Radio Access Selection Approaches in Heterogeneous Wireless Networks

Melhem El Helou, Marc Ibrahim, Samer Lahoud, Kinda Khawam

▶ **To cite this version:**

HAL Id: hal-01018230
https://hal.science/hal-01018230

Submitted on 3 Jul 2014

Radio Access Selection Approaches in Heterogeneous Wireless Networks

Melhem El Helou*†, Marc Ibrahim*, Samer Lahoud† and Kinda Khawam‡

*Saint-Joseph University, ESIB, Campus des Sciences et Technologies, Mar Roukoz, Lebanon
†University of Rennes 1, IRISA, Campus de Beaulieu, 35042 Rennes, France
‡University of Versailles, PRISM, 45 Avenue des Etats-Unis, 78035 Versailles, France

Abstract—Along with the rapid growth of mobile broadband traffic, multiple radio access technologies (RATs) are being integrated and jointly managed. To optimize heterogeneous network performance, efficient Common Radio Resource Management (CRRM) mechanisms need to be defined. This paper tackles the access technology selection — a key CRRM functionality — and proposes a hybrid approach that combines benefits from both network-centric and user-centric methods. Network information, that is periodically broadcasted, assists mobile users in their decisions. By broadcasting appropriate decisional information, the network tries to globally control users decision in a way to meet operator objectives. On the other hand, mobiles also integrate their needs and preferences to select their access technology so as to maximize their own utility. In comparison with other RAT selection techniques, including network-centric, hybrid and user-centric methods, simulation results prove the efficiency of our hybrid approach in enhancing resource utilization and maximizing user satisfaction.

Index Terms—Radio access technology selection, heterogeneous wireless networks, hybrid approach.

I. INTRODUCTION

To cope with the rapid growth of mobile broadband traffic, different radio access technologies (*e.g.,* HSPA, LTE, WiFi and WiMAX) are being integrated and jointly managed. So as to optimize resource utilization, while enhancing user experience (Always Best Connected concept [1]), efficient Common Radio Resource Management (CRRM) mechanisms need to be defined. Typically, when a new or a handover session arrives, a decision must be made as to what technology it should be associated with. This is known as the Radio Access Technology (RAT) selection, a key CRRM functionnality.

In order to consider operator objectives, including efficient exploitation of radio resources, network-centric schemes have been proposed: network elements collect necessary measurements and information. They take selection decisions transparently to end-users in a way to enhance heterogeneous network performance [2]–[6]. However, to reduce network complexity, signaling and processing load, user-centric methods have also gained in importance: based on their individual needs and preferences, rational users select their access technology in a way to selfishly maximize their payoff (utility) [6]–[12]. Because individual users have no information on the global network state (*i.e.,* network load conditions), user-centric approaches are known for their potential inefficiency.

In this article, we propose a hybrid decision method that combines benefits from both network-centric and user-centric

approaches. Network information is designed to assist mobile users in their decisions: mobiles make their selection decision based on their individual needs and preferences as well as on the cost and partial QoS parameters signaled by the network. To maximize user experience, we present a satisfaction-based Multi-Criteria Decision-Making (MCDM) method. In comparison with existing MCDM methods, namely Simple Additive Weighting (SAW) [10], Multiplicative Exponent Weighting (MEW) [10], Grey Relational Analysis (GRA) [11] and Technique for Order Preference by Similarity to Ideal Solution (TOPSIS) [12], our algorithm meets the exact needs of mobile users (*e.g.,* traffic class, throughput demand, cost tolerance), thus avoiding oversized and undersized alternatives.

A particular attention is then addressed to the network to make sure it broadcasts appropriate decisional information, so as to enhance resource utilization while individual users are maximizing their own utility. We therefore introduce the slope tuning policy to dynamically derive what to signal to mobiles.

The rest of this paper is organized as follows: Section II describes our hybrid decision framework. Our satisfaction-based decision-making method is presented in section III. Section IV introduces our tuning policy. Simulation parameters and results are discussed in section V. Section VI concludes the document.

II. HYBRID DECISION FRAMEWORK

A. Network topology

Consider a heterogeneous wireless network composed of N_T radio access technologies. The user-perceived signal-to-noise ratio ($SNR^x, x = 1, ..., N_T$) determines the user modulation and coding scheme, and subsequently its instantaneous peak rate (*i.e.,* its perceived throughput when connected alone to RAT x). Since practically the set of achievable instantaneous peak rates is not continuous, RAT x cell is logically divided into N_Z^x concentric rings (Fig. 1). Users in ring Z_k^x, $k = 1, ..., N_Z^x$, with SNR^x between δ_k^x and δ_{k-1}^x, have a peak rate of D_k^x when connected to RAT x. The user peak rate in RAT x, namely D^x, is then expressed as a function of the user-perceived SNR as follows:

$$D^x = \begin{cases} 0 & if \ SNR^x < \delta_{N_Z^x}^x, \\ D_{N_Z^x}^x & if \ \delta_{N_Z^x}^x \leq SNR^x < \delta_{N_Z^x-1}^x, \\ ... \\ D_1^x & if \ \delta_1^x \leq SNR^x < \delta_0^x = \infty. \end{cases} \quad (1)$$

where $\delta^x_{N^x_Z}$ is the minimum SNR that allows transmission, at the lowest bit rate $D^x_{N^x_Z}$, given a target error probability.
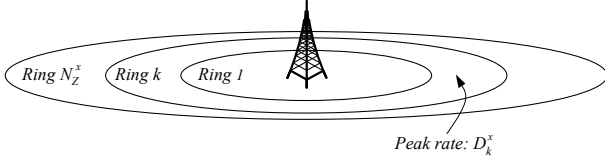

Fig. 1. A cell divided into N_Z concentric rings

B. Network resources

In RAT x, the radio resource is divided into elementary resource units (RU). Typically, in OFDM(A)-based technologies (*e.g.,* LTE and WiMAX technologies), resource units are defined as OFDM symbols (one-dimensional allocations) or OFDMA slots (two-dimensional allocations: m subchannels by n OFDMA symbols). However, in CDMA-based technologies (*e.g.,* HSPA technology), codes, power and allocation time are treated as RUs.

C. Network information

Network information is periodically sent to all mobile users using the logical communication channel (*i.e.,* radio enabler) proposed by the IEEE standard 1900.4 [13]. In our work, this information is assumed to implicitly integrate operator objectives, guiding users decision. It may be static or variable so as to dynamically optimize short- or long-term network performance.

When a new or a handover session arrives, the mobile decodes the decisional information, evaluates available alternatives, and selects the technology that best suits it.

In this setting, we assume that the network information provides cost and some QoS parameters: they can be seen as incentives to join available alternatives.

- Cost parameters: Because flat-rate pricing strategies waste resources, result in network congestion and thus degrade network performance [14], they are not optimal in supporting QoS. A volume-based model is therefore proposed: mobile users are charged based on the amount of traffic they consume; in our work, *costs* are defined on a per kbyte basis.
- QoS parameters: The number of resource units (RUs) that need to be allocated to future arrivals are broadcasted:
  - Mobiles are guaranteed an average minimum number of RUs, denoted by n_{min}.
  - They also have priority to occupy up to an average maximum number of RUs, denoted by n_{max}.

The network load conditions and capacity are, however, masked. In fact, n_{min} and n_{max} reveal the operator intention to serve future arrivals: they do not exclusively reflect the load conditions, but also other potential operator objectives.

Since the smallest allocation unit (*i.e.,* RU) may be different from one technology to another, there is a need to homogenize the QoS information. QoS parameters are then expressed as throughputs: d_{min} and d_{max} instead of n_{min} and n_{max}. Yet, because perceived throughputs highly depend on radio conditions (or equivalently on adopted modulation types and FEC coding rates), d_{min} and d_{max} are derived *for the most robust modulation and coding scheme (i.e.,* as perceived by users in ring Z_{N_Z}). Consequently, when evaluating available alternatives, mobiles should combine their individual radio conditions with the provided QoS parameters: for that they multiply d_{min} and d_{max} with a given modulation and coding gain, denoted by $g(M, C)$.

D. RAT selection

For each incoming session, the network proposes one or more alternatives, which are the available access technologies. For each alternative (a), the network broadcasts the three parameters: $d_{min}(a)$, $d_{max}(a)$, and $cost(a)$. From the user point of view, these parameters are the decision criteria that will be used to evaluate the different access technologies. As in all multi-criteria decision making methods, the mobile has to define a utility function that will be computed for all of the available alternatives. This utility is obtained after normalizing and weighting the decision criteria.

The particularity of our RAT selection process resides in the normalization step that takes into account the traffic class and throughput demand as detailed in the next section. Such approach overcomes some limitations of classical methods due to undersized and oversized decisions.

III. SATISFACTION-BASED DECISION METHOD

A. Normalization and Traffic classes

As stated above, the normalization of decision criteria $d_{min}(a)$, $d_{max}(a)$, and $cost(a)$ depends on the session traffic class and throughput demand. For traffic class c and alternative a, the normalization is a mapping of $d_{min}(a)$, $d_{max}(a)$, and $cost(a)$ to $\hat{d}^c_{min}(a)$, $\hat{d}^c_{max}(a)$, and $\widehat{cost}^c(a)$ respectively.

In our work, we define two traffic classes : streaming, and elastic classes. Before we give the normalizing functions for each traffic class, let us note that $\hat{p}^c(a), p \in \{d_{min}, d_{max}, cost\}$, can be viewed as the satisfaction of a class c session with respect to criterion p for alternative a:

- Streaming sessions ($c = S$): since designed to support real-time variable bit rate services (*e.g.,* MPEG-4 video service), streaming sessions are fairly flexible and usually characterized by a minimum, an average and a maximum bandwidth requirement. Their throughput satisfaction is therefore modelled as an S-shaped function (Fig. 2(a)):

$$\hat{d'}^S(a) = 1 - \exp(\frac{-\alpha(\frac{d'(a).g(M,C)}{R_{av}})^2}{\beta + (\frac{d'(a).g(M,C)}{R_{av}})}) \qquad (2)$$

where $d' = \{d_{min}, d_{max}\}$.
R_{av} represents session needs: an average throughput demand. α and β are two positive constants to determine the shape of the S-shaped function.
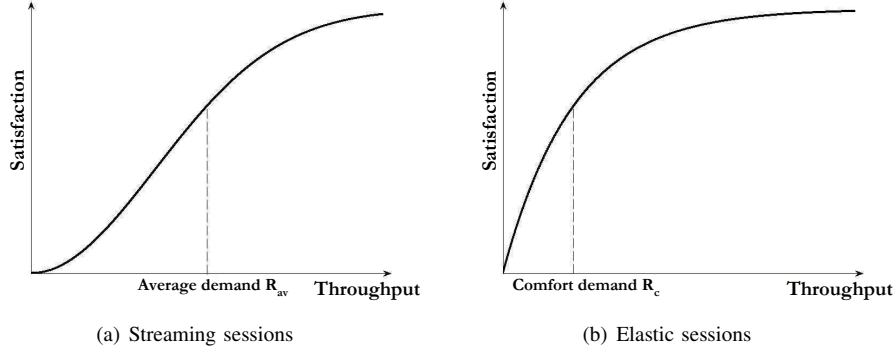
(a) Streaming sessions   (b) Elastic sessions

Fig. 2. Throughput satisfaction function forms

- Elastic sessions ($c = E$): since designed to support traditional data services (*e.g.,* file transfer, email and web traffic), elastic sessions adapt to resource availability (*i.e.,* load conditions), requiring no QoS guarantees. Thus, d_{min} is completely ignored. Moreover, the satisfaction with respect to d_{max} has a concave shape (Fig. 2(b)): the satisfaction increases slowly as the throughput exceeds the comfort throughput demand R_c of the user (*i.e.,* the mean throughput beyond which, user satisfaction exceeds 63% of maximum satisfaction).

$$\hat{d}_{max}^E(a) = 1 - \exp(-\frac{d_{max}(a).g(M,C)}{R_c}) \qquad (3)$$

The monetary cost satisfaction is, however, modelled as a Z-shaped function (Fig. 3): the slope of the satisfaction curve increases rapidly with the cost.

$$\widehat{cost}^c(a) = \exp(-\frac{cost(a)^2}{\lambda^c}), c \in \{S, E\} \qquad (4)$$

λ^c represents the cost tolerance parameter: a positive constant to determine the shape of the Z-shaped function.
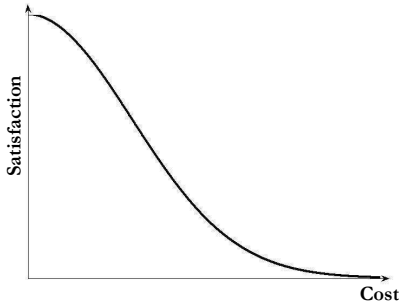

Fig. 3. Monetary cost satisfaction function form

B. User Profile and Utility Function

The user profile defines the cost tolerance parameter and the weights that a given session will apply to normalized criteria. More precisely, the user profile is the set of vectors $(\lambda^c, w_{d_{min}}^c, w_{d_{max}}^c, w_{cost}^c), c \in \{S, E\}$, where w_p^c is the weight of $\hat{p}^c, p \in \{d_{min}, d_{max}, cost\}$. The utility function of a class c session for alternative a is defined by :

$$U^c(a) = w_{d_{min}}^c.\hat{d}_{min}^c(a) + w_{d_{max}}^c.\hat{d}_{max}^c(a) + w_{cost}^c.\widehat{cost}^c(a)$$
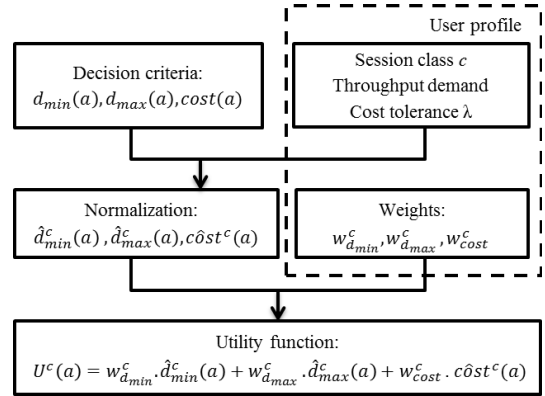
Figure 4 summarizes the decision process.

Fig. 4. Satisfaction-based multi-criteria decision process

- For each alternative (*a*), the mobile combines its radio conditions with the QoS parameters signaled by the network: it multiplies $d_{min}(a)$ and $d_{max}(a)$ with a given modulation and coding gain to determine its perceived QoS parameters, as provided by the network.
- Then, based on the user needs (*i.e.,* traffic class *c*, throughput demand and cost tolerance λ), it computes the normalized decision criteria: $\hat{d}_{min}^c(a)$, $\hat{d}_{max}^c(a)$ and $\widehat{cost}^c(a)$.
- Next, it combines the user preferences (*i.e.,* $w_{d_{min}}^c$, $w_{d_{max}}^c$ and w_{cost}^c) to the normalized decision criteria, so as to compute the weighted normalized criteria: $w_{d_{min}}^c.\hat{d}_{min}^c(a)$, $w_{d_{max}}^c.\hat{d}_{max}^c(a)$ and $w_{cost}^c.\widehat{cost}^c(a)$.
- Finally, it computes the utility function for each alternative (*a*) and selects the alternative with the highest score.

By broadcasting appropriate decisional information, the network tries to globally control users decision in a way to enhance resource utilization. On the other hand, mobiles make their decisions so as to maximize their own satisfaction. Selection decisions take then into account both user needs and preferences and operator objectives. Network complexity and processing load are, however, reduced.

IV. THE SLOPE TUNING POLICY

Because mobile users also rely on their needs and preferences to select their best alternative, the network does not

completely control individual decisions. However, by dynamically tuning its broadcasted information, the network tries to globally influence users decision in a way to enhance resource utilization.

When a RAT dominates all the others (*i.e.,* provides higher QoS parameters for the same cost or the same QoS parameters for a lower cost), common radio resources are inefficiently utilized causing performance degradation. In fact, mobile users would select the dominant alternative, leading to unevenly distributed traffic load. While a technology is overcrowded, the others are almost unexploited. This inefficiency is very similar to that of the user-centric approaches. To avoid it, QoS parameters, signaled by the network, needs to be modulated as a function of the load conditions.

Actually, as technologies are progressively loaded, QoS parameters (*i.e.,* d_{min} and d_{max}) are gradually tuned. When the operator bandwidth guarantees — identified as a generic load factor — exceed a predefined threshold S_1, these parameters are linearly reduced down to zero, as shown in Fig. 5. The slope helps to well respond to traffic load fluctuations.
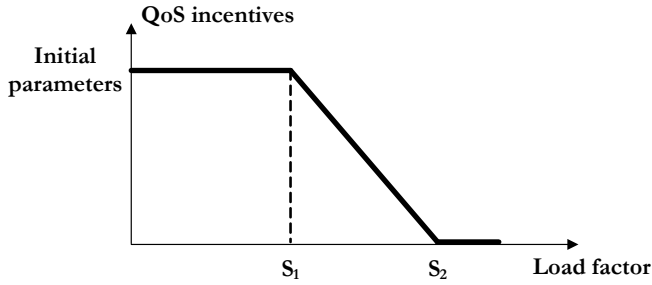

Fig. 5.   Bandwidth guarantees reduction — Slope tuning

In order to reduce network complexity and processing load (one of the drawbacks of the network-centric approaches), the proposed policy is basic and simple. Yet, it helps to efficiently distribute traffic load over the available RATs and thus to better utilize radio resources.

V. Performance Evaluation

For illustration, we consider a heterogeneous wireless network composed of Mobile WiMAX and LTE radio access technologies. They are supposed to utilize a channel bandwidth of 5 and 10 MHz respectively. For economical and technical reasons, WiMAX base station and LTE Evolved Node B are assumed to be co-localized, having the same coverage. As a consequence, and since both technologies use OFDMA on the downlink, their cell may be divided into overlayed concentric rings.

For the sake of simplicity, users are of two types: users with good radio conditions adopting the 64-QAM: 3/4 (4.5 bits/symbol) and users with bad radio conditions adopting the 16-QAM: 1/2 (2 bits/symbol). Their peak rates (*i.e.,* their perceived throughput when connected alone to WiMAX and LTE technologies) are reported in Table I.

We compare our hybrid approach with other network-centric, hybrid and user-centric methods. To evaluate selection decisions, network and users utilities are introduced. The

RAT	64-QAM: 3/4	16-QAM: 1/2
Mobile WiMAX (5 MHz)	16.6 Mb/s	7.4 Mb/s
LTE (10 MHz)	33.5 Mb/s	14.9 Mb/s

TABLE I
PEAK RATES IN MOBILE WIMAX AND LTE

network utility reflects operator objectives: it is defined as the total offered throughput. However, the users utility reflects the average user-perceived satisfaction: it depends on their needs and preferences and thus take into account both QoS and cost considerations.

Radio resources are actually allocated using fair time scheduling. Yet, when our hybrid method is employed, mobiles are first provided with their minimum guaranteed throughput given by d_{min}. Then, fair time scheduling is used to provide them with up to their maximum throughput given by d_{max}. The remaining resources may afterwards be equitably shared (*i.e.,* after receiving their maximum throughput, all mobiles have the same priority leading to fair time scheduling).

Mobiles arrive in sequence and are uniformly ready either to pay for better performances, or to sacrifice within limits their service quality seeking to save up money. When users decision needs to be evaluated, or typically when their perceived satisfaction is to be computed, a set of cost tolerance parameter and QoS and cost weights is used according to user preferences (cf. Table II).

Set No.	λ	w_{QoS}	w_{cost}
1	60	0.7	0.3
2	45	0.3	0.7

TABLE II
COST TOLERANCE PARAMETER AND QoS AND COST WEIGHTS

For comparison purposes, six different RAT selection techniques are considered:

- Peak rate maximization: Mobile users have no information on the global network state. Based on their radio conditions, they select the RAT that offers them the best peak rate.
- Instantaneous rate maximization: Mobiles are assumed to know the exact numbers of users that are connected to available technologies. Assuming that fair time scheduling is employed, they select the RAT that offers them the best throughput. Their estimated throughput in RAT *x*, $\overline{D^x}$, at the time of selection, is computed as:

$$\overline{D^x} = \frac{D^x}{1 + N^x} \qquad (5)$$

where D^x represents the user peak rate when connected to RAT x and N^x represents the number of users that are connected to RAT x at the time of selection.

- Satisfaction-based using peak rate (SB - PR): Using their peak rates, mobiles adopt the Satisfaction-based multi-criteria decision-making method to select their best RAT. In order to evaluate the different technologies, the provided QoS parameters in Eq. 2 and 3 are replaced with the peak rate that mobiles can achieve when connected to these technologies.

- Satisfaction-based using instantaneous rate (SB - IR): Mobiles use the Satisfaction-based multi-criteria decision-making method to select the RAT that maximizes their expected utility. In Eq. 2 and 3, the provided QoS parameters are replaced with the average throughput that mobiles estimate to obtain (cf. Eq. 5).
- Exhaustive search: The network considers all possible associations involving all users. It finally selects the combination that optimizes its own utility. Actually, it assigns mobiles to either WiMAX or LTE technologies in a way to maximize the total offered throughput. This is known to be the optimal method with respect to operator objectives: it leads to the highest network utility.
- Our hybrid approach: The network periodically sends decisional information (*i.e.,* cost and QoS parameters) to assist mobile users in their decisions. QoS parameters are signaled with the following thresholds: $S_1 = 0.3$ and $S_2 = 0.8$. Thereby, a network is considered to be low-loaded when its load factor is below 0.3. Initial d_{min} and d_{max} are then signaled (cf. Table III). Yet, when its load factor exceeds 0.8, a network is considered to be highly loaded, providing no QoS guarantees.

RAT	d_{min} **(Mb/s)**	d_{max} **(Mb/s)**	$cost$ **(unit/kB)**
Mobile WiMAX	1	1.5	4
LTE	1.5	2	6

TABLE III
INITIAL QOS AND COST PARAMETERS

When using the peak rate maximization and the SB - PR methods, mobiles select their RAT without any network assistance. Decisions are then user-centric. However, when employing the instantaneous rate maximization and the SB - IR methods, load conditions signaled by the network assist mobile users in their decisions. The latter two methods are thus considered to be hybrid. Finally, when adopting the exhaustive search method, decisions are network-centric since they are made by the network transparently to end-users.

Since in practice telecom operators will not reveal neither the exact numbers of users that are connected to their RATs nor the scheduling algorithm they adopt, the instantaneous rate maximization and the SB - IR methods are not realistic. Yet, they serve to illustrate the gain from masking network load conditions and only signaling cost and some QoS parameters in order to enhance resource utilization.

A. Streaming sessions

We assume that streaming sessions have an average long-term throughput of 1 Mb/s. So as to improve content quality, they can furthermore benefit from throughputs up to 1.5 Mb/s (*i.e.,* R_{av} = 1 Mb/s and R_{max} = 1.5 Mb/s).

When our proposed hybrid approach is used, the cost tolerance parameter and the weights that are assigned to the decision criteria (*i.e.,* d_{min}, d_{max} and $cost$) are put forward in Table IV. When profile no. 1 is assigned to users that are ready to pay for better performances, profile no. 2 is attributed to those that seek to save up money.

Profile No.	λ	$w_{d_{min}}$	$w_{d_{max}}$	w_{cost}
1	60	14/30	7/30	0.3
2	45	0.2	0.1	0.7

TABLE IV
USER PROFILES FOR STREAMING SESSIONS

Figures 6 and 7 respectively show the network utility and the average user utility as a function of the total throughput demand.

The network utility, defined as the total offered throughput, generally increases with the total throughput demand. Yet, when a RAT gets overloaded, its offered throughput stagnates and no longer increases with additional throughput demand.
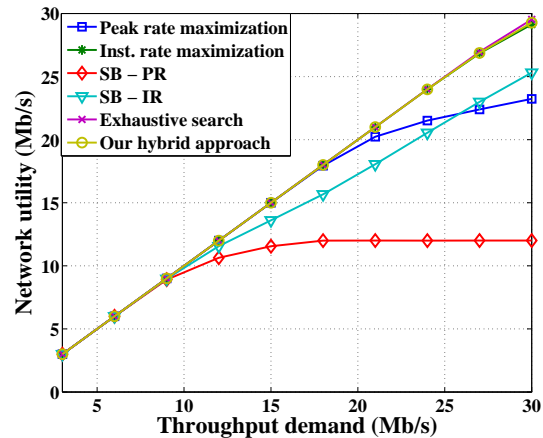

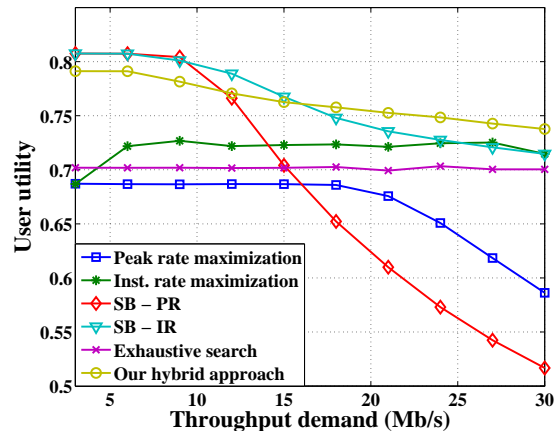Fig. 6. Network utility: Streaming sessions scenario

Fig. 7. User utility: Streaming sessions scenario

When the SB - PR method is used, all users select the mobile WiMAX technology (*i.e.,* Mobile WiMAX is their best trade-off between cost and QoS decision criteria). Regardless of user preferences and radio conditions, mobile WiMAX is expected to provide mobile users with the highest utility. Since mobiles use their peak rate in estimating their utility, their decisions do not depend on network load conditions. As a result, mobiles continue to select the WiMAX technology even when it gets overloaded.

At low traffic load, mobile WiMAX can meet users QoS needs, while charging them less. When users benefit from

throughputs up to their R_{max} and pay less, they have the highest utility (*i.e.,* satisfaction). However, when WiMAX gets loaded, it becomes no longer able to fulfill users QoS needs. Typically, at medium and high traffic load, WiMAX becomes saturated leading to a significant decrease of the user-perceived throughput below R_{av} (cf. Fig. 6). As a consequence, user-perceived satisfaction will also dramatically decrease (cf. Fig. 7).

Furthermore, when the peak rate maximization method is adopted, all users select the LTE technology. Actually, independently of their modulation and coding scheme, mobiles can achieve the best peak rate when connected to the LTE technology. Here again, their decisions do not change with network load conditions. As a consequence, at high traffic load, user-perceived throughput goes below R_{max}. Yet, it continues to be greater than R_{av}.

On the other hand, since LTE charges more than WiMAX does, mobile users experience the lowest satisfaction level at low traffic load. As a matter of fact, when all RAT selection techniques meet users QoS needs, the peak rate maximization method assign all users to the LTE technology, thus charging them more. At high traffic load, because user-perceived throughput decreases, their experienced utility also diminishes.

Moreover, when the SB - IR method is employed, users combine their needs and preferences with network load conditions to select their best RAT. As a consequence, at low traffic load and regardless of their radio conditions, all users select the mobile WiMAX technology: their QoS needs are perfeclty met while paying less. This leads to the highest user-perceived utility, as in the case of the SB - PR method. However, when the mobile WiMAX gets loaded, users may start to join the LTE technology according to their radio conditions and preferences (*i.e.,* their willingness to pay for better performances). Precisely, based on their modulation and coding scheme, as well as on their cost tolerance parameter and QoS and cost weights (cf. Table II), users estimate the utility they can obtain in both of the available RATs. They then select the technology with the highest expected utility. In fact, users with bad radio conditions that are ready to pay for better performances are the first to start to join the LTE technology. Besides, users with good radio conditions that seek to save up money are the last to start to join the LTE technology.

Consequently, since users are not proportionally distributed over the two RATs, mobile WiMAX gets overloaded before the LTE. Thus, the growth rate of the network utility decreases as the total throughput demand increases (cf. Fig. 6). This means that the average user-perceived throughput decreases. Yet, it remains greater than R_{av}. When some users start to join LTE and so pay more, and others that are connected to WiMAX start to perceive lower throughputs, the average user satisfaction also decreases as the total throughput demand increases (cf. Fig. 7).

Furthermore, our hybrid approach and the instantaneous rate maximization method perfectly meet users QoS needs, even at high traffic load. Their network utility, as depicted in Fig. 6, is so close to that of the exhaustive search method, known to be the optimal one with respect to resource utilization. Yet, as shown in Fig. 7, our hybrid approach provides the highest user utility.

In fact, when the instantaneous rate maximization method is used, mobiles select the RAT that offers them the best throughput. Therefore, a kind of load balancing is achieved: Mobile WiMAX and LTE are similarly occupied with respect to their maximum capacity. As a result, the network utility can likely follow the throughput demand increase. On the other hand, when our hybrid approach is employed, the network modulates the broadcasted QoS parameters as a function of its load conditions. It tries to push future arrivals to less loaded RATs, thus enhancing resource utilization. By integrating their needs and preferences, mobiles can avoid oversized decisions and so improve their perceived-satisfaction. Typically, at low traffic load, when both RATs can perfectly meet users QoS needs, mobile WiMAX will be preferred since it charges less. This explains why, when using our hybrid method, user utility is constantly higher than when adopting the instantaneous rate maximization method. The latter ignores user preferences (*i.e.,* its willingness to pay for better performances or to save up money) and mainly deals with load balancing. However, because the proportion of users that are connected to the LTE technology is almost constant and the user-perceived throughput is always close to R_{max}, user utility hardly changes as a function of the total throughput demand. On the other side, when using our hybrid method, since the proportion of users that are connected to the LTE increases with the total throughput demand, the average user utility decreases since LTE charges more than WiMAX. Yet, it always remains greater than that of the instantaneous rate maximization method.

Moreover, when using the exhaustive search method, the network involves all users at each decision epoch: it considers all possible combinations and selects the one that maximizes its own utility. Since user needs and preferences are ignored, and RATs are not statistically similarly occupied, this network-centric method provides the lowest user utility amongst the instantaneous rate maximization method and our hybrid approach. As a matter of fact, the network seeks to optimize its own utility, regardless of user preferences. In other words, when different combinations lead to the same network utility, they are assumed equivalent. The one that better distributes mobiles over the two RATs has no priority. As a result, statistically, the proportion of users that are connected to the LTE is higher than those of the instantaneous rate maximization and our hybrid methods, leading to lower user-perceived satisfaction (cf. Fig. 7).

To conclude, so as to illustrate the gain from masking network load conditions and only signaling cost and some QoS parameters, we compare our hybrid approach with the SB - IR one. Actually, when using our hybrid method, we can push users to LTE long before WiMAX really gets overloaded. By reducing the broadcasted QoS parameters in WiMAX, typically with $S_1 = 0.3$ and $S_2 = 0.8$, future arrivals are encouraged to join LTE well in advance in comparison with

the SB - IR scenario. Thereby, sessions are better distributed over the two RATs, leading to higher network utility as shown in Fig. 6.

At low traffic load, both methods perfectly meet users QoS needs. Yet, since the proportion of users that are connected to the most expensive RAT (*i.e.,* LTE) is higher when our hybrid approach is used, user-perceived satisfaction is lower than that of the SB - IR method. However, at high throughput demand, because future arrivals start to join LTE much earlier than the SB - IR case, WiMAX is on average less loaded when using our hybrid approach. As a consequence, WiMAX can better serve its on-going sessions leading to higher user-perceived throughput. Therefore, although mobiles may pay more (*i.e.,* the proportion of users that are connected to LTE is higher), they experience significantly better performances leading to higher satisfaction (Fig. 7). After all, by dynamically tuning QoS parameters, the network enhances resource utilization while mobiles maximize their satisfaction.

B. Elastic sessions

Elastic sessions adapt to resource availability. Their needs are expressed as comfort throughput, denoted by R_c. We assume in the following that R_c is related to the user willingness to pay and thus imposed by the user profile (cf. Table V). Typically, when users are ready to pay for better performances, they have a comfort throughput of 1.25 Mb/s. Yet, when they seek to save up money, they are content with a comfort throughput of 0.75 Mb/s.

Profile No.	λ	$w_{d_{min}}$	$w_{d_{max}}$	w_{cost}	R_c **(Mb/s)**
1	60	0	0.7	0.3	1.25
2	45	0	0.3	0.7	0.75

TABLE V
USER PROFILES FOR ELASTIC SESSIONS

We respectively depict in figures 8 and 9 the network utility and the average user utility as a function of the total number of users denoted by N_{total}.
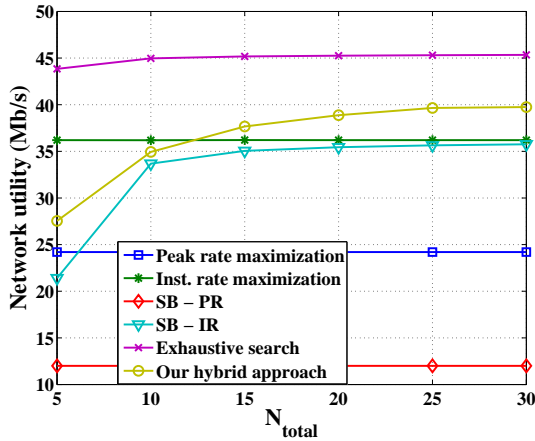

Fig. 8.   Network utility: Elastic sessions scenario

When connected alone to a RAT, an elastic session can occupy all of the available resources. However, when several
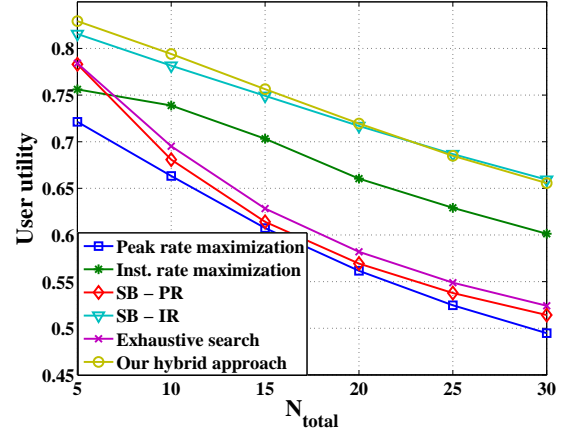

Fig. 9.   User utility: Elastic sessions scenario

sessions are present, they all share these resources. As a result, the network utility, defined as the total offered throughput, do not usually change as a function of the total number of users N_{total} (cf. Fig 8). Yet, the average user-perceived throughput is reduced.

As in the case of streaming sessions, when the SB - PR method is used, all users are connected to the mobile WiMAX technology regardless of the network load conditions. As shown in Fig. 8, the total offered throughput (*i.e.,* the network utility) is close to 12 Mb/s independently of N_{total}: it actually corresponds to the weighted average total throughput taking into account both users with good and bad radio conditions. However, the average user-perceived throughput linearly decreases with N_{total}, leading to a significant decrease of the user-perceived satisfaction (cf. Fig. 9).

Moreover, when the peak rate maximization method is adopted, all users select the LTE technology. The network utility is then, on average, higher than that of the SB - PR method. As a consequence, user-perceived throughput is also higher. But, since all users are connected to the most expensive RAT (*i.e.,* LTE), the satisfaction improvement with respect to the perceived throughput criterion fails to offset the satisfaction decrease with respect to the cost criterion. This leads to a lower user-perceived satisfaction in comparison with the SB - PR case (cf. Fig. 9).

Furthermore, when the exhaustive search method is employed, optimal resource utilization is achieved as shown in Fig. 8. Yet, the average user utility is not that interesting. First, when assigning mobiles to the available RATs, this network-centric method do not consider user preferences. It actually ignores user willingness to pay for better performances or to save up money, and only seeks to maximize the network offered throughput. Second, in order to better exploit the available resources, only few users with good radio conditions may be assigned to LTE. The majority, with bad and also good radio conditions, will be connected to mobile WiMAX, all competing for the same resources. As a result, few users connected to LTE will have excellent throughputs that far outweigh their R_c. The others will experience relatively low throughputs that may be well below their R_c. This asso-

ciation optimizes the total offered throughput, but not the user-perceived satisfaction (cf. Fig. 9). Actually, because of its concave form, even the satisfaction with respect to the throughput criterion is not maximized.

In comparison with the exhaustive search method, mobiles are better distributed over the two RATs when the instantaneous rate maximization method is adopted. In fact, users select the RAT that offers them the best throughput, leading to a kind of load balancing as in the streaming case. As a result, mobiles with equivalent radio conditions will have close throughputs regardless of their access technology. Since even users with bad radio conditions may be connected to LTE, the network utility is on average lower than that of the exhaustive search method known to be the optimal one. However, because on average perceived throughputs better meet user needs (*i.e.,* their R_c), the user utility is significantly higher than that of the exhaustive search approach.

On the other hand, when the SB - IR method is used, mobile users combine their needs and preferences with the network load conditions so as to select their best RAT. At low traffic load (typically for N_{total} = 5), more users select the mobile WiMAX technology in comparison with the instantaneous rate maximization method. When WiMAX can meet user needs very well, it charges them less. Occasionally, based on the current load conditions, a user with bad radio conditions, that is ready to pay for better performances, would select the LTE technology. As N_{total} increases, more users including those with good radio conditions start to join LTE, leading to higher network utility. The latter remain almost constant at medium and high load conditions. On average, it is slightly lower than that of the instantaneous rate maximization method. Yet, since selection decisions take into account user needs and preferences, typically their cost considerations, the user utility is significantly better than that of the instantaneous rate maximization method.

Lastly, by masking network load conditions and only signaling some cost and QoS parameters, our hybrid approach drives users decision in a way to enhance resource utilization. At low traffic load, more users typically those with bad radio conditions, that are ready to pay, select LTE. This leads to a higher network utility in comparison with the SB - IR method where, as explained before, users may occasionally join LTE (cf. Fig. 8). As a result, and although users pay on average more, they experience higher satisfaction since they have quite better throughput.

As N_{total} increases, QoS parameters are reduced with S_1 = 0.3 and S_2 = 0.8. As a consequence, future arrivals are encouraged to join LTE much earlier than the SB - IR case. However, users with good radio conditions that seek to save up money are the last to start to join LTE. In comparison with the SB - IR method, most users that are connected to WiMAX have good radio conditions, and more users with good and bad radio conditions are connected to LTE. This leads to higher total offered throughput, as shown in Fig. 8. Yet, the user utility is so close to that of the SB - IR scenario, since users having better performances pay on average more.

VI. CONCLUSION

In this paper, we address the access technology selection in heterogeneous wireless networks and proposes a hybrid decision method. Cost and QoS parameters, periodically signaled by the network, assists mobile users in their decisions. Our proposed solution combines benefits from both network-centric and user-centric approaches: it considers operator objectives as well as user needs and preferences, without unduly complicating the network. In comparison with different RAT selection techniques, including network-centric, hybrid and user-centric approaches, simulation results prove the efficiency of our hybrid method in enhancing resource utilization and maximizing user satisfaction. In the streaming sessions scenario, it optimizes the total offered throughput and maximizes the average user utility (except at low traffic load, where the non-realistic SB - IR method provides higher user satisfaction). Also, in the elastic sessions scenario, our hybrid approach significantly enhances resource utilization and maximizes user utilities in comparison with various user-centric and hybrid methods. Furthermore, compared with the exhaustive search method, known to be the optimal one with respect to resource utilization, our hybrid approach provides significantly higher user satisfaction.

REFERENCES

[1] E. Gustafsson and A. Jonsson, "Always Best Connected," *IEEE Wireless Communications Journal*, February 2003.
[2] A. Sang, X. Wang, M. Madihian, and R. D. Gitlin, "Coordinated Load Balancing, Handoff/Cell-Site Selection, and Scheduling in Multi-Cell Packet Data Systems," *Wireless Networks Journal*, January 2008.
[3] M. Ibrahim, K. Khawam, and S. Tohme, "Network-Centric Joint Radio Resource Policy in Heterogeneous WiMAX-UMTS Networks for Streaming and Elastic traffic," in *Proc. IEEE WCNC*, April 2009.
[4] M. C. Lucas-Estañ, J. Gozalvez, and J. Sanchez-Soriano, "Integer Linear Programming Optimization of Joint RRM Policies for Heterogeneous Wireless Systems," *Computer Networks Journal*, January 2012.
[5] J. Gozalvez, M. C. Lucas-Estañ, and J. Sanchez-Soriano, "Joint Radio Resource Management for Heterogeneous Wireless Systems," *Wireless Networks Journal*, May 2012.
[6] 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 ICC*, June 2011.
[7] F. Moety, M. Ibrahim, S. Lahoud, and K. Khawam, "Distributed Heuristic Algorithms for RAT Selection in Wireless Heterogeneous Networks," in *Proc. IEEE WCNC*, April 2012.
[8] O. Ercetin, "Association Games in IEEE 802.11 Wireless Local Area Networks," *IEEE Transactions on Wireless Communications*, Dec. 2008.
[9] D. Niyato and E. Hossain, "Dynamics of Network Selection in Heterogeneous Wireless Networks: An Evolutionary Game Approach," *IEEE Transactions on Vehicular Technology*, May 2009.
[10] E. Stevens-Navarro and V. Wong, "Comparison between vertical handoff decision algorithms for heterogeneous wireless networks," in *Proc. IEEE VTC*, May 2006.
[11] Q. Song and A. Jamalipour, "Network Selection in an Integrated Wireless LAN and UMTS Environment Using Mathematical Modeling and Computing Techniques," *IEEE Wireless Communications Journal*, June 2005.
[12] I. Chamodrakas and D. Martakos, "A Utility-Based Fuzzy TOPSIS Method for Energy Efficient Network Selection in Heterogeneous Wireless Networks," *Applied Soft Computing Journal*, July 2012.
[13] "IEEE Standard for Architectural Building Blocks Enabling Network-Device Distributed Decision Making for Optimized Radio Resource Usage in Heterogeneous Wireless Access Networks," *IEEE 1900.4-2009.*
[14] R. Edell and P. Varaiya, "Providing internet access: what we learn from index," *IEEE Network Journal*, September-October 1999.