A Hybrid Approach for Radio Access Technology Selection in Heterogeneous Wireless Networks

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

A Hybrid Approach for Radio Access Technology Selection in Heterogeneous Wireless Networks

Melhem El Helou*†, Samer Lahoud*, Marc Ibrahim† and Kinda Khawam‡
*University of Rennes 1, IRISA, Campus de Beaulieu, 35042 Rennes, France
†Saint-Joseph University, ESIB, Campus des Sciences et Technologies, Mar Roukoz, Lebanon
‡University of Versailles, PRISM, 45 avenue des Etats-Unis, 78035 Versailles, France

***Abstract*—In heterogeneous wireless networks, different radio access technologies are integrated and may be jointly managed. To optimize composite network performance and capacity, 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 decision framework to dynamically integrate operator objectives and user preferences. Mobile users make their selection decision based on their needs and preferences as well as on the cost and QoS information signaled by the network. Appropriate decisional information should then be derived so that the network better utilizes its radio resources, while mobile users maximize their own utility. We thus present two tuning policies, namely the staircase and the slope tuning policies, to dynamically modulate this information. Simulation results illustrate the gain from using our tuning policies in comparison with a static one: they lead to better network performance, larger operator gain and higher user satisfaction.**

***Index Terms*—Radio access technology selection, heterogeneous wireless networks, hybrid decision-making approach.**

I. Introduction

Multiple radio access technologies (RATs), such as IEEE 802.11 WLANs, mobile WiMAX, HSPA+ and LTE, are being integrated to form a heterogeneous wireless network. This cost-effective solution provides high capacity and global service coverage. However, radio resources need to be jointly managed. Typically, when a new or a handover session arrives, a decision must be made as to what technology it should be associated with. Robust decisions inevitably help to enhance resource utilization and user satisfaction.

So as 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. In [1] and [2], network selection is formulated as an optimization problem. The *best* assignment is derived to optimize the associated objective function, defined as a heterogeneous network performance metric: perceived throughput in [1] and service time in [2]. In [3] and [4], Semi-Markov Decision Process (SMDP) is proposed to find the optimal access policy that maximizes the long-term reward function. In [5], a fuzzy neural methodology is used to jointly decide of the network association and the bandwidth allocation. A reinforcement signal is also generated to optimize the decision-making process: the means and standard deviations of the input and output bell-shaped membership functions are adjusted accordingly.

However, to reduce network complexity, signaling and processing load, mobile-terminal-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). Since their payoff does not only depend on their own decision, but also on the decisions of other mobiles, game theory is widely adopted as a theoretical decision-making framework ([2], [4] and [6]). Players (*i.e.,* the individual users) will try to reach a mutually agreeable solution, or equivalently, a set of strategies they will unlikely want to change. Also, in [7] and [8], *Simple Additive Weighting* (SAW), *Multiplicative Exponent Weighting* (MEW), *Grey Relational Analysis* (GRA) and *Technique for Order Preference by Similarity to Ideal Solution* (TOPSIS) are used as multi-criteria decision-making methods. Individual users combine their QoS parameters (*e.g.,* instantaneous peak rates), calculate decision metrics, and select their access technology accordingly. Because individual users have no information on the global network state (*i.e.,* network load conditions), mobile-terminal-centric approaches are known for their potential inefficiency.

In this article, we propose a hybrid decision method that combines benefit from both network-centric and mobile-terminal-centric approaches. The network information, that is periodically broadcasted, assists 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 information signaled by the network. A particular attention is then addressed to the network to make sure it broadcasts appropriate decisional information so as to better exploit its radio resources, while individual users are maximizing their own utility. We thus present two tuning policies, namely the staircase and the slope tuning policies, to dynamically derive what to signal to the mobiles. Our hybrid framework may be naturally integrated into Self-Organizing Networks (SON) [9].

The rest of this paper is organized as follows: Section II describes our hybrid decision framework. Section III presents our tuning policies. Our system model is detailed in section IV. Section V discusses simulation parameters and results. Section

VI concludes the document.

II. Hybrid decision framework

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 [10]. This information implicitly integrates operator objectives. 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.

A. Network information

The network information provides cost and some QoS parameters: they can be seen as incentives to join available alternatives.

- The cost parameters: Because flat-rate pricing strategies waste resources [11], result in network congestion and thus degrade network performance [12], 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.
- The QoS parameters: The number of radio resource units (RRUs) (*e.g.,* OFDM symbols or OFDMA slots) that need to be allocated to future arrivals are broadcasted:
  - Mobiles are guaranteed an average minimum number of RRUs, denoted by n_{min}.
  - They also have priority to occupy up to an average maximum number of RRUs, denoted by n_{max}.

  The network loading 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 loading conditions, but also other potential operator objectives (*e.g.,* energy consumption).

  Since the smallest allocation unit (*i.e.,* RRU) may be different from one technology to another, there is a need to homogenize the QoS information. The 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.*

  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.

Although QoS parameters are provided, our decision framework is independent of local resource allocation schemes. First, enough RRUs are allocated to meet all of the operator commitments (*i.e.,* the minimum guaranteed throughput given by d_{min}). Then, any priority scheduling algorithm (including opportunistic schemes) could be adopted to allocate to each session up to its maximum prioritized throughput given by d_{max}. The remaining resources may afterwards be equitably granted to all sessions.

B. Technology 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's point of view, the network parameters are decision criteria that will be used by the mobile to rank the access technologies. For that, the mobile has to adopt a multi-criteria decision making method: it defines a utility function that will be computed for all available alternatives. This utility is obtained after normalizing and weighting the decision criteria.

III. Tuning policies

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 broadcasting appropriate decisional information, the network tries to globally influence users decision in a way to satisfy its own objectives.

In our work, we focus on efficient resource utilization: the network information is dynamically derived in a way to enhance heterogeneous network performance. On the other hand, mobile users make their decisions so as to maximize their own utility.

When a radio access technology 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 mobile-terminal-centric approaches. To remedy it, the QoS information, signaled by the network, needs to be modulated as a function of the loading conditions.

In this section, we present two tuning policies, namely the staircase and the slope tuning policies, to dynamically derive the QoS information. In order to reduce network complexity and processing load (one of the drawbacks of the network-centric approaches), our policies are basic and simple. Yet, they help to efficiently distribute traffic load and thus to better utilize radio resources.

A. Staircase tuning

Initial QoS parameters (*i.e.,* d_{min} *and* d_{max}) are specified. When the operator bandwidth guarantees — identified as a generic load factor — exceed a predefined threshold S_1, these parameters are reduced in the corresponding technology following a step function, as shown in Fig. 1. Yet, when S_2 is reached, they are set to zero.

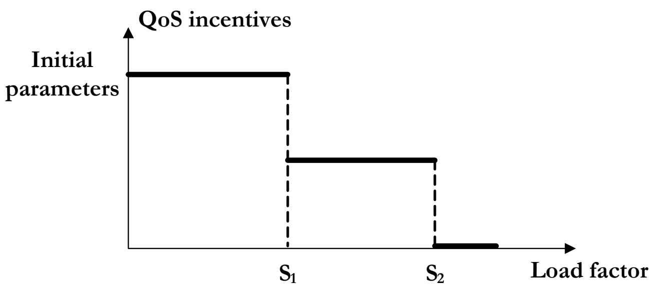


Fig. 1. Bandwidth guarantees reduction — Staircase tuning

B. Slope tuning

As technologies are progressively loaded, the QoS parameters are gradually tuned. When S_1 is reached, these parameters are linearly reduced down to zero, as shown in Fig. 2. The slope helps to better respond to traffic load fluctuations.

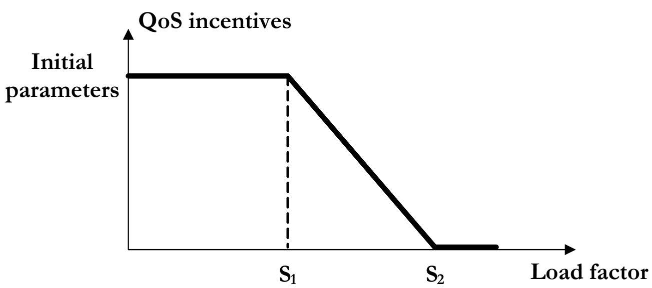


Fig. 2. Bandwidth guarantees reduction — Slope tuning

When the QoS parameters are dynamically modulated, future arrival decisions are pushed to less loaded technologies, thus enhancing long-term network performance.

IV. System model

The radio resource is divided into multiple radio resource units (RRU), thus compatible with OFDM-based technologies (*e.g.,* LTE and WiMAX technologies). In the time domain, transmissions are further organized into (radio) frames of length 10 ms. At each scheduling epoch, resource units are allocated to individual users based on their priority and current needs. Actually, before any scheduling is applied, the minimum guaranteed resource units (the operator guaranteed commitments) are directly allocated. Then, the Weighted Fair Queueing is adopted to share out the remaining resources; grants are however limited to d_{max}. Session priorities are based on the cost they pay for one unit of traffic. Finally, when all active sessions have been allocated enough resources as to meet their d_{max}, the remaining resources are equitably distributed (according to the Round Robin service discipline).

Mobile users arrive sequentially. The total number of arrivals is limited to N_{total}; it sets the traffic load. Their sojourn time is considered to be much greater in comparison with the simulation time. Consequently, the system dynamics will then slow down until a pseudo-stationary regime is attained, where all measurements are performed. Results are validated through extensive simulations.

After they decode cost and QoS parameters, mobiles adopt a satisfaction-based decision making method to evaluate and then rank the different alternatives. 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 $\hat{cost}^c(a)$ respectively.

In our work, we define three traffic classes : inelastic, streaming, and elastic classes. Before we give the normalizing functions for each traffic class, we 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:

- Inelastic sessions ($c = I$): since designed to support constant bit rate circuit emulation services, inelastic sessions require stringent and deterministic bandwidth guarantees. Thus, d_{max} should not have any impact on the final decision. Besides, the satisfaction with respect to d_{min} has a step shape (Fig. 3(a)): mobiles expect to be satisfied when d_{min} is greater or equal to their fixed throughput demand R_f; otherwise, they are not satisfied.

$$\hat{d}^I_{min}(a) = \begin{cases} 0 & if\ d_{min}(a) < R_f \\ 1 & if\ d_{min}(a) \geq R_f \end{cases} \tag{1}$$

- 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. 3(b)):

$$\hat{d'}^S(a) = 1 - \exp \frac{-\alpha(\frac{d'(a)}{R_{av}})^2}{\beta + (\frac{d'(a)}{R_{av}})} \tag{2}$$

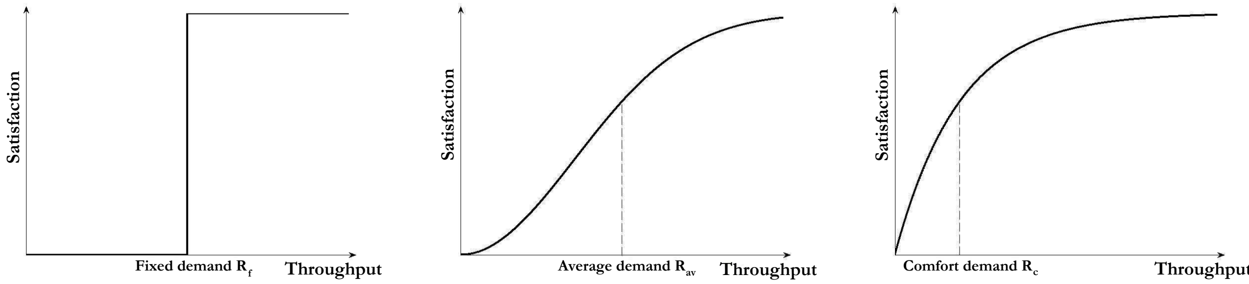
  where $d' = \{d_{min}, d_{max}\}$.
  R_{av} represents session needs: an average throughput demand. α and β are two positive constants to determine the shape of the sigmoid function.
- Elastic sessions ($c = E$): since designed to support traditional data services (*e.g.,* file transfer, email and web 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. 3(c)): 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).

$$\hat{d}^E_{max}(a) = 1 - \exp - \frac{d_{max}(a)}{R_c} \tag{3}$$

The monetary cost satisfaction is, however, modelled as a Z-shaped function for the three traffic classes (Fig. 4): the slope of the satisfaction curve increases rapidly with the cost.

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

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



(a) Inelastic sessions (b) Streaming sessions (c) Elastic sessions

Fig. 3. Throughput satisfaction function forms

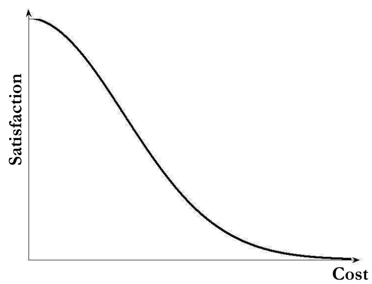


Fig. 4. Monetary cost satisfaction function form

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 $(\lambda^c, w^c_{d_{min}}, w^c_{d_{max}}, w^c_{cost}), c \in \{I, S, E\}$, where w^c_p 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 :

$$U^c(a) = w^c_{d_{min}}.\hat{d}^c_{min}(a) + w^c_{d_{max}}.\hat{d}^c_{max}(a) + w^c_{cost}.\hat{cost}^c(a)$$

The Figure 5 summarizes the decision process:

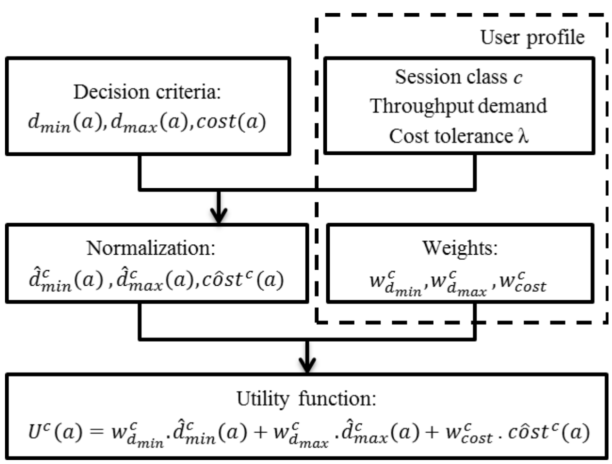


Fig. 5. Satisfaction-based multi-criteria decision process

- 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: $\hat{d}^c_{min}(a)$, $\hat{d}^c_{max}(a)$ and $\hat{cost}^c(a)$.
- Next, it combines the user preferences (*i.e.,* $w^c_{d_{min}}$, $w^c_{d_{max}}$ and w^c_{cost}) to the normalized decision criteria, so as to compute the weighted normalized criteria: $w^c_{d_{min}}.\hat{d}^c_{min}(a)$, $w^c_{d_{max}}.\hat{d}^c_{max}(a)$ and $w^c_{cost}.\hat{cost}^c(a)$.
- 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. The selection decisions take then into account both the user needs and preferences and the operator objectives. Network complexity and processing load are, however, reduced.

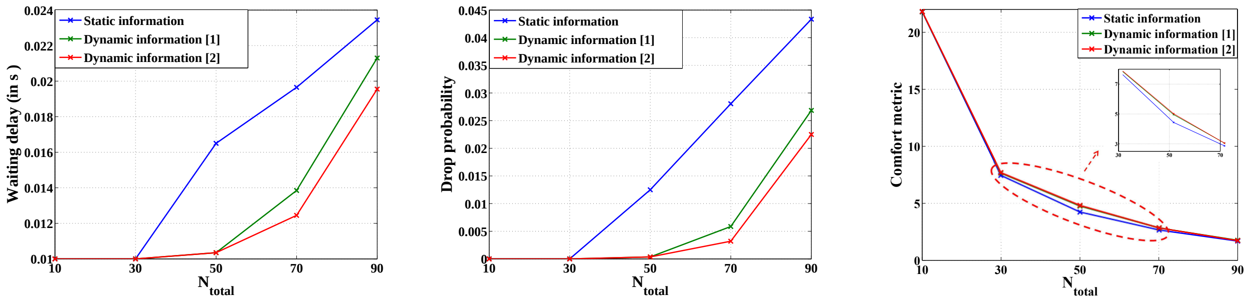
V. Simulation parameters and results

For illustration, we consider two access technologies each with a capacity of 70 Mb/s. Each is assumed to propose three different service classes, namely Premium, Regular and Economic. Initial QoS and cost parameters are depicted in Table I.

Service class	d_{min} **(Mb/s)**	d_{max} **(Mb/s)**	$Cost$ **(unit/kB)**
Premium	1	1.35	6
Regular	0.7	1	4
Economic	0.35	0.7	2

TABLE I
Initial QoS and cost parameters

After they arrive, mobiles are uniformly associated with a user profile. Detailed user profiles are presented in Table II. Inelastic and streaming session needs are respectively expressed as fixed (*i.e.,* R_f) and average long-term throughput (*i.e.,* R_{av}). We assume that the set of possible throughput demands is given by $D = \{0.5, 1, 1.5, 2\}$ Mb/s. Inelastic sessions generate packets according to a deterministic distribution, whereas streaming sessions generate packets according to a poisson process. In our work, we fix delay constraints for the latter session types. A maximum delay requirement of 100 ms is fixed. Since resources are limited, some packets may miss their deadline; they will be dropped as they are no longer useful. Furthermore, elastic session needs are expressed as comfort throughput (*i.e.,* R_c). We suppose that the set of possible comfort throughputs is given by $C = \{0.75, 1.25\}$



(a) Mean packet delay for RT sessions (b) Mean packet drop probability for RT sessions (c) Mean comfort metric for elastic sessions

Fig. 6. Experienced QoS

Mb/s. While inelastic and streaming sessions uniformly choose one of the possible throughput demands (regardless of the user cost tolerance parameter), we assume in the following that elastic session comfort throughput is related to the user willingness to pay and thus imposed by the user profile.

Profile No.	Traffic class	λ	$w_{d_{min}}$	$w_{d_{max}}$	w_{cost}
1	Inelastic	60	0.7	0	0.3
2	Streaming	60	14/30	7/30	0.3
3	Elastic	60	0	0.7	0.3
4	Inelastic	25	0.3	0	0.7
5	Streaming	25	0.2	0.1	0.7
6	Elastic	25	0	0.3	0.7

TABLE II
DETAILED USER PROFILES

We also assume that mobiles are uniformly associated with a set of modulation and coding gains. These multiplicative factors reflect the user radio conditions in the different technologies and are supposed to remain constant in time. Two sets of gains are considered and reported in Table III.

Set No.	RAT 1	RAT 2
1	1.5	1.5
2	2	1

TABLE III
MODULATION AND CODING GAIN

When the two access technologies provide the same QoS parameters, users that are associated with set no. 2 would select RAT 1: they expect to have better radio conditions and thus to perceive higher throughputs in RAT 1. All other alternatives (proposed by RAT 2) are subsequently dominated. Also, users that are associated with set no. 1 randomly join their access technology, since they expect to perceive similar throughputs in the two available technologies. This situation leads to unevenly distributed traffic load. However, when the network information is dynamically modulated according to the staircase or to the slope tuning policies, the QoS parameters are changed in a way to drive future arrivals to the less loaded RAT: loaded technologies provide lower QoS parameters and thus push future users to less loaded technologies.

When staircase policy is adopted, reduced QoS parameters are presented in Table IV.

To analyze long-term network performance, six major key performance indicators are defined: mean delay, packet loss rate (for inelastic and streaming sessions), comfort metric (for elastic sessions), throughput, operator gain and perceived satisfaction level.

Service class	d_{min} (Mb/s)	d_{max} (Mb/s)
Premium	0.5	0.7
Regular	0.35	0.5
Economic	0.2	0.5

TABLE IV
REDUCED QOS PARAMETERS (STAIRCASE POLICY)

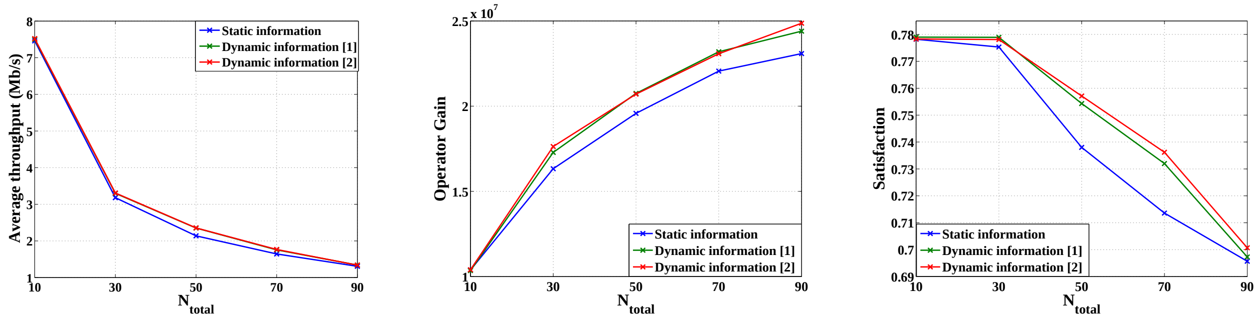
A. Performance Results

The proposed tuning policies are compared with the static one. In the latter case, initial QoS parameters are maintained fixed, except when the access technology is no longer able to guarantee to future arrivals the initial QoS parameters.

In the following, we assume that S_1 and S_2 are respectively set to 0.5 and 0.9 times the access technology capacity. Before S_1, the network provides constant QoS parameters. After S_2, QoS incentives are no longer provided to future arrivals: the network keeps a margin of about 10% of the RAT capacity to provide on-going sessions with more than their minimum guaranteed throughputs.

1) Real-time sessions: Because real-time (RT) sessions (*i.e.,* inelastic and streaming sessions) require tight delay constraints, access technologies should meet their throughput demands. However, users with a demand of 2 Mb/s may suffer: even the Premium guarantees may be lower than their throughput demand. When the access technology is highly loaded, the resource scheduler will not be able to provide them with more than their minimum guaranteed throughputs, thus leading to packet loss. So as to reduce the packet drop probability, we should avoid that a technology gets overloaded long before the others. User decisions should then be driven.

Figures 6(a) and 6(b) respectively show the mean waiting delay and the packet drop probability as a function of the total number of arrivals. When the slope intervention policy denoted as Dynamic information [2] is adopted, it best responds to traffic load fluctuations and thus provides a shorter delay, a lower drop probability and a subsequently better overall QoS level. On the other hand, the staircase intervention policy denoted as Dynamic information [1] is disadvantageous when



(a) Mean perceived throughput (b) Operator gain (c) Perceived user satisfaction

Fig. 7. Operator gain and global network performance

all technologies have exceeded their S_1: while load conditions are critical, RAT 1 is once again privileged until the operator guarantees exceed S_2 (i.e., until RAT 1 no longer provides QoS guarantees to future arrivals). Yet, real-time sessions performance are always significantly enhanced in comparison with the static scenario (denoted as Static information).

2) Elastic sessions: We define the comfort metric as the ratio of the perceived throughput to the comfort throughput. When the network information is dynamically variable, sessions are better distributed over the two technologies. More RRUs are then on average allocated to on-going sessions. Typically, elastic sessions would experience higher throughput and subsequently higher comfort metric, as shown in Fig. 6(c). However, at low traffic load (since tuning policies are not yet triggered) and at high traffic load (since all technologies becomes similarly occupied regardless of the tuning policy), performance enhancement is not that significant for elastic sessions.

3) Operator gain and global network performance: When tuning policies are triggered, QoS parameters are reduced. To benefit from the same initial bandwidth guarantees, mobile users may have to select a higher priority service class, and thus have to pay more. Also because fewer real-time packets are dropped (cf. Fig. 6(b)) and more elastic packets are served (cf. Fig. 6(c)), users consume on average a larger amount of traffic (Fig. 7(a)) and once again pay more. We illustrate in Fig. 7(b) the average operator gain. When operators dynamically intervene, they gain more.

We depict in Fig. 7(c) the average user-perceived satisfaction. Although mobiles may pay more, we notice a higher satisfaction when tuning policies are implemented. Higher costs are then justified since users benefit from significantly better performances. At low traffic load, tuning policies are not yet triggered. Equivalent performances, costs and subsequently satisfactions are intuitively observed. However, at very high traffic load, the performance gain over the static scheme begins to reduce; henceforth, it slightly offsets the cost considerations, leading to close user satisfaction.

VI. Conclusion

In this article, we address the access technology selection in heterogeneous wireless networks. We first propose a hybrid decision framework: the cost and QoS information, signaled by the network, assists mobile users in their decisions. Our proposed approach takes into account both the user needs and preferences and the operator objectives, without unduly complicating the network. We further present two tuning policies, namely the staircase and the slope tuning policies, to adjust the decisional information in a way to enhance resource utilization, while individual users are maximizing their own satisfaction. In comparison with the static scheme, performance results show that our tuning policies enhance network performance, provide larger operator gain and higher user satisfaction. Since it best responds to traffic load fluctuations, the slope tuning policy has proved to be an efficient strategy that enhances resource utilization.

References

[1] 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.

[2] 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.

[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] 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.

[5] L. Giupponi, R. Agustí, J. Pérez-Romero, and O. Sallent, “A framework for JRRM with resource reservation and multiservice provisioning in heterogeneous networks,” *Mobile Networks and Applications Journal*, December 2006.

[6] D. Niyato and E. Hossain, “Dynamics of network selection in heterogeneous wireless networks: An evolutionary game approach,” *IEEE Transactions on Vehicular Technology*, May 2009.

[7] E. Stevens-Navarro and V. Wong, “Comparison between vertical handoff decision algorithms for heterogeneous wireless networks,” in *Proc. IEEE VTC*, May 2006.

[8] Q. Song and A. Jamalipour, “A network selection mechanism for next generation networks,” in *Proc. IEEE ICC*, may 2005.

[9] 3GPP TS 32.521, “Telecommunication management; Self-Organizing Networks (SON) Policy Network Resource Model (NRM) Integration Reference Point (IRP); Requirements,” March 2010.

[10] “IEEE standard for architectural building blocks enabling network-device distributed decision making for optimized radio resource usage in heterogeneous wireless access networks,” *IEEE Std 1900.4-2009*.

[11] R. Edell and P. Varaiya, “Providing internet access: what we learn from index,” *IEEE Network Journal*, September-October 1999.

[12] F. Zhang, Y. Yan, and A. Ahmad, “Pricing for efficient usage in wired and wireless networks,” in *Int. Telecommunications Network Strategy and Planning Symposium (NETWORKS)*, June 2004.