

8. Global Comparisons

In this final section, extensive simulations are conducted in a dynamic setting to compare the heuristic algorithms **R** and **PR**, the replicator dynamics algorithm and the optimal centralized approach **O**. The probability vectors are respectively [0 0 0 0 0 0.1 0.3 0.6] in LTE and [0.7 0.3 0 0 0 0 0 0] in WiFi. The arrival of users follows a Poisson distribution and each user leaves the system after downloading a file whose size follows an exponential distribution of mean 5 MBytes. It is worth mentioning that the replicator dynamics was altered to cope with a dynamic scenario where the number of active users is variable. In fact, when convergence is reached, the algorithm is stuck in the attained equilibrium as the equations where θ is updated in (26) cannot change any more. Therefore, whenever an equilibrium is attained, users are forced to slightly drive away from it $(\theta_k - \epsilon, 1 - \theta_k + \epsilon)$ and replay the game to be able to *sense* any new arrival. Two versions of the modified game are proposed: the first one, namely Replicator 1, where θ is updated every 0.1s and the second one, namely Replicator 2, where θ is updated at a much faster pace of 1ms. Results are portrayed in Figure 10 for various arrival rates λ.

The enhanced optimal algorithm **O** generates clearly the best performing state of the network when compared with both heuristics **PR** and **R**, and the

Table 2: Relative Change in Cost

$ArrivalRate$	PR	R	$Replicator2$
0.9	840%	25.13×10^2%	173%
1.0	11.26×10^2%	13.33×10^5%	415%

replicator dynamics algorithm. For the game theoretic algorithm, we can see that Replicator 2 is much more efficient than Replicator 1 as the algorithm has time to converge thanks to the frequent strategy updates. Furthermore, Replicator 2 has the best performance right after the optimal algorithm and succeeds in keeping the cost low even for high arrival rates. However, this improvement comes at the cost of signaling and convergence delay. The **R** and **PR** algorithms are one-shot schemes that give instant results without weighing down the system with any signaling messages and still with efficient results.

The **PR** algorithm has the worst performance as long as the arrival rate remains below 0.8. Above this threshold, the system is very crowded and the **R** algorithm performs very badly because all users will favor the RAT that gives them the highest peak rate and will find themselves stuck in an overcrowded RAT. The probabilistic decision in **PR** tempers this edge effect. This is further highlighted when the system reaches the limit of stability through the relative change in cost for the **R**, **PR** and the Replicator 2 algorithms in comparison with the optimal approach reported in table 2. In particular, we can see that the **R** algorithm has very mediocre performances.

9. Conclusion

Undeniably, ubiquitous radio access remains the essential backbone for supporting the ever increasing demand for bandwidth. Operators will profit from the abundance of diverse air interfaces in the same operating area and put forward advanced RAT selection mechanisms. The stringent performance targets and the novel flat architecture of beyond 4G networks have triggered a new trend in RRM favoring lightweight distributed schemes to costly centralized schemes. Hence, users are intelligent terminals that can discover the radio environment and connect to available RATs aiming at minimizing their own cost. In the first part of this work, we compared the centralized optimal RAT selection policy against two heuristic distributed RAT selection algorithms. By studying different realistic scenarios based on users density and spatial distribution, we identified which heuristic algorithm to favor based on the scenario at hand. In the second part of the paper, the centralized approach is compared against distributed game-theoretic RAT selection algorithm. In this competitive environment, resorting to non-cooperative game theory is natural in order to obtain optimal RAT selection. We characterized the Nash equilibriums of the RAT selection game and put forward a decentralized algorithm based on replicator dynamics to achieve those equilibriums. Although the game distributed approach delivers more efficient results in comparison with the distributed heuristic approach, this improvement comes at the cost of signaling burden and convergence delay. The distributed heuristic approach is a one-shot algorithm that gives instant results without weighing down the system with any signaling messages.

References

[1] NGMN Alliance, 5G White Paper, February 2015.

[2] S. Sesia, I. Toufik, and M. Baker, LTE - The UMTS Long Term Evolution: From Theory to Practice, John Wiley & Sons, Ltd, 2009.

[3] J. G. Andrews, S. Buzzi, W. Choi, S. V. Hanly, A. Lozano, A. C. K. Soong, and J. C. Zhang, What Will 5G Be, in *IEEE J. Selected Areas in Commun.*, vol. 32, no. 6, June 2014.

[4] ADVA et al., Horizon 2020 Advanced 5G Network Infrastructure for Future Internet PPP Industry Proposal (Draft Version 2.1), 2013.

[5] METIS D6.6. Final Report on the METIS 5G System Concept and Technology Roadmap. ICT-317669 METIS Deliverable 6.6, Version 1, April 2015.

[6] F. Rebecchi, M. Dias de Amorim, V. Conan, A. Passarella, R. Bruno, and M. Conti, Data Offloading Techniques in Cellular Networks: A Survey, in *IEEE Commun. Surveys and Tutorials*, vol. 17, no. 2, Nov. 2014.

[7] IEEE Standard for Local and metropolitan area networks Part 16: Air Interface for Fixed Broadband Wireless Access Systems, IEEE Std., 802.16-2004 standard.

[8] Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications, ANSI/IEEE Std 802.11, 1999.

[9] J. Andrews, and A. Ghosh, Fundamentals of WiMAX: Understanding Broadband Wireless Networking, Prentice Hall Pub. Date: February 27, 2007.

[10] IEEE standard for local and metropolitan area networks - part 21: Media independent handover (2009).

[11] K. Taniuchi, *et al.*, IEEE 802.21: Media independent handover: features, applicability, and realization, 2009.

[12] D. Barth, O. Bournez, O. Boussaton, and J. Cohen, Distributed Learning of Wardrop Equilibria, In 7th International Conference on Unconventional Computation. Lecture Notes in Computer Science, 2008.

[13] D. Kumar, E. Altman, and J.-M. Kelif, User-Network Association in a WLAN-UMTS Hybrid Cell: Global & Individual Optimality, INRIA, Rapport de recherche RR-5961, 2006.

[14] E. Stevens-Navarro, Y. Lin, and V. W. S. Wong, An MDP-based Vertical Handoff Decision Algorithm for Heterogeneous Wireless Networks, IEEE Trans. on Vehicular Technology, in press, 2008.

[15] M. Ibrahim, K. Khawam, and S. Tohme, Network-centric joint radio resource policy in heterogeneous wimax-umts networks for streaming and elastic traffic, in IEEE Wireless Communications and Networking Conference (WCNC), April 2009.

[16] G.H. Carvalho, *et al.*, A joint call admission control-based approach for initial RAT selection in HetNets, International Conference on Computer, Information and Telecommunication Systems (CITS), 2013.

[17] I. Tsompanidis, A. Zahran, and C. Sreenan, Towards utility-based resource management in heterogeneous wireless networks, MobiArch '12 Proceedings of the seventh ACM international workshop on Mobility in the evolving internet architecture, 2012.

[18] J. Choque, R. Aguero, E-M Hortiguela and L. Munoz, Optimum Selection of Access Networks within Heterogeneous Wireless Environments Based on Linear Programming Techniques, Mobile Networks and Management Lecture Notes of the Institute for Computer Sciences, Social Informatics and Telecommunications Engineering V. 68, pp 135-149, 2011.

[19] A. Kumar, R.K. Mallik, and R. Schober, A Probabilistic Approach to Modeling Users' Network Selection in the Presence of Heterogeneous Wireless Networks, IEEE Transactions on Vehicular Technology, vol.63, no.7, pp.3331,3341, 2014.

[20] O. Eng Hwee and J. Y. Khan. On optimal network selection in a dynamic multi-RAT environment. Communications Letters, IEEE, 14(3):217-219, 2010.

[21] L. Pirmez, J. C. Carvalho, Jr., F. C. Delicato, F. Protti, L. F. R. C. Carmo, P. F. Pires, and M. Pirmez. SUTIL - Network selection based on utility function and integer linear programming. Computer Networks, 54:2117-2136, September 2010.

[22] D. Kumar, and E. Altman, User-Network Association in an 802.11 WLAN and 3G UMTS Hybrid Cell: Individual Optimality, Proceedings of IEEE Sarnoff Symposium, Princeton, USA, 2007.

[23] P. Coucheney, C. Touati, and B. Gaujal, Fair and efficient user-network association algorithm for multitechnology wireless networks. INFOCOM, 2009.

[24] S.V. Saboji, C. B. Akki, Congestion-aware Proactive Vertical Handoff Decision Using Coalition Game, International Journal of Soft Computing and Engineering (IJSCE) ISSN: 2231-2307, Volume-1, Issue-6, January 2012.

[25] K. Radhika and A. Venugopal Reddy, Vertical Handoff Decision using Game Theory Approach for Multi-mode Mobile Terminals in Next Generation Wireless Networks, International Journal of Computer Applications (0975 - 8887) Volume 36- No.11, December 2011.

[26] E. Aryafar, A. Keshavarz-Haddad, M. Wang, and C. Mung, RAT selection games in HetNets, IEEE INFOCOM 2013.

[27] R. Trestian, O. Ormond, G-M. Muntean, Game Theory-Based Network Selection: Solutions

and Challenges, IEEE Communications Surveys & Tutorials, vol.14, no.4, pp.1212,1231, 2012.

[28] 3GPP TR 25.814 V7.1.0, Physical Layer Aspects for Evolved Universal Terrestrial Radio Access (UTRA) (Release 7), 2006.

[29] H.Holma and A.Toskala, LTE for UMTS: OFDMA and SC-FDMA based radio access, John Wiley & Sons, 2009

[30] M. Ibrahim, K. Khawam, A. Samhat, S. Tohm, Analytical Framework For Dimensioning Hierarchical WiMax-WiFi Networks, Elsevier, Computer Networks, 2009.

[31] K. Khawam, M. Ibrahim, M. Abdennebi, D. Marinca, S. Tohm, Analytical Modelling in 802.11 Ad Hoc Networks, Elsevier, Computer Networks, 2011.

[32] B. Sklar, *Rayleigh Fading Channels in Mobile Digital Communication Systems*, IEEE Communications Magazine, Vol. 35 Iss.7, 1997.

[33] T. Bonald, A. Proutire, *Wireless Downlink Data Channels: User Performance and Cell Dimensioning*, MobiCom'03, 2003.

[34] E. Gustafsson and A. Jonsson, Always Best Connected, IEEE Wireless Communications Journal, February 2003.

[35] K. Khawam, M. Ibrahim, J. Cohen, S. Lahoud, and S. Tohme, Individual vs. global radio resource management in a hybrid broadband network, in Proc. IEEE Int. Conf. on Communications (ICC), June 2011.

[36] M. Heusse, F. Rousseau, G. Berger-Sabbatel, and A. Duda, Performance anomaly of 802.11b, IEEE Infocom 2003.

[37] F. Kelly, Charging and Rate Control for Elastic Traffic, in *European Trans. Telecommunications*, vol. 8, no. 1, 1997.

[38] China Mobile Research Institute, C-RAN: The Road Towards Green RAN. White Paper, ver. 2, 2011.

[39] J. B. Rosen, Existence and uniqueness of equilibrium points for concave n-person games, Econometrica, vol. 33, 1965

[40] J. Hofbauer, and K. Sigmund, Evolutionary game dynamics. Bulletin of the American Mathematical Society, 4:479–519, 2003.

[41] F. Moety, M. Ibrahim, S. Lahoud, and K. Khawam, Distributed heuristic algorithms for RAT selection in wireless heterogeneous networks, In IEEE WCNC 2012.

[42] S. Chieochan, E. Hossain, and J. Diamond, Channel assignment schemes for infrastructure-based 802.11 WLANs: A survey, Communications Surveys & Tutorials, IEEE, vol.12, no.1, pp.124,136, 2010.