

Multiscale Fairness and Its Application to Resource Allocation in Wireless Networks

Eitan Altman, Konstantin Avrachenkov, Sreenath Ramanath

► **To cite this version:**

Eitan Altman, Konstantin Avrachenkov, Sreenath Ramanath. Multiscale Fairness and Its Application to Resource Allocation in Wireless Networks. Jordi Domingo-Pascual; Pietro Manzoni; Sergio Palazzo; Ana Pont; Caterina Scoglio. 10th IFIP Networking Conference (NETWORKING), May 2011, Valencia, Spain. Springer, Lecture Notes in Computer Science, LNCS-6641 (Part II), pp.225-237, 2011, NETWORKING 2011. <10.1007/978-3-642-20798-3_17>. <hal-01597970>

HAL Id: hal-01597970

<https://hal.inria.fr/hal-01597970>

Submitted on 29 Sep 2017

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Multiscale Fairness and its Application to Resource Allocation in Wireless Networks

Eitan Altman, Konstantin Avrachenkov, and Sreenath Ramanath

INRIA, 2004 Route des Lucioles, BP 93, 06902 Sophia Antipolis, France
{eitan.altman,konstantin.avrachenkov,
sreenath.ramanath}@sophia.inria.fr
<http://www-sop.inria.fr/>

Abstract. Fair resource allocation is usually studied in a static context, in which a fixed amount of resources is to be shared. In dynamic resource allocation one usually tries to assign resources instantaneously so that the average share of each user is split fairly. The exact definition of the average share may depend on the application, as different applications may require averaging over different time periods or time scales. Our main contribution is to introduce new refined definitions of fairness that take into account the time over which one averages the performance measures. We examine how the constraints on the averaging durations impact the amount of resources that each user gets.

Keywords: Resource allocation; Multiscale fairness; α -fairness; T -scale fairness; Networking

1 Introduction

Let us consider some set S of resource that we wish to distribute among I users by assigning user i a subset S_i of it. We shall be interested in allocating subsets of the resource fairly among the users. The set S may actually correspond to one or to several resources. We shall consider standard fairness criteria for sharing the resources among users. We shall see, however, that the definition of a resource will have a major impact on the fair assignment.

We associate with each user i a measurable function x_i that maps each point in S to some real number. Then, we associate with each i a utility u_i which maps all measurable subsets S_i to the set of real numbers. We shall say that S is a resource if $u_i(S_i)$ can be written for each $S_i \subset S$ as

$$u_i(S_i) = f \left(\int_{S_i} x_i(s) ds \right)$$

As an example, consider I mobiles that wish to connect to a base station between 9h00 and 9h10 using a common channel. The time interval is divided into discrete time slots whose number is N . Assume that the utility for each mobile s of receiving a subsets \mathcal{N}_i of slots depend only on the number of slots N_i it receives. Then the set of N slots is considered to be a resource.

Next assume that if mobile i receives the channel at time slot t then it can transmit at a throughput of X_t^i . Assume that the utility of user i is a function of the total throughput it has during this fraction of an hour. Then again the N slots are considered as a resource.

We adopt the idea that fair allocation should not be defined in terms of the object that is split but in terms of the utility that corresponds to the assignments. This is in line with the axiomatic approach for defining the Nash bargaining solution for example. With this in mind, we may discover that the set of N slots cannot always be considered as a resource to be assigned fairly. Indeed, a real time application may consider the N slots as a set of n resources, each containing $B = N/n$ consecutive slots. A resource may correspond to the number of time slots during a period of 100 msec. The utility of the application is defined as a function of the instantaneous rate, i.e. the number of slots it receives during each period of 100 msec. (With a playout buffer that can store 100 msec of voice packets, the utility of the mobile depends only on how many slots are assigned to it during 100 msec and not which slots are actually assigned to it.)

Related work: Our work is based on the α -fairness notion introduced in [1]. This, as well as other fairness notions can be defined through a set of axioms, see [2]. This paper is inspired by several papers which already observed or derived fairness at different time-scales [3–8]. However, we would like to mention that the T -scale fairness (a unifying generalization of long- and short- term fairness) and multiscale fairness are new concepts introduced in the present work.

Structure of the paper: In Section 2, we introduce a resource sharing model which is particularly suitable for wireless applications. We also define several fairness criteria. In Sections 3, we apply these new concepts to study spectrum allocation in fading channels. Section 4 concludes the paper and provides avenues for future research.

2 Resource Sharing model and fairness definitions

Consider n mobiles located at points x_1, x_2, \dots, x_n , respectively. We assume that the utility U_i of mobile i depends on its location x_i and on the amount of resources s_i it gets.

Let \mathbf{S} be the set of assignments; an assignment $s \in \mathbf{S}$ is a function from the vector x to a point in the n -dimensional simplex. Its i th component, $s_i(x)$ is the fraction of resource assigned to mobile i .

Definition 1. An assignment s is α -fair if it is a solution of

$$Z(x, s, \alpha) := \max_s \sum_i Z_i(x_i, s_i, \alpha) \text{ such that,} \\ \sum_i s_i = 1, s_i \geq 0 \quad \forall i = 1, \dots, n \quad (1)$$

$$\text{where, } Z_i(x_i, s_i, \alpha) := \frac{(U_i(x_i, s_i))^{1-\alpha}}{1-\alpha} \text{ for } \alpha \neq 1 \text{ and} \\ Z_i(x_i, s_i, \alpha) := \log(U_i(x_i, s_i)) \text{ for } \alpha = 1$$

We shall assume throughout that U_i is non-negative, strictly increasing and is concave in s_i . Then for any $\alpha > 0$, $Z_i(x_i, s_i, \alpha)$ is strictly concave in s_i . We conclude that $Z(x_i, s_i, \alpha)$ is strictly concave in s for any $\alpha > 0$ and therefore there is a unique solution $s^*(\alpha)$ to (2).

Definition 2. [1] We call $Z_i(s_i, \cdot, \alpha)$ the fairness utility of mobile i under s_i , and we call $Z(s, \cdot, \alpha)$ the instantaneous degree of α -fairness under s .

In applications, the state X will be random, so that the instantaneous amount of resource assigned by an α -fair allocation will also be a random variable. Thus, in addition to instantaneous fairness we shall be interested in the expected amount assigned by being fair at each instant.

Definition 3. We call $E[Z(s, X, \alpha)]$ the expected instantaneous degree of α -fairness under s .

In Section 2.1 we introduce the expected long-term fairness in which the expected amount of resource is assigned fairly.

Definition 4. We say that a utility is linear in the resource if it has the form:

$$U_i(x_i, s_i) := s_i q_i(x_i).$$

For example, consider transmission between a mobile source and a base station, and assume

(i) that the base station is in the origin ($x = 0$) but at a height of one unit, whereas all mobiles are on the ground and have height 0. Thus, the distance between the base station and a mobile located on the ground at point x is $\sqrt{1 + |x|^2}$.

(ii) that the Shannon capacity can be used to describe the utility. If the resource that is shared is the frequency then the utility has the linear form:

$$U(C, x) := Cq(x); \text{ with } q(x) = \log \left(1 + \frac{P(x^2 + 1)^{-\beta/2}}{\sigma^2} \right)$$

2.1 Fairness over time: Instantaneous Versus Long term α -fairness

Next we consider the case where $x_i(t)$, $i = 1, \dots, n$, may change in time.

Definition 5. We define an assignment to be instantaneous α -fair if at each time t each mobile is assigned a resource so as to be α -fair at that instant.

Consider the instantaneous α -fair allocation and assume that time is discrete. We thus compute the instantaneous α -fair assignment over a period of T slots as the assignment that maximizes (for $\alpha \neq 1$)

$$\sum_{i=1}^n \frac{(U_i(x_i(t), s_i(t)))^{1-\alpha}}{1-\alpha} \text{ for every } t = 1, \dots, T.$$

This is equivalent to maximizing

$$\sum_{t=1}^T \sum_{i=1}^n \frac{(U_i(x_i(t), s_i(t)))^{1-\alpha}}{1-\alpha}. \quad (2)$$

For $\alpha = 1$, we replace

$$\frac{(U_i(x_i(t), s_i(t)))^{1-\alpha}}{1-\alpha} \text{ by } \log[U_i(x_i(t), s_i(t))]$$

The optimization problem (2) corresponds to the α -fair assignment problem in which there are nT players instead of n players, where the utility of player $i = kn + j$ ($k = 0, \dots, T-1, j = 1, \dots, n$) is defined as

$$\mathcal{U}_i(x_i, s_i) = U_j(x_j(k+1), s_j(k+1)).$$

Remark 1. Thus the expected instantaneous fairness criterion in the stationary and ergodic case regards assignments at different time slots of the same player as if it were a different player at each time slot!

Note that when considering the proportional fair assignment, then the resulting assignment is the one that maximizes $\prod_{i=1}^n \prod_{t=1}^T U_i(x_i(t), s_i(t))$.

Definition 6. Assume that the state process $X(t)$ is stationary ergodic. Let λ_i be the stationary probability measure of $X(0)$. The long term α -fairness index of an assignment $s \in \mathbf{S}$ of a stationary process $X(t)$ is defined as

$$\bar{Z}_\lambda(s) := \sum_{i=1}^n \bar{Z}_\lambda^i(s); \text{ with } Z_\lambda^i(s) = \frac{\left(E_\lambda [U_i(X_i(0), s_i(X(0)))] \right)^{1-\alpha}}{1-\alpha}.$$

An assignment s is long-term α -fair if it maximizes $Z_\lambda(s)$ over $s \in \mathbf{S}$.

As we see, instead of attempting to have a fair assignment of the resources at every t , it is the expected utility in the stationary regime that one assigns fairly according to the long-term fairness. Under stationarity and ergodicity conditions on the process $X(t)$ this amounts in an instantaneous assignment of the resources in a way that the time average amount allocated to the users are α -fair.

2.2 Fairness over time: T -scale α -fairness

Next we define fairness concepts that are in between the instantaneous and the expected fairness. They are related to fairness over a time interval T . Either continuous time is considered or discrete time where time is slotted and each slot is considered to be of one time unit. Below, we shall understand the integral to mean summation when ever time is discrete.

Definition 7. The T -scale α -fairness index of $s \in \mathbf{S}$ is defined as

$$Z_T(s) := \sum_{i=1}^n Z_T^i(s); \text{ with } Z_T^i = \frac{\left[\frac{1}{T} \int_0^T U_i(X_i(t), s_i(X(t))) dt \right]^{1-\alpha}}{1-\alpha}.$$

The expected T -scale α -fairness index is its expectation. An assignment s is T -scale α -fair if it maximizes $Z_T(s)$ over $s \in \mathbf{S}$.

Definition 8. The T -scale expected α -fairness index of $s \in \mathbf{S}$ is defined as

$$Z_T(s) := \sum_{i=1}^n Z_T^i(s); \text{ with } Z_T^i = \frac{\left[\frac{1}{T} \int_0^T E[U_i(X_i(t), s_i(X(t)))] dt \right]^{1-\alpha}}{1-\alpha}$$

Assume that the state processes is stationary ergodic. Then for any assignment $s \in \mathbf{S}$ we would have by the Strong Law of Large Numbers:

$$\lim_{T \rightarrow \infty} \frac{1}{T} \int_0^T U_i(X_i(t), s_i(X(t))) dt = E_\lambda [U_i(X_i(0), s_i(X(0)))]$$

P-a.s. Hence, for every i and s , we have P-a.s.

$$\begin{aligned} \lim_{T \rightarrow \infty} Z_T^i(s) &= \lim_{T \rightarrow \infty} \frac{\left[\frac{1}{T} \int_0^T U_i(X_i(t), s_i(X(t))) dt \right]^{1-\alpha}}{1-\alpha} \\ &= \frac{(E_\lambda [U_i(X_i(0), s_i(X(0)))]^{1-\alpha}}{1-\alpha} = \bar{Z}_\lambda^i(s). \end{aligned}$$

Assume that U_i is bounded. Then Z_T^i is bounded uniformly in T . The bounded convergence then implies that

$$\lim_{T \rightarrow \infty} E[Z_T^i(s)] = \bar{Z}_\lambda^i(s). \quad (3)$$

Theorem 1. Assume that the convergence in (3) is uniform in s . Let $s^*(T)$ be the T -scale α fair assignment and let s^* be the long term α -fair assignment. Then the following holds:

- $s^* = \lim_{T \rightarrow \infty} S^*(T)$
- For any $\epsilon > 0$, s^* is an ϵ -optimal assignment for the T -scale criterion for all T large enough.
- For any $\epsilon > 0$, $s^*(T)$ is an ϵ -optimal assignment for the long term fairness for all T large enough.

Proof. According to [9], any accumulation point of $s^*(T)$ as $T \rightarrow \infty$ is an optimal solution to the problem of maximizing \bar{Z}_T over S . Due to the strict concavity of \bar{Z}_T in s it has a unique solution and it coincides with any accumulation point of $s^*(T)$. This implies the first statement of the theorem. The other statements follow from Appendices A and B in [9].

2.3 Fairness over different time scales: Multiscale fairness

We consider real time (RT) and non-real time (NRT) traffic. Resource allocation policy for RT traffic is *instantaneous-fair*, while for the NRT traffic, it is *expected-fair*. The available resources are divided amongst the RT and NRT traffic so as to guarantee a minimum quality of service (QoS) requirement for the RT traffic and to keep service time as short as possible for the NRT traffic.

The real time traffic would like the allocation to be instantaneously α -fair. For $\alpha > 0$, this guarantees that at any time it receives a strictly positive allocation.

The non-real time traffic does not need to receive at each instant a positive amount of allocation. It may prefer the resources to be assigned according to the T -scale α -fair assignment where T may be of the order of the duration of the connection. Moreover, different non real time applications may have different fairness requirements. For instance, bulk FTP transfer can prefer fairness over time scale longer than a time scale for some streaming application.

In order to be fair, we may assign part (say half) of the resource according to the instantaneous α -fairness and the rest of the resources according to the T -scale α -fairness. We thus combine fairness over different time scales.

We may now ask how to choose what part of the resource would be split according to the instantaneous assignment and what part according to the T -scale assignment. We propose to determine this part using the same α -fair criterion.

Specifically we define the multiscale fairness as follows:

Definition 9. *The multiscale α -fairness index of $s \in \mathbf{S}$ is defined as*

$$Z_{T_1, \dots, T_n}(s) := \sum_{i=1}^n Z_{T_i}^i(s); \text{ with } Z_{T_i}^i = \frac{\left[\frac{1}{T_i} \int_0^{T_i} U_i(X_i(t), s_i(X(t))) dt \right]^{1-\alpha}}{1-\alpha}$$

The expected multiscale α -fairness index is its expectation. An assignment s is multiscale α -fair if it maximizes $Z_{T_1, \dots, T_n}(s)$ over $s \in \mathbf{S}$. We also say that multiscale α -fair assignment is (T_1, \dots, T_n) -scale fair assignment.

3 Application to spectrum allocation in fading channels

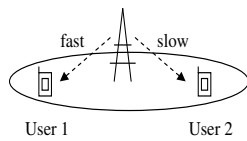


Fig. 1. Spectrum allocation in random fading channels

We consider a fast-changing and a slowly-changing user (Fig. 1), whose channels are modeled by the Gilbert model. The users can be either in a good or in a bad state. The dynamics of the users is described by a Markov chain $\{Y_i(t)\}_{t=0,1,\dots}$ $i = 1, 2$, with the transition matrix and stationary distribution as:

$$P_i = \begin{bmatrix} 1 - \epsilon_i \alpha_i & \alpha_i \\ \beta_i & 1 - \epsilon_i \beta_i \end{bmatrix}; \quad \pi_i = \begin{bmatrix} \frac{\beta_i}{\alpha_i + \beta_i} & \frac{\alpha_i}{\alpha_i + \beta_i} \end{bmatrix}.$$

Let $\epsilon_1 = 1$ and $\epsilon_2 = \epsilon$. Note that the parameter ϵ does not have an effect on the stationary distribution, but, it influences for how long the slowly-changing user stays in some state. The smaller ϵ , the more seldom the user changes the states.

We assume that state 1 is a bad state and state 2 is a good state. Let h_{ij} represent the channel gain coefficient of user i in channel state j . The utility (achievable throughput via Shannon capacity) of user i in state j is given by

$$U_{ij} = s_{ij} \log_2 \left(1 + \frac{|h_{ij}|^2 p_i}{\sigma^2} \right)$$

where s_{ij}, p_i is the resource allocation and power that corresponding to user i .

First, we would like to analyze T -scale fairness and to see the effect of the time scale on the resource allocation. Specifically, we consider the following optimization criterion

$$\sum_{i=1}^2 \frac{1}{1-\alpha} \left[\frac{1}{T} \sum_{t=0}^T U_i(t) \right]^{1-\alpha} \rightarrow \max_{s_1, s_2} \quad (4)$$

with $U_i(t) = s_i(t) q_{i, Y_i(t)}$ and $s_1(t) + s_2(t) = 1$.

Let us consider several options for the time horizon T :

Instantaneous fairness. If we take $T = 1$ we obtain the instantaneous fairness. Namely, the criterion (4) takes the form

$$\frac{1}{1-\alpha} [U_1^{1-\alpha}(0) + U_2^{1-\alpha}(0)] \rightarrow \max_{s_1, s_2}$$

The solution of the above optimization problem is given by

$$s_i(0) = \frac{q_{i, Y_i(0)}^{(1-\alpha)/\alpha}}{q_{1, Y_1(0)}^{(1-\alpha)/\alpha} + q_{2, Y_2(0)}^{(1-\alpha)/\alpha}}$$

This allocation results in the following expected throughputs

$$\theta_1 = \sum_{i,j} \frac{q_{1,i}^{1/\alpha}}{q_{1,i}^{(1-\alpha)/\alpha} + q_{2,j}^{(1-\alpha)/\alpha}} \pi_{1,i} \pi_{2,j}, \theta_2 = \sum_{i,j} \frac{q_{2,j}^{1/\alpha}}{q_{1,i}^{(1-\alpha)/\alpha} + q_{2,j}^{(1-\alpha)/\alpha}} \pi_{1,i} \pi_{2,j}. \quad (5)$$

Mid-term fairness. Let us take the time horizon as a function of the underlying dynamics time parameter ϵ , that is $T = T(\epsilon)$, satisfying the following conditions: (a) $T(\epsilon) \rightarrow \infty$ and (b) $T(\epsilon)\epsilon \rightarrow 0$. The condition (a) ensures that

$$\frac{1}{T(\epsilon)} \sum_{t=0}^{T(\epsilon)} 1\{Y_1(t) = i\} \rightarrow \pi_{1,i}, \quad \text{as } \epsilon \rightarrow 0,$$

and the condition (b) ensures that

$$\frac{1}{T(\epsilon)} \sum_{t=0}^{T(\epsilon)} 1\{Y_2(t) = i\} \rightarrow \delta_{Y_2(0), i}, \quad \text{as } \epsilon \rightarrow 0.$$

This follows from the theory of Markov chains with multiple time scales (see e.g., [10]). It turns out to be convenient to take the following notation for the resource allocation: We denote by $s(t)$ the allocation for the fast-changing user and by $1-s(t)$ the resource allocation for the slowly-changing user. Thus, we have $s_1(t) = s(t)$ and $s_2(t) = 1-s(t)$. We denote by $\bar{s}_{i,j} = E[s(t)|Y_1(t) = i, Y_2(t) = j]$. We note that since the fast-changing user achieves stationarity when $T(\epsilon) \rightarrow \infty$ we are able to solve (4) in stationary strategies. Then, the criterion (4) takes the form

$$\frac{1}{1-\alpha} \left[(\pi_{1,1}q_{1,1}\bar{s}_{1,Y_2(0)} + \pi_{1,2}q_{1,2}\bar{s}_{2,Y_2(0)})^{1-\alpha} + ((1-\pi_{1,1}\bar{s}_{1,Y_2(0)} - \pi_{1,2}\bar{s}_{2,Y_2(0)})q_{2,Y_2(0)})^{1-\alpha} \right] \rightarrow \max_{\bar{s}_{1,Y_2(0)}, \bar{s}_{2,Y_2(0)}}$$

The above nonlinear optimization problem can be solved numerically. The expected throughputs in the mid-term fairness case are given by

$$\begin{aligned} \theta_1 &= (\pi_{1,1}q_{1,1}\bar{s}_{1,1} + \pi_{1,2}q_{1,2}\bar{s}_{2,1})\pi_{2,1} + (\pi_{1,1}q_{1,1}\bar{s}_{1,2} + \pi_{1,2}q_{1,2}\bar{s}_{2,2})\pi_{2,2}, \\ \theta_2 &= (1 - \pi_{1,1}\bar{s}_{1,1} - \pi_{1,2}\bar{s}_{2,1})q_{2,1}\pi_{2,1} + (1 - \pi_{1,1}\bar{s}_{1,2} - \pi_{1,2}\bar{s}_{2,2})q_{2,2}\pi_{2,2}. \end{aligned} \quad (6)$$

Long-term fairness. In the case of long-term fairness we set $T = \infty$ which results in the following criterion

$$\frac{1}{1-\alpha} [E[U_1]^{1-\alpha} + E[U_2]^{1-\alpha}] \rightarrow \max_{s_1, s_2}$$

Due to stationarity, we can solve the above optimization problem over sequences in stationary strategies. Namely, we have the following optimization problem

$$\begin{aligned} \frac{1}{1-\alpha} \left[((\pi_{1,1}\pi_{2,1}\bar{s}_{1,1} + \pi_{1,1}\pi_{2,2}\bar{s}_{1,2})q_{1,1} + (\pi_{1,2}\pi_{2,1}\bar{s}_{2,1} + \pi_{1,2}\pi_{2,2}\bar{s}_{2,2})q_{1,2})^{1-\alpha} \right. \\ \left. + ((\pi_{2,1} - \pi_{1,1}\pi_{2,1}\bar{s}_{1,1} - \pi_{1,2}\pi_{2,1}\bar{s}_{2,1})q_{2,1} \right. \\ \left. + (\pi_{2,2} - \pi_{1,1}\pi_{2,1}\bar{s}_{1,1} - \pi_{1,2}\pi_{2,2}\bar{s}_{2,2})q_{2,2})^{1-\alpha} \right] \rightarrow \max_{\bar{s}_{1,1}, \bar{s}_{1,2}, \bar{s}_{2,1}, \bar{s}_{2,2}} \end{aligned}$$

The expected throughputs in the long-term fairness case are given by

$$\begin{aligned} \theta_1 &= (\pi_{1,1}\pi_{2,1}\bar{s}_{1,1} + \pi_{1,1}\pi_{2,2}\bar{s}_{1,2})q_{1,1} + (\pi_{1,2}\pi_{2,1}\bar{s}_{2,1} + \pi_{1,2}\pi_{2,2}\bar{s}_{2,2})q_{1,2} \\ \theta_2 &= (\pi_{2,1} - \pi_{1,1}\pi_{2,1}\bar{s}_{1,1} - \pi_{1,2}\pi_{2,1}\bar{s}_{2,1})q_{2,1} \\ &\quad + (\pi_{2,2} - \pi_{1,1}\pi_{2,1}\bar{s}_{1,1} - \pi_{1,2}\pi_{2,2}\bar{s}_{2,2})q_{2,2} \end{aligned} \quad (7)$$

Let us also consider the expected instantaneous fairness which is given by the criterion

$$\frac{1}{1-\alpha} [E[U_1^{1-\alpha}(t)] + E[U_2^{1-\alpha}(t)]] \rightarrow \max_{s_1, s_2}$$

which is equivalent to

$$\frac{1}{1-\alpha} \left[\sum_{ij} \pi_{1,i}\pi_{2,j} \int_0^1 (sq_{1,i})^{1-\alpha} dF_{ij}(s) \right]$$

$$+ \sum_{ij} \pi_{1,i} \pi_{2,j} \int_0^1 ((1-s)q_{2,j})^{1-\alpha} dF_{ij}(s) \Big] \rightarrow \max_{F_{ij}}$$

where $F_{ij}(s)$ is the distribution for $s(t)$ conditioned on the event $\{Y_1(t) = i, Y_2(t) = j\}$. The above criterion is maximized by

$$F_{ij}(s) = \begin{cases} 0, & \text{if } s < q_{1,i}^{(1-\alpha)/\alpha} / (q_{1,i}^{(1-\alpha)/\alpha} + q_{2,j}^{(1-\alpha)/\alpha}), \\ 1, & \text{if } s \geq q_{1,i}^{(1-\alpha)/\alpha} / (q_{1,i}^{(1-\alpha)/\alpha} + q_{2,j}^{(1-\alpha)/\alpha}). \end{cases}$$

Thus, we can see that the expected instantaneous fairness criterion is equivalent to instantaneous fairness.

Multiscale fairness: Next, let us consider multiscale fairness over time. Specifically, (T_1, T_2) -scale fairness is defined by the following criterion

$$\frac{1}{1-\alpha} \left[\left(\frac{1}{T_1} \sum_{t=0}^{T_1} U_1(t) \right)^{1-\alpha} + \left(\frac{1}{T_2} \sum_{t=0}^{T_2} U_2(t) \right)^{1-\alpha} \right] \rightarrow \max_{s_1, s_2}$$

In this particular example, there are 6 possible combinations of different time scales. It turns out that in this example only the $(1, \infty)$ -scale fairness gives a new resource allocation. The other combinations of time scales reduce to some T -scale fairness. Thus, let us first consider the multiscale fairness when we apply instantaneous fairness to the fast-changing user and long-term fairness to the slowly-changing user. The $(1, \infty)$ -scale fairness corresponds to the following optimization criterion

$$\frac{1}{1-\alpha} [U_1(0)^{1-\alpha} + E[U_2(t)]^{1-\alpha}] \rightarrow \max_{s_1, s_2}$$

which is equivalent to

$$\frac{1}{1-\alpha} [(q_{1,Y_1(0)}(\bar{s}_{Y_1(0),1}\pi_{2,1} + \bar{s}_{Y_1(0),2}\pi_{2,2}))^{1-\alpha} + (q_{2,1}(1 - \bar{s}_{Y_1(0),1})\pi_{2,1} + q_{22}(1 - \bar{s}_{Y_1(0),2})\pi_{2,2})^{1-\alpha}] \rightarrow \max_{\bar{s}_{Y_1(0),1}, \bar{s}_{Y_1(0),2}}$$

The expected throughputs in the $(1, \infty)$ -scale fairness case are given by

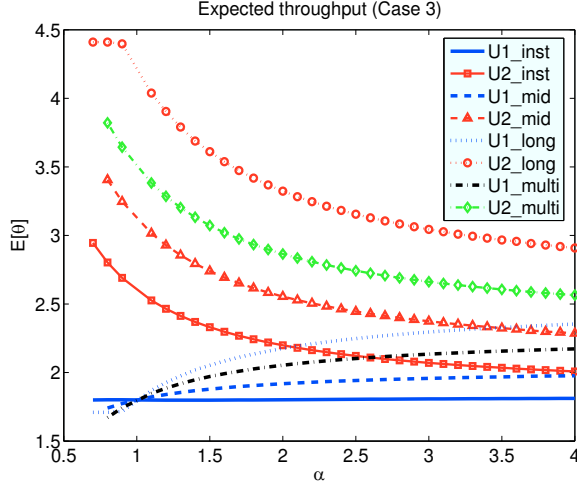
$$\begin{aligned} \theta_1 &= (q_{1,1}(\bar{s}_{1,1}\pi_{1,1}\pi_{2,1} + \bar{s}_{1,2}\pi_{1,1}\pi_{2,2}) + q_{1,2}(\bar{s}_{2,1}\pi_{1,2}\pi_{2,1} + \bar{s}_{2,2}\pi_{1,2}\pi_{2,2})), \\ \theta_2 &= (q_{2,1}(1 - \bar{s}_{1,1})\pi_{2,1} + q_{22}(1 - \bar{s}_{1,2})\pi_{22})\pi_{1,1} \\ &\quad + (q_{2,1}(1 - \bar{s}_{2,1})\pi_{2,1} + q_{22}(1 - \bar{s}_{2,2})\pi_{22})\pi_{1,2}. \end{aligned}$$

As we have mentioned above, the other combinations of time scales reduce to some T -scale fairness. In particular, $(1, T(\epsilon))$ -fairness reduces to the instantaneous fairness, $(T(\epsilon), \infty)$ -fairness reduces to long-term fairness, and $(T(\epsilon), 1)$ -, $(\infty, 1)$ - and $(\infty, T(\epsilon))$ -fairness all reduce to mid-term fairness.

Let us consider a numerical example. The parameters are given in Table 1. We consider three typical cases. The first case corresponds to the symmetric

Table 1. Case 1,2 & 3: Shannon capacity (q)/probability(π)

	Case-1		Case-2		Case-3	
	state-1 (bad)	state-2 (good)	state-1 (bad)	state-2 (good)	state-1 (bad)	state-2 (good)
User-1	2/0.2	8/0.8	3/0.1	9/0.9	3/0.9	9/0.1
User-2	2/0.2	8/0.8	1/0.3	7/0.7	1/0.3	7/0.7

**Fig. 2.** Throughput(θ) as a function of α for instantaneous, mid-term, long-term and $(1, \infty)$ -scale fairness criteria (Case 3).

scenario. In the second case, the fast-changing user has in general better channel conditions. In the third scenario the slowly-changing user (user 2) is more often in the good channel state than the fast-changing user (user 1).

We plot the expected throughput of the mobiles for various fairness criteria for case-3 in Fig. 2. Plots and explanation for case-1 and case-2 are provided in [11].

In the third scenario, the second user always gets better share in terms of throughput. This is expected as the second user spends on average more time in a channel with good state and the long or short term throughput is the principal component of the optimization criteria. It is natural that long term fairness gives the best efficiency for both types of users. However, we note that the $(1, \infty)$ -scale fairness provides better control in term of fairness. The $(1, \infty)$ -scale fairness based allocation provides the second best efficiency after the long term fairness based allocation. Thus, we conclude that multiscale fairness provides good sensitivity to the variation of the fairness parameter and at the same time good performance in expected throughput. Below we shall see that the multiscale fairness has another good property with respect to variance of the throughput. It is curious to observe that in this example the instantaneous fairness does not really help fast-changing user.

Coefficient of variation: We compute the coefficient of variation for short-term, mid-term, long-term and multiscale fairness. For this, we first compute

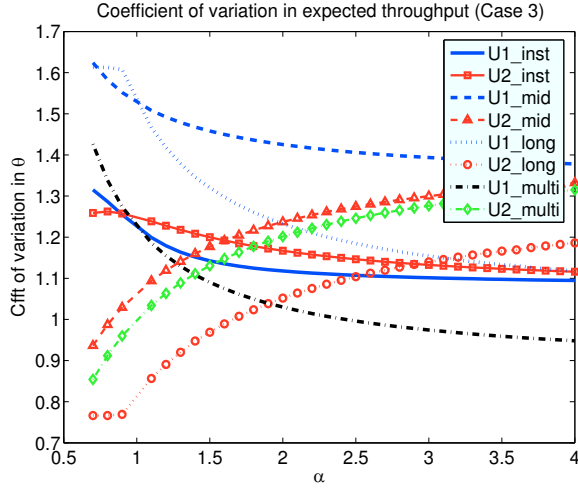


Fig. 3. Coefficient of variation in expected throughput as a function of α for instantaneous, mid-term, long-term and $(1, \infty)$ -scale fairness criteria (Case 3).

the second moment of the throughput and then find the ratio of the standard deviation to its mean. For any user i , the coefficient of variation $\Gamma_i = \frac{\sqrt{\mathbf{E}[\theta_i^2]}}{\mathbf{E}[\theta_i]}$.

In Fig 3, we plot the coefficient of variation in throughput for the various fairness criteria considered above. It is very interesting to observe that except the $(1, \infty)$ -scale fairness criterion all the other fairness criteria behave similarly with respect to the coefficient of variation. Only in the case of $(1, \infty)$ -scale fairness the coefficient of variation decreases for short-term fairness oriented user. This is a very desirable property of the multiscale fairness as a short-term fairness oriented user is typically a user with a delay sensitive application. Also, here as in the case of the throughput, the multiscale fairness provides at the same time good fairness and efficiency, now efficiency in terms of overall variance.

4 Conclusion and Future Research

We have introduced T -scale fairness and multiscale fairness. The notion of T -scale fairness allows one to address in a flexible manner requirements of emerging applications (like You Tube) which demand quality of service requirement between strict real time traffic and best effort traffic. The notion of multiscale fairness allows one to use a single optimization criterion for resource allocation when different applications are present in the network. We have compared the new fairness notions with previously known criteria of instantaneous and long-term fairness criteria. We have illustrated the new notions by their application in wireless networks. Specifically, we have considered spectrum allocation when users with different dynamics are present in the system. We have demonstrated that the multiscale fairness provides a versatile framework for resource allocation. In the near future we plan to investigate in detail how multiscale fairness criterion allocates resources when a number of applications with different QoS requirements are present in the network. It is also interesting to investigate T -scale fairness in the non-stationary regime.

Acknowledgement

This work was done in the framework of the INRIA and Alcatel-Lucent Bell Labs Joint Research Lab on Self Organized Networks and the Ecoscells project.

References

1. J. Mo and J. Walrand, "Fair End-to-end Window-based Congestion Control", *IEEE/ACM Transactions on Networking*, vol. 8, no. 5, pp. 556-567, Oct. 2000.
2. T. Lan, D. Kao, M. Chiang and A. Sabharwal, "An Axiomatic Theory of Fairness for Resource Allocation", In Proc. of IEEE INFOCOM'10, March 2010.
3. E. Altman, K. Avrachenkov and B. J. Prabhu, "Fairness in MIMD Congestion Control Algorithms", In Proc. of IEEE INFOCOM'05, March, 2005, Miami, USA.
4. E. Altman, K. Avrachenkov and A. Garnaev, "Generalized Alpha-Fair Resource Allocation in Wireless Networks", In Proc. of 47th IEEE Conference on Decision and Control, Cancun, Mexico, Dec. 9-11, 2008.
5. E. Altman, K. Avrachenkov and A. Garnaev, "Alpha-Fair Resource Allocation under Incomplete Information and Presence of a Jammer", In Proc. of NETCOOP'09, also in LNCS, vol. 5894, pp. 219-233, 2009.
6. M. Bredel and M. Fidler, "Understanding Fairness and its Impact on Quality of Service in IEEE 802.11", Arxive, 2008.
7. V. Ramaiyan, A. Kumar and E. Altman, "Fixed Point Analysis of Single Cell IEEE 802.11e WLANs: Uniqueness and Multistability", *IEEE/ACM Transactions on Networking*, vol. 16, no. 5, pp. 1080-1093, Oct. 2008.
8. F.P. Kelly, A.K. Maulloo and D.K.H. Tan, "Rate control for communication networks: shadow prices, proportional fairness and stability", *Journal of the Operational Research Society*, vol. 49, no. 3, pp. 237-252, 1998.
9. M. Tibball, A. Lombardi, O. Pourtallier and E. Altman, "Continuity of optimal values and solutions for control of Markov chains with constraints", *SIAM J. Control and Optimization*, vol. 38, no. 4, pp. 1204-1222, 2000.
10. J. Filar, H.A. Krieger and Z. Syed, "Cesaro limits of analytically perturbed stochastic matrices", *Lin. Alg. Appl.*, vol. 353, pp. 227-243, 2002.
11. E. Altman, K. Avrachenkov, S. Ramanath, "Multiscale Fairness and its Application to Dynamic Resource Allocation in Wireless Networks", INRIA research report number RR-7783. Available at <http://hal.inria.fr/inria-00515430/en/>