Fairness in Non-convex Systems

Corinne Touati, Hisao Kameda, Atsushi Inoie University of Tsukuba, Japan {corinne, kameda, inoie}@osdp.cs.tsukuba.ac.jp

Abstract—In general, the set of users utilities is bounded because of the limitation of resources. There may exist many Pareto optimal points in the set of users utilities. For selecting a Pareto optimum point, a family of fairness criteria, that contains the max-min fairness and a parameterized family of fairness (by Mo and Walrand), has been proposed and examined in some concrete networking contexts that result in specific convex utility sets. We newly examine general compact (closed and bounded) utility sets which include the specific utility sets as special cases. We first prove that each of the family of fairness criteria gives a unique fair (Pareto optimum) point if the utility set is convex. We find, however, counter-examples where each of the family of fairness criteria gives multiple fair points if the utility set is not convex. We propose an extention of the family of fairness criteria such that each of them gives only a unique fair point regardless of whether the utility set is convex or not, to which we give proofs. By using a specific load balancing model, we illustrate the counter-examples and how each criterion of our extended fair family gives a unique fair point.

Index Terms— Fairness, convex and non-convex systems, Pareto optimality, max-min fairness, proportional fairness, mathematical programming / optimization, load balancing, distributed computer systems.

I. INTRODUCTION

In a multi-user system, it is generally not possible to optimize simultaneously the performance of every user. Instead, we consider the Pareto optima that are the points of a system such that there exists no other state where all users have simultaneously better benefits. Thus, we may not have any absolute preference among Pareto optima. The weighted sum optimization (which maximizes the weighted sum of the users' utilities with different weights) is Pareto optimal. The well-known overall optimization [18], [9], [10], [7] maximizes the total utility over all users, and can be seen as a particular case of weighted sum optimization.

There exist a priori innumerably many Pareto-optimal situations. The choice of the one to achieve can be controversial among users and a selection criterion is fairness.

In networking contexts, various fairness concepts have been already proposed [2], [13], [8], [14], [15]. Among

them are the well-known max-min fairness (adopted by the ABR - Available Bit Rate service of the ATM - Asynchronous Transfer Mode) and the proportional fairness (achieved by the Vegas version of the TCP protocol [11]). Before them, mathematicians and economists have developed several fairness criteria, one of the oldest being proposed by Nash [16]. It is commonly referred to as the Nash Bargaining Solution and coincides with the proportional fairness introduced by Kelly when some simplifying assumptions are made [19], [20]. In the context of congestion control, Mo and Walrand [15] introduced a parameterized formulation of fairness, which was later extended by considering concave and increasing utility functions in the context of bandwidth allocation [19], [20]. The utility functions represent the satisfaction perceived by the user (typically function of the delay or of the capacity). The interest of this family is that it covers the max-min fair solution, the Nash Bargaining Solution and the overall optimization for particular values of the parameter. It hence enables the provider of a system to finely tune the solution according to its trade-off choice between fairness and global efficiency. However, the formulation and study constrained itself to the case where the utility functions are increasing and concave, which leads to convex utility sets.

In reality, there exist many systems where the set of achievable utilities (also called the "bargaining set" in cooperative game theory) is not convex. In such systems, defining fairness becomes more challenging, which is the main question addressed by this paper. In game theory, some recent research raised the question of how to define fairness in non-convex systems [5], [21]. These studies focus, however, on an axiomatic definition while we are interested here in an optimization formulation.

Additionally, load balancing systems are of great interest since they are a means of efficiently sharing resources among users. Yet, although some research on how to implement fairness has been conducted [17], there is still a lack of theoretical formulation of fairness in this field, which motivates our research. As a special case, we hence consider a model consisting of two servers shared by two users or two sets of users. As the utilities of users we consider the inverse of the delays (or equivalently the mean response time) that they observe. For this system, the utility sets are not convex.

The contributions of this paper are of two kinds. On the one hand, rather than studying fairness in a specific networking environment, we choose to examine it in a general theoretical framework. In particular, we extend the definition of max-min fairness and of the Mo and Walrand fair family to any system. For this, we first show that in convex systems a unique max-min fair solution exists while in general utility sets the max-min fair criterion leads to a finite set of equivalent points. These points are equivalent in the sense that they only differ from the user's labels. Hence, in each system, a "tie-breaking" rule needs to be agreed upon. We also show that, when the utility set is not convex, the existing formulation of the fair family has an infinity of solution points. We then propose a new definition of the family that assures the uniqueness of the solution. The new family corresponds to the previously defined one when the set is convex and we prove that it always gives Pareto optimal solutions. In addition to these results, we show that, when implementing the fair family, one has to make sure of using utility and not costs sets.

On the other hand, we apply the fair family in a load balancing system. We show that in spite of the simplicity of the model, this system can serve as a support to the study of fairness in non-convex systems. We study the behavior of the fairness objectives through numerical experiments. In particular, we observe that the fair points can be located on convex and on concave segments of the Pareto border and that they rapidly converge to the max-min fair solution.

The rest of this paper is organized as follows. We first present the max-min fairness and its properties in non-convex systems (Section II), and then present the fair family also both in convex and non-convex systems (Section III). We then point out that, for any given system, one must carefully define the utilities (Section IV). Section V introduce the load balancing model while Section VI presents shows some numerical results. The last section finally concludes the article.

II. MAX-MIN FAIRNESS

We consider a system in which we want to share efficiently and fairly a given resource among n users. We treat in this section the well known "max-min" fairness. In the next section, we focus on a general fair family that covers, in particular, the max-min criterion.

Max-min fairness has been widely studied in special cases of convex and compact utility systems (see for

instance [2], [3], [4], [8], [12], [13], [14], [22]). We propose here to study the criterion on a general framework covering in any compact utility sets (convex or not).

We assume that to each n-dimensional allocation vector corresponds a utility vector $\boldsymbol{u} = (u_1, \ldots, u_n)$ that represents the relative satisfaction perceived by the users with this allocation. We denote by $\mathcal{U} \subset \mathbb{R}^n_+ \setminus \{0\}$ the achievable utilities, that is, the achievable utilities are strictly positive.

In the following, we suppose that \mathcal{U} is compact (for real sets, this amounts to suppose that it is closed and bounded).

A. Definition

1) Pareto optimality: Let us first recall the definition of Pareto optimality over a set U:

$$\mathbf{\Pi} = \left\{ \boldsymbol{u} \in \mathcal{U} \middle| \begin{array}{c} \forall \boldsymbol{v} \in \mathcal{U}, \exists i, v_i > u_i \Rightarrow \\ \exists j, v_j < u_j, i, j \in \{1, ..., n\} \end{array} \right\}.$$
(1)

The Pareto set is therefore included in the upper right border of the utility set, which is why we also refer to it as the *Pareto border*. Studying the Pareto set is of interest since it represents the points that cannot be dominated in terms of utilities. Therefore, fair criteria should naturally reach Pareto efficient equilibria.

2) Max-Min fairness: Max-min fairness is, roughly speaking, the allocation in which the minimal utility is maximized, in a recursive manner. More formally, let us define the associated permutation of a point:

Definition 1: For any point u of \mathbb{R}^n , we define an ordering permutation p a permutation of $\{1, ..., n\}$ (that is to say a bijection from $\{1, ..., n\}$ to itself) such that $u_{p(1)} \leq u_{p(2)} \leq ... \leq u_{p(n)}$. The vector $\overrightarrow{u} = (\overrightarrow{u_i})_i = (u_{p(i)})_i$ is the ordered representation of u.

Definition 2: We define a *partial order* relation \prec by :

$$oldsymbol{u} \prec oldsymbol{v} \Leftrightarrow \exists k \in \{1, ..., n\}, \left\{ egin{array}{c} orall i < k, \overrightarrow{u_i} = \overrightarrow{v_i} \ \overrightarrow{u_k} < \overrightarrow{v_k}. \end{array}
ight.$$

Following [3], we say that $u \in U$ is *max-min fairer* than v if $v \prec u$.

Let us consider for instance two points of coordinates u = (7,3,6,3) and v = (5,4,6,3). Then $\overrightarrow{u} = (3,3,6,7)$ and $\overrightarrow{v} = (3,4,5,6)$. We compare \overrightarrow{u} and \overrightarrow{v} coordinate by coordinate, starting by their smaller element. As $\overrightarrow{u_1} = \overrightarrow{v_1}$ and $\overrightarrow{u_2} < \overrightarrow{v_2}$ then $u \prec v$ regardless of the values of the bigger elements. We then extend the definition of the max-min fair solution of [3] to any system (convex or not):

Definition 3: A max-min fair point is maximal for the relation \prec .

We show that, similarly to what is known in special networking convex cases [2], [3], [4], [8], [12], [13], the max-min fair points are always efficient:

Lemma 1: A max-min fair point is Pareto optimal.

Proof: Suppose that \boldsymbol{u} is not Pareto optimal. Then $\exists \boldsymbol{v} \in \mathcal{U}, \forall i, u_i \leq v_i$, and $\exists k, u_k < v_k$, and hence $\boldsymbol{u} \prec \boldsymbol{v}$.

We define the following class of equivalence:

Definition 4: Two points u and v are said utilitysymmetrical, and we write $u \sim v$ if $\overrightarrow{u} = \overrightarrow{v}$ (or equivalently if neither $u \prec v$ nor $v \prec u$).

Similarly, we say that a utility set \mathcal{U} is *symmetrical* if $\forall u \in \mathcal{U}$, and $\forall p$ permutation of $\{1, ..., n\}$ the point v defined by $\forall i \in \{1, ..., n\}, v_i = u_{p(i)}$ is also an element of \mathcal{U} .

B. Convex utility sets

When supposing the utility sets \mathcal{U} to be convex, we have the following properties:

Proposition 1: If the utility set \mathcal{U} is convex, a unique max-min fair point exists.

Proof: Let u and v be two max-min fair points, $u \neq v$. If \mathcal{U} is convex, we consider $z \in \mathcal{U}$ such that $z = \frac{u+v}{2}$. As $u \neq v$, let k such that $u_k = \min\{u_i | u_i \neq v_i\}$. If $u_k < v_k$ then $z_k > u_k$ and $u \prec z$ and u is not max-min fair. Else, if $\exists l, v_l < v_k$ then $v \prec z$. If not, then $v_k < z_k$ and hence $v \prec z$ and v is not max-min fair.

Proposition 2: If the utility set \mathcal{U} is convex and symmetrical, then the max-min fair point u^* satisfies: u_2 - $u_1^* = u_2^* = \dots = u_n^*$.

Proof: Suppose that \boldsymbol{u} is max-min fair and that $u_1 \exists i, j, i \neq j$, such that $u_i < u_j$. Define $\boldsymbol{v} \sim \boldsymbol{u}$ by $v_i = u_j$, $v_j = u_i$ and $\forall k \notin \{i, j\}, v_k = u_k$ and $\boldsymbol{z} = \frac{\boldsymbol{u} + \boldsymbol{v}}{2}$. Then $\forall k \notin \{i, j\}, z_k = u_k$ and $z_i > u_i$. Hence $\boldsymbol{u} \prec \boldsymbol{z}$.

These results are similar to the ones known in the special cases of convex sets [2], [3], [4], [8], [12], [13], [14], [15], [19], [20].

C. General case

In this section, we do not assume the convexity of the utility set.

Proposition 3: Let \mathcal{U} be a compact and non-empty utility set. If u is a max-min fair point, then v is max-min fair if and only if $u \sim v$.

Proof: Let u and v be max-min fair. Then $v \not\prec u$ and $u \not\prec v$ and hence $\overrightarrow{u} = \overrightarrow{v}$.

Conversely, let u be max-min fair and v such that $u \sim v$. Let $z \in U$. Then either $z \prec u$ (which implies $z \prec v$) or $z \sim u$ (which leads to $z \sim v$). Hence v is max-min fair.

Hence, there can exists up to n! max-min fair point in a non-convex system. This result is illustrated in the case n = 2 by the points u and v in Fig. 1.

Let us recall that fair criteria were originally defined, in the context of game theory, as the unique solution satisfying a set of axioms among which is the symmetry axiom, stating that the solution does not depend on specific labels [5]. The only difference between two utility-symmetrical points is the mapping of the solution to the identity of the users. It has therefore no impact on the total performance of the system or the fairness of the solution. Utility-symmetrical points are, hence, from the fair and optimal point of view, identical. Hence, on a system to system basis, a "tie-breaking" rule as to be decided upon, which is not needed when restricting \mathcal{U} to be convex.

Remark 1: We can observe an interesting property. In the case of convex utility functions, the max-min fair solution is, roughly speaking, the Pareto optimal point for which the utilities are as close as possible, minimizing the distance of the Pareto border to the line of equal utilities $(u_1 = u_2 = ... = u_n)$. This property does not hold with non-convex utility sets. Let us consider for instance the point $a = (a_1, a_2)$ and $b = (b_1, b_2)$ of the Fig. 2, in a two users case. We see that $b_2 = a_1$ but $b_1 > a_2$ and so $a \prec b$ (from Def. 2). Hence, *b* is max-min fairer than *a* although it is located further from the line $u_1 = u_2$.

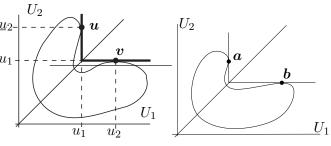


Fig. 1. Two points can be Fig. 2. Max-min point *b* has equally max-min fair. less homogeneous utilities than some other Pareto optima.

D. Summary of the section

In this section, we have extended the definition of max-min fairness to any utility set. We have shown that when the utility set is convex and compact, a unique max-min fair point exists. When the set is not convex, there exists a set of equivalent max-min fair points. These points are equivalent in the sense that they correspond to a re-ordering of the users' labels. Hence, on any system, a "tie-breaking rule" needs to be agreed upon. In Section V, we provide an example of such rule.

III. GENERAL FAIR FAMILY

In networking contexts, most of the research of fair equilibria has been conducted in the case of resource sharing with increasing and concave utility functions, which leads to convex utility sets. Recently, in the context of congestion control, Mo and Walrand [15] proposed a general and simple uniform description (parameterized by parameter α) of a wide family of fair criteria, including in particular proportional fairness and maxmin fairness. Touati et al. [19], [20] then generalized it in order to express it in term of utility rather than on the resource itself. Yet, the previous formulation restrained itself to the case where the utility functions are increasing and convex. In this section, we propose to extend the fair family to any system.

A. Definition

Consider the solution of $\max_{u \in \mathcal{U}} F_{\alpha}(u)$ where

$$F_{\alpha}(\boldsymbol{u}) = \begin{cases} \frac{1}{1-\alpha} \sum_{i} \left(\beta_{i} u_{i}\right)^{1-\alpha} \text{if } \alpha \geq 0, \alpha \neq 1, \\ \sum_{i} \log\left(u_{i}\right) & \text{if } \alpha = 1. \end{cases}$$
(2)

In this formulation, α is called the fairness parameter and $\beta_i > 0, i \in \{1, ..., n\}$ are the weights associated to the users' utility functions. The case $\alpha = 0$ corresponds to the weighted sum of the utilities of the users. When $\alpha \rightarrow 1$ the criterion converges to a Nash bargaining solution or proportional fair point.

In the following, and without loss of generality, we suppose that $\forall i, \beta_i = 1$. (This amounts in considering the utilities $v \in \mathcal{V}$ with $\forall i, v_i = \beta_i u_i$. Obviously, the set \mathcal{V} is convex if and only if \mathcal{U} is convex.)

Our study rely on the following lemmas:

Lemma 2: Let $\mathcal{V} \subset \mathbb{R}^n_+$ be a compact and non-empty and G a strictly concave and continuous function from \mathcal{V} to \mathbb{R} . Then, there exists $\tilde{G} \in \mathbb{R}$ and $\mathcal{S}_{\tilde{G}} \subset \mathbb{R}^n$ such that

$$\tilde{G} = \max_{\boldsymbol{v} \in \mathcal{V}} G(\boldsymbol{v}) \text{ and } \mathcal{S}_{\tilde{G}} = \{ \boldsymbol{v} \in \mathcal{V} | G(\boldsymbol{v}) = \tilde{G} \} \neq \emptyset.$$
(3)

Moreover, if G is increasing with respect to each of its variables, then the points of $S_{\tilde{G}}$ are Pareto optimal with respect to the set \mathcal{V} .

Proof: As \mathcal{V} is compact and G is continuous, then $G(\mathcal{V})$ is compact (see, e.g. [23]), which gives Eq. (3).

Suppose that $u \in S_{\tilde{G}}$ is not Pareto optimal. Then, there exists $v \in \mathcal{V}$ such that $\forall i \in \{1, ..., n\}, v_i \geq u_i$, and that $\exists j \in \{1, ..., n\}, v_j > u_j$. Then $G(v_1, ..., v_n) \geq$ $G(u_1, ..., u_{j-1}, v_j, u_{j+1}, ..., u_n) > G(u_1, ..., u_n)$. Hence, $G(v) > \tilde{G}$ which is impossible. in Eq. (2) is continuous, strictly concave and strictly increasing with any of its n variables.

Proof: On $\mathbb{R}^+ \setminus \{0\}$, we define:

$$f_{\alpha}: x \mapsto \begin{cases} \frac{1}{1-\alpha} x^{1-\alpha} & \text{if } \alpha \neq 1, \alpha > 0\\ \log(x) & \text{if } \alpha = 1. \end{cases}$$

Obviously, f_{α} is strictly increasing and concave. Hence the result, by summation over the *n* variables.

B. Convex utility sets

We extend the formulation of Eq. (2) to convex and compact utility sets. We first prove the uniqueness of the solutions.

Lemma 4: Consider the notations of Lemma 2. If \mathcal{V} is convex then $\mathcal{S}_{\tilde{G}}$ is a singleton.

Proof: Suppose that $\exists u, v \in S_{\tilde{G}}, u \neq v$. Then $G(u) = G(v) = \tilde{G}$. Consider $z = \frac{u+v}{2}$. As \mathcal{V} is convex, then $z \in \mathcal{V}$. As G is strictly concave, then $G(z) > \tilde{G}$ which contradicts the definition of \tilde{G} .

We can now apply the result to the optimization function F_{α} . From Lemma 3 and 4, we obtain:

Theorem 1: If the utility set \mathcal{U} is convex and compact, then, for each $\alpha > 0$, there exists a unique point \tilde{u} solution of $F(\tilde{u}) = \max_{u \in \mathcal{U}} F_{\alpha}(u)$ and it is Pareto optimal. *Remark 2:* Obviously, when $\alpha = 0$, the function F_0 is linear and therefore not strictly concave. Therefore, if the utility set is not strictly concave, several solutions may be obtained. This is actually a well-known result: the global optimization is not unique.

Finally, we note that, similarly to what is obtained with linear utilities in flow control, as α grows to infinity the solution S_{α} converges to the one given by the maxmin fair criterion [15]. In the following subsection, we propose to study the general framework of all compact utility sets.

C. Extension of the fair family to general utility sets

Definition 5: For any utility set \mathcal{U} and $\alpha > 0$, we define $\tilde{F}_{\alpha} = \max_{\boldsymbol{u} \in \mathcal{U}} F_{\alpha}(\boldsymbol{u})$ and $\mathcal{S}_{\tilde{F}_{\alpha}} = \{\boldsymbol{u} \in \mathcal{U} | F_{\alpha}(\boldsymbol{u}) = \tilde{F}_{\alpha}\}$. For clarity of presentation, we note in the following \mathcal{S}_{α} instead of $\mathcal{S}_{\tilde{F}_{\alpha}}$.

From Lemma 2 and 3, we have:

Proposition 4: For any compact utility set \mathcal{U} , F_{α} exists and S_{α} is not empty. Additionally, S_{α} is included in the Pareto border.

When the set \mathcal{U} is not convex, cases where the optimization problem have several solutions (even infinitely many) exist. (Concrete examples will be given in Section VI.) More precisely, if \mathcal{U} is a subset of \mathbb{R}^n ,

 S_{α} is a set of dimension up to n-1. We therefore see that the formulation of Eq. (2) does not constitute a good general fairness objective when the utility sets are not convex.

a) General algorithm: One interesting property of the set of solutions S_{α} is that it is Pareto optimal. We therefore propose to consider an algorithm that determines exactly one point of S_{α} . As shown in Section II, with general utility sets, the max-min fairness criterion gives a finite set of equivalent points. We therefore propose to consider, as a fair solution associated to the parameter α , the max-min fair solution of the set S_{α} . We then show, in the rest of this subsection, how to determine the max-min fair point of a (non-convex) set. We propose to then apply this method to the set S_{α} , for any value of α .

b) Determining the max-min fair solution: We propose to determine the max-min fair point of the system by using the fair family. The following proposition characterizes the set of solutions corresponding to an infinite value of α :

Proposition 5: The solution S_{∞} corresponding to an infinite value of α is the set of points u such that:

$$\mathcal{S}_{\infty} = \{ \boldsymbol{u} \in \boldsymbol{\Pi} | \overrightarrow{u_1} = \max\{ \overrightarrow{v_1}, \boldsymbol{v} \in \mathcal{U} \} \},$$

with Π the Pareto border.

Proof: From Prop. 4, these points are Pareto optimal. Consider $\alpha > 1$ and $u \in S_{\alpha}$. As $\alpha > 1$, we have:

$$\overrightarrow{u_1}^{1-\alpha} \geq \overrightarrow{u_2}^{1-\alpha} \geq \ldots \geq \overrightarrow{u_n}^{1-\alpha}$$

with \vec{u} the ordered representation of u (as defined in Def. 1). Define $m = \min\{i \in \{2, ..., n\}, \ \vec{u_i}^{1-\alpha} < \vec{u_1}^{1-\alpha}\}$. As α grows to infinity, then $\forall i \geq m, \ (\vec{u_i}/\vec{u_1})^{1-\alpha} \to 0$, and therefore $F(\alpha) \simeq (m-1) \frac{\vec{u_1}^{1-\alpha}}{1-\alpha}$.

We illustrate this result in Fig. 3, which is an enhancement of Fig. 2 for a two users case. All the points of the thick line correspond to S_{∞} and hence are identically fair for the Mo and Walrand extended family. Therefore, computing S_{∞} is not sufficient to determine the max-min fair solution.

Yet, we can show an interesting characterization of the max-min points by the fair family:

Theorem 2: A point $u \in \mathcal{U}$ is max-min fair if and only if

$$\exists \alpha^*, \forall \alpha > \alpha^*, \boldsymbol{u} \in \mathcal{S}_{\alpha}$$

Proof: See Appendix.

Consider again Fig. 3. For a given $\alpha \ge 0$ and $K \in \mathbb{R}$, we consider the curve $\mathcal{F}_{\alpha,K}$ of points of solution \boldsymbol{u} of $F_{\alpha}(\boldsymbol{u}) = K$. For $\alpha = 0$, $\mathcal{F}_{0,K}$ has the shape of the dashed line. As α increases, the concavity of the curve increases (its shape is given by the curved dashed line for an intermediate value). Finally, when α is infinite, the shape of $\mathcal{F}_{\infty,K}$ is given by the thick line. Simultaneously, when K increases, $\mathcal{F}_{\alpha,K}$ gets further from the origin. Hence, we see that, although $\boldsymbol{a} \in \mathcal{S}_{\infty}$ and $\boldsymbol{b} \in \mathcal{S}_{\infty}$, for any finite value of α , we have $F(\boldsymbol{b}) > F(\boldsymbol{a})$.

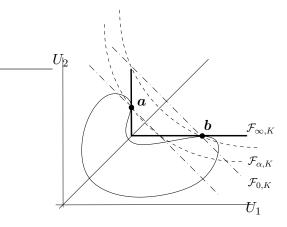


Fig. 3. Some solutions of \tilde{F}_{∞} are not acceptable.

Hence, in a non-convex system, the class of max-min fair points is given by the intersection of the solution of the fair family for large values of α .

D. Procedure to obtain a unique fair point for the Mo and Walrand fairness and for the Max-min fairness

In this section, we give the procedures to obtain a unique fair point corresponding to: a) Max-min fairness and b) the Mo and Walrand solution.

Let \mathcal{U} a compact utility set (not necessarily convex). Consider the family F_{α} as defined in Eq. (2). Our procedure is as follows:

- 1) Do: $\mathcal{V} := \mathcal{U}, i := 1.$
- 2) If computing max-min fairness (a), then choose a randomly large value of α_1 . Else (b), take $\alpha_1 = \alpha$.
- 3) Repeat:

• compute
$$\begin{cases} \tilde{F}_{\alpha_i} = \max_{\boldsymbol{u} \in \mathcal{V}} F_{\alpha_i}(\boldsymbol{u}) \\ S_{\alpha_i} = \{ \boldsymbol{u} \in \mathcal{V} | F_{\alpha_i}(\boldsymbol{u}) = \tilde{F}_{\alpha_i} \} \end{cases}$$

• Do: $\mathcal{V} := S_{\alpha_i}, i := i+1$

• Randomly choose a large value α_i ($\alpha_i \neq \alpha_j, j < i$)

until all points of \mathcal{V} are utility-symmetrical (as defined in Def. 4)

4) If V has more than one element, then apply the tie-breaking rule. Examples of tie-breaking rules are vast. For example, one could choose to give priority to users with low indices. Another solution would be to randomly choose one of the equivalent points. In Section VI, we give a possible rule in load balancing systems.

The convergence of the procedure is assured by Theorem 2. We note that the procedure would require at most n iterations. In practice, however, this number could be much lower. Also, F_{α} rapidly converges to F_{∞} and in practice, it is sufficient to consider values of the parameter above 20.

Remark 3: To check whether all the elements of a non-empty set \mathcal{V} are utility-symmetrical, one can proceed as follows:

- Set i := 0
- Choose $oldsymbol{x} \in \mathcal{V}$
- Remove x from \mathcal{V}
- While i = 0 do:

If $\mathcal{V} = \emptyset$ then i = 1, else

- Choose $y \in \mathcal{V}$
- if $x \prec y$ or $y \prec x$ then i = 2
- Remove y from \mathcal{V} .
- If i = 1 (respectively i = 2), all the elements are utility-symmetrical (respectively they are not).

E. Summary of the section

We have shown in this section that, for any convex system and any fairness parameter, the Mo and Walrand formulation gives a unique solution. On the other hand, when the utility set is not supposed convex, we have seen that the set of solution S_{α} can have infinitely many optimal elements (we provide a concrete example in Section VI). We therefore proposed to extend the fair family so as to choose, for any solution set, the maxmin fair solution of S_{α} . However, in a general nonconvex system, computing the max-min fair solution is not a simple problem. We hence provided a method to determine the max-min fair solution of a general set by using the optimization of a simple concave function (Eq. (2)).

IV. A NOTE ON COST MINIMIZATION PROBLEMS

It is sometimes useful to consider a set of costs, rather than of utilities. The cost functions could be, for instance, the delay experienced for each flow of a system. We show in this section that the *minimization* of the objective (2) over the *cost* set is not acceptable as a fair criterion. A concrete example will be given in the context of load balancing in Remark 4 in Section V.

A. Definition

We can define the Pareto border Π' of a cost set similarly as of a utility set (Eq. (1)) by reversing the inequalities. Let us denote by C the set of costs. Then, it seems natural to adapt the fair family to find the set of solutions T_{α} in C such as to *minimize*

$$H_{\alpha}(\boldsymbol{c}) = \begin{cases} \frac{1}{1-\alpha} \sum_{i} \left(\beta_{i} c_{i}\right)^{1-\alpha} \text{if } \alpha \geq 0, \alpha \neq 1, \\ \sum_{i} \log\left(c_{i}\right) & \text{if } \alpha = 1. \end{cases}$$
(4)

B. Validity

The following proposition shows that when α grows to infinity, the solution given by the minimization of H_{α} is such that it minimizes the smaller cost, which is not max-min fair.

Proposition 6: The solution \mathcal{T}_{∞} corresponding to an infinite value of α is the set of points x such that:

$$\mathcal{T}_{\infty} = \{ oldsymbol{c} \in oldsymbol{\Pi'} | \overrightarrow{c_1} = \min\{d_1^{'}, oldsymbol{d} \in oldsymbol{\Pi'} \} \}$$

with Π' the Pareto border for the cost set C.

Proof: Similarly to Lemma 2, we can show that these points are Pareto optimal. Let c be in the set of solutions T_{α} . For $\alpha > 1$, we have:

$$\overrightarrow{c_1}^{1-\alpha} \geq \overrightarrow{c_2}^{1-\alpha} \geq \ldots \geq \overrightarrow{c_n}^{1-\alpha}$$

Let us define $m \in \mathbb{N}^+$ by $m = \max\{i, i \in \{1, ..., n\}, \overrightarrow{c_i}^{1-\alpha} < \overrightarrow{c_1}^{1-\alpha}\}$. As α grows to infinity then $(\overrightarrow{c_i}/\overrightarrow{c_1})^{1-\alpha} \to 0$ and hence $H_{\alpha}(c) \sim \frac{\overrightarrow{c_1}^{1-\alpha}}{1-\alpha} \sum_{i=1}^m \beta_i$.

V. LOAD BALANCING MODEL

A. Load balancing model

We consider the simple distributed computer system represented in Fig. 4. It consists of two servers (computers), labelled 1 and 2, and two flows of demands ϕ_1 and ϕ_2 arriving from users 1 and 2 at servers 1 and 2, respectively (similar to the system studied in [6]). A fraction x_i $(0 \le x_i \le \phi_i, i \in \{1, 2\})$ of a flow of jobs is forwarded from server i to the other server $j \ (\neq i)$. We denote by x the vector (x_1, x_2) and by l_1 and l_2 , respectively, the resulting loads on nodes 1 and 2. Then $\forall i, j \in \{1, 2\}, i \neq j, \quad l_i = \phi_i - x_i + x_j.$ We assume that node i has an exponential service time with mean $1/\mu_i$ $(i \in \{1, 2\})$. Then, the mean delay at server i under the load of rate β_i is given by $(\mu_i - \beta_i)^{-1}$. For simplicity, we assume that forwarding a job requires a fixed delay t. Therefore the delay experienced by the flow arriving from the user i, can be written: for $i, j \in \{1, 2\}, (i \neq j),$

$$T_{i}(\boldsymbol{x}) = \frac{1}{\phi_{i}} \left[\frac{\phi_{i} - x_{i}}{\mu_{i} - \phi_{i} + x_{i} - x_{j}} + x_{i} \left(t + \frac{1}{\mu_{j} - \phi_{j} + x_{j} - x_{i}} \right) \right].$$
 (5)

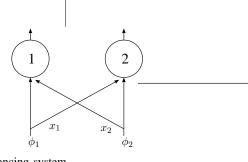


Fig. 4. Load balancing system

A difficulty in its analysis is that the performance of each user depends on the allocations of others. To write it more formally, for each i ($i \in \{1, 2\}$), T_i depends on x_j with $j \neq i$.

1) Set of constraints: We denote by \mathcal{X} the set of constraints on (x_1, x_2) . The forwarding rates should be positive or null and bounded by the arrival flow. Also, the service time at each server should not grow without bounds: $\mu_i - \phi_i + x_i - x_j > 0, \forall i, j \in \{1, 2\}, i \neq j$. Hence, we have

$$\mathcal{X} = \left\{ m{x} \in \mathbb{R}^2 \ \Big| \ egin{array}{c} 0 \leq x_i \leq \phi_i, \ i \in \{1,2\} \ ext{and} \ \phi_1 - \mu_1 < x_1 - x_2 < \mu_2 - \phi_2 \end{array}
ight\},$$

and the necessary condition (NC) for the system to be feasible is $\mu_1 + \mu_2 > \phi_1 + \phi_2$. The set \mathcal{X} is clearly convex.

When $x_1 - x_2 \rightarrow \mu_2 - \phi_2$ or $x_1 - x_2 \rightarrow \phi_1 - \mu_1$, the values of T_i and T_2 grow to infinity. We then approximates \mathcal{X} by:

$$\mathcal{X}_{\varepsilon} = \left\{ \boldsymbol{x} \in \mathbb{R}^2 \middle| \begin{array}{c} 0 \le x_i \le \phi_i, \ i \in \{1, 2\} \text{ and} \\ \phi_1 - \mu_1 + \varepsilon \le x_1 - x_2 \le \mu_2 - \phi_2 - \varepsilon \end{array} \right\}$$

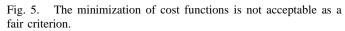
with ε a positive constant. Then, we can allow $T_i, i \in \{1, 2\}$ to have arbitrarily large values by properly choosing the value of ε . Additionally, for any $\varepsilon > 0$, the set $\mathcal{X}_{\varepsilon}$ is compact.

2) Utilities: As a utility function, we consider the inverse of the delay (which is homogeneous to a capacity allocation): $U_i(\mathbf{x}) = 1/T_i(\mathbf{x})$. The utility set is then

$$\mathcal{U}_{\varepsilon} = \{ \boldsymbol{u} = (u_1, u_2) | \exists \boldsymbol{x} \in \mathcal{X}_{\varepsilon}, u_i = U_i(\boldsymbol{x}), i \in \{1, 2\} \}.$$

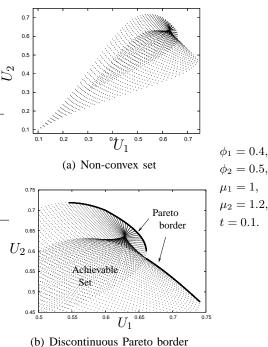
Obviously $\forall \boldsymbol{u} \in \mathcal{U}_{\varepsilon}, \forall i \in \{1, 2\}, u_i > 0.$

Remark 4: Although it would seem natural to consider the utility $U'_i(T(\boldsymbol{x})) = -T(\boldsymbol{x})$, we note that it would amount to consider the cost minimization family H of Section IV, which is not acceptable. This result is illustrated on Fig. 5. We see how the solutions of H_{α} diverge from the max-min fair point as α increases, and hence how H_{α} is not acceptable as a fair family.



B. A non-convex utility set

As illustrated by Fig. 6, the utility set is not necessarily convex. As a result, the Pareto set may be locally concave or discontinuous.

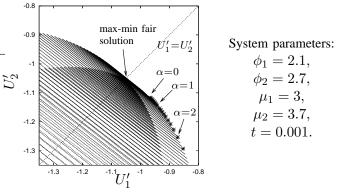


(b) Discontinuous Faleto bo

Fig. 6. Utility set

Compactness: The utility set is compact, as it is the image of a compact set $(\mathcal{X}_{\varepsilon})$ by continuous functions $(U_1$ and $U_2)$.

In this section, we have described a simple load balancing system. We have seen that, in spite of the simplicity of the model, the utility set can be non-convex. In the following section, we apply our general fairness criterion to this system.



VI. FAIR LOAD BALANCING

In this section, we study the properties of the fair (and optimal) points in the load balancing problem described in the previous section.

In the following subsection, we present how we applied the fair family to the system. Section VI-B then present some numerical results we obtained.

A. Computational method

Since the utility set is not necessarily convex, we apply the procedure of the extended fair family defined in Section. III-D. In the general case (of n users), the construction needs at most n iterations. Hence, in this simple load balancing system, we solve the optimization problem of Eq. (2) at most n = 2 times.

To compute the fair solutions, we fix the value of ε of $\varepsilon = 0.001$ and proceed as follows:

• 1st iteration: For given values of the parameters $(\mu_1, \mu_2, \phi_1, \phi_2, t \text{ and } \alpha)$, we set up our program so as to compute

$$\widetilde{F_{\alpha}} = \max\{F_{\alpha}(\boldsymbol{x}) | \boldsymbol{x} \in \mathcal{X}_{\varepsilon}\}$$
(6)

and S_{α} with a 0.1% accuracy. That is, we determine the set :

$$\mathcal{S}_{\alpha}^{*} = \{ \boldsymbol{x} \in \mathcal{X}_{\varepsilon} | F_{\alpha}(\boldsymbol{x}) \ge 0.999 \widetilde{F_{\alpha}} \}.$$
(7)

• 2^{nd} iteration: Since n = 2, then S^*_{α} is a set of dimension 1. When it does not consist of a single point, we consider α' , with α' sensibly larger than α and then solve

$$\operatorname{Fsol}_{\alpha'} = \max\{F_{\alpha'}(\boldsymbol{x}), \boldsymbol{x} \in \mathcal{S}_{\alpha}^*\}, \quad (8)$$

and the unique solution

$$\overline{u^{\alpha}}$$
 such that $\overline{u^{\alpha}} \in \mathcal{S}^*_{\alpha}$ and $F_{\alpha'}(\overline{u^{\alpha}}) = \widetilde{\mathrm{Fsol}}_{\alpha'}^{\prime}$.
(9)

Tie-breaking rule: We recall that for each value of the parameter, the fair family gives a class of solution point. In this case, there is at most 2! = 2 equivalent solution points (similar to *a* and *b* of Fig. 1). We hence need to set up a "tie-breaking rule":

If u = (a, b) and v = (b, a) with a < b are equivalently fair for some value of the parameter α , then the chosen solution point is:

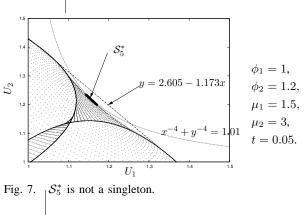
•
$$\boldsymbol{u}$$
 if $U_1((0,0)) < U_2((0,0))$,

• v otherwise.

B. Numerical results

1) Example and interest of the method: Fig. 7 represents an example where the Mo and Walrand fair criterion does not give a unique solution, that is, the set of the fair solutions, S_5^* , has multiple points (note the statement given after Prop. 4). This shows the appeal of our method. ($\phi_1 = 1$, $\phi_2 = 1.2$, $\mu_1 = 1.5$, $\mu_2 = 3$, t = 0.05 and $\alpha = 5$.)

We illustrate in Figs. 7 and 8 the construction of our fair solution. As S_5^* is not reduced to a single element (Fig. 7), we compute $\widetilde{\text{Fsol}}_{20}$ (Eq. (8)) and the unique solution \overline{u}^5 (Eq. (9)) shown in Fig. 8.



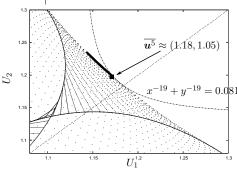


Fig. 8. Extended fairness: solution.

2) *Remarks:* We provide here some remarks drawn from numerical results.

Set S_{α}^* : As previously mentioned, the set S_{α}^* is of dimension 1. It is either:

- a single point (Fig. 9 with $\alpha = 0, 1, 2, ..., 70$)
- a portion of curve, either continuous (Fig. 7) or not (Fig. 10).

Convergence For very large values of α , the program becomes unstable, due to accumulation of round-off errors. Hence, in the experiments, we limited the value of α to be less than 100, which might be, for special values of the parameters, not sufficient for reaching the max-min fair point. Yet, as the construction of the α -fair point considers the max-min fair point of a restricted part of the Pareto set, the experiments show that in practice

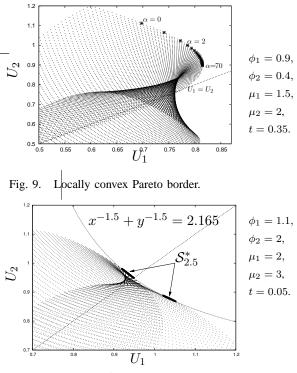


Fig. 10. The set S^*_{α} can be discontinuous.

the family rapidly converges to the max-min fair point as α grows.

Tie-breaking rule Finally, we note that in our experiments, we did not face any situation where we had to make use of the tie-breaking rule. Although in theory utility-symmetrical fair points exists, we expect them to be rarely found in practice.

VII. CONCLUSION

In this paper, we have studied in a general framework a family of points parameterized by a parameter α which enables us to finely tune the tradeoff between global efficiency and fairness. Our starting point was a family originally introduced in the context of fair bandwidth allocation. We have shown that this family can be used in the much wider range of systems in which the utility set is convex and compact. More precisely, we have shown that in that case the fair family has a unique solution.

In addition, when the utility set is not convex, we have shown that there exists a class of equivalent max-min fair points. As a given optimization function does not necessarily have a unique extremum, the fair family may have infinitely many solution points and defining fairness becomes more challenging. We have therefore defined a new fair family, that coincides with the previous one when the utility functions are supposed convex and which gives a unique solution for any (compact) utility set. This research hence opens the road for deeper study on fairness with general utility sets. Non-convex systems exist in a variety of networking situations. We have exhibited a simple static load balancing model in which the utility set is not convex. The system consists of two identical servers (computers) with their own arrival and own queue. We have assumed that the delay induced by forwarding a flow is fixed, and have considered as a utility function the inverse of the total delay experienced by the flow. We have seen that in spite of the simplicity of the model, the utility set can be non-convex and the Pareto border continuous or not. Through numerical results, we examined the parameterized fairness objectives in the load balancing system.

APPENDIX

Proof of Theorem 2 1) (sufficiency) Let v be maxmin fair and $u \in \mathcal{U}$. If u is max-min fair, then $\overrightarrow{u} = \overrightarrow{v}$ and so $F_{\alpha}(u) = F_{\alpha}(v)$.

If u is not max-min fair, then $u \prec v$. Define $m = \min\{i \in \{1, ..., n\} | \overrightarrow{u_i} < \overrightarrow{v_i}\}$. Then $\overrightarrow{u_i} = \overrightarrow{v_i}$ for i = 1, ..., m. Define also $m' = \min\{i > m | \overrightarrow{v_m} < \overrightarrow{v_i}\}$. Then $\overrightarrow{v_m} = \overrightarrow{v_{m+1}} = ... = \overrightarrow{v_{m'-1}} < \overrightarrow{v_{m'}}$ and $\overrightarrow{v_j} > \overrightarrow{v_m}, j = m', ..., n$.

We then want to show that $\exists \varepsilon$ such that $\forall \alpha > \varepsilon, F_{\alpha(\boldsymbol{v})} > F_{\alpha(\boldsymbol{u})}$.

We have: • $\sum_{j=1}^{m-1} \left(\frac{\overrightarrow{v_j}}{\overrightarrow{v_m}}\right)^{1-\alpha} = \sum_{j=1}^{m-1} \left(\frac{\overrightarrow{u_j}}{\overrightarrow{v_m}}\right)^{1-\alpha}.$ • $\sum_{j=m'}^{n} \left(\frac{\overrightarrow{v_j}}{\overrightarrow{v_m}}\right)^{1-\alpha} \xrightarrow{\alpha \to \infty} 0. \text{ Therefore } \exists \varepsilon_1 > 1 \text{ such } \\ \text{that } \alpha > \varepsilon_1 \Rightarrow \sum_{j=m'}^{n} \left(\frac{\overrightarrow{v_j}}{\overrightarrow{v_m}}\right)^{1-\alpha} < 1.$ • $\left(\frac{\overrightarrow{u_m}}{\overrightarrow{v_m}}\right)^{1-\alpha} \rightarrow +\infty, \text{ then } \exists \varepsilon_2 > 1 \text{ such that } \alpha > \\ \varepsilon_2 \Rightarrow \left(\frac{\overrightarrow{u_m}}{\overrightarrow{v_m}}\right)^{1-\alpha} > m' - m + 1.$ • Since $\overrightarrow{v_m} = \overrightarrow{v_{m+1}} = \dots = \overrightarrow{v_{m'-1}} < \overrightarrow{v_{m'}}, \text{ then } \\ \sum_{i=m}^{m'-1} \frac{\overrightarrow{v_j}}{\overrightarrow{v_m}} = m' - m$

Finally, for all $\alpha > \max\{\varepsilon_1, \varepsilon_2\}$: $\sum_{j=1}^n \left(\frac{\overrightarrow{u_j}}{\overrightarrow{v_m}}\right)^{1-\alpha} = \sum_{j=1}^{m-1} \left(\frac{\overrightarrow{v_j}}{\overrightarrow{v_m}}\right)^{1-\alpha} + \sum_{j=m'}^n \left(\frac{\overrightarrow{u_j}}{\overrightarrow{v_m}}\right)^{1-\alpha} + m' - m > \sum_{j=1}^n \left(\frac{\overrightarrow{v_j}}{\overrightarrow{v_m}}\right)^{1-\alpha}$ and so $F_{\alpha}(\boldsymbol{v}) > F_{\alpha}(\boldsymbol{u})$. Hence $\forall \boldsymbol{u} \in \mathcal{U}$ not max-min fair, $\exists \varepsilon, \forall \alpha > \varepsilon, F_{\alpha}(\boldsymbol{v}) > F_{\alpha}(\boldsymbol{u})$.

2) (necessity) The proof of necessity goes similarly as that of sufficiency. Let $v \in U$ such that $\exists \varepsilon_1$ such that $\forall \alpha > \varepsilon_1, v \in S_{\alpha}$. Let $u \in U, u \neq v$. Suppose that $v \prec$ u. Then define m and m' with $m = \min\{i | \vec{u_i} > \vec{v_i}\},\$ and $m' = \min\{i | \overrightarrow{u_{m'}} > \overrightarrow{u_m}\}$. Then as before,

$$\sum_{j=1}^{m-1} \left(\frac{\overrightarrow{v_j}}{\overrightarrow{u_m}}\right)^{1-\alpha} = \sum_{j=1}^{m-1} \left(\frac{\overrightarrow{u_j}}{\overrightarrow{u_m}}\right)^{1-\alpha}$$

As $\overrightarrow{u_m} > \overrightarrow{v_m}$, then $\left(\frac{\overrightarrow{v_m}}{\overrightarrow{u_m}}\right)^{1-\alpha} \to \infty$ and therefore $\exists \varepsilon_2$ such that $\forall \alpha > \varepsilon_2, \left(\frac{\overrightarrow{v_m}}{\overrightarrow{u_m}}\right)^{1-\alpha} > m' - m + 1$. As

$$\sum_{j=m'} \left(\frac{u_j}{\overrightarrow{u_m}}\right) \to 0, \text{ then } \exists \varepsilon_3, \sum_{j=m'} \left(\frac{u_j}{\overrightarrow{u_m}}\right) < 1$$

Finally, for any $\alpha > \max\{\varepsilon_1, \varepsilon_2, \varepsilon_3\}, F_{\alpha}(\boldsymbol{u}) > F_{\alpha}(\boldsymbol{v}),$ which contradicts $v \in S_{\alpha}$.

REFERENCES

- [1] Aubin, J-P (1998). Optima and Equilibria: An Introduction to Nonlinear Analysis, 2nd Ed.. Springer-Verlag. Berlin.
- [2] Bertsekas, D. and R. Gallager (1992). Data Networks, 2nd Ed.. Prentice-Hall. Englewood Cliffs.
- [3] Bonald, T. and L. Massoulié (2001). Impact of fairness on internet performance. In: Proc. Sigmetrics Conf., 82-91.
- [4] Le Boudec, J-Y. Le (2000). Rate Adaptation, Congestion Control and Fairness: A Tutorial. Ecole Polytechnique Féd. de Lausanne.
- [5] Conley, J and S. Wilkie (1996) An Extension of the Nash Bargaining Solution to Nonconvex Problems. Games and Economic behavior 13, 26-38.
- [6] Kameda, H., E. Altman, T. Kozawa and Y. Hosokawa (2000). Braess-like paradoxes in distributed computer systems. IEEE Trans. Automatic Control 45(9), 1687-1691.
- [7] Kameda, H., J. Li, C. Kim and Y. Zhang (1997). Optimal Load Balancing in Distributed Computer Systems. Springer.
- [8] Kelly, F. P. (1997). Charging and rate control for elastic traffic. European Transactions on Telecommunications 8, 33–37.
- [9] Kim, C. and H. Kameda (1990). An algorithm for optimal static load balancing in distributed computer systems. IEEE Trans. Comput. 41, 381–384.
- [10] Li, J. and H. Kameda (1998). Load balancing problems for multiclass jobs in distributed/parallel computer systems. IEEE Trans. Comput. 47, 322-332.
- [11] Low, S., L. Peterson and L. Wang. (2001) Understanding TCP Vegas: a duality model. In: Proc. of ACM SIGMETRICS, 226-235.
- [12] Massoulié, L. and J. Roberts (1999). Bandwidth sharing: objectives and algorithms. In: Proc of IEEE INFOCOM 3, 1395-1403.
- [13] Maulloo, A., F. P. Kelly and D. Tan (1998). Rate control in communication networks: Shadow prices, proportional fairness and stability. J. of the Operational Res. Soc. 49, 237-252.
- [14] Mazumdar, R., L. G. Mason and C. Douligeris (1991). Fairness in network optimal flow control: Optimality of product forms. IEEE Trans. Communications 39(5), 775-782.
- [15] Mo, J. and J. Walrand (2000). Fair end-to-end window-based congestion control. IEEE/ACM Trans. Networking 8(5), 556-567.
- [16] Nash, J. (1950). The Bargaining Problem. Econometrica 18, 155-162.
- [17] Sabin, G., V. Sahasrabudhe and P. Sadayappan (2004). On Fairness in Distributed Job Scheduling Across Multiple Sites. In: Proc. of IEEE Int. Conf. on Cluster Computing, 35-44.
- [18] Tantawi, A. N. and D. Towsley (1985). Optimal static load balancing in distributed computer systems. J. ACM 32(2), 445-465.

- [19] Touati, C., E. Altman, and J. Galtier (2001). On fairness in bandwidth allocation. Inria Research Report RR 4269.
- [20] Touati, C., E. Altman, and J. Galtier (2002). Utility Based Fair Bandwidth Allocation. In: Proc. of IASTED NPDPA, 126-131.
- [21] Xu, Y and N. Yoshihara (2005). Alternative characterizations of three bargaining solutions for nonconvex problems. COE/RES Discussion Paper Series, No.126.
- [22] Yaïche, H., R. R. Mazumdar and C. Rosenberg (2000). A game theoretic framework for bandwidth allocation and pricing in broadband networks. IEEE/ACM Trans. on Net. 8(5), 667-677.
- [23] Zakon, E. (2004). Mathematical Analysis, Vol I, ISBN 1-931705-02-X.