Optimizing Network Information for Radio Access Technology Selection

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

Optimizing Network Information for Radio Access Technology Selection

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—The rapid proliferation of radio access technologies (*e.g.,* HSPA, LTE, WiFi and WiMAX) may be turned into advantage. When their radio resources are jointly managed, heterogeneous networks inevitably enhance resource utilization and user experience. In this context, we tackle the Radio Access Technology (RAT) selection and propose a hybrid decision framework that integrates operator objectives and user preferences. Mobile users are assisted in their decisions by the network that broadcasts cost and QoS parameters. By signaling appropriate decisional information, the network tries to globally control users decision in a way to meet operator objectives. Besides, mobiles combine their needs and preferences with the signaled network information, and select their access technology so as to maximize their own utility. Deriving network information is formulated as a Semi-Markov Decision Process (SMDP). We show how to dynamically optimize long-term network reward, aligning with user preferences.

Index Terms—Radio access technology selection, Semi-Markov Decision Process, hybrid decision-making approach.

I. INTRODUCTION

Devised with the vision of heterogeneity, next-generation networks integrate and jointly manage various radio access technologies (RATs). To exploit the potential of this convergence, efficient RAT selection techniques need to be defined. In the recent few years, many network-centric and mobile-terminal-centric approaches have been proposed. Typically, in [1]–[4], a Semi-Markov Decision Process (SMDP) is used to model the RAT selection problem. The network, independently of end-users, finds an optimal access policy that maximizes its long-term reward function (*i.e.,* its expected utility calculated over an infinitely long trajectory of the Markov chain). Selection decisions then meet operator objectives, without necessarily aligning with user preferences. However, to enhance user experience, multi-criteria decision-making methods are presented in [5]–[8]. Based on their needs and preferences (*e.g.,* traffic requirements, QoS-maximizing, cost- or energy-minimizing preferences), rational users select their access technology in a way to selfishly maximize their utility. Yet, because mobiles do not have a global vision of the network, their decisions are eventually in no one long-term interest.

In this article, we introduce an SMDP-based hybrid method combining benefits from both network-centric and mobile-terminal-centric approaches. Mobile users are assisted in their decisions by the network that broadcasts cost and QoS parameters. Two mutually dependent decision-making problems are thus brought into play. The first one, on the network side,

consists in deriving appropriate network information so as to guide users decision in a way to meet operator objectives. The second one, where individual users combine their needs and preferences with the signaled network information, consists in selecting the radio access technology to be associated with so as to maximize user utility. As a consequence, RAT selection dynamically integrates operator objectives and user needs and preferences.

The basic idea of our hybrid approach was first presented in [9], where intuitive tuning policies are introduced to dynamically derive network information as a function of the load conditions. In the present contribution, deriving network information is formulated as a Semi-Markov Decision Process [10]. Our goal is to dynamically optimize the long-term network reward, by trying to globally control users decision. Unlike previous studies [1]–[4] and since the network does not completely control individual decisions, transitions between the states of the SMDP do not only depend on network actions, arrival and departure rates, but also on user needs and preferences. The decision-making on the mobile side, using a multi-criteria decision-making method, is implicitly modeled as probabilistic transition rates.

The rest of this paper is organized as follows: The network model is introduced in section II. Section III presents our hybrid decision framework. The SMDP and the policy iteration algorithm are described in section IV. Numerical results are analyzed in section V. Section VI concludes the document.

II. NETWORK MODEL

A. Network Topology

Consider a heterogeneous wireless network composed of two OFDM(A)-based radio access technologies. In RAT x, $x \in \{x_1, x_2\}$, the user-perceived signal-to-noise ratio SNR^x determines the user modulation and coding scheme (mod^x, cod^x), and therefore its instantaneous peak rate (*i.e.,* its perceived throughput when connected alone to RAT x). Since practically the set of possible modulation and coding schemes is limited, RAT x cell is logically divided into N_Z^x zones (*e.g.,* concentric rings). Users in zone Z_k^x, $k = 1, ..., N_Z^x$, are supposed to have an SNR^x between δ_k^x and δ_{k-1}^x, and then to use ($mod^x(k)$, $cod^x(k)$) as a modulation and coding scheme. Although our solution adapts to different deployment scenarios, we focus on the more realistic and cost effective one where the two RATs base stations are co-localized (*i.e.,*

geographical cells overlap). The intersection of their respective zones thus leads to N_Z heterogeneous zones.

B. Network Resources

The radio resource is divided into elementary resource units (RUs), requiring both a time and a frequency dimension. When connected to RAT x in zone Z_k, $k = 1, ..., N_Z$, the maximum amount of bits $m^x(k)$ transmitted on a RU depends on $(mod^x(k), cod^x(k))$ and is given by:

$$m^x(k) = N_{os} \cdot N_{sc} \cdot log_2(mod^x(k)) \cdot cod^x(k) \cdot (1 - BLER) \quad (1)$$

where N_{os} and N_{sc} respectively represent the number of OFDM symbols and subcarriers per RU, and BLER the block error rate obtained as a function of the user-perceived SNR. In our work, and since we are interested in the large time scale radio conditions, an average SNR value per zone is considered for all RUs.

In the time domain, resources are further organized into frames of length T_f^x. When N_{ru} resource units per frame are allocated to a user in zone Z_k connected to RAT x, its perceived throughput $\overline{D}$ is expressed as follows:

$$\overline{D} = \frac{N_{ru} \cdot m^x(k)}{T_f^x} \quad (2)$$

C. Traffic Model

In our work, the number of traffic classes N_C is two. Users belong to either streaming ($c = 1$) or elastic ($c = 2$) traffic classes. Class c arrivals in zone Z_k follow a Poisson process of rate $\Lambda(k, c)$. We assume that streaming sessions have an average long-term throughput of R_{av}. Yet, to improve their content quality, they can benefit from throughputs up to R_{max}. Their duration is considered to be exponentially distributed with a mean of $1/\mu_1$.

Elastic sessions, however, adapt to resource availability. Their needs are expressed as comfort throughput denoted by R_c, and their size is assumed to be exponentially distributed with a mean of L bytes. Note that their service rate μ_2 also depends on their perceived throughputs.

III. HYBRID DECISION FRAMEWORK

A. Network Information

Using the logical communication channel (*i.e.,* radio enabler) proposed by the IEEE 1900.4 standard [11], network information is periodically sent to all mobile users. In our work, this information is assumed to implicitly integrate operator objectives, guiding users decision.

When a new or a handover session arrives, the mobile decodes network information, evaluates and then ranks available RATs.

The network is fully described by the exact numbers of class c users, in zone Z_k, that are connected to RAT x. Yet, in our work, only cost and partial QoS parameters are sent to mobiles. This significantly reduces signaling load. Furthermore, by masking RAT load conditions, even QoS parameters may be tuned so that mobiles decisions are consistent with generic

operator objectives (*e.g.,* enhance resource utilization, reduce network energy consumption).

In this setting, cost and QoS parameters signaled by the network are seen as incentives to join available RATs:

- Cost parameters: As flat-rate pricing strategies are proved to waste resources, and thus not optimal in supporting QoS, a volume-based model is 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: 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}. However, and since RUs may have different descriptions in the different RATs (*i.e.,* different N_{os} and N_{sc}), we express QoS parameters as throughputs: d_{min} and d_{max} instead of n_{min} and n_{max}. They are derived for the most robust modulation and coding scheme (*i.e.,* as perceived by users in zone Z_{N_Z}). Consequently, when evaluating available RATs, 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 $g(M, C)$.

B. RAT Selection

For RAT x, the network broadcasts the three parameters: $d_{min}(x)$, $d_{max}(x)$, and $cost(x)$. Using the satisfaction-based multi-criteria decision-making method [12], mobiles compute a utility function for each of the available RATs, and select the one with the highest score. This utility depends on user radio conditions, needs and preferences (*i.e.,* traffic class, throughput demand, QoS-maximizing and cost-minimizing preferences) as well as on the cost and QoS information sent by the network.

In our work, when $cost(x)$ is maintained fixed, $d_{min}(x)$ and $d_{max}(x)$ are dynamically tuned trying to globally control users decision. Let N_I^x be the number of possible $(d_{min}(x), d_{max}(x))$ couples that may be signaled to incite mobile users to join RAT x. In the next section, selecting the $(d_{min}(x), d_{max}(x))$ couple to be broadcasted for each RAT x is formulated as a Semi-Markov Decision Process (SMDP). The goal is to dynamically optimize the long-term discounted network reward, while mobiles maximize their own utility.

IV. SEMI-MARKOV DECISION PROCESS

At each user arrival or departure, signaled network information may vary. The SMDP is then used to dynamically derive QoS parameters in a way that optimizes the long-term network reward. We first start by defining the state space, actions, state dynamics, and reward functions. Then, using the *policy iteration* algorithm, we find the optimal solution.

A. States of the SMDP

A state of RAT x is the $(N_Z \times N_C \times N_I^x)$-tuple $n^x(t)$ for $\{k = 1, ..., N_Z, c = 1, ..., N_C, i = 1, ..., N_I^x\}$:

$$n^x(t) = (n^x(k, c, i, t)),$$

where $n^x(k,c,i,t)$ is a stochastic process representing the number of class c users in zone Z_k, having the ith ($d_{min}(x)$, $d_{max}(x)$) couple, at time t. In the remaining, we assume stationarity, and thus omit t.

To protect ongoing sessions, an admission control is performed: new arrivals may join RAT x with the ith ($d_{min}(x)$, $d_{max}(x)$) couple to the extent that RAT x available resources are enough to meet their d_{min}, while not compromising the QoS guarantees of ongoing ones. Consequently, the set of admissible states in RAT x is:

$$\mathcal{N}_a^x = \left\{ n^x \in \mathbb{N}^{N_Z \times N_C \times N_I^x} \mid \right.$$

$$\left. \sum_{k=1}^{N_Z} \sum_{c=1}^{N_C} \sum_{i=1}^{N_I^x} n^x(k,c,i) \cdot n_{min}^x(i) \leq n_{total}^x \right\} \quad (3)$$

where $n_{min}^x(i)$ is the number of RUs necessary to guarantee the d_{min} of the ith QoS parameters couple, and n_{total}^x is the total number of RUs used for data transmission in RAT x.

Let the ($N_Z \times N_C \times N_I^{x_1} + N_Z \times N_C \times N_I^{x_2}$)-tuple $s = (n^{x_1}, n^{x_2})$ be the state of the heterogeneous network defined as the concatenation of RAT x_1 and RAT x_2 substates. The state space $\mathcal{S}$ of the network is then defined as:

$$\mathcal{S} = \left\{ s = (n^{x_1}, n^{x_2}) \mid n^{x_1} \in \mathcal{N}_a^{x_1}, n^{x_1} \in \mathcal{N}_a^{x_2} \right\}$$

B. Action Set

In each state s, an action is taken by the network: QoS incentives to join each RAT are derived. An action a is the quadruple defined by $a = (d_{min}(x), d_{max}(x))$, $x \in \{x_1, x_2\}$, where $d_{min}(x)$ and $d_{max}(x)$ represents the QoS parameters of RAT x as perceived by users in zone Z_{N_Z}. Based on their needs (*e.g.*, traffic class, throughput demand) and preferences, as well as on their modulation and coding scheme (*i.e.*, geographical position), users act differently to these actions.

Obviously, $N_I^{x_1} \cdot N_I^{x_2}$ actions are possible. However, given a state $s = (n^{x_1}, n^{x_2})$, not all actions are feasible. We then denote by $\mathcal{A}$ the set of all possible actions and by $\mathcal{A}(s) \subset \mathcal{A}$ the subset of feasible actions in state s.

When both RATs provide no QoS incentives (*i.e.*, $d_{min}(x_1) = d_{max}(x_1) = d_{min}(x_2) = d_{max}(x_2) = 0$), new arrivals are blocked and can not join any RAT.

C. State Dynamics

As the network does not completely control individual decisions, transitions between the states of the SMDP do not only depend on network actions, arrival and departure rates, but also on user needs and preferences. The decision-making on the mobile side, using a multi-criteria decision-making method, actually appears as probabilistic transition rates.

Let $p^x(k,c,a)$ represent the probability that class c users in zone Z_k select RAT x, when action a is adopted. As action a may be blocking, $p^{x_1}(k,c,a) + p^{x_2}(k,c,a)$, $\forall k,c$, is not necessarily equal to one: it can be either zero or one. Transition rates $T(s,s',a)$ between states $s = (n^{x_1}, n^{x_2})$ and s' are then expressed as:

$$\begin{cases} \Lambda(k,c)\, p^{x_1}(k,c,a) & \text{if } s' = (n^{x_1} + e^{x_1}(k,c,i), n^{x_2}) \\ \Lambda(k,c)\, p^{x_2}(k,c,a) & \text{if } s' = (n^{x_1}, n^{x_2} + e^{x_2}(k,c,i)) \\ n^{x_1}(k,c,i)\, \mu_c^{x_1}(s) & \text{if } s' = (n^{x_1} - e^{x_1}(k,c,i), n^{x_2}) \\ n^{x_2}(k,c,i)\, \mu_c^{x_2}(s) & \text{if } s' = (n^{x_1}, n^{x_2} - e^{x_2}(k,c,i)) \\ 0 & \text{Otherwise} \end{cases} \quad (4)$$

where $e^x(k,c,i)$ is defined as a ($N_Z \times N_C \times N_I^x$)-tuple containing all zeros except for the $(k,c,i)^{th}$ element, that is equal to one, and new arrivals join RAT x with the ith QoS parameters couple proposed by action a. Hence, for example, when a class c user in zone Z_k joins RAT x_1, with the ith QoS parameters couple proposed by action a, the network moves to state $s' = (n^{x_1} + e^{x_1}(k,c,i), n^{x_2})$.

The state dynamics can equivalently be characterized by the state transition probabilities $p(s,s',a)$ of the embedded chain:

$$p(s,s',a) = T(s,s',a) \cdot \tau(s,a) \quad (5)$$

where $\tau(s,a)$ is the expected sojourn time for each state-action pair, defined as follows:

$$\left\{ \sum_x \sum_k \sum_c [\Lambda(k,c)p^x(k,c,a) + \sum_i n^x(k,c,i)\mu_c^x(s)] \right\}^{-1} \quad (6)$$

D. Reward Function

To formulate optimization objectives, let $r(s,a)$ denote the permanence reward earned by the network in state s, when action a is adopted. Unlike the impulsive reward, received upon transitions, the permanence reward represents the benefit and penalty continuously received by the network whilst in state s (*i.e.*, it is actually defined on a per unit time basis). In our work, $r(s,a)$ is expressed as the sum of a network utility $N(s,a)$ and a blocking cost $B(s,a)$.

The network utility is given by:

$$N(s,a) = \sum_x \sum_k \sum_c u^x(c) n^x(k,c,i) R^x(k,c,i) \quad (7)$$

where $u^x(c)$ is the class c utility earned per unit time in RAT x, and $R^x(k,c,i)$ represents the data rate of class c users in zone Z_k, that have joined RAT x with the ith ($d_{min}(x)$, $d_{max}(x)$) couple. In fact, mobiles are first provided with their minimum guaranteed throughput given by $d_{min} \cdot g(M,C)$. Then, fair time scheduling is used to provide them with up to their maximum throughput given by $d_{max} \cdot g(M,C)$. 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).

Furthermore, the blocking cost reflects the penalty of rejecting future arrivals. $B(s,a)$ is thus proportional to the arrival rates in blocking states, and is expressed as follows:

$$B(s,a) = -b \cdot \sum_k \sum_c \Lambda(k,c)(1 - \sum_x p^x(k,c,a)) \quad (8)$$

where b is the cost per unit time inflicted on the network for blocking a new arrival.

E. Uniformization

Before we proceed to solving the SMDP problem (*i.e.,* determining the action the network should take in each state) using the *policy iteration* algorithm, the continuous-time Markov Decision Process should be transformed into a discrete-time Markov chain. This can be done using uniformization.

The time is thereby discretized into intervals of constant duration τ, that is smaller than the expected sojourn times of all states: $0 \leq \tau < \tau(s, a), \forall s \in \mathcal{S}$.

Transition probabilities are then modified as follows:

$$\begin{cases} \bar{p}(s, s', a) = p(s, s', a) \frac{\tau}{\tau(s,a)} & \text{for } s' \neq s \\ \bar{p}(s, s', a) = 1 - \sum_{s' \neq s} \bar{p}(s, s', a) & \text{Otherwise} \end{cases} \quad (9)$$

where $\bar{p}(s, s', a)$ represents the probability that the network moves from state s to s' within τ, when action a is adopted.

Moreover, the reward is also modified as follows: $\bar{r}(s, a) = r(s, a)\tau$, where $\bar{r}(s, a)$ is the reward earned for a time τ.

F. Policy Iteration Algorithm

A policy π is a mapping from $\mathcal{S}$ to $\mathcal{A}$. $\pi(s)$ represents the action to take in state s. Let $H_\pi(s) = s, s_1, s_2, ..., s_n, ...$ be a trajectory of the Markov chain, when policy π is adopted. The long-term discounted reward $dr(H_\pi(s))$ of state s is the discounted sum of the rewards earned on that trajectory (that starts from s), and is expressed as follows:

$$\bar{r}(s, \pi(s)) + \psi \bar{r}(s_1, \pi(s_1)) + ... + \psi^n \bar{r}(s_n, \pi(s_n)) + ...$$

where ψ is the discounting factor ($0 < \psi < 1$). In our work, the value function of state s, denoted by $V_\pi(s)$, is set as the expected value of $dr(H_\pi(s))$ over all possible trajectories.

Our goal is to find an optimal policy π_{opt}, that maximizes the expected long-term discounted reward of each state s:

$$V_{\pi_{opt}}(s) \geq V_\pi(s), \qquad \forall s, \pi$$

We therefore use the following *policy iteration* algorithm:

- Step 0 (Initialization): We choose an arbitrary stationary policy π.
- Step 1 (Value Determination): Given the current policy π, we solve the following system of linear equations to calculate the discounted value function V_π of all states:

$$V_\pi(s) = \bar{r}(s, \pi(s)) + \psi \sum_{s' \in \mathcal{S}} \bar{p}(s, s', \pi(s)) V_\pi(s')$$

- Step 2 (Policy Improvement): When any improvement is possible, we update the current policy π. For each $s \in \mathcal{S}$, we find:

$$\hat{\pi}(s) = \arg \max_{a \in \mathcal{A}(s)} \left\{ \bar{r}(s, a) + \psi \sum_{s' \in \mathcal{S}} \bar{p}(s, s', a) V_\pi(s') \right\}$$

- Step 3 (Convergence test): If $\hat{\pi} = \pi$, the algorithm is stopped with $\pi_{opt} = \pi$. Otherwise, we go to step 1.

V. NUMERICAL RESULTS

The numerical results were obtained using Matlab on IGRIDA[1]. For illustration, we consider a heterogeneous wireless network composed of mobile WiMAX ($x = W$) and LTE ($x = L$) RATs. They are supposed to utilize a channel bandwidth of 3 and 5 MHz respectively. For the sake of simplicity, the cell is assumed divided into two zones (*i.e.,* $N_Z = 2$). While users with good radio conditions (*i.e.,* in zone 1) are considered adopting the $(64 - QAM, 3/4)$ modulation and coding scheme, users with bad radio conditions (*i.e.,* in zone 2) are supposed to employ the $(16 - QAM, 1/2)$ one. Their peak rates (*i.e.,* their perceived throughput when connected alone to mobile WiMAX and LTE technologies) are depicted in Table I.

RAT	64-QAM: 3/4	16-QAM: 1/2
Mobile WiMAX (3 MHz)	9.9 Mb/s	4.4 Mb/s
LTE (5 MHz)	16.6 Mb/s	7.4 Mb/s

TABLE I
PEAK RATES IN MOBILE WiMAX AND LTE

We assume that class c arrivals are uniformly distributed over the two zones and follow a Poisson distribution of rate $\Lambda_c = \Lambda$ (*i.e.,* $\Lambda(k, c) = \Lambda/N_Z, \forall k, c$). However, to analyze more finely network performance, different cell arrival rates will be considered.

Moreover, streaming sessions are supposed to have the following parameters: $R_{av} = 1$ Mb/s, $R_{max} = 1.5$ Mb/s and $1/\mu_1 = 45$ s. R_c of elastic sessions are further considered related to user preferences. On the one side, when users are ready to pay for better performances, they have a comfort throughput of 1.25 Mb/s. On the other side, when they seek to save up money, they are content with a comfort throughput of 0.75 Mb/s. Nevertheless, their session size L is set to 5 Mbytes.

Cost parameters are maintained fixed: $cost(W) = 4$ and $cost(L) = 6$. Yet, QoS parameters are dynamically tuned trying to globally control users decisions. For RAT x, three possible $(d_{min}(x), d_{max}(x))$ couples belonging to I^x may be signaled (*i.e.,* $N_I^W = N_I^L = 3$). In this work, the following I^W and I^L sets are considered: $I^W = \{(0, 0), (0.5, 1), (1, 1.5)\}$ Mb/s and $I^L = \{(0, 0), (0.75, 1.25), (1.5, 2)\}$ Mb/s.

The probabilities $p^x(k, c, a)$, that class c users in zone Z_k select RAT x when action a is adopted, are calculated according to the satisfaction-based multi-criteria decision-making method we have introduced in [12]. They mainly depend on user preferences, traffic class and throughput demand. Note that the probability class c users are ready to pay for better performances is assumed equal to 0.5.

Besides, and since we suppose that no RAT is preferred for any traffic class, $u^x(c), \forall x, c$, are set to one. The network utility then comes down to the sum of user-perceived throughputs. Furthermore, so as to enlarge the number of states involved in the value function, ψ is fixed at 0.99.

---

[1]A computing grid available to research teams at IRISA/INRIA in Rennes, France

For comparison purposes, the staircase tuning policy [9] is also investigated. The highest QoS parameters are first signaled. Next, when the operator bandwidth guarantees – identified as a generic load factor – exceed a predefined threshold S_1, these parameters are reduced for the corresponding RAT following a step function, as shown in Fig. 1. However, when S_2 is reached, they are set to zero. Future arrivals are thereby pushed to the less-loaded RAT.
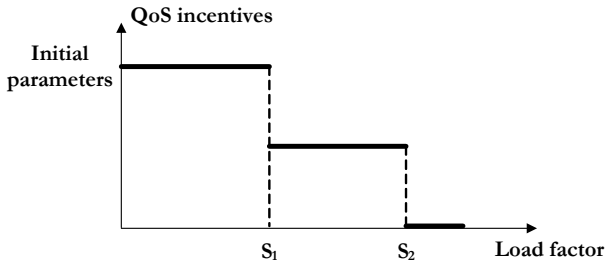

Fig. 1. QoS parameters reduction using the Staircase policy

A. Performance Evaluation

Figure 2 illustrates the average reward as a function of the cell arrival rate Λ, for different blocking costs. When b is zero, the reward function is reduced to the network utility representing the total offered throughput. Otherwise, it also includes a penalty term proportional to the blocking cost b and the cell arrival rate.
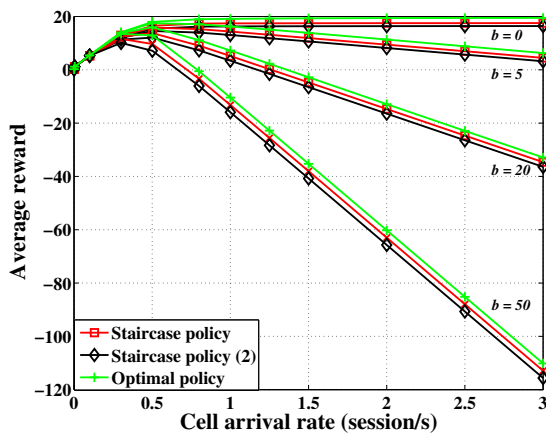

Fig. 2. The impact of the blocking cost on the average reward

At low arrival rate, no blocking occurs leading to similar rewards regardless of b. The reward function, reduced to the network total throughput, then increases with the cell arrival rate. Yet, as the latter increases further, or equivalently, when the average number of simultaneous sessions augments, network resources are almost always exploited, and not enough may be allocated to future arrivals. Therefore, the blocking probability (i.e., the long-term fraction of time spent in a blocking state) also increases. Moreover, and since the penalty term is proportional to the cell arrival rate, the reward function received by the network whilst in a blocking state is as reduced as the arrival rate increases. For all these reasons, the average reward decreases more when the cell arrival rate increases, except for b equal to zero. In fact, when b is zero, the average reward stagnates at high arrival rate. It represents the long-term

sum of user-perceived throughputs. Otherwise, the average reward obviously decreases with increasing blocking costs. We further note that the optimal policy always outperforms the staircase one with $S_1 = 0.35$ and $S_2 = 0.85$, denoted as Staircase policy (2). However, when S_1 and S_2 are carefully set to 0.3 and 0.95, the Staircase policy provides an average reward that is closer to the optimal one.
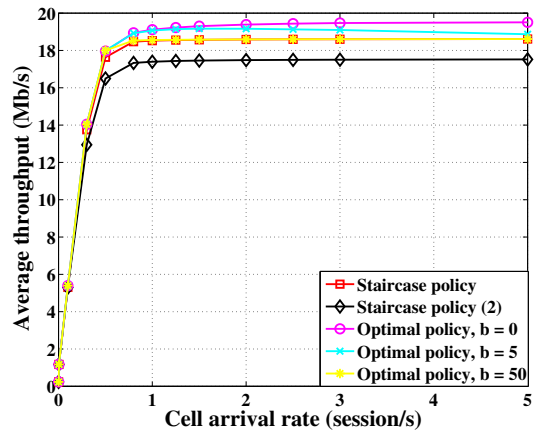

Fig. 3. Network total throughput

Besides, the impact of the penalty term on the reward function, and thereafter on the optimal policy, strongly depends on the blocking cost b. On the one hand, the higher b the more the network avoids blocking actions, even if at the expense of the network utility. On the other hand, the lower b, the more the network tries to maximize its total throughput, even if leading to more blocking states. We respectively depict in figures 3 and 4 the network total throughput and the percentage in number of blocking states as a function of the cell arrival rate. The optimal policy is illustrated for different b values. Particularly, when b is zero, the network total throughput, but also the percentage of blocking states, are maximized. The blocking cost b may therefore be tuned to control optimization objectives. Further, the performance of the staircase policy depends on S_1 and S_2. When S_1 and S_2 are respectively set to 0.35 and 0.85, a lower throughput is achieved in comparison with when $S_1 = 0.3$ and $S_2 = 0.95$. Actually, when these thresholds are carefully chosen, the staircase policy can provide quite similar performances as the optimal one ($b = 50$). They both effectively avoid blocking actions and guide users decisions. In the remaining, we only consider the case where $S_1 = 0.3$ and $S_2 = 0.95$.

It is worth noting that for a given b, when the cell arrival rate is different, the state dynamics and penalty terms are also different. This may lead to dissimilar optimal policies. Thus, and as shown in Fig. 4, the percentage in number of blocking states first increases with the cell arrival rate. Then, when the latter increases further, for b different from zero, this percentage decreases as the penalty term becomes relatively very significant.

Moreover, the blocking probability $\mathbb{P}_b$ depends not only on the number of blocking states, but mostly on the stationary distribution achieved by the different policies (i.e., on the
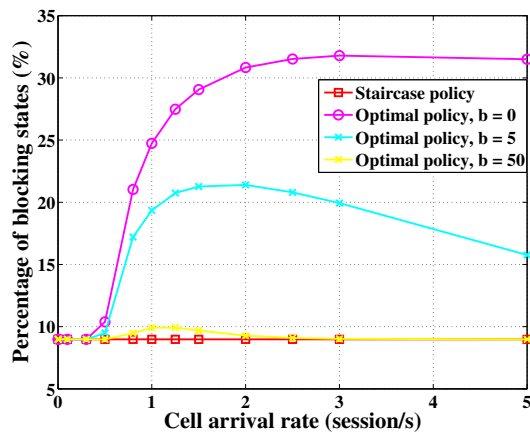
Fig. 4. Percentage in number of blocking states

long-term fraction of time spent in the different states). In the following, to efficiently analyze the impact of the blocking cost on $\mathbb{P}_b$, we separately consider streaming and elastic sessions.

The service time of elastic sessions depends both on their size assumed to be exponentially distributed with a mean of 5 Mbytes, and on their perceived throughputs. As shown before, the lower b, the higher the network total throughput leading to lower average service times. When the optimal policy is adopted (*i.e.,* the actions are fixed to the optimal ones), the SMDP may be reduced to a Markov chain, where departure rates increases with decreasing blocking costs. As a result, for a given cell arrival rate, the lower b, the lower the long-term number of simultaneous sessions. This also means that, although the lower b the higher the percentage of blocking states, the long-term fraction of time spent in these states is reduced as b is low. Accordingly, the lower b, the lower $\mathbb{P}_b$ for elastic sessions as illustrated in Fig. 5.
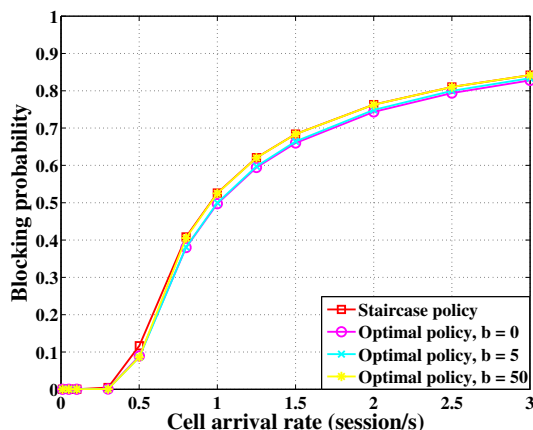

Fig. 5. Blocking probability for elastic sessions

Nevertheless, the service time of streaming sessions exclusively depends on their duration, considered to be exponentially distributed with a mean of 45 s. Thereby, maximizing the network total throughput will not reduce average service times. Consequently, as the number of blocking states increases with decreasing b, the blocking probability for streaming sessions also increases (cf. Fig. 6). The long-term fraction of time spent in all blocking states will actually be higher. Here again, for both traffic classes, the performance of the staircase policy

with carefully chosen S_1 and S_2 is comparable to the optimal one ($b = 50$).
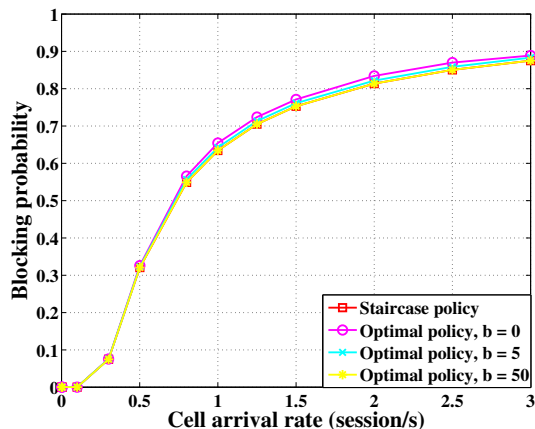

Fig. 6. Blocking probability for streaming sessions

VI. CONCLUSION

In this paper, a Semi-Markov Decision Process is used to dynamically derive QoS information for RAT selection in heterogeneous wireless networks. Through numerical results, we demonstrate the network ability to globally control users decisions in a way to maximize its long-term reward. We also show how the blocking cost may be tuned to control optimization objectives, aligning with user needs and preferences. Besides, we prove that the intuitive and logical staircase policy, with carefully chosen S_1 and S_2 thresholds, provides very close performances to the optimal one.

REFERENCES

[1] 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.
[2] J. P. Singh, T. Alpcan, P. Agrawal, and V. Sharma, "A Markov Decision Process based Flow Assignment Framework for Heterogeneous Network Access," *Wireless Network Journal*, February 2010.
[3] X. Zhang, H. Jin, X. Ji, Y. Li, and M. Peng, "A Separate-SMDP Approximation Technique for RRM in Heterogeneous Wireless Networks," in *Proc. IEEE WCNC*, April 2012.
[4] L. Zhu, F. Yu, B. Ning, and T. Tang, "Cross-Layer Handoff Design in MIMO-Enabled WLANs for Communication-Based Train Control Systems," *IEEE Journal on Selected Areas in Communications*, 2012.
[5] F. Bari and V. C. Leung, "Automated Network Selection in a Heterogeneous Wireless Network Environment," *IEEE Network Journal*, January - February 2007.
[6] L. Wang and D. Binet, "Mobility-Based Network Selection Scheme in Heterogeneous Wireless Networks," in *Proc. IEEE VTC*, April 2009.
[7] O. E. Falowo and H. Anthony Chan, "Dynamic RAT Selection for Multiple Calls in Heterogeneous Wireless Networks Using Group Decision-Making Technique," *Computer Networks Journal*, March 2012.
[8] Q.-T. Nguyen-Vuong, N. Agoulmine, E. Cherkaoui, and L. Toni, "Multicriteria Optimization of Access Selection to Improve the Quality of Experience in Heterogeneous Wireless Access Networks," *IEEE Transactions on Vehicular Technology*, May 2013.
[9] M. El Helou, S. Lahoud, M. Ibrahim, and K. Khawam, "A Hybrid Approach for Radio Access Technology Selection in Heterogeneous Wireless Networks," in *Proc. EW*, April 2013.
[10] M. L. Puterman, *Markov Decision Processes*. (John Wiley, 1994).
[11] "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*.
[12] M. El Helou, S. Lahoud, M. Ibrahim, and K. Khawam, "Satisfaction-based Radio Access Technology Selection in Heterogeneous Wireless Networks," in *Proc. WD*, November 2013.