Control of the Trade-Off between Resource Efficiency and User Fairness in Wireless Networks Using Utility-Based Adaptive Resource Allocation

Emanuel B. Rodrigues and Fernando Casadevall, Technical University of Catalonia

ABSTRACT

This work addresses the fundamental problem of the trade-off between resource efficiency and user fairness in wireless networks that use opportunistic radio resource allocation. The concept of managing the trade-off by controlling the system fairness index is applied. In order to do that, two adaptive utility-based resource allocation frameworks consisting of subcarrier assignment and power allocation algorithms are proposed. These frameworks are named utility-based alpha-rule and beta-rule, and are suitable for non-real-time and real-time services, respectively. Not only can both frameworks be designed to work as well-known classic policies found in the literature, but also as adaptive policies, which are able to meet a desired system fairness target. System level simulations show that the proposed frameworks are powerful tools to the network operator, since they can decide in which trade-off point of the efficiency-fairness plane they want to operate the system.

INTRODUCTION

Wireless communications are characterized by the scarcity of radio resources, such as time slots, subcarriers, codes, power, or modulation and coding schemes (MCSs). Due to this reason, accurate (optimal) usage of the available resources becomes mandatory. Opportunistic radio resource allocation (RRA) algorithms were proposed to tackle this fundamental problem. The term "opportunistic" means that the resources will be dynamically allocated based on users' instantaneous channel state information (CSI). The key idea is to allocate more resources to the users with better channel conditions, which leads to higher resource utilization and system capacity. However, an opportunistic strategy benefits users closer to the base station (BS), that is, those with the highest signal-to-noise ratio (SNR), which can cause starvation of users with worse channel conditions. This can severely degrade some users' experience as a result of unfair resource allocation and increased variability in the scheduled rate and delay. Moreover, long delays in scheduling of packets coming from bad channels can cause severe degradation in the performance of the overall system for higher-layer protocols, such as the Transport Control Protocol (TCP). On the other hand, schemes that provide absolute system fairness deal with the worst case scenario, penalizing users with better conditions and reducing system efficiency. Consequently, the trade-off between maximal capacity and fairness is one of the most fundamental issues in wireless systems. From a network operator perspective, it is very important to use the channel efficiently because the available radio resources are scarce and the revenue must be maximized. From the users' point of view, it is more important to have fair resource allocation such that they are not in a starvation/outage situation and their quality of service (QoS) requirements are guaranteed. Then the question is, how can the network operator manage this trade-off? RRA algorithms based on utility theory are some of the best candidates to answer this question, as shown in the following sections. The next presents the state of the art in management of the trade-off between resource efficiency and user fairness and highlights the novel contributions of the article. General utility-based resource allocation frameworks suitable for non-real-time (NRT) and real-time (RT) services are described, and we then show particular parametric RRA frameworks that can be designed to work as well-known classic RRA policies or dynamically adjusted according to the network opertator's objectives. The impact of these utility-based RRA frameworks on the aforementioned trade-off is evaluated by means of extensive system-level simulations. Finally, the conclusions are drawn.

MANAGEMENT OF THE TRADE-OFF BETWEEN RESOURCE EFFICIENCY AND USER FAIRNESS

The objective of this work is to study the trade-off between system resource efficiency and user fairness in the ambit of the medium access con-

0163-6804/11/$25.00 © 2011 IEEE

trol (MAC) layer (L2) and propose RRA frameworks able to balance these two opposing factors in scenarios with NRT or RT services.

In the present work, the assumed concept of fairness is based on QoS. It is well known that the inherent characteristics and transmission requirements of RT traffic differ from those of NRT data traffics. RT services, such as voice over IP (VoIP) and videoconference, require low and bounded delay, while NRT services, such as the web and File Transfer Protocol (FTP), are not delay-sensitive but require overall high throughput. Due to these factors, QoS metrics that can be used as fairness indicators in scenarios with RT or NRT services are delay and throughput, respectively. Among the articles that have proposed RRA algorithms to cope with this trade-off in an NRT scenario, two main approaches can be highlighted: cross-layer packet scheduling (PSC) [1] and utility-theory-based resource allocation [2–6].

Most of the PSC algorithms found in the literature that effect a compromise between efficiency and throughput-based fairness among NRT flows are based on the Proportional Fairness (PF) concept [7]. As a generalization of the PF criterion, we can highlight the weighted α-proportional fairness PSC algorithm, which is also known as the alpha-rule and was initially proposed by [1].

A more general class of RRA algorithms is based on utility fairness. Utility fairness is defined with a utility function that composes the optimization problem, where the objective is to find a feasible resource allocation that maximizes the utility function specific to the fairness concept used. There is a general family of utility functions that was evaluated in [3, 4] that includes the weighted α-proportional fairness algorithm as a special case.

On the other hand, the trade-off between efficiency in resource usage and user delay-based fairness is much less studied in a scenario with RT services. To the best of our knowledge, only [8] has investigated this trade-off in detail. The authors concluded that channel-aware opportunistic schedulers cause big rate and delay variability, which can lead to unfair situations frequently.

The utility fairness concept is used in this article to propose two generalized parametric RRA frameworks suitable for NRT and RT services, respectively, that can balance efficiency and fairness in wireless systems according to the network operator's interest. These frameworks are composed of subcarrier assignment and power allocation algorithms, and can be designed to work as well-known classic RRA policies by adjusting only one parameter in their corresponding parametric structures. Previous work has used this approach in an NRT scenario (alpha-rule) to propose only PSC algorithms [3, 4]. However, the present work uses the alpha-rule to propose novel utility-based power allocation algorithms based on multi-level waterfilling. We call the RRA policy proposed in this article utility-based alpha-rule. Furthermore, as far as we are concerned, this is the first work to use the utility fairness concept to propose an RRA framework suitable for RT services. A utility optimization problem based on the Head-Of-Line (HOL) delay is proposed and a closed-form solution is found, which we call utility-based beta-rule. These frameworks are powerful tools for the cellular operators, who have the possibility of dynamically choosing which RRA strategy is more convenient for their interests at a given instant, for example the maximization of capacity, fairness or even satisfaction (see [9]).

Previous work proposed and evaluated parametric solutions that can provide different levels of compromise between resource efficiency and fairness by varying a controlling parameter [3, 8]. However, they only evaluate static trade-offs; the controlling parameter is not adapted during network operation. The present work goes beyond and presents a novel criterion to adapt the controlling parameter of the RRA framework: a feedback control loop that meets a system fairness target. The system throughput- or delay-based fairness is calculated using a fairness index based on the general fairness function proposed by [10]. This new idea states that the trade-off between efficiency and fairness can be managed by adaptively controlling a system fairness index. That is, the network operator sets a system fairness target, and the RRA frameworks adapt their utility functions dynamically in order to operate at the desired trade-off point. In this way, a network operator will be able to answer the following question: which network performance in terms of capacity can be expected under the constraint of, say, 90 percent fairness?

UTILITY-BASED RESOURCE ALLOCATION FOR OFDMA-BASED CELLULAR NETWORKS

Utility theory is a powerful tool that can be used to design RRA algorithms able to achieve different levels of fairness in the resource allocation process [2, 5].

The concept of managing the trade-off by controlling the system fairness index is general and can be applied to any wireless system in which the users compete for centralized network resources. As a proof-of-concept we consider the orthogonal frequency-division multiple access (OFDMA) system, which is the chosen multiple access scheme for next-generation broadband cellular networks.

The utility-based RRA frameworks for OFDMA systems designed in this work run in a distributed manner in each cell and are split into dynamic subcarrier assignment (DSA) and adaptive power allocation (APA) algorithms. We consider two optimization problems suitable for NRT or RT services: maximization of the total utility with respect to the throughput or the head-of-line (HOL) delay, respectively. The two utility functions considered in this work are $U_i(T_i[n])$ and $V_j(d_j^{\text{hol}}[n])$. The throughput $T_i[n]$ of the ith NRT user is the average data rate calculated by means of a smoothing exponential filtering of the instantaneous data rate $r_i[n]$ (see [4] for more details). The HOL delay $d_j^{\text{hol}}[n]$ of the jth RT user is the time the oldest packet in the user buffer has to wait to gain access to the wire-

The concept of managing the trade-off by controlling the system fairness index is general and can be applied to any wireless system in which the users compete for centralized network resources. As a proof of concept we consider the OFDMA system, which is the chosen multiple access scheme for next-generation broadband cellular networks.

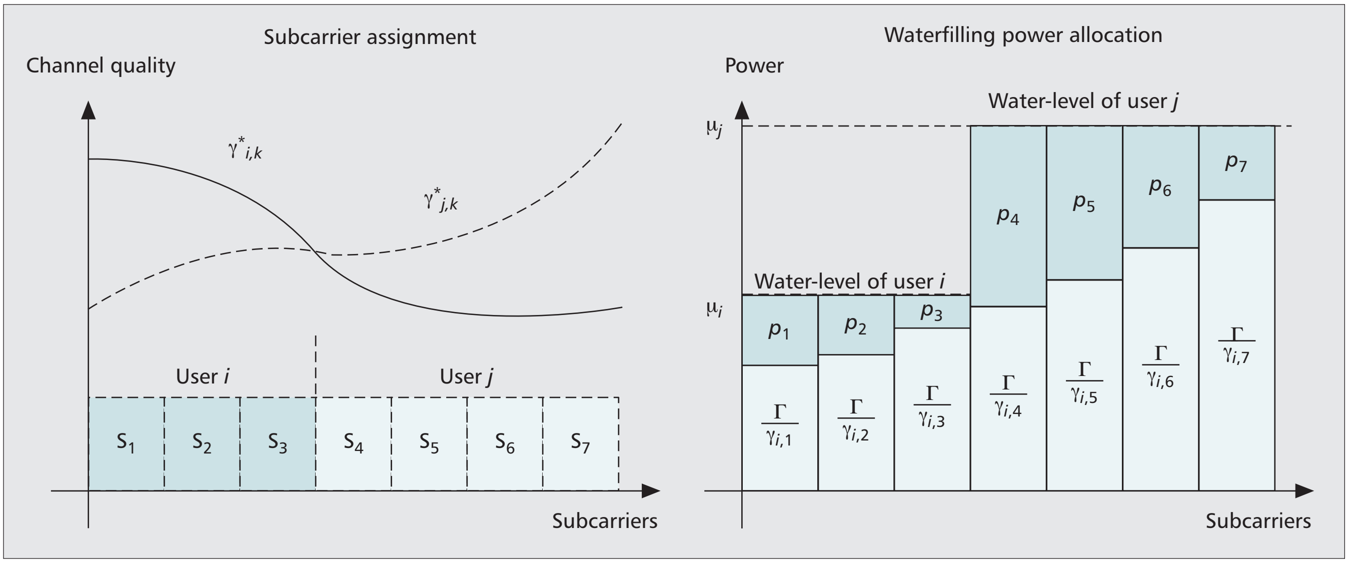

Figure 1. *Utility-based resource allocation.*

less channel. More information about the use of HOL delay in resource allocation problems can be found in [11, 12]. Two main constraints are assumed in the aforementioned optimization problems: a given subcarrier can be assigned to only one user; and the sum of all subcarriers' power must be equal to or lower than the total BS transmit power.

The optimum solutions for the two aforementioned optimization problems are still open problems. The majority of the suboptimum solutions proposed in the literature are based on the problem-splitting technique, which splits the problems into two stages: DSA and APA. In the present work, we also use this technique, as explained in the following. It was shown in [6, 13] that the separate problems can be simplified, and the resulting optimization problems are in the form of a weighted sum rate maximization [14] whose weights are adaptively controlled by the marginal utilities $U_i'(T_i[n])$ and $V_j'(d_j^{\text{hol}}[n])$, which are the derivatives of the correspondent utility functions with respect to the throughput and HOL delay, respectively.

The weighted sum rate maximization problem has a linear objective function, which greatly simplifies the corresponding algorithms. Considering equal power allocation among the subcarriers, the DSA problem has a closed form solution [6, 13]. The mobile terminal (MT) $m(k, n)$ is chosen to transmit on the kth subcarrier in the nth transmission time interval (TTI) if it satisfies the condition given by Eq. 1:

$$m(k,n) = \underset{j\in\mathcal{M}}{\arg\max}\left\{w_j\left(T_j,d_j^{hol}\right)\cdot c_{j,k}[n]\right\} \qquad (1)$$

where $\mathcal{M}$ is the set of all MTs in the cell, $c_{j,k}[n]$ denotes the instantaneous achievable transmission efficiency of the jth MT on the kth subcarrier (Shannon capacity) assuming equal power allocation per subcarrier, and $w_j(T_j, d_j^{\text{hol}})$ is a weight factor of the jth MT that can be based on throughput and/or HOL delay depending on whether the jth MT uses an NRT or RT service.

Assuming that the DSA was already done, the optimal power allocation of the weighted sum rate maximization problems has a solution in the form of a utility-based multilevel waterfilling [6, 15]:

$$p_k^*[n] = \left[\mu\cdot w_j\left(T_j,d_j^{hol}\right) - \frac{\Gamma}{\gamma_{j,k}[n]}\right]^+ \qquad (2)$$

where $[x]^+ \triangleq$, max $(0, x)$, $p_k^*[n]$ is the current optimal power allocated to the kth subcarrier belonging to the jth MT ($k \in \mathcal{K}_j$), $\Gamma/\gamma_{j,k}[n]$ is the inverse of the effective channel-to-noise ratio (CNR, i.e., channel quality) of the kth subcarrier assigned to the jth MT at the nth TTI, μ is a non-negative variable that represents the water level of the waterfilling problem, and $w_j(T_j, d_j^{\text{hol}})$ is the same weight factor used in Eq. 1. The constant Γ is called SNR gap, which indicates the difference between the theoretical limit and the SNR needed to achieve a certain data transmission rate for a practical system [5].

The weight factors in Eq. 1 and Eq. 2 depend on the marginal utilities and the service class the MT belongs to. They are given by $w_j^{\text{nrt}} = U_j'(T_j[n-1])$ and $w_j^{\text{rt}} = |V_j'(d_j^{\text{hol}}[n])|/T_j[n-1]$ for NRT and RT services, respectively (see [6] and [13] for more details).

Figure 1 explains how the utility-based RRA proposed in this article works. Consider a scenario in which two NRT users i and j compete for seven subcarriers, where the former has better channel conditions than the latter ($\gamma_{i,k} > \gamma_{j,k}, \forall k$). The channel qualities $\gamma_{i,k}^*$ and $\gamma_{j,k}^*$ plotted in the figure are utility-scaled versions of their original channel qualities $\gamma_{i,k}$ and $\gamma_{j,k}$, respectively; that is, $\gamma_{i,k}^* = w_i^{\text{nrt}}\cdot\gamma_{i,k}$ and $\gamma_{j,k}^* = w_j^{\text{nrt}}\cdot\gamma_{j,k}$. According to Eq. 1, subcarriers $k = 1 \ldots 3$ are assigned to user i and subcarriers $k = 4 \ldots 7$ are assigned to user j. Notice that if the utility-based weights w_i^{nrt} and w_j^{nrt} were not used, all sub-carriers would have been assigned to user i, who originally had better channel conditions ($\gamma_{i,k} > \gamma_{j,k}, \forall k$). Thus, the utility-based weights provided a fairer resource allocation. Notice that the same reasoning can be applied for RT services with the correspondent weights.

The utility-based waterfilling power allocation is also described in Fig. 1. Let us assume more general waterfilling expressions such as $p_k = (\mu a_k - b_k)^+$, where a_ks and b_ks are arbitrary positive numbers [15]. According to Eq. 2, we have $a_k = w_i^{\mathrm{nrt}}$ and $b_k = (\Gamma/\gamma_{i,k})$ for $k = 1 \ldots 3$; and $a_k = w_j^{\mathrm{nrt}}$ and $b_k = (\Gamma/\gamma_{j,k})$ for $k = 4 \ldots 7$. This characterizes multilevel waterfilling. In Fig. 1 the water-levels for each user are given by $\mu_i = \mu w_i^{\mathrm{nrt}}$ and $\mu_j = \mu w_j^{\mathrm{nrt}}$. Comparing the classic and multi-level waterfilling cases, one can notice that the utility-based weights added to the problem a new kind of QoS-based prioritization among users, which did not exist in the classic waterfilling allocation with a single water level. In the utility-based APA, the users that have higher weights will have more power available to their subcarriers. The same reasoning can be applied for RT services using the correspondent weights.

ADAPTIVE RESOURCE ALLOCATION FRAMEWORKS

Table 1 summarizes the features of the two parametric RRA frameworks proposed in this article. These frameworks rely on the use of two families of utility functions: $U_j\ (T_j[n])$ and $V_j\ (d_j^{\mathrm{hol}}\ [n])$. The former is based on throughput and is suitable for NRT services, while the latter is based on HOL delay and is suitable for RT services. These functions provide several degrees of throughput- and delay-based fairness depending on the value of the non-negative parameters $\alpha \in [0,\infty)$ and $\beta \in [0,\infty)$, respectively.

As can be observed in Table 1, $U_j\ (T_j\ [n])$ is a family of concave and increasing utility functions, which represents that the satisfaction of the NRT MTs increases when their throughput increases. On the other hand, $V_j\ (d_j^{\mathrm{hol}}\ [n])$ is a family of concave and decreasing utility functions, which shows that the utility of the RT MTs decreases when their HOL delay increases. Remember that the user utility-based weights w_j^{nrt} or w_j^{rt} used in Eq. 1 and Eq. 2 are directly proportional to the marginal utility, which is the derivative of the respective utility function. The values of the weights for different RRA policies are also shown in Table 1. As explained earlier, these weights play an important role in the DSA and APA algorithms, as can be observed in Eq. 1 and Eq. 2, respectively. One can clearly see that the higher the user weight, the higher the priority of this user to get a subcarrier and the higher the amount of power reserved to the subcarriers assigned to him (Fig. 1). The weights presented in Table 1 also show that MTs experiencing poor QoS (low throughput or high HOL delay) will have higher priority in the resource allocation process. Furthermore, the higher the values of α and β, the higher the priority of the user in bad conditions. Therefore, one can conclude that when α and β have higher values, the MTs with poorest QoS are benefited, and so the fairness in the system becomes stricter.

We will show in the following that, depending on the value of the fairness controlling parameters α and β, the general utility-based RRA frameworks presented in Table 1 can be designed to work as different RRA policies, achieving different performances in terms of resource efficiency and throughput or delay-based fairness.

UTILITY-BASED ALPHA-RULE FOR NON-REAL-TIME SERVICES

The generalized RRA framework suitable for NRT services presented in Table 1 joins in a unified structure the following classic policies: Max-Rate (MR) [5] (linear utility function with $\alpha = 0$), PF [2, 6] (logarithmic function with $\alpha = 1$), Max-Min Fairness (MMF) [2, 4] (exponential function with $\alpha \rightarrow \infty$), or any hybrid among these policies. Furthermore, the flexibility of this framework allows the proposal of a fairness-adaptive policy based on throughput, which is described below.

Adaptive Throughput-Based Fairness — We propose in this article the Adaptive Throughput-Based Fairness (ATF) policy, which is an adaptive version of the utility-based alpha-rule. It aims to achieve an efficient trade-off between resource efficiency and throughput-based fairness planned by the network operator in a scenario with NRT services. This is done by means of the adaptation of the fairness controlling parameter α in the utility function $U_j\ (T_j\ [n])$ presented in Table 1. The user priority in resource allocation is very sensitive to the value of α, so small values are sufficient to provide desired fairness degrees on the ATF DSA and APA algorithms.

The ATF policy is based on the definition of a fairness index ϕ_j^{nrt}, which is based on throughput and calculated for each NRT MT in the cell. The user fairness index changes with time and is defined as $\phi_j^{\mathrm{nrt}}\ [n] = T_j\ [n-1]/T_j^{\mathrm{req}}$, where T_j^{req} is the throughput requirement of the jth MT. Next, a fairness index for the whole cell comprising all NRT flows is defined by $\Phi^{\mathrm{nrt}}\ [n] = (\sum_{j=1}^{M} \phi_j^{\mathrm{nrt}}\ [n])^2 / (M \cdot \sum_{j=1}^{M} (\phi_j^{\mathrm{nrt}}\ [n])^2)$, where M is the number of MTs in the cell. This proposed Cell Fairness Index (CFI) is a particularization of the well-known Jain's fairness index proposed by Jain et al. in [10]. The general Jain's fairness function is independent of the allocation metric being used. In our case, the allocation metric is given by $\phi_j^{\mathrm{nrt}}\ [n]$. Notice that $1/M \leq \Phi^{\mathrm{nrt}}\ [n] \leq 1$. A perfect fair allocation is achieved when $\Phi^{\mathrm{nrt}}\ [n] = 1$, which means that the throughput allocated to all MTs are equally proportional to their throughput requirements (all user fairness indexes are equal). The unfairest allocation occurs when $\Phi^{\mathrm{nrt}}\ [n] = 1/M$, which means that all resources were allocated to only one MT. It is important to notice that the fairness calculation procedure presented above is general in the sense that different classes of NRT users with different throughput requirements can be contemplated.

The objective of the ATF policy is to assure a strict throughput-based fairness distribution among the MTs, i.e. the instantaneous CFI Φ^{nrt} $[n]$ must be kept around a planned value $\Phi_{\mathrm{target}}^{\mathrm{nrt}}$. Therefore, the ATF policy adapts the parameter α in the utility-based alpha-rule weight in order to achieve the desired operation point. Aiming this objective, the new value of the parameter α

Comparing the classic and multilevel waterfilling cases, one can notice that the utility-based weights added to the problem a new kind of QoS-based prioritization among users, which did not exist in the classic waterfilling allocation with a single water-level.

The Adaptive Delay-Based Fairness policy is proposed in this work as the counterpart of the ATF policy presented earlier. The difference is that ADF is the adaptive version of the utility-based beta-rule and is suitable for RT services.

is calculated using a feedback control loop of the form:

$$\alpha[n] = \alpha[n-1] - \eta_{\text{nrt}} \cdot \left(\Phi^{\text{nrt}}_{\text{filt}}[n] - \Phi^{\text{nrt}}_{\text{target}}\right) \qquad (3)$$

where $\Phi^{\text{nrt}}_{\text{filt}}[n]$ is a filtered version of the CFI $\Phi^{\text{nrt}}[n]$ using an exponential smoothing filtering, which is used to suppress short-run fluctuations and smooth time series with slowly varying trends; $\Phi^{\text{nrt}}_{\text{target}}$ is the desired value for the CFI; and the parameter η_{nrt} is a step size that controls the adaptation speed of the parameter α.

The ATF policy is an iterative and sequential process. At each TTI, the following sequence of actions is taken:

- α is calculated according to Eq. 3
- The utility-based weight factor w_j^{nrt} is calculated according to Table 1
- The DSA algorithm is executed according to Eq. 1
- The APA algorithm is executed according to Eq. 2

This process is executed indefinitely. After some iterations (TTIs), the ATF policy reaches a stable convergence of the fairness pattern defined by the target CFI. The simplicity of the ATF policy makes it a robust and reliable way to control the trade-off between resource efficiency and throughput-based fairness among NRT flows. Keeping the cell fairness around a planned target value, the network operator can have a more strict control of the network QoS and also have a good prediction about the performance in terms of system capacity.

UTILITY-BASED BETA-RULE FOR REAL-TIME SERVICES

Some classic policies suitable for RT services can also be formulated using the parametric RRA

Policies	α **or** β	**Weights**[a]	**Characteristics**
NRT services — utility-based alpha-rule — $U_i\left(T_j[n]\right) = \frac{T_j[n]^{1-\alpha}}{1-\alpha}$			
MR	0	1	High resource efficiency and low throughput-based fairness
PF	1	$\frac{1}{T_j[n-1]}$	Static trade-off between resource efficiency and throughput-based fairness
MMF	$\alpha \to \infty$	$\lim_{\alpha\to\infty} \frac{1}{T_j[n-1]^{\alpha}}$	Low resource efficiency and high throughput-based fairness
ATF	adaptive	$\frac{1}{T_j[n-1]^{\alpha}}$	Dynamic trade-off between resource efficiency and throughput-based fairness
RT Services — Utility-Based Beta-Rule — $V_j\left(d_j^{\text{hol}}[n]\right) = \frac{-\left(d_j^{\text{hol}}[n]\right)^{1+\beta}}{1+\beta}$			
PF	0	$\frac{1}{T_j[n-1]}$	High resource efficiency and low delay-based fairness
M-LWDF	1	$\frac{d_j^{\text{hol}}[n]}{T_j[n-1]}$	Static trade-off between resource efficiency and delay-based fairness
FIFO	$\beta \to \infty$	$\lim_{\beta\to\infty} \frac{d_j^{\text{hol}}[n]^{\beta}}{T_j[n-1]}$	Low resource efficiency and high delay-based fairness
ADF	Adaptive	$\frac{d_j^{\text{hol}}[n]^{\beta}}{T_j[n-1]}$	Dynamic trade-off between resource efficiency and delay-based fairness

[a] The utility-based weights w_j^{nrt} or w_j^{rt} used in Eqs. 1 and 2 that are suitable for NRT and RT services, respectively.

Table 1. *Features of the proposed parametric RRA frameworks.*

framework presented in Table 1, utility-based beta-rule. These policies are: PF [2, 6] (linear utility function with $\beta = 0$), Modified Largest Weighted Delay First (M-LWDF) [11] (quadratic function with $\beta = 1$), and First In First Out (FIFO) [12] (exponential function with $\beta \to \infty$). Furthermore, the flexibility of this framework allows not only the formulation of any hybrid among these policies, but also the proposal of a fairness-adaptive policy based on HOL delay, which is described in the following.

Adaptive Delay-Based Fairness — The Adaptive Delay-Based Fairness (ADF) policy is proposed in this work as the counterpart of the ATF policy presented earlier. The difference is that ADF is the adaptive version of the utility-based beta-rule and is suitable for RT services, as explained in Table 1.

The ADF policy is based on the definition of a fairness index ϕ_j^{rt}, which is based on the HOL delay and calculated for each RT MT in the cell. The user fairness index is defined as $\phi_j^{rt}[n] = d_j^{req}/d_j^{hol}[n]$, where d_j^{req} is the delay requirement of the jth MT. Normally, this requirement is the same for all flows of the same type and is equal to the delay budget of the RT service (maximum time that a packet can spend in the buffer before being discarded). The fairness index for the whole cell comprising all RT flows $\Phi^{rt}[n]$ is calculated in the same way as the CFI $\Phi^{nrt}[n]$ for the NRT scenario presented earlier. The difference now is that $\phi_j^{rt}[n]$ replaces $\phi_j^{nrt}[n]$.

The objective of ADF is to keep the CFI $\Phi^{rt}[n]$ around a planned value Φ_{target}^{rt}, ensuring a strict delay-based fairness distribution among the MTs. In order to do that, the parameter β in the utility-based beta-rule is adapted by the ADF policy so that the desired operation point is achieved. The new β values are calculated using a feedback control loop of the same form presented in Eq. 3. The difference is that the variables related to the scenario of RT services must be used: $\Phi_{filt}^{rt}[n]$ is a filtered version of the CFI $\Phi^{rt}[n]$ using an exponential smoothing filtering; Φ_{target}^{rt} is the desired value for the CFI; and the parameter η_{rt} is a step size that controls the adaptation speed of the parameter β. ADF also consists of an iterative and sequential process, as described for the ATF policy earlier. In the ADF case, the players are the parameter β and the utility-based weight factor w_j^{rt}.

Again, the ADF policy offers a great flexibility to the network operator because it can use the adaptive beta-rule to work on a desired operation point of the trade-off plane between resource efficiency and delay-based fairness among RT flows.

SIMULATION RESULTS

The adaptive RRA frameworks presented earlier were evaluated by means of system level simulations, which took into account the main characteristics of an OFDMA system. The main simulation parameters are depicted in Table 2.

A good way to evaluate the trade-off between resource efficiency and user fairness is plotting a 2D plane between total cell throughput (capacity) and Cell Fairness Index (CFI) (see earlier sections and [10]). Figures 2 and 3 present the planes built from the simulations of the utility-based alpha-rule and beta-rule frameworks on scenarios with 16 active NRT flows and 105 active RT flows, respectively. In each set of simulations, two approaches were assessed: Joint and Equal Power Allocation (EPA). In the former, both DSA and APA algorithms use a given RRA policy, for example PF. The latter approach means that the chosen policy is used only on the DSA algorithm, while on the power allocation step the total transmission power is equally divided among the subcarriers.

Figure 2 shows the performance comparison among the policies regulated by the utility-based alpha-rule proposed in this work, including the classic ones (MR, PF and MMF), which are indicated as single markers, and the new adaptive policy ATF, which is indicated as solid (Joint) and dashed (EPA) lines. One can clearly see the static behavior of the classic policies on the efficiency-fairness plane. MMF is able to provide maximum throughput-based fairness at the expense of low system capacity, while MR is the most efficient on the resource usage but provides an unfair throughput distribution among users. The PF policy appears as a fixed trade-off between MMF and MR, with intermediate throughput-based fairness and system capacity.

The ATF policy, which controls the parameter α adaptively according to (3) in order to achieve a desired Cell Fairness Target (CFT), is able to cover the whole path between the classic policies in the efficiency-fairness plane. Notice in the ATF curves that the fairness targets set in the simulations (0.2, 0.4, 0.6, 0.8 and 1.0) are always met. One can observe that the performance of the ATF policy for very low fairness region (CFT = $1/J$, where J is the number of MTs) converges to the performance of the MR policy, as expected. In this way, it can be concluded that the ATF policy can adaptively adjust the utility-based RRA framework presented in Table 1 in order to provide a dynamic trade-off between resource efficiency and throughput-based fairness.

It can also be seen in Fig. 2 that there is not a considerable advantage in using an adaptive power allocation for the problem and scenario considered in this work. The Joint approach presents a small gain in cell throughput for the same CFIs compared with the EPA approach. This gain is due to the faster convergence of the parameter α when both DSA and APA algorithms are used. Furthermore, α stabilizes in lower values when APA is used, which yields higher cell throughput. This small gain comes at the expense of higher computational cost, which has to be taken into account by the network operator.

The RRA framework suitable for RT services, which is regulated by the utility-based beta-rule proposed in this article, is evaluated in Fig. 3. This figure shows the performance of three classic policies (PF, M-LWDF and FIFO) and the new adaptive policy ADF. As expected, the classic policies present a tradeoff between resource efficiency, illustrated as the total cell throughput, and user fairness, represented by the CFI based on HOL delay. PF uses the radio

MMF is able to provide maximum throughput-based fairness at the expense of low system capacity, while MR is the most efficient on the resource usage but provides an unfair throughput distribution among users.

PF uses the radio resources more efficiently but does not present so high delay-based fairness values, while FIFO provides maximum fairness in the delay distribution but is very inefficient in the resource usage. M-LWDF presents a good static trade-off, with cell throughput as high as the one presented by PF.

resources more efficiently but does not present so high delay-based fairness values, while FIFO provides maximum fairness in the delay distribution but is very inefficient in the resource usage. M-LWDF presents a good static trade-off, with cell throughput as high as the one presented by PF.

On the other hand, for the same CFIs, the adaptive ADF policy provides equal or better cell throughput than the classic policies. Furthermore, the results demonstrate that it is able to meet successfully the CFTs defined in the simulations (0.5, 0.6, 0.7, 0.8, 0.9 and 1.0). The fairness target of 0.5 is approximately met because of the structure of the utility-based beta-rule, which delimitates the action and corresponding

Parameter	Value
Number of cells	1
Maximum BS transmission power	1 W
Cell radius	500 m
MT speed	static
Carrier frequency	2 GHz
Number of subcarriers	192
Subcarrier bandwidth	15 kHz
Path loss[a]	$L = 128.1 + 37.6 \log_{10} d$
Log-normal shadowing standard dev.	8 dB
Small-scale fading	Typical Urban (TU)
AWGN power per subcarrier	–123.24 dBm
BER requirement	10^{-6}
Link adaptation	Continuous using effective Shannon capacity formula
Transmission time interval (TTI)	0.5 ms
NRT traffic model	Full buffer
Throughput filtering time constant	1000
Minimum α value	0
Maximum α value	10
ATF control time window	0.5 ms
ATF fairness target (Φ^{nrt}_{target})	Variable
ATF step size (η_{nrt})	0.1
ATF filtering time constant	10
RT traffic model	Packets of 32 bytes with interarrival time of 2 ms
HOL delay filtering time constant	100
Minimum β value	0
Maximum β value	10
ADF control time window	0.5 ms
ADF fairness target (Φ^{rt}_{target})	Variable
ADF step size (η_{rt})	0.1
ADF filtering time constant	10
RT delay budget	100 ms
Simulation time span	5 s
Number of independent simulation runs	30

[a] *d* is the distance to the BS in kilometers.

Table 2. *Simulation parameters.*

performance of ADF inside the range of PF and FIFO, the two possible extremes of the trade-off between resource efficiency and delay-based fairness considering the RRA framework suitable for RT services.

Similarly to the case of the ATF policy in Fig. 2, the ADF Joint policy shows a small gain in system capacity compared with the EPA case. This indicates that a controllable trade-off can be properly achieved with less computational complexity, by applying the ADF policy only to the DSA algorithm and next applying equal power allocation among the sub-carriers.

It is also interesting to evaluate aspects related to cell coverage and user QoS when the system fairness is controlled by the adaptive utility-based policies proposed in this work, namely ATF and ADF. Fig. 4 depicts the user throughput and HOL delay when the ATF EPA and ADF EPA policies are used in the scenarios with NRT and RT services, respectively. The impact of the cell coverage is assessed by considering different groups of MTs: the ones with best propagation conditions (10 percent of the MTs with smallest path loss + shadowing losses, i.e. located close to the BS), and the ones with worst propagation conditions (10 percent of the MTs with highest path loss + shadowing losses, i.e. located in the cell edge).

Regarding the ATF policy, when the cell fairness target is set to the minimum value (CFT = 1/*J*) and ATF is configured to work as an opportunistic policy, we have that the worst MTs are in outage, i.e. they have no chance to transmit. On the other hand, the best users have an excess of QoS in terms of throughput. As the CFT increases, the ATF takes resources from the "rich" and give them to the "poor," improving a little the QoS of the worst users. Notice that high fairness (i.e., similar performance between the best and worst groups) comes at the expense of capacity loss (trade-off between efficiency and fairness). Taking into account the ADF policy, when a small CFT has to be guaranteed, there is a remarkable difference in the delay-based QoS of the best and worst user groups. In order to achieve higher CFTs, ADF performs a controllable QoS degradation of the best users in favor of the rest of the users, but not necessarily for the worst users. Certainly, as can be seen in Fig. 4, this resource reallocation is not translated directly in QoS improvement of the worst users, but at least network congestion, i.e. excessive delay of many flows, is avoided for a longer period of time. However, the delay-based fairness implemented by ADF must be carefully managed when high fairness targets are desired. In this situation, the packet delays of both groups approach the RT delay budget, and attention must be paid to not allow excessive packet losses due to high delays.

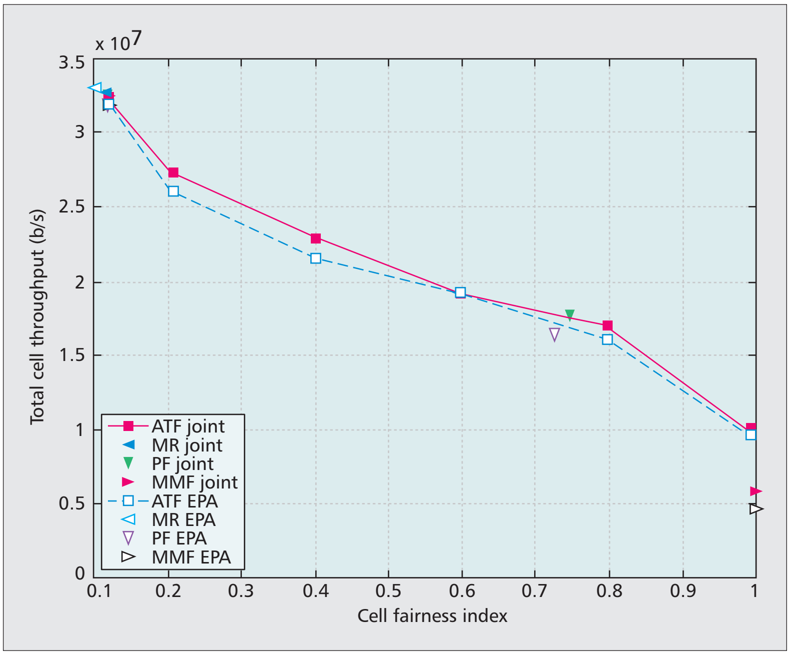

Figure 2. *Evaluation of the trade-off between resource efficiency and throughput-based fairness for NRT services.*

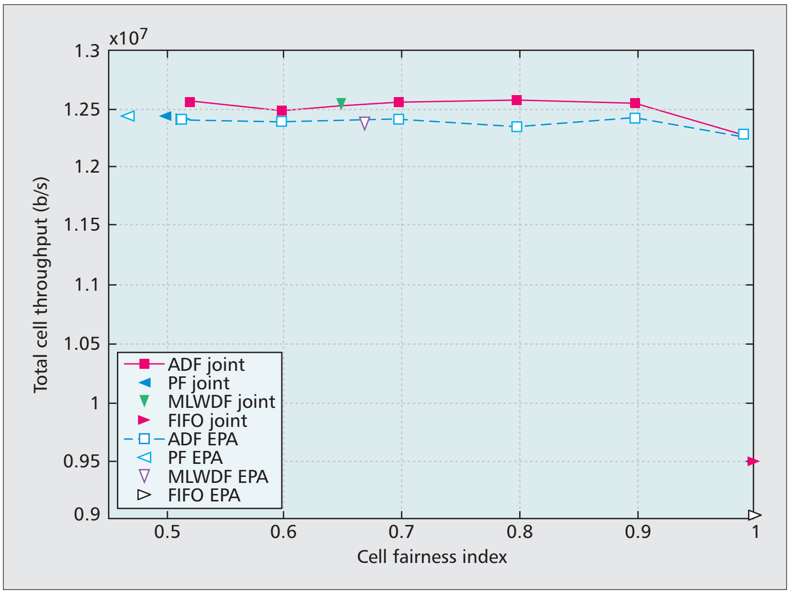

Figure 3. *Evaluation of the trade-off between resource efficiency and delay-based fairness for RT services.*

CONCLUSIONS

This work addresses the fundamental problem of the trade-off between resource efficiency and user fairness in wireless networks that use opportunistic radio resource allocation. Using Utility Theory concepts, two adaptive RRA frameworks comprised of dynamic subcarrier assignment and adaptive power allocation algorithms are proposed in this work. The first one is based on the utility-based alpha-rule and is suitable for NRT services, while the second one is based on the utility-based beta-rule and is suitable for RT services. Both frameworks can be designed to work not only as well-known classic policies found in the literature, but also as new adaptive policies, which are able to meet a desired cell fairness target.

Simulation results show that the adaptive policies of both frameworks outperform the classic policies and are able to operate on any desired trade-off point of their respective efficiency-fairness planes. This is a remarkable

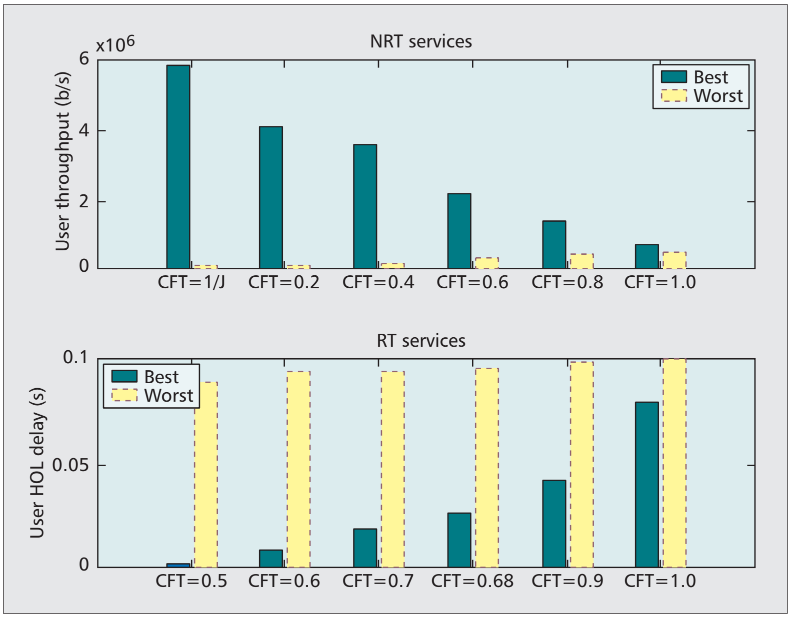

Figure 4. *User throughput and user HOL delay for NRT and RT services, respectively.*

strategic advantage to the network operators, because they can now control the aforementioned trade-off and decide in which point on the plane they want to operate. It can also be concluded that it is sufficient to apply the utility-based rules only to the DSA algorithm. Equal power allocation among subcarriers provides almost the same result as the utility-based APA with much less computational complexity.

Finally, it should be highlighted that the conceptual RRA frameworks proposed in this work can be used in any wireless system with minor adjustments. Furthermore, the application of these frameworks in a multicell scenario is straightforward, since they are based on policies that use only local BS information.

ACKNOWLEDGMENT

The authors wish to acknowledge the activity of the Network of Excellence in Wireless Communications NEWCOM++ of the European Commission (contract no. 216715) that motivated this work. Emanuel B. Rodrigues has a Ph.D. scholarship support by the Improvement Coordination of Superior Level People (CAPES), Brazil.

REFERENCES

[1] J. Mo and J. Walrand, "Fair End-to-End Window-based Congestion Control," *IEEE/ACM Trans. Net.*, vol. 8, no. 5, Oct. 2000, pp. 556–67.

[2] F. Kelly, "Charging and Rate Control for Elastic Traffic," *Euro. Trans. Commun.*, vol. 8, 1997, pp. 33–37.

[3] A. Sang *et al.*, "A Flexible Downlink Scheduling Scheme in Cellular Packet Data Systems," *IEEE Trans. Wireless Commun.*, vol. 5, no. 3, Mar. 2006, pp. 568–77.

[4] E. B. Rodrigues *et al.*, "Resource Allocation and Packet Scheduling in OFDMA-based Cellular Networks," *Proc. 4th Int'l. Conf. Cognitive Radio Oriented Wireless Networks and Communications — CrownCom*, June 2009, pp. 1–6.

[5] G. Song and Y. G. Li, "Cross-Layer Optimization for OFDM Wireless Networks — Part I: Theoretical Framework," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, Mar. 2005, pp. 614–24.

[6] G. Song and Y. G. Li, "Cross-Layer Optimization for OFDM Wireless Networks — Part II: Algorithm Development," *IEEE Trans. Wireless Commun.*, vol. 4, no. 2, Mar. 2005, pp. 625–34.

[7] P. Viswanath, D. N. C. Tse, and R. Laroia, "Opportunistic Beamforming Using Dumb Antennas," *IEEE Trans. Info. Theory*, vol. 48, no. 6, June 2002, pp. 1277–94.

[8] R. Srinivasan and J. S. Baras, "An Analysis of Delay-Constrained Opportunistic Scheduling for Cellular Wireless Systems," *Center for Satellite and Hybrid Communication Networks — CSHCN*, Tech. Rep. TR 2004-2, 2004.

[9] E. B. Rodrigues and F. Casadevall, "Adaptive Radio Resource Allocation Framework for Multi-User OFDM," *Proc. IEEE VTC Spring*, Apr. 2009, pp. 1–6.

[10] R. Jain, D. Chiu, and W. Hawe, "A Quantitative Measure of Fairness and Discrimination for Resource Allocation in Shared Computer Systems," DEC Research, tech. rep. TR-301, Sept. 1984.

[11] M. Andrews *et al.*, "Providing Quality of Service Over A Shared Wireless Link," *IEEE Commun. Mag.*, vol. 32, no. 2, June 2001, pp. 150–54.

[12] S. Shakkottai and A. L. Stolyar, "Scheduling Algorithms for a Mixture of Real-Time and Non-Real-Time Data in HDR," *Proc. 17th Int'l. Teletraffic Congress — ITC*, 2001, pp. 793–804.

[13] H. Lei *et al.*, "A Packet Scheduling Algorithm Using Utility Function for Mixed Services in the Downlink of OFDMA Systems," *Proc. IEEE VTC Fall*, Sept. 2007, pp. 1664–68.

[14] L. M. C. Hoo *et al.*, "Multiuser Transmit Optimisation for Multicarrier Broadcast Channels: Asymptotic FDMA Capacity Region and Algorithms," *IEEE Trans. Commun.*, vol. 52, no. 6, June 2004, pp. 922–30.

[15] D. P. Palomar and J. R. Fonollosa, "Practical Algorithms for a Family of Waterfilling Solutions," *IEEE Trans. Sig. Proc.*, vol. 53, no. 2, Feb. 2005, pp. 686–95.

BIOGRAPHIES

EMANUEL BEZERRA RODRIGUES (emanuel@tsc.upc.edu) received B.Sc. and M.Sc. degrees in electrical engineering from the Federal University of Ceará (UFC), Brazil, in 2001 and 2004, respectively. He worked in the Wireless Telecom Research Group (GTEL-UFC) from 2001 to 2007 and did a three-month internship at Ericsson Research at Linköping, Sweeden, in 2004. He has just finished his Ph.D. studies at the Signal Theory and Communications Department (TSC) of the Universitat Politécnica de Catalunya (UPC), Spain. His main research interests are radio resource management and QoS control for wireless communication systems.

FERNANDO J. CASADEVALL (ferranc@tsc.upc.edu) received his Engineer of Telecommunication and Dr.Eng. degrees from UPC, Spain, in 1977 and 1983 respectively. In 1978 he joined UPC, where he was an associate professor from 1983 to 1991. He is currently a full professor in the Signal Theory and Communications Department. After graduation he was concerned with equalization techniques for digital fiber optic systems. He has also been working in the field of digital communications with particular emphasis on digital radio and its performance under multipath propagation conditions. In the last 15 years, he has mainly been concerned with the performance analysis and development of digital mobile radio systems. In particular his research interest include cellular and personal communication system, multipath transceiver design (including software radio techniques), mobility and radio resources management, and end-to-end QoS issues. During the last 10 years he has participated in more than 30 research-projects funded by both public and private organizations. In particular, he actively participated in 10 research projects funded by the European Commission, being Project Manager of three of them: ARROWS, EVEREST, and AROMA (see http://www.gcr.tsc.upc.edu for details). He has published around 100 technical papers in both international conferences and magazines, most of them IEEE publications. He has also been a Technical Program Committee member of different international IEEE supported conferences as well as a reviewer of several IEEE magazines. From October 1992 to January 1996 he was responsible for the Information Technology Area in the National Agency for Evaluation and Forecasting (Spanish National Research Council).