Generalized Nash Bargaining Solution for bandwidth allocation[1]

Corinne Touati, Eitan Altman, Jérôme Galtier

INRIA Sophia-Antipolis, 2004 route des Lucioles - BP 93 - 06902 Sophia-Antipolis Cedex, France

Abstract

For over a decade, the Nash bargaining solution (NBS) concept from cooperative game theory has been used in networks to share resources fairly. Due to its many appealing properties, it has recently been used for assigning bandwidth in a general topology network between applications that have linear utility functions. In this paper, we use this concept for allocating the bandwidth between applications with general concave utilities. Our framework includes in fact several other fairness criteria, such as the max-min criteria. We study the impact of concavity on the allocation and present computational methods for obtaining fair allocations in a general topology, based on a dual Lagrangian approach and on Semi-Definite Programming.

Key words: Nash Bargaining, bandwidth allocation, fairness.

1 Introduction

Fair bandwidth assignment has been one of the important challenging areas of research and development in networks supporting elastic traffic. Indeed, *Max-min* fairness has been adopted by the ATM forum for the Available Bit Rate (ABR) service of ATM [2]. Although the max-min fairness has some optimality properties (Pareto optimality), it has been argued that it favors too much long [1] connections and does not make efficient use of available bandwidth. In contrast, the concept of *proportional fairness* (of the throughput assignment) has been proposed by Kelly [17, 18]. It gives rise to a more efficient solution in terms of network resources by providing more resources to shorter [1]

[1] A very preliminary and shorter version of the fairness criterion was presented in the conference NPDPA [27] and preliminary numerical results using SDP programming appeared in [28].

Preprint submitted to Elsevier Science *12 December 2005*

connections.

Although the object that is shared fairly seems to be a very specific one (the throughput), it is shown in [17, 18] that in fact, the starting point for obtaining (weighted) proportional fairness of the throughput can be any general (concave) utility function per connection.

As opposed to this approach, we wish to use a fairness concept that is directly defined in terms of the users' utilities rather than of the throughputs they are assigned to. Yet, as in weighted proportional fairness, it would be desirable to express this concept as the solution of a utility maximization problem, since it makes it possible to use recent algorithms for utility maximization in networks, along with decentralized implementations [16, 15, 19].

The Nash Bargaining Solution (NBS) is a natural framework that allows us to define and design fair assignment of bandwidth between applications with different concave utilities and has already been used in networking problems [22, 29]. It is characterized by a set of axioms that are appealing in defining fairness. As already recognized in [18] and later in [29], proportional fairness agrees with the NBS if the object that is shared fairly is the throughput and if the user's minimum required rate is zero. We are interested in the NBS since it can been seen as a natural extension of the proportional fairness criterion which is probably the most popular fairness notion in network design today: it appears widely in the Internet world (indeed, it is shown in [20] that some versions the TCP congestion control protocol achieves proportional fair bandwidth sharing) as well as in wireless communications ([3]).

We use the NBS to study the fairness of an assignment where connection n has a concave utility over an interval $[MR_n, PR_n]$. It thus has a minimum rate requirement MR_n and does not need more than PR_n. Utility functions with such features have been identified in [26] for representing some real time applications such as voice and video, and in the case that $MR_n = 0$, for elastic traffic.

We briefly mention some other useful concepts for fairness or for allocating resources. One of the properties defining the Nash bargaining solution has had some criticism (an axiom stating that the solution is not affected by reducing the domain) since it implies that a player does not care how much other players have given up. Two alternative notions of fair sharing have thus been introduced with the same other properties of the Nash bargaining solution, but with a variation of the above property: the modified Thomson solution and the Raiffa-Kalai-Smorodinsky solution. A unified treatment of the Nash solution as well as of these two has been introduced in [7] for two players and extended

[1] The terms *long* and *short* refer here to the distance, counted in the number of hops or of saturated links for instance, as we will explain later.

in [11] for the multiperson case. These concepts have been applied to Internet pricing in [6]. Yet another "fair" concept for sharing resources is the Aumann-Shapley solution for cooperative game, which has desirable properties such as Pareto optimality. Haviv [12] proposes this approach to allocating congestion costs in a single node under various queueing disciplines.

We study in this paper the way the concavity of the utilities affect the bandwidth assignment according to the NBS, as well as according to a generalized version of the proportional fairness (in which the utilities that correspond to the different connections are fairly allocated).

Contribution of the paper : In previous work on Nash bargaining in networks, only linear utilities have been studied. We introduce in this paper quadratic utility functions. It is a sufficiently large class in order to represent utilities with various degrees of concavity. At the same time they are sufficiently simple for computation purposes. We indeed propose in the paper a Lagrangian (see [29]) approach as well as a reduction to Semi Definite Programming (SDP) which allows to use a large variety of open source (and other) libraries to compute efficiently the fair solution in realistic networks. We indicate the complexity of the solution by analyzing matrices that are used in the SDP formulation. We demonstrate how various fairness concepts can be handled by the proposed SDP method. We study the way the concavity of the utilities affect the bandwidth assignment according to the NBS, as well as according to a generalized version of the proportional fairness (in which the utilities that correspond to the different connections are fairly allocated). We illustrate numerically the bandwidth allocation for the Cost network [1].

Structure of the paper : After briefly introducing the notations used in the article (Sec. 2), we present the fair allocations in Sec.3 and focus on the NBS in Sec. 4 and its properties. We then propose a new fair family that covers for special cases the previously defined fairness criteria, including in particular the Nash, the Thomson and the Kalai-Smorondinsky solutions in Sec. 5 and propose a quadratic approximation for the utility of each connection, which allows us to parameterize the degree of concavity of the utility function using a single parameter. We use this approximation to further analyze the impact of concavity of utilities on the resulting assignment. Quadratic utilities can be viewed as a second order approximation (in terms of a Taylor series) of any sufficiently smooth concave utility function. We then present in Sec. 7 a Lagrangian approach which allows us to implement a decentralized protocol for the bandwidth allocation. We finally present in Sec. 8 a novel alternative approach using Semi Definite Programming (SDP) and some numerical results in Sec. 9

2 Definitions and notations

In the following, we denote by *connection* a "source-destination" pair. For a given network and set of connections, there exists a infinite set of feasible allocations. We are interested in those that are optimal (as we will define in Subsection 2.2) and satisfying the system's constraints (Subsection 2.3). The fair allocation we will present all belong to this set. Let us first summarize in Subsection 2.1 the notations used in this article.

2.1 Notations

We summarize in Table 1 the different notations used in the article. We suppose that to each connection n is associated a utility function f (on which we will give more insight in Section 4.1) and has a minimum and a maximum requirement in terms of bandwidth, that we denote MR_n and PR_n respectively.

$\|ens\|$	number of elements of set *ens*,	X	set of feasible allocations,
n	a connection,	$\mathcal{N}$	set of connections,
x_n	bandwidth allocated to connection n,	$\vec{x}$	vector of allocations (of size $\|\mathcal{N}\|$),
MR_n	minimal bandwidth requirements of connection n,	PR_n	maximal bandwidth required for connection n,
l	a directed link,	$\mathcal{L}$	set of links in the network,
C_l	capacity of link l,	$\vec{C}$	vector of links capacities.

Table 1
Notations

We also define some orders among feasible allocations (that is to say satisfying the system's constraints, as defined in Section 2.3). We write:

- $\vec{x} \leq \vec{y}$ if $\forall i \in 1, \ldots, N, x_i \leq y_i$
- $\vec{x} \leq_{lg} \vec{y}$ if either $\vec{x} = \vec{y}$ or $\exists i, i > 0, \forall j \in 1, \ldots, i-1, x_j = y_j$ and $x_i < y_i$ (lexicographic order)
- $\vec{x} \preceq \vec{y}$ if $\sigma(\vec{x}) \leq_{lg} \sigma(\vec{y})$ with $\sigma(\vec{x})$ and $\sigma(\vec{y})$ the ordered versions of $\vec{x}$ and $\vec{y}$[2] .

2.2 Efficiency

An allocation is said to be Pareto optimal or Pareto efficient if it is impossible to increase the allocation of a connection without strictly decreasing another

[2] that is to say that $\forall i$, $\sigma(\vec{x})_i$ is the i^{th} smaller element of x.

one. In other words, an allocation Pareto efficient is maximal in the sense of $\leq$. In the networking context, this amounts in saying that each connection goes through at least one saturated link. In general, if $|\mathcal{N}|$ is the number of connections, the set of Pareto optimal points is a subset of size $|\mathcal{N}| - 1$. The fair allocations we will present are all Pareto optimal.

2.3 Allocations constraints.

We briefly introduce here the constraints in the allocations vectors. They are of two types: the ones associated to the users and those due to the routing policy.

User constraints As previously mentioned, we suppose that each connection is associated to a minimum and a maximum requirement, in term of bandwidth. We therefore have a system of $|\mathcal{N}|$ independent linear inequalities $\forall n \in \mathcal{N}, MR_n \leq x_n \leq PR_n$.

Routing policy We consider the case of both fixed-routing and fractional-routing. In fixed routing, we define a matrix A of size $|\mathcal{L}| \times |\mathcal{N}|$ that specifies the links that the packets of each connection will go through:

$$A_{n,l} = \begin{cases} 1 \text{ if connection } n \text{ goes through link } l, \\ 0 \text{ otherwise.} \end{cases}$$

In some cases, we do not want to fill the links but keep some empty space, call *left-over capacity*. For instance, an operator may wish to use only up to 90% of the link in order to let some other traffic join the system. Therefore we suppose that we can only use the quantity $(1 - \kappa_l)C_l$ on each link l. Let κ be the vector of size $|\mathcal{L}|$ of leftover capacities. The capacity constraints can then be written:

$$A\vec{x} \leq (1 - \vec{\kappa})\vec{C}. \tag{1}$$

In the case of fractional routing, each packet transmitted by a source can follow its own path. We denote by $\phi_n^{(u,v)}$ the flow of connection n in the link from u to v. Note that $\phi_n^{(v,u)}$ exists and is a distinct variable. V is the set of nodes and for $u \in V$, $N(u)$ is the set of nodes connected to u. Finally, $s(n)$ (respectively $d(n)$) is the source (destination) of connection n. We have

4 types of constraints for each connection n:

$$\begin{array}{ll} x_n = \sum_{v \in N(s(n))} \phi_n^{(s(n),v)} - \sum_{u \in N(d(n))} \phi_n^{(u,s(n))} & \text{balance of flows from source } s(n), \\ x_n = \sum_{u \in N(d(n))} \phi_n^{(u,d(n))} - \sum_{v \in N(s(n))} \phi_n^{(d(n),v)} & \text{balance of flows to destination } d(n), \\ \forall u \neq \{s(n), d(n)\}, \sum_{v \in N(u)} \phi_n^{(v,u)} = \sum_{v \in N(u)} \phi_n^{(u,v)} & \text{Kirchhoff or conservation law,} \\ \forall u, v, \phi_n^{(u,v)} \geq 0. & \text{positive or null flows.} \end{array}$$

Let us introduce $r_n^{s(n)} = -r_n^{d(n)} = 1$, and $r_n^u = 0$ for u different from $s(n)$ and $d(n)$. If we can use the fraction $(1 - \kappa_l)C_l$ on each link $l = \{u, v\}$ then the capacity constraints are:

$$\forall l = \{u, v\} \in \mathcal{L}, \qquad \sum_{n \in \mathcal{N}} \phi_n^{(u,v)} \leq (1 - \kappa_l)C_l$$

Note that in fact this equation applies both to (u, v) and (v, u), which means that the capacity constraints will be verified on each orientation of $\{u, v\}$.

$$\text{Finally, } \forall n \in \mathcal{N}, \begin{cases} \forall u \in V, & x_n r_n^u = \sum_{w \in N(u)} \phi_n^{(u,w)} - \sum_{w \in N(u)} \phi_n^{(w,u)}, \\ \forall \{u, v\} \in \mathcal{L}, & \phi_n^{(u,v)} \geq 0, \\ \forall l = \{u, v\} \in \mathcal{L}, & \sum_{n \in \mathcal{N}} \phi_n^{(u,v)} \leq (1 - \kappa_l)C_l. \end{cases}$$

We can notice that in both cases, the constraints are linear.

3 Fair allocations

In this section, we review the criteria commonly used for fairly sharing resources in networks. Each of them can be found in both a simple and a weighted version. The latter was introduced in order to let users express the relative value of their traffic. Then, a connection with weight k may be equivalent to k connections of weight 1.

We illustrate the following definitions on the classical linear network represented in Figure 1. It consists of $|\mathcal{L}|$ links shared by $|\mathcal{L}| + 1$ connections (numbered from 0 to $|\mathcal{L}|$) such that :

- each connection n, $n \neq 0$ goes only through link n,
- connection 0 uses all links.

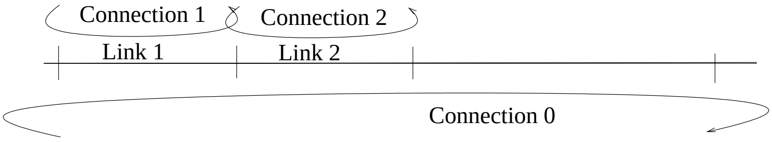


Fig. 1. Linear network

Note that, for any Pareto optimal allocation, the global revenue can be written:

$$R_{glob} = \sum_{0 \leq n \leq |\mathcal{L}|} x_n = (1 - |\mathcal{L}|)x_0 + \sum_{l \in \mathcal{L}} C_l. \tag{2}$$

3.1 Maximization of global throughput

This criterion consists of the maximization of the total allocated bandwidth $R_{glob} = \max \sum_{n \in \mathcal{N}} x_n$.

A major drawback of this criterion is that it can lead to situations in which the allocation is null for one or more connections, as we can see in the linear example (indeed, from Equation 2 the criterion imposes that $x_0 = 0$). This is why it is not considered as fair.

A variant, in which we assign to each user a utility function (we will see in Section 4.1 how this can be justified in the case of networks) has been developed. Roughly speaking, the utility function represents the satisfaction of a user from its allocation. This can actually be the perceived quality of an audio or video signal. Some algorithms ([16], [15]) were then developed to maximize the sum of the users' utilities, which is called *social-welfare* optimization.

An opposite approach is the max-min criterion that seeks to allocate the resources in the most homogeneous way, while being Pareto optimal.

3.2 Max-Min fairness

Several equivalent (as proved in [4]) definitions can be found. Let us first define the bottleneck of a connection.

Definition 1 *A link l is a bottleneck of connection n if:*

- *It is fully used,*
- *The allocation of connection n is the greatest among all connections using link l.*

We then have the following definitions:

Definition 2 *An allocation is max-min fair if and only if each source has a bottleneck.*

Definition 3 *The max-min allocation is maximal for the order $\preceq$.*

Definition 4 *The max-min fair allocation is such that any increase of the allocation for one connection would be at the expense of another connection whose allocation was already smaller. In other words, if $\vec{x}$ is the max-min allocation then $\forall \vec{y}, \vec{y} \neq \vec{x}$, $y_n > x_n \Rightarrow \exists s, x_s \leq x_n, y_s < x_s$.*

In the linear example, the allocation of connection 0 will be half of the smaller link capacity: $x_0 = C_{l_{min}}/2$. The total revenue is then: $R_{max-min} = (1 - \frac{|\mathcal{L}|}{2})C_{l_{min}} + \sum_{l \in \mathcal{L}} C_n$.

The max-min fairness has been adopted by the ATM forum for the Available Bit Rate (ABR) service of ATM ([2]). It has been argued that max-min fairness gives to much allocation to *long* connections (in our case the connection 0) and does not efficiently utilize bandwidth. Therefore, a new concept has been introduced to Kelly: the proportional fairness.

3.3 Proportional fairness

It has been defined by two equivalent ways (as proved in [4]).

Definition 5 *The proportional fair assignment maximizes $\max_{\vec{x} \in X} \sum_{n \in \mathcal{N}} ln(x_n)$.*

Definition 6 *Proportional fairness is the unique allocation $\vec{x}$ such that for any other allocation $\vec{x}'$, we have $\sum_{n \in \mathcal{N}} \frac{{x'}_n - x_n}{x_n} \leq 0$.*

In the linear network, $x_0 = 1/\sum_{n \in \mathcal{N}} \frac{1}{C_n - x_0}$. Furthermore, if the links are identical with capacity C, then $x_0 = \dfrac{C}{|\mathcal{L}| + 1}$. We note that, as expected: $x_{0_{\text{global opt}}} < x_{0_{\text{prop fair}}} < x_{0_{\text{max-min}}}$. Moreover $R_{\text{global opt}} \geq R_{\text{prop fair}} \geq R_{\text{max-min}}$. The compromise between fairness and efficiently was the primary reason for the success of proportional fairness.

Its weighted version can be written. An allocation $\vec{x} \in X^{wpf}$ is weighted proportional fair if for any other allocation $\vec{x}^* \in X$ we have ([17, 18]): $\sum_{n \in \mathcal{N}} w_n \frac{x_n^* - x_n}{x_n} \leq 0$. Equivalently: $x^{wpf} = \max_{\vec{x} \in X} \sum_{x \in \mathcal{N}} w_n \ln(x_n) = \max_{\vec{x} \in X} \prod_{n \in \mathcal{N}} x_n^{w_n}$.

The congestion control mechanism based on linear increased-multiplicative decreased window flow control can lead to proportionally fair allocations under certain conditions ([18]). Unfortunately this result does not hold for all versions of TCP, and in particular to TCP Reno. But it has been proved that TCP Vegas actually achieve proportional fairness [20].

3.4 Potential delay minimization

This criterion, introduced in [21], is defined by: $\min_{\vec{x} \in X} \sum_{n \in \mathcal{N}} \frac{1}{x_n}$. It is a minimization problem of an inverse function of the bandwidth allocated, that is to say the transfer time.

3.5 Relation between the different fairness criteria

Mo and Walrand ([23]) recently showed that these criteria can be written as a single optimization function. Let $\vec{x}_\alpha$ be the allocation solution of:

$$\vec{x}_\alpha = \max_{\vec{x} \in X} \sum_{n \in \mathcal{N}} \frac{{x_n}^\alpha}{1-\alpha} \tag{3}$$

With $\alpha \geq 0$, $\alpha \neq 1$. Then, for specific values of α, $\vec{x}_\alpha$ corresponds to the different fairness criteria previously mentioned : the global maximization of the throughput ($\alpha = 0$), proportional fairness ($\alpha \to 1$), potential delay minimization ($\alpha = 2$) or max-min fairness ($\alpha \to \infty$)

Figure 3.5 represents the allocation to connection 0 when α grows from zero to infinity in the case of the linear network with equal link capacities. We clearly see the influence of the parameter α. As α increases, the differences between the allocation of the connections vanish, thus continuously deriving from a global optimization equilibrium to a max-min allocation.

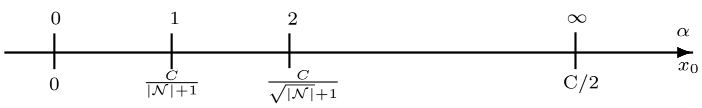


Fig. 2. Allocation to connection 0 with α

4 Nash criterium

The Nash Bargaining Solution (NBS) [24] concept for fair allocation is frequently used in cooperative game theory. It is defined by a set of axioms that game theorists find natural to require in seeking fair allocations. These axioms deal with utilities associated to users, which is the natural interest of this approach compared to the previously seen fairness criteria.

The NBS has already been used (under a simplified version) in networks of general topologies to propose a fair share of resources ([29, 22]). But in both cases, the utility functions were linear. It has been shown that in fact, the utility functions can be any concave function.

Several fairness criteria were defined in cooperative game theory (see [6] and references herein). We consider here the Nash concept, since it can be seen, as previously mentioned, as a generalization of the widely studied proportional fairness.

Let us suppose that a finite number of perfectly rational individuals can collaborate in order to get mutual benefit. We further suppose that they can compare their satisfaction from the possession of the objects of the bargaining. We can then associate the users to a utility function, which is, obviously, not unique: if u is such a function, then $au + b$ is an equivalent one (for $a, b \in \mathbb{R}, a > 0$). In the case where the players cannot find an agreement, the game ends at the "disagreement point", characterized by a certain utility, u^0.

Let $X \subset \mathbb{R}^n$ denote the set of possible strategies. It is a convex closed and non empty set. The utility functions, $f_i : X \to \mathbb{R}, i = 1, \dots, k$, are supposed to be upper bounded functions. The set of achievable utilities, U, $U \subset R^k$ such that $U = \{u \in \mathbb{R}^k | x \in X, u = (f_1(x), ..., f_k(x))\}$ is non empty, convex and closed and $u^0 \in \mathbb{R}^k$ is the utility from which the players accept to bargain. Finally, we denote by U^0 the set $U^0 = \{u \in U | u^0 \leq u\}$, the subset of U in which the players achieve more than their minimum requirements. Similarly, we define $X_0 = \{x \in X | \forall i, f_i(x) \geq u_i^0\}$.

Definition 7 *A mapping $S : (U, u^0) \to \mathbb{R}^n$ is said to be an NBP (Nash bargaining point) if:*

*(1) it guarantees the **minimum required performances**: $S(U, u^0) \in U^0$.*
*(2) $S(U, u^0)$ is **Pareto optimal**.*
*(3) It is **linearly invariant**, i.e. the bargaining point is unchanged if the performance objectives are affinely scaled. More precisely, if $\phi : \mathbb{R}^n \to \mathbb{R}^n$ is a linear map such that $\forall i, \phi_i(v) = a_i v_i + b_i$, then $S(\phi(u), \phi(u^0)) = \phi(S(U, u^0))$.*
*(4) S is **symmetric**, i.e. does not depend on the specific labels. Hence, con-*

nections with the same minimum performance u_i^0 and the same utilities will have the same performances.

(5) *S is* ***not affected by reducing the domain*** *if a solution to the problem with the larger domain can be found on the restricted one. More precisely, if $V \subset U$, and $S(U, u^0) \in V$ then $S(U, u^0) = S(V, u^0)$.*

Definition 8 *The point $u^* = S(U, u^0)$ is called the Nash Bargaining Point and $f^{-1}(u^*)$ is the set of Nash Bargaining Solutions.*

We have the equivalent optimization problem:

Theorem 9 *[29, Thm. 2.1, Thm 2.2] and [10]. Let the utility functions f_i be concave, upper-bounded, defined on X which is a convex and compact subset of $\mathbb{R}^n$. Let J be the set of users able to achieve a performance strictly superior to their initial performance, i.e. $J = \{j \in \{1, ..., N\} | \exists x \in X_0, \text{ s.t. } f_j(x) > u_j^0\}$. Assume that $\{f_j\}_{j \in J}$ are injective. Then there exists a unique NBP as well as a unique NBS x that verifies $f_j(x) > u_j(x), j \in J$, and is the unique solution of the problem:*

$$\max \prod_{j \in J} (f_j(x) - u_j^0), \quad x \in X_0. \tag{4}$$

Equivalently, it is the unique solution of $\max \sum_{j \in J} \ln(f_j(x) - u_j^0), \quad x \in X_0$.

In 1991, [22] adapted the Nash bargaining solution to Jackson networks. Some years after, Yaïche et al. [29] adapted it for bandwidth allocation in networks. In their works however, they restricted themselves to linear utility functions.

We can note that the NBS corresponds to proportional fairness in the case where the utility functions are linear and where the MR_i are null.

Remark 10 (Other fairness criteria defined through a set of axioms) *The last axiom (5) of the Nash Bargaining Point has suffered some criticisms, as it does not take into account how much the other players have given up. Two other interesting fairness criteria were then defined when modifying this last axiom, namely the Raiffa-Kalai-Smorondinsky and the Thomson (or "utilitarian choice rule") solution [8]. We do not treat here in more detail these two criteria as they correspond respectively of special cases of max-min fairness (Section 3.2) and the maximization of the global throughput (Section 3.1) when considering the utility of the applications.*

4.1 Utility functions in networking

The interest of the Nash bargaining concept, as opposed to the previous ones used so far in the telecommunication context is to take into account the different interests the users have for the shared resource [3]. In networks context, the bandwidth does not have the same value for different users. For instance, a user consulting his or her emails does not have the same needs that another one using phone over IP. The utility functions represent the impact of the bandwidth allocation on the perceived quality. We show in the following the shapes of the utility functions for different types of application. Our discussion is qualitative and inspired from the work of Shenker [26]. For numerical results, for instance on audio communications, the reader may refer to [14, 9]. We illustrate the different shapes of the utility functions on Figure 3.

Elastic applications have no real-time requirements and no rate constraints. Typical examples are file transfer or email. Their utility function is concave increasing without a minimum required rate.

"Delay adaptive" or "rate adaptive" applications have soft real-time requirements. Typical examples are voice or video over IP. In such applications, the compression rate of data is computed as a function of the quantity of available resource. The utility functions that we use to represent these applications are slightly different than those in [26]. In [26], the utility is strictly positive for any non zero bandwidth and tends to zero when the bandwidth does. We consider in contrast that the utility equals zero below a certain value, as in [29]. Indeed, in many voice applications, one can select the transmission rate by choosing an appropriate compression mechanism and existing compression software have an upper bound on the compression, which implies a lower bound on the transmission rate for which a communication can be initiated, which we denote MR. Thus, a maximum compression rate is associated with the lower acceptable quality for the user. If there is no sufficient bandwidth, the connection is not initiated. This kind of behavior generates utility functions that are null for bandwidth below MR and which are not differentiable at the point $(MR, 0)$. Similarly, it is useless to allocate a bandwidth greater than a certain threshold PR because the perceived gain for a human being will not be noticeable. As an example, for voice transmission, we usually consider throughputs in the range $[16, 40]$ kb/s. A user to whom we would allocate a throughput of 200 kb/s would not have a better quality feeling than that if its throughput was halved.

4.2 Proposed approximation scheme: quadratic utility functions

The utility functions of both "elastic traffic" and "delay adaptive" applications have a minimum value MR_n below which they equals zero (in the former case,

[3] Note, however, that in some cases are implemented *service classes*. They are so far limited to pricing purposes and the underlying idea is usually to give priority to some packets at the buffers.

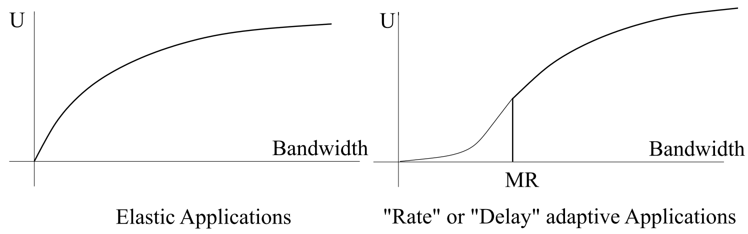


Fig. 3. Utility function in networking

$MR_n = 0$). As explained by Nash, the equilibrium point should not depend on the chosen representation of utility function. Therefore, we assume that $f_n(MR_n) = 0$. Beyond MR_n the function is concave and increasing with the bandwidth. We propose to approximate such a utility function with a parabola with parameters that depend on the applications (see Fig. 4). The general equation has the form:

$$f_n(x_n) = c_n - a_n(x_n - b_n)^2. \tag{5}$$

Note that the utility function is defined only until the point (PR_n, fPR_n), so

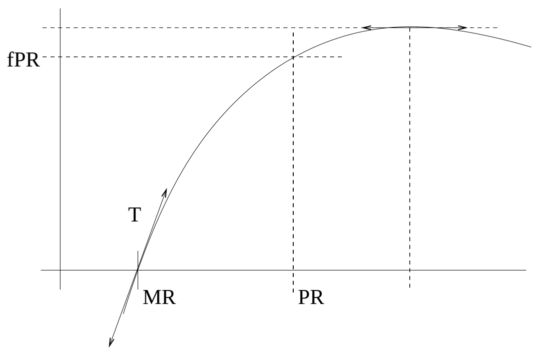


Fig. 4. Quadratic utility functions

we may ignore the whole right part of the parabola (and in particular, the part in which the function decreases).

We introduce:

- T_n the tangent of the utility function at the point $(MR_n, 0)$,
- fPR_n the utility value at point PR_n.

f_n can be equally be defined by a_n, b_n, c_n or through the equations $f_n(MR_n) = 0$, $f_n(PR_n) = fPR_n$ and $f'_n(MR_n) = T_n$. We should note that, since PR_n is in the increasing part of the function, we have $\frac{1}{2}T_i(PR_i - MR_i) \leq fPR_i \leq T_i(PR_i - MR_i)$. We thus define the *concavity of the utility*, β_n, through

$$fPR_n = T_n \cdot \beta_n \cdot (PR_n - MR_n).$$

Note that $1/2 \leq \beta_n \leq 1$ and the smaller β_n is, the more concave is the utility. The limit $\beta_n = 1$ is the linear case (studied in [29]).

We can therefore equivalently use in the following the parameters a_n, b_n c_n or MR_i, T_i and β_i linked by the equations:

$$a_n = T_n \frac{1-\beta_n}{PR_n - MR_n},\ b_n = \frac{PR_n - (2\beta_n - 1)MR_n}{2(1-\beta_n)} \text{ and } c_n = \frac{T_n}{4}\frac{PR_n - MR_n}{1-\beta_n}.$$

5 Proposed fairness scheme

As previously mentioned, we propose to apply the classic fairness criteria used in networks to the utility functions of applications rather than to the throughput. We saw that, by applying this philosophy to the proportional fairness criterium, we obtain the Nash bargaining solution, historically defined by a set of axioms rather than an optimization problem.

We have presented a family of fairness criteria, parameterized by a real α ($\alpha > 0, a \neq 1$) that allows an network operator to choose an equilibrium between fairness and resource utilization.

When applied to the utilities of the connections, the optimization problem becomes:

$$\max_{n \in \mathcal{N}} \frac{[f_n(x_n)]^{1-\alpha}}{1-\alpha}, \text{if } a > 0, a \neq 1 \text{ and } \prod f_n(x_n) \text{ if } \alpha = 1 \tag{6}$$

subject to the problem constraints.

We also propose to approximate utility functions by quadratic parameterized functions. Note that, as in our approximation $f(MR) = 0$, the proposed criterion coincides with the NBS when $\alpha \to 1$. The advantage of using quadratic utilities is that it allows one to approximate the concave real utilities drawn from experiments with functions that are simple to handle through a small set of parameters. Also, it allows us to easily analyze the influence of these parameters on the obtained equilibrium. Of particular interest is the influence of the concavity.

5.1 Property: influence of the concavity

We study the impact of concavity on the NBS. Consider two differentiable functions f and g defined on the same interval $(MR, PR]$ where both are strictly positive on $]MR, PR]$.

Definition 11 *We say that f is more concave than g if for every $x \in (MR, PR]$ the relative derivative of f is smaller than or equal to that of g, i.e. $f'(x)/f(x) \leq$*

$g'(x)/g(x)$. If f or g are not differentiable at x, one could require instead that the same relation holds for the supergradients: if $\hat{f}(x)$ is the largest supergradient of f at x and $\hat{g}(x)$ is the smallest supergradient of g at x, then we require $\hat{f}(x)/f(x) \leq \hat{g}(x)/g(x)$.

Motivated by (4), we say that :

Definition 12 *An assignment $\vec{x}$ is more fair in the sense of NBS than an assignment $\vec{y}$ if $\prod_{n\in\mathcal{N}}(f_n(x_n) - f_n(MR_n)) \geq \prod_{n\in\mathcal{N}}(f_n(y_n) - f_n(MR_n))$.*

Consider the case in which $f_n(MR_n) = 0$. Consider 2 connections with utilities f and g as above competing for the bandwidth of a single link of capacity C. If we had ignored the utilities of the connections, we would have assigned them an equal bandwidth (according to the original proportional allocation), which we denote by $x = C/2$.

Proposition 13 *If two connections are competing for the bandwidth of a single link, the fair assignment has the property that more bandwidth is assigned to the less concave function.*

PROOF. We show that by transferring bandwidth from the connection with the more concave utility (say f) to the other one, we improve the fairness (assuming this does not violate the MR and PR constraints) in the sense of the NBS. Indeed, we have

$$g(x+\epsilon)f(x-\epsilon) = g(x)f(x)\left(1 + \epsilon\left[\frac{g'(x)}{g(x)} - \frac{f'(x)}{f(x)}\right] + o(\epsilon)\right).$$

We conclude that there is some ϵ_0 s.t. $\forall\epsilon < \epsilon_0$, $g(x+\epsilon)f(x-\epsilon) > g(x)f(x)$. Hence we strictly improve the fairness by transferring an amount of ϵ_0 to the connection with less concave utility. By further increasing this amount, we shall eventually reach a local maximum (since our function is continuous over a compact interval). This will be a global maximum since (4) is a maximization problem of a concave function over a convex set. $\diamond$

Example Let two connections of bandwidth x_1 and x_2 sharing a single link of capacity C. Their utility functions f and g are represented in Figure 5.
For $x \geq 0$, $f(x) = 3x\mathbb{1}_{0\leq x\leq 1} + (2+x)\mathbb{1}_{x>1}$, and $g(x) = 2x$.
Then $f'(x)/f(x) = \begin{cases} x^{-1} & \text{for } x \in [0,1[, \\ (2+x)^{-1} & \text{for } x \geq 1, \end{cases}$ whereas $\forall x, g'(x)/g(x) = x^{-1}$.
(At $x = 1$, f is not differentiable but its supergradients at that point constitute the set $[1/3, 1]$). Thus f is more concave than g. We assume that $PR_1 + PR_2 >$

C. The NBS is the argument of

$$\zeta(C) = \max f(x)g(C-x) = \max\left(\max_{x\in[0,1]} h(x), \max_{x>1} k(x)\right)$$

$$\text{with } h(x) = 6x(C-x), \text{ and } k(x) = 2(x+2)(C-x).$$

Proposition 14 *The NBS is depicted in Fig 6. We distinguish 3 regions:*
(i) $C < 2$, *where* $\zeta(C) = 3c^2/2$ *and the NBS is* $(x_1^*, x_2^*) = (C/2, C/2)$,
(ii) $2 \leq C < 4$, *where* $\zeta(C) = 6(C-1)$ *and* $(x_1^*, x_2^*) = (1, C-1)$,
(iii) $C \geq 4$, *where* $\zeta(C) = 2(c/2+1)^2$ *and* $(x_1^*, x_2^*) = (C/2-1, C/2+1)$.

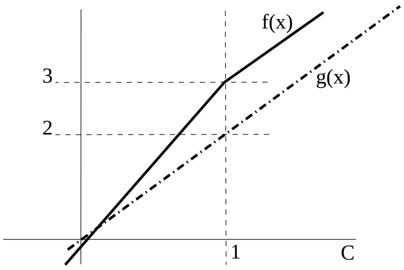


Fig. 5. Utility functions.

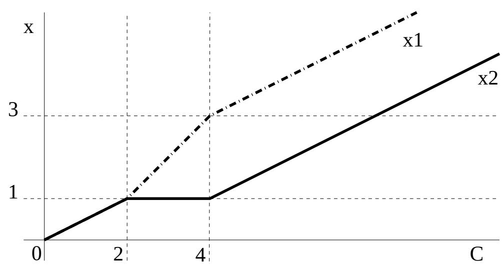


Fig. 6. NBS of two connections sharing a link.

We see in this example that the least concave function receives at least as much as the other one, and the difference increases with C. It is impressive to note that there is a region in which an increase in the capacity benefits only to one connection. The example illustrates the power of the NBS approach: the original proportional fairness, or even weighted proportional fairness, would assign a proportion of the capacity to each connection that does not vary as we increase the capacity, since it is insensitive to the utilities. In contrast, utility sensitive fairness concepts allocate the bandwidth in a dynamic way: the proportion assigned to each connection is a function of the capacity.

For the experimental part, we propose to study 2 different fairness scenarios, that we present in the rest of this section.

5.2 Possible optimizations

We propose two types of optimizations, called *connection-aware* and *network-aware*. The idea in the first is to optimize the allocation to the connections and in the former to better utilize the network resources.

Connection-aware optimization We associate to each connection a utility function and, once the minimal requirements are satisfied, fairly allocate the extra bandwidth among the users. The $\vec{\kappa}$ vector is given and the optimization problem is $\max \sum_{n\in\mathcal{N}} \frac{f_n(x_n)^{1-\alpha}}{1-\alpha}$ with the system's constraints previously mentioned.

Network-aware optimization We apply the fairness concepts to the remaining bandwidth available in the links. Indeed, it can be interesting to maximize the bandwidth available in a link in order to use it for instance for some other traffic type (this is of particular interest if the routing policies for connections cannot be dynamically modified).

Therefore, $\vec{x}$, the bandwidth allocation to connections, is given. We then consider the fair allocation of remaining bandwidth, that is to say of vector $\vec{\kappa}$. The problem can be written $\max \sum_{l\in\mathcal{L}} \frac{f_l(\kappa_l)^{1-\alpha}}{1-\alpha}$. We can note that the allocation vector being given, the optimization problem is of interest only is we consider the fractional-routing. Indeed, if the routing is fixed, then the quantity available in each link is fixed as well.

We make the following assumption in the rest of the article:

Hypothesis 15 *The network has sufficient capacity to satisfy all the minimum requirements.*

In the next section we propose an example of explicit rate computation. This analytic study is possible, however, only in the case of simple network topologies and rely on specific assumptions.

6 Explicit computation of rates

In this section, we study the fair allocations in the linear network and its generalization, the grid network.

6.1 The linear network

We again consider the linear network. We use the obvious associated routing and consider in the following the connection-aware optimization with fixed routing constraints.

We recall that the fairness equilibria is obtained through the optimization of:

$\frac{1}{1-\alpha}\sum_{n\in\mathcal{N}}(f_n(x_n))^{1-\alpha}$ if $\alpha \neq 1$, and $\prod_{n\in\mathcal{N}} f_n(x_n)$otherwise with the constraints:

$$\begin{cases} \forall n \in \mathcal{N}, MR_n \leq x_n \leq PR_n & \text{(user's constraint)}, \\ \forall n \in \mathcal{L}, x_0 + x_n \leq C_n & \text{(capacity constraints)}. \end{cases} \tag{7}$$

Note that in this network all the links can be saturated: $\forall n \in \mathcal{L}, x_0 + x_n = C_n$.

We make two significant assumptions. First, that each link has the same capacity C. Secondly, we suppose that each of the connections $n, n \neq 0$ has the same utility function: $\forall n \in \mathcal{L}, a_n = a_1, b_n = b_1, c_n = c_1$. Thus the fair allocation will assign the same bandwidth to each connection $n, n \neq 0$.

Therefore, by denoting by x the bandwidth allocated to connection 0, the fair allocation problem becomes the maximization of (8) under the constraints (7).

$$\begin{cases} g(x) = \frac{1}{1-\alpha}\left[(f_0(x))^{1-\alpha} + |\mathcal{L}|(f_1(C-x))^{1-\alpha}\right] & \text{if } \alpha \neq 1 \\ h(x) = f_0(x)(f_1(C-x))^{|\mathcal{L}|} & \text{otherwise.} \end{cases} \tag{8}$$

Remark 16 *The problem is of interest only if:*

$$\begin{cases} MR_0 + MR_1 \leq C & \textit{(feasible system, Hypothesis 15)}, \\ PR_0 + PR_1 > C & \textit{(not all maximum demands can be satisfied)}. \end{cases}$$

Solution of the linear problem. By differentiating (8) we obtain:

$$\begin{cases} a_0(x-b_0)f_1(C-x)^\alpha = |\mathcal{L}|a_1(C-x-b_1)f_0(x)^\alpha & \text{if } \alpha \neq 1, \\ a_0(x-b_0)f_1(C-x)^{|\mathcal{L}|} = |\mathcal{L}|a_1(C-x-b_1)f_1(C-x)^{|\mathcal{L}|-1}f_0(x) & \text{otherwise.} \end{cases}$$

We note that for $\alpha = 1$, $f_1(C-x) = 0 \Rightarrow h(x) = 0$. Therefore $f_1(C-x) \neq 0$ and the bandwidth x associated to connection 0 satisfies, for all α:

$$a_0(x-b_0)(f_1(C-x))^\alpha = |\mathcal{L}|a_1(C-x-b_1)(f_0(x))^\alpha. \tag{9}$$

Limits and asymptotic analysis

Proposition 17 *When $|\mathcal{L}|$ grows to infinity, the allocation of connection* 0 *tend to* $\max(MR_0, C - b_1)$.

The proof is given in Appendix. Note that the limit does not depend on the concavity of the utility of connection 0. We show in Figure 6.1 how the system converge to x_{lim} as $\mathcal{L}$ grows to infinity in the case of the NBS ($\alpha = 1$) and $C - b_1 \geq MR_0$.

We further refine the analysis of the limit when $|\mathcal{L}|$ becomes large:

Proposition 18 *If $C - b_1 < MR_0$ then x is such that*

$$x - MR_0 \sim Z \text{ with } \quad Z^\alpha \sim_{l \in \mathcal{L}} \frac{1}{|\mathcal{L}|} \left(\frac{1}{2(b_0 - MR_0) \cdot a_0} \right)^{\alpha - 1} \frac{f_1(C - MR_0)^\alpha}{f_1'(C - MR_0)}.$$

$$\text{Otherwise, } x = C - b_1 + Z + o(1/|\mathcal{L}|) \text{ with } \quad Z = \frac{1}{|\mathcal{L}|} \frac{c_1^\alpha}{2a_1} \frac{f_0'(C - b_1)}{f_0(C - b_1)^\alpha}.$$

PROOF. If $C - b_1 < MR_0$, then, as $|\mathcal{L}| \to \infty$, the left hand side of (9) tends to a non-null constant: $\lim_{|\mathcal{L}| \to \infty} a_0(x - b_0)(f_1(C - x))^\alpha = a_0(MR_0 - b_0)(f_1(C - MR_0)^\alpha)$.

We now examine the right hand side of (9). It can be written as:

$$\begin{aligned} |\mathcal{L}| a_1 (C - x - b_1) f_0(x)^\alpha &= |\mathcal{L}| a_1 (C - b_1 - MR_0 - z)(c_0 - a_0(MR_0 + z - b_0)^2)^\alpha \\ &= |\mathcal{L}| a_1 (C - b_1 - MR_0 - z)(2a_0(b_0 - MR_0)z - a_0 z^2)^\alpha \\ &\sim_{|\mathcal{L}| \to \infty} |\mathcal{L}| a_0^\alpha a_1 2^\alpha (C - b_1 - MR_0)(b_0 - MR_0)^\alpha z^\alpha. \end{aligned}$$

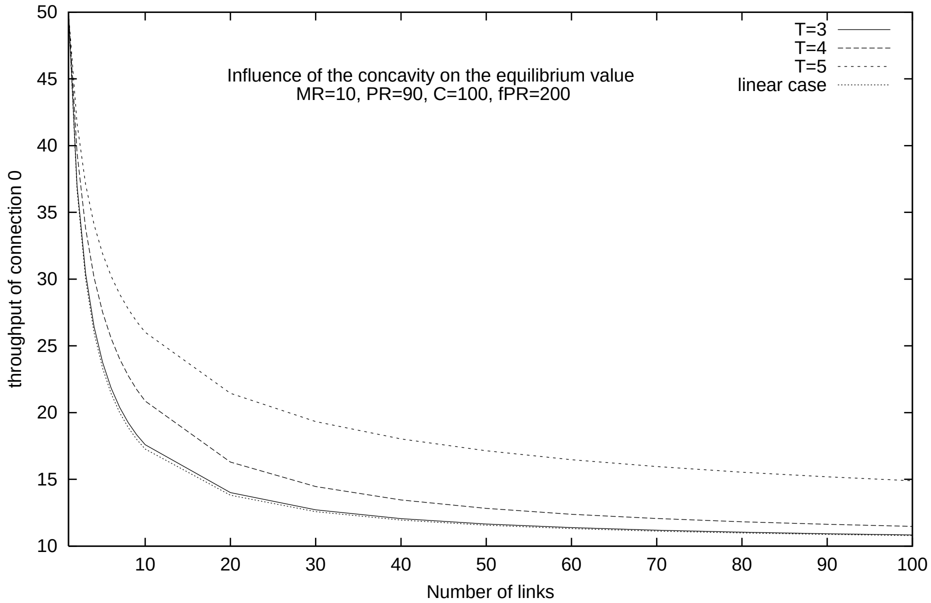
which yields (18) by substituting the appropriate expressions.

If $MR_0 \leq C - b_1$, as $|\mathcal{L}| \to \infty$, the left hand side tends to the constant $a_0(C - b_1 - b_0)c_1^\alpha$. The right hand side is: $-|\mathcal{L}| a_1 z \Big(c_0 - a_0(C - b_1 - b_0 + z)^2 \Big)$. Hence the result. $\diamond$

Remark 19 *In (18), we can check the asymptotes for special cases (and for $\alpha = 1$):*

- *If $\beta_1 \to 1$ then $Z = \dfrac{C - MR_0 - MR_1}{|\mathcal{L}|}$ (linear case, already obtained in [29]).*
- *If $\beta_1 \to 1/2$ then $Z = \dfrac{C - MR_0 - MR_1}{2|\mathcal{L}|} \left[1 - \dfrac{PR_1 - MR_1}{C - MR_0 - PR_1} \right]$.*

We studied the case of a linear network with two kinds of connections. In the case where $\alpha = 1$ (NBS), we note (from (9)) that the problem is a third order



polynomial and can therefore be explicitly solved. If $\alpha = 0$ (global optimization), the allocation is given by the solution of a first order polynomial. In any other case, only numerical results can be used. Still, we can make the limits of the allocations explicit for any value of α, and give an equivalent when the number of links becomes large.

6.2 Grid network

This network is the natural generalization of the linear network. It consists of $K \times L$ capacity links with K horizontal routes and L vertical routes as shown in Figure 7. We focus here on the fixed routing policy.

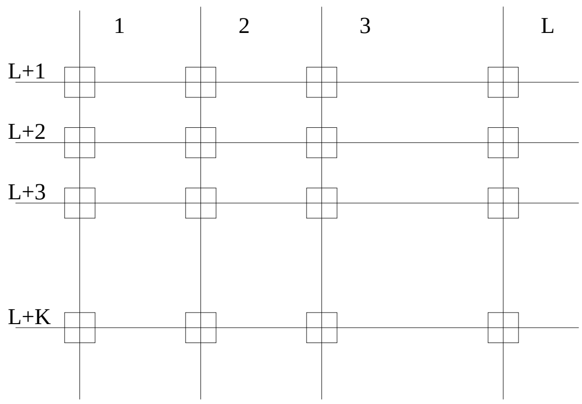


Fig. 7. A grid network.

As in the previous example, we also suppose that, for each $i \in \{1, \ldots L\}$,$j \in \{1, \ldots K\}$, $MR_i + PR_{L+j} \geq C_{i,L+j}$ and $PR_i + MR_{L+j} \geq C_{i,L+j}$. We further suppose that the links have equal capacity. We suppose that all the horizontal connections (respectively vertical) have the same utility function f_h (respectively f_v). We can then conclude that all the horizontal connections (vertical) will get the same throughput x ($x_v = C - x$). We then wish to maximize, in the case of the NBS:

$$\prod_{n \in [1:L] \bigcup [L+1, L+K]} f_n(x_n) = (f_h(x))^K * (f_v(C-x))^L. \tag{10}$$

Proposition 20 *In the grid network, if* $C - b_1 < MR_0$*,* x *verifies:* $x - MR_h \sim Z$ *with:* $Z = \frac{K}{L}\left(\frac{1}{2(b_h - MR_h)a_h}\right)^{\alpha-1} \frac{f_v(c-MR_0)}{f_v'(c-MR_0)}$*. Otherwise,* $x = C - b_1 + Z + o(K/L)$ *with* $Z = \frac{K}{L}\frac{c_v^\alpha}{2a_v}\frac{f_h'(c-b_v)}{f_h'(c-b_v)}$*.*
Note that if $L = K$ *and* $f_h = f_v$ *we obtain* $x = C/2$*.*

PROOF. Maximizing Equation 10 is similar to maximizing $f_h(x) * (f_v(C - x))^{L/K}$. This is equivalent to the linear problem by substituting $|\mathcal{L}|$ to L/K. $\diamond$

7 Lagrangian method

The Lagrangian method was proposed by [29] to obtain the NBS for the special case of linear utility function. It has the advantage of having distributed implementations. We generalize below this approach to the quadratic utility, for which the linear case can be recovered by taking $\beta \to 1$.

Unfortunately, the resulting decentralized iterative algorithm of the problem proposed in [29] is not fully satisfactory for quadratic utilities, as we will see in the following.

7.1 Lagrangian multipliers

We now use the Kuhn-Tucker conditions for (6) to obtain an alternative characterization of the NBS in terms of the corresponding Lagrange multipliers under the hypothesis of fixed routing and the connection aware optimization.

Proposition 21 *Under the hypothesis that* $\forall l \in \mathcal{L}, \sum a_{ln} MR_n < C_l$ *(this amounts in saying that no link is saturated*[4] *), the NBS is characterized by:*

[4] This assumption does not constraint our problem. If a link is saturated when considering the minimal demands of the connections, then we rewrite the system by suppressing this link and the connections that were going through it. The bandwidth

$\exists \mu_l \geq 0, l \in \mathcal{L}$ *such that* $\forall n \in \mathcal{N}$, *we have*

$$x_n = \min \left(PR_n, MR_n + \left(\sum_{l \in \mathcal{L}} \mu_l a_{l,n} \right)^{-1} + \right.$$
$$\left. \tfrac{1}{2} \cdot \frac{PR_n - MR_n}{1 - \beta_n} \times \left[1 - \sqrt{1 + \frac{4 \left(\frac{1-\beta_n}{PR_n - MR_n} \right)^2}{\left(\sum_{l \in \mathcal{L}} \mu_l a_{l,n} \right)^2}} \right] \right).$$

PROOF. Under the assumption $\sum_{n \in \mathcal{N}} a_{ln} MR_n < C_l$, the set of possible solutions of (6) is non-empty, convex and compact. The constraints (as defined in Subsection 2.3) are linear in x_n and $f(x) = \sum_{n \in \mathcal{N}} \ln f_n(x_n)$ is C^1. Therefore the first order Kuhn-Tucker conditions are necessary and sufficient for optimality. The Lagrangian associated with (6) is

$$\mathcal{L}(x, \lambda, \delta, \mu) = f(x) - \sum_{n \in \mathcal{N}} \lambda_n (MR_n - x_n) - \sum_{n \in \mathcal{N}} \delta_n (x_n - PR_n) - \sum_{l \in \mathcal{L}} \mu_l ((Ax)_l - C_l).$$

For $n \in \mathcal{N}$, $\lambda_n \geq 0$ and $\delta_n \geq 0$ are the Lagrange multipliers associated with the constraints $x_n \geq MR_n$ and $x_n \leq PR_n$ respectively. $\mu_l \geq 0, l \in \mathcal{L}$ are the Lagrange multipliers associated with the capacity constraints. The first order optimality conditions are thus: $\forall n \in \mathcal{N}$,

$$\frac{f'_n(x_n)}{f_n(x_n)} = \delta_n + \sum_{l \in \mathcal{L}} \mu_l A_{l,n} - \lambda_n, \quad (x_n - MR_n)\lambda_n = 0, \quad (x_n - PR_n)\delta_n = 0,$$

$$\text{and } \forall l, l \in \mathcal{L}, ((A\vec{x})_l - C_l)\mu_l = 0$$

Moreover, $\sum a_{ln} MR_n < C_l$ implies that $\forall n, \lambda_n = 0$ as in [29], and either $x_n = PR_n$ or $\delta_n = 0$, which yields the conclusion. $\diamond$

Remark 22 *As $\beta \to 1$ we obtain the solution of [29] corresponding to linear utility:* $x_n = \min \left(PR_n, MR_n + \left[\sum_{l \in \mathcal{L}} \mu_l a_{l,n} \right]^{-1} \right)$.

$\mu_l, l \in \mathcal{L}$ represent the implied cost associated with the network link l. They represents the marginal cost of a rate unit allocated for any connection crossing link l.

---

that these connections were using is then a new capacity constraint on the other links of the network.

Once we have explicitly expressed the NBS in terms of the Lagrange multipliers, we can actually solve the NBS using the dual problem in which we compute the Lagrange multipliers. The dual problem is :

$$\max_{\mu \in \mathbb{R}_+^{\mathcal{L}}} d(\mu) \quad \text{with} \quad d(\mu) = \min_{\vec{x} \in X} L(\vec{x}, \mu) = L(\overline{x}, \mu)$$

if we denote by $\overline{x}$ the optimal allocation. The vector $\overline{x} = \overline{x_1, x_2 ... x_n}$ is the NBS. We now use the result obtained in the primal, and, for a given vector μ, we note for each connection n: $x_n(\mu) = g_n(\sum\limits_{l \in \mathcal{L}} \mu_l \cdot a_{l,n})$.

$$\text{With } g_n(p) = \begin{cases} PR_n & \text{if } p \leq \dfrac{2\beta_n - 1}{\beta_n} \dfrac{1}{PR_n - MR_n}, \\ \\ MR_n + \dfrac{1}{p} + \dfrac{PR_n - MR_n}{2(1-\beta_n)} \left[1 - \sqrt{1 + \left(\frac{2}{p} \frac{1-\beta_n}{PR_n - MR_n}\right)^2}\right] & \text{otherwise.} \end{cases}$$

We obtain for each $\mu \in \mathbb{R}^{\mathcal{L}}$:

$$d(\mu) = \sum_{n \in \mathcal{N}} -\ln(f_n(g_n(\sum_{l \in \mathcal{L}} \mu_l a_{l,n}))) + \sum_{l \in \mathcal{L}} \mu_l \sum_{n \in \mathcal{N}} a_{l,n} g_n\Big(\sum_{l \in \mathcal{L}} \mu_l a_{l,n}\Big) - \sum_{l \in \mathcal{L}} C_l \mu_l.$$

The idea is then to use a appropriate step γ and choose at each iteration:

$$\mu_l^{(k+1)} = \max\left(0, \mu_l^{(k)} + \gamma \frac{\partial d}{\partial \mu_l}\right). \tag{11}$$

But, if the utilities are not linear, then the partial derivative of d depends on their parameters. Therefore, the dual cannot be used to obtain a fully distributed algorithm anymore since every link needs to know the utility functions and allocations of all the connections of the system.

The Lagrangian relaxation let us introduce a optimization problem where each user computers its allocation x_n

$$x_n = g_n(\sum \mu_l a_{l,n})$$

while the network computes the link prices μ_l iteratively from Equation 11.

In the case of proportional fairness ($\alpha = 1$) with linear utility functions, the dual formulation can be used to obtain a decentralized algorithm. Unfortunately, the use of quadratic utilities require the knowledge in the network of all the connections and parameters of connections, which makes this method not suitable for large systems.

8 A Semi-Definite Programming (SDP) solution

In this section, we propose an alternative centralized method for solving the general fairness problem (6). It uses a mathematical program called Semi-Definite Program, which can be solved in polynomial time in theory and is tractable in practice. One can then use public domain programs to solve SDP [5]. SDP solves the minimization problem of a linear combination of variables (given by the scalar product of a vector L and the vector of variables) subject to a constraint of positive semi-definiteness (psd) of some general symmetric matrix P whose entries are either variables or constants. More details on SDP can be found in [5]. We study the complexity of our method and show that the proposed construction of the SDP matrix allows us to analyze large networks.

In the first subsection, we see how SDP was defined as an extension from linear programming and then recall some basic results of linear algebra. We describe the general shape of our SDP matrices to solve our optimization problem (Subsection 8.2), and the details for the different values of α (Subsection 8.3). We finally present a detailed description for a simple example (Subsection 8.4).

8.1 From linear programming to SDP

Definition 23 (Linear programming) *A linear program (LP) has the form:*

$$\min\{c^T x | Ax \geq b\}$$

where x is the unknown vector, c is a given coefficient vector, $A \in M_{m,n}$ is a $m \times n$ matrix of constraints and b is the vector of constraints.

We can extend linear to conic programming. In any Euclidian Space E we can define convex pointed cones ([13], [25]). Then, each cone $\mathcal{K}$ induce a partial order in E, that we denote $\geq_{\mathcal{K}}$: $a \geq_K b \Leftrightarrow a - b \geq_K 0 \Leftrightarrow a - b \in \mathcal{K}$.

[5] see http://www.cs.nyu.edu/cs/faculty/overton/sdppack/sdppack.html

Definition 24 (Conic programming) *Let $\mathcal{K}$ be a convex pointed cone of E, $c \in \mathbb{R}^n$ an objective vector, $b \in E$ and A a linear mapping* $A : \begin{cases} x \mapsto Ax \\ \mathbb{R}^n \to E. \end{cases}$. *Then, the optimization problem* $\min_x\{c^T x | Ax \geq_K b\}$ *is a conic problem.*

Semi Definite Programming refers to the problems associated to $\mathcal{K} = S_m^+$ the cone of semi-definite positive matrices in the Euclidian space $E = S^m$ the set of symmetric matrices of size $m \times m$.

Definition 25 *A symmetric matrix A of size $m \times m$ is said semi-definite positive and we note $A \succeq 0$ if $\forall x \in \mathbb{R}_m, x^T Ax \geq 0$. Equivalently, all its eigenvalues*[6] *are positive or null.*

Definition 26 (Semi-Definite Programming) *A semi-definite programming problem is an optimization problem of the form*

$$\min c^T.vec(X) \quad \textit{such that} \begin{cases} A.vec(X) = b \\ X \succeq 0. \end{cases}$$

with $X \in S_n^+$, $c \in \mathbb{R}^{n^2}$, $b \in \mathbb{R}^m$ and $A \in M_{m,n^2}$. [7]

The following are results of linear algebra needed to understand our solving method. The proofs can be found in [13].

Proposition 27 (Real symmetrical matrices.) *All the eigenvalues of a real symmetric matrix are real.*

Proposition 28 (Symmetric matrices of size 2) *Let a, b and c be real positive numbers. Then, the matrix* $M = \begin{pmatrix} m_{1,1} & m_{1,2} \\ m_{2,1} & m_{2,2} \end{pmatrix}$ *is positive semi-definite (psd) if and only if* $m_{1,1}m_{2,2} \geq M_{1,2}^2$.

Proposition 29 (Symmetric matrices of size 1) *A matrix of size 1 (a scalar) is positive if and only if its unique element is positive or null.* $M = (m_{1,1}) \succeq 0 \Leftrightarrow m_{1,1} \geq 0$.

---

[6] The eigenvalues are the solutions of the polynomial $\det(A - \lambda I)$.

[7] If X is the matrix $X = (x_{i,j}), 1 \leq i, j \leq n$ then $vec(X)$ is the vector $(x_{1,1}, \ldots, x_{1,n}, x_{2,n}, \ldots x_{n,n})$.

Proposition 30 *A bloc diagonal matrix*[8] *is psd if and only if all its blocs are psd.*

8.2 General solving method with SDP

The general idea of SDP is to transform the original maximization of a *function* into a minimization problem of some new *variable* (or more generally of a linear combination of variables) subject to a constraint of positive semi-definiteness (psd) of some general matrix P. The positiveness (psd) of the matrix will:

- insure the system constraints,
- replace the objective function (of (6)) by a single variable.

We will see that in our case, we can construct the matrix with blocs of size 1 and 2. Note that the order of the blocs in the matrix is not relevant. Note further that the number of variables cannot be known a priori. Indeed, the creation of the matrix and of the blocs induce the creation of several intermediate variables that we will precise in the following.

Finally, the entries we will give to the SDP solver are the matrix and a vector of variable. In our case, this vector will be composed of "0" at each but one entry, corresponding to the final variable we want to maximize (this entry will contain the value "-1").

In the following two paragraphs, we explicit the blocs expressing the system constraints and those that enable us to transform the users utility functions into variables.

System constraints As seen in Subsection 2.3 the system constraints are linear. From Proposition 29 these constraints can be expressed via matrices of size 1 (scalar matrices).

We give in this paragraph the constraints representing the system with fixed routing (the fractional routing is similar). The capacity constraints are given by the positivity of $|\mathcal{L}|$ matrices, corresponding to the $|\mathcal{L}|$ links. Indeed, we recall that these constraints are $\forall l \in \mathcal{L}, (Ax)_l \leq C_l$. Then, it is sufficient to introduce $|\mathcal{L}|$ scalar matrices $capa^l$, such that: $\forall l \in \mathcal{L}, capa^l_{1,1} = C_l - (Ax)_l$.

---

[8] A bloc diagonal matrix has the form $A = \begin{pmatrix} \boxed{A_1} & & & 0 \\ & \boxed{A_2} & & \\ & & \ddots & \\ 0 & & & \boxed{A_n} \end{pmatrix}$.

The users constraints are given by: $\forall n \in \mathcal{N}, MR_n \leq x_n \leq PR_n$. We therefore introduce $2 \times |\mathcal{N}|$ scalar matrices *util* so that $util^i_{1,1} = PR_i - x_i$ if $i \leq |\mathcal{N}|$ and $util = x_i - MR_i$ if $|\mathcal{N}| < i \leq 2|\mathcal{N}|$.

Relations between the variables The initial variables are the bandwidth allocated to connections $x_n, n \in \mathcal{N}$. We introduce the intermediate variables $w_n, n \in \mathcal{N}$ with the following proposition:

Proposition 31 *Let x_n be a positive number and f_n its utility function, as defined in Equation 5. We introduce $F_n = \begin{pmatrix} -\dfrac{w_n - c_n}{a_n} & x_n - b_n \\ x_n - b_n & 1 \end{pmatrix}$. Then $F_n \succeq 0 \Leftrightarrow w_n \leq f_n(x_n)$.*

By using the blocs $F_n, n \in \mathcal{N}$, we replace the initial variables $x_n, n \in \mathcal{N}$ by $w_n, n \in \mathcal{N}$ and instead of maximizing a function of x_n, we maximize a function of w_n. Indeed, by distinguishing three cases depending of the sign of $\alpha - 1$, we can prove that $\forall \alpha \geq 0, \frac{1}{1-\alpha} \sum w_n^{1-\alpha} \leq \frac{1}{1-\alpha} \sum f_n(x_n)^{1-\alpha}$. We can note, from this simple example that we replaced the utility functions by a constraint of psd of a matrix. Let us suppose we have $\mathcal{N}$ variables y_n such that $\forall n \in \mathcal{N}, y_n \leq \frac{w_n^{1-\alpha}}{1-\alpha}$ (for a given α). Then, the psd of scalar matrix END, defined by $END_{1,1} = (\sum_{k \in \mathcal{N}} y_k - z)$ insures $z \leq \sum \frac{w_n^{1-\alpha}}{1-\alpha}$ and maximizing z leads to the desired optimization since $z \leq \sum_n y_n \leq \frac{w_n^{1-\alpha}}{1-\alpha} \leq \frac{f_n(x_n)^{1-\alpha}}{1-\alpha}$.

We provide, in the following subsection a method to construct matrices whose psd constraint will insure that (we omit the "n" subscript):

$$y \leq w^{1-\alpha} \quad \text{if } 1 - \alpha > 0, \qquad y \geq w^{1-\alpha} \quad \text{otherwise.}$$

8.3 Different values of α

The first paragraph deals with the case $\alpha \neq 1$ whereas second concerns the most complex case of the NBS ($\alpha = 1$).

8.3.1 Case $\alpha \neq 1$

Remark 32 (Case $\alpha = 2$) *The positiveness of matrix $H = \begin{pmatrix} w & 1 \\ 1 & y \end{pmatrix}$ ensures that $y \geq 1/w$.*

We use an idea of Nemirovski to provide a resolution method with a good approximation for any value of $\alpha > 0$ with $\alpha \neq \{1, 2\}$.

If $0 < \alpha < 1$, we have $1 - \alpha > 0$. Therefore, it is sufficient to provide one or several matrices whose psd will ensure that $y \leq w^{1-\alpha}$.

Proposition 33 (Case $0 < \alpha < 1$**)** *Then* $1 - \alpha < 1$ *and*

$$\forall \varepsilon > 0, \exists p \in \mathbb{N}, k \in \{0, \ldots, 2^p - 1\}, |(1-\alpha) - k/2^p| \leq \varepsilon.$$

Let $w, y \in \mathbb{R}^+$*. It is possible, using SDP constraints, to bound* y *and* w *by the relation* $y \leq w^{k/2^p}$*.*

PROOF. Let $c_1, \ldots, c_p$ a serie of 0/1 integers, such that $k = \sum_{i=1}^{p} c_i 2^{i-1}$. We note $y_0 = 1$ and submit $y_1, \ldots, y_p$ to the following constraints:

$$\begin{pmatrix} y_{i-1} & y_i \\ y_i & w \end{pmatrix} \text{ if } c_i = 1 \text{ and } \begin{pmatrix} y_{i-1} & y_i \\ y_i & 1 \end{pmatrix} \text{ if } c_i = 0.$$

Then obviously $y_i^2 \leq y_{i-1} w^{c_i}$, and $y_p \leq w^{k/2^p}$. Hence the result, by setting $y_p = y$. ◇

If $\alpha > 1$ we aim at finding matrices those psd constraint will ensure that $y \geq w^{1-\alpha}$.

Proposition 34 (Case $1 < \alpha < 2$**)** *Since* $0 < -(1-\alpha) < 1$ *we have:*

$$\forall \varepsilon > 0, \exists p \in \mathbb{N}, k \in \{0, \ldots, 2^p - 1\}, |-(1-\alpha) - k/2^p| \leq \varepsilon.$$

Let $w, y \in \mathbb{R}^+$*. It is possible, using SDP constraints, to bound* w *and* y *by the relation* $w \geq y^{-k/2^p}$*.*

PROOF. Let c be an intermediate variable. Using Proposition 33, one can set $c \leq y^{\beta}$. Also, one can write: $\begin{pmatrix} y & 1 \\ 1 & w \end{pmatrix} \succeq 0$ which leads to $yw \geq 1$. Then $yw^{\beta} \geq 1$. ◇

The following proposition covers the cases $\alpha \in [2; +\infty[$.

Proposition 35 (Case $2 < \alpha$**)** *We have* $0 < \frac{1}{1-\alpha} < 1$*. Then*

$$\forall \varepsilon > 0, \exists p \in \mathbb{N}, k \in \{0, \ldots, 2^p - 1\}, \left|\frac{1}{1-\alpha} - k/2^p\right| \leq \varepsilon.$$

Let $y, w \in \mathbb{R}^+$. It is possible, using SDP constraints to bound w et y by the relation $y \geq w^{2^p/k}$.

PROOF. Similarly to the proof of Proposition 33, we obtain $wy^\beta \geq 1$. $\diamond$

8.3.2 Computing the NBS (case $\alpha = 1$)

We propose in this section a method to compute the NBS (case $\alpha = 1$). The method used is slightly different than from the other cases since we don't use the intermediate variables $y_n, n \in \mathcal{N}$. We directly exhibit a serie of matrices whose psd insure that $z \leq \prod_n w_n$. The result for the NBS relies on the following:

Proposition 36 *Let z and $w_1, \ldots, w_n$ be real positive numbers. Then, using SDP constraints, it is possible to bound these numbers by the relation:*

$$z^{2^{\lceil \log_2(N) \rceil}} \leq \prod_{i=1}^{N} w_i.$$

PROOF. Let p be the smallest integer such that $2^p \geq N$. We construct a family of real positive variables $z_{i2^k+1,(i+1)2^k}$ with $1 \leq k \leq p$ and $i \in \{0, \ldots, 2^{p-k} - 1\}$ satisfying the constraints expressed by the following $2^p - 1$ matrices:

$$\begin{pmatrix} z_{2i2^{k-1}+1,(2i+1)2^{k-1}} & z_{i2^k+1,(i+1)2^k} \\ z_{i2^k+1,(i+1)2^k} & z_{(2i+1)2^{k-1}+1,(2i+2)2^{k-1}} \end{pmatrix} \succeq 0,$$

where we denote $z_{jj} = w_j$ for $j \in \{1, \ldots, n\}$ and $z = z_{1,2^p}$. Then we obtain constraints of the forms $z_{1,2}^2 \leq w_1 w_2$, $z_{3,4}^2 \leq y_3 y_4$, $z_{1,4}^2 \leq z_{1,2} z_{3,4}$ and finally $z_{1,2^p}^{2^p} \leq \prod_i w_i$. $\diamond$

To solve the NBS problem, we propose to add artificial connections so that the total number of connections is of the form 2^p and use the previous proposition. We also create fictitious links that these connections will use so that they do not modify the share of the others. It is important to bound the bandwidth allocated to these connections so that the share obtained by SDP does not grow without control that would provoke an error. As these fictitious connections use their own link, no matter the value of their upper bound PR_i, we will obtain $x_i = PR_i$.

The following proposition shows that the blocs of the matrix are of size at most 2 and that their number remains reasonable (of the order of $\mathcal{O}(\mathcal{N} + \mathcal{L})$). The SDP approach can hence offer simple and fast solutions for dimensioning purposes or for studying existing system, and is suitable for the study of large networks.

Proposition 37 *We stress that the proposed construction will lead to at most* $6|\mathcal{N}| - 7$ *variables,* $4|\mathcal{N}| - 5$ *blocks of size 2, and* $4|\mathcal{N}| + |\mathcal{L}| - 4$ *blocks of size 1.*

PROOF. Let p be the smallest integer such that $2^p \geq |\mathcal{L}|$ and $q = 2^p$. Then, one can check that, our method will lead to: $q + \sum_{i=0}^{p} q2^{-i} = 3.2^p - 1$ variables, $q + \sum_{i=1}^{p} q2^{-i} = 2^{p+1} - 1$ blocks of size 2 and $l + 2.2^p = 2^{p+1} + l$ blocks of size 1. In the worst case, we have $|\mathcal{L}| = 2^{p-1} + 1$, hence the result. $\diamond$

8.4 A simple example of NBS computation with fixed routing

We finally illustrate in this sub-section the previous results with the construction of the matrix for the computation of the NBS on a simple network. We consider a network with $|\mathcal{L}| = 4$ links and $|\mathcal{N}| = 3$ connections. The routing matrix is:

$$A = \begin{pmatrix} 1 & 0 & 1 \\ 0 & 1 & 0 \\ 0 & 1 & 1 \\ 1 & 0 & 0 \end{pmatrix}.$$

Recall that element $a_{i,j}$ equals 1 if and only if connection j uses link i. In our SDP program, we add an artificial connection so that the total number of connections has the form 2^p with $p \in \mathbb{N}$ as explained in 8.3.1. We suppose that this extra connection uses its own link and therefore does not modify the NBS of our problem. A possible network associated with this problem is represented in Figure 8.

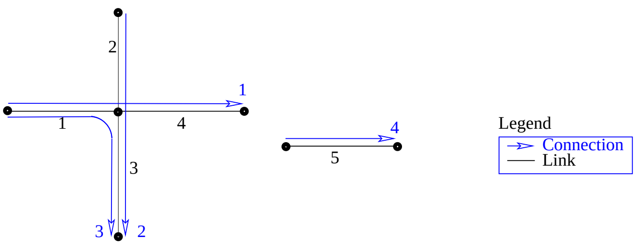


Fig. 8. Simple computation of the NBS.

The first four blocs of the matrix link the variables x_n with their utility w_n:

$$\mathrm{MAT}_{1,n} = \begin{pmatrix} -\frac{w_n - c_n}{a_n} & x_n - b_n \\ x_n - b_n & 1 \end{pmatrix}$$

The following matrices link the w_i variables together to obtain a single variable that SDP will maximize (from Proposition 36).

$$\mathrm{MAT}_{2,1} = \begin{pmatrix} w_1 & y_{1,2} \\ y_{1,2} & w_2 \end{pmatrix}, \mathrm{MAT}_{2,2} = \begin{pmatrix} w_3 & y_{3,4} \\ y_{3,4} & w_4 \end{pmatrix}, \mathrm{MAT}_{2,3} = \begin{pmatrix} y_{1,2} & y_{1,4} \\ y_{1,4} & y_{3,4} \end{pmatrix}.$$

The positiveness of these matrices implies that $(y_{1,4})^4 \leq (y_{1,2})^2 \cdot (y_{3,4})^2 \leq w_1 \cdot w_2 \cdot w_3 \cdot w_4$. Then, maximizing the single variable $y_{1,4}$ will lead to the required optimization of $\prod_n (c_n - a_n(x_n - b_n)^2)$.

We incorporate the (linear) constraints of the problem. The constraints $(Ax)_l \leq C_l$ lead to the declaration of L matrices that are, in our example:

$$\mathrm{MAT}_{3,1} = (C_1 - (x_1 + x_3)), \quad \mathrm{MAT}_{3,2} = (C_2 - x_2), \quad \mathrm{MAT}_{3,3} = (C_3 - (x_2 + x_3))$$

$$\text{and } \mathrm{MAT}_{3,4} = (C_4 - x_1).$$

Finally, the constraints $x_i \leq PR_i$ and $x_i \geq MR_i$ are reflected by 8 scalar matrices:

$$\mathrm{MAT}_{3,4+n} = (PR_n - x_n), \quad \text{and} \quad \mathrm{MAT}_{3,8+n} = (x_n - MR_n), \quad 1 \leq n \leq 4.$$

We can notice that the values PR_4 and MR_4 corresponding to the artificial connection are not important since the connection is independent of the others and SDP will give $x_4 = PR_4$.

The entries we should give to the SDP algorithm are the matrix obtained by concatenation of the blocs we described and the vector L. As we want to maximize $y_{1,4}$, the vector should be of the form: $L = (0, 0, 0, 0, 0, 0, 0, 0, 0, 0, -1)$ (or an equivalent one if we renumber the variable in a different order). We can notice that we have in this simple example 11 variables: 4 allocation variables x_n, 4 utility variables w_n and 3 intermediate variables $y_{1,2}$, $y_{3,4}$ and $y_{1,4}$.

We present in the last section some numerical results obtained with the csdp solver. The matrices we used were obtained according to the method we previously explained.

9 Numerical experiments

We implemented the SDP approach using a Matlab program on a SUN ULTRA 1 computer to obtain the fair shares. We first tested our program on the same linear network example for which we had explicit expressions, and the results completely agreed. We then considered two more complex networks which we describe below. The computation time (including the display part) in both cases was less than a minute.

The utility functions are chosen identical for each connection (although the program can handle different parameters without increasing the complexity) and are $MR = 10$, $PR = 80$, $a = 1/490$, $b = 745$ and $c = 1102.5$. The bandwidth parameters and the allocations are given in percentage of the total capacity of the link. For each network, we present a figure showing the set of links and a figure representing the bandwidth allocated to each connection. Finally, in the connection-aware case, the links have the same capacity $C = 100$, while in the case of network-aware optimization experiments, the link sizes and the connections requirements are different one from another.

9.1 NBS for a small network with fixed routing

We consider the network represented in Figure 9. It contains $|\mathcal{L}| = 10$ links and $|\mathcal{N}| = 11$ connections. The routing is characterized by matrix A:

$$A = \begin{pmatrix} 0&1&0&0&0&0&0&1&0&0&0 \\ 0&1&0&0&0&0&0&1&1&0&0 \\ 0&1&0&0&0&0&0&0&1&0&0 \\ 0&0&0&0&0&0&0&0&1&0&0 \\ 1&0&0&1&0&0&0&0&0&1&0 \\ 1&0&0&1&1&0&0&0&0&1&0 \\ 1&0&0&0&1&0&0&0&0&1&1 \\ 0&0&1&0&1&0&1&0&0&1&1 \\ 0&0&1&0&0&0&0&0&0&0&1 \\ 0&0&0&0&0&1&1&0&0&1&0 \end{pmatrix}$$

The SDP solution is given in Figure 10, in which each connection is represented by a line whose gray scale represent its allocation. We can note that the SDP formulation required to introduce 36 variables and the SDP matrix was of size 104 (31 matrices of size 2, and 42 of size 1).

We can note that all the links cannot be saturated simultaneously even although the solution is Pareto optimal. This is for instance the case of links 1,

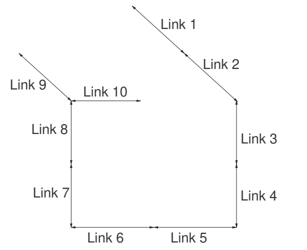


Fig. 9. Small network with fixed routing.

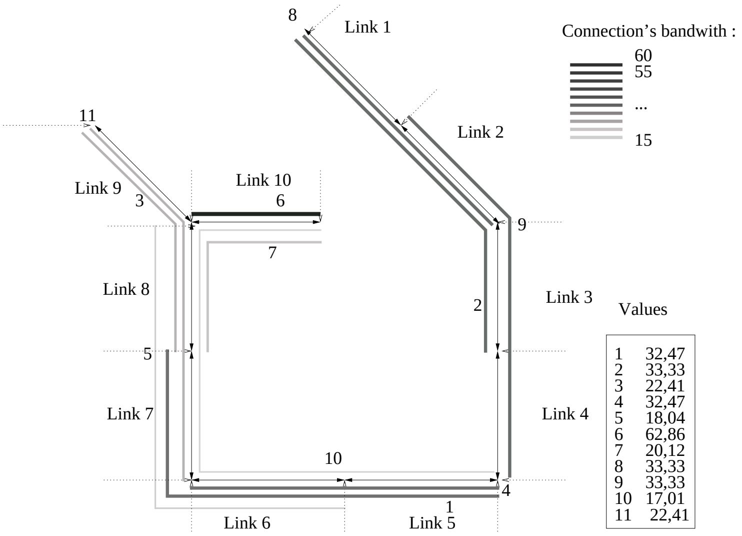


Fig. 10. First network: solution

3 and 4. Moreover, we can distinguish 2 independent systems:

- The one formed by the links 1 to 4 and used by exclusively by connections 2, 8 and 9,
- The system that consists of links 5 to 11, used by the other connections.

Note that in the first system, only one link is saturated. This link being used by the three connections, each of them received one third of its capacity (since they have the same utility function).

The second system is more complex and we cannot determine, without the help of computers the different allocations of the connections. Still, the numerical results agree with our expectations since connection 6, using only one link, obtain the highest allocation (almost 63% of the link capacity), whereas connection 10, crossing the maximum number of links, obtain the smaller allocation (17% of the link capacity). In the case of max-min optimization these differences would have been, of course, smaller.

Finally, let us recall that the critical element in the allocation of a connection is the set of saturated links and not the total number of links crossed by a connection. Thus, in our example although connection 3 crosses more links than connection 11, it receives the same allocation. Indeed, the extra link it goes through is link 7, which is not saturated. The same observation holds for connections 1 and 4.

We then considered the COST network represented in Figure 11. It contains 11 nodes representing major European cities. We present two optimization schemes, corresponding respectively to the connection-aware optimization with fixed routing and to the network-aware optimization with fractional routing.

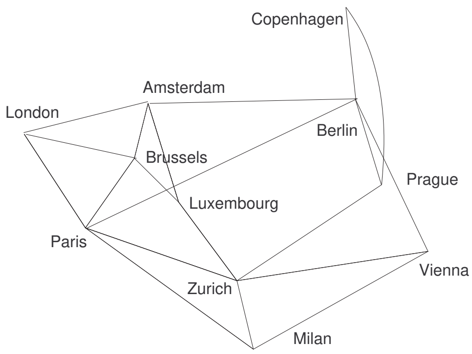


Fig. 11. COST network: links.

NBS in the case of connection-aware optimization with fixed routing. We considered in the simulation the 30 connections having the higher forecast demands [9]. The solution obtained in the case of the NBS is given in Figure 12 and the results are summarized in Table 2. The routing was chosen arbitrarily in order to minimize the number of links crossed by each connection.

Pa - Mi	47.34	Mi - Vi	63.00	Pr - Co	80.00	Zu - Pa - Lo	21.87	Zu - Pr - Be	50.00
Pa - Lo	33.93	Pr - Be	50.00	Be - Am	27.11	Pa - Zu - Vi	25.48	Zu - Lu - Br	35.79
Pa - Be	80.00	Lo - Br	80.00	Zu - Vi	55.06	Be - Am - Lu	27.11	Lo - Am - Be	23.73
Pa - Br	43.66	Lo - Am	76.27	Vi - Be	63.00	Mi - Vi - Be	37.00	Pa - Br - Am	28.42
Pa - Zu	33.19	Am - Br	49.54	Mi - Pa - Br	27.93	Be - Am - Br	22.04	Mi - Zu - Lu - Am	28.42
Mi - Zu	71.58	Co - Be	80.00	Mi - Pa - Lo	24.74	Zu - Lu - Am	35.79	Vi - Zu - Pa - Lo	19.46

Legend: **Co** : Copenhagen, **Be**: Berlin, **Am**: Amsterdam, **Lo**: London, **Br**: Brussels, **Pa**: Paris, **Lu**: Luxembourg, **Pr**: Prague, **Zu**: Zurich, **Vi**: Vienna, **Mi**: Milan.

Table 2
Bandwidth allocation in COST networks.

The solution involved adding extra 65 intermediate variables, and the size of the psd matrix was 215 (63 matrices of size 2, and 89 matrices of size 1). As in the previous example, the connections having the higher allocations

[9] We withdrew the connections whose demands would be lower than 2.5 Gb/s according to experiments dating from 1993.

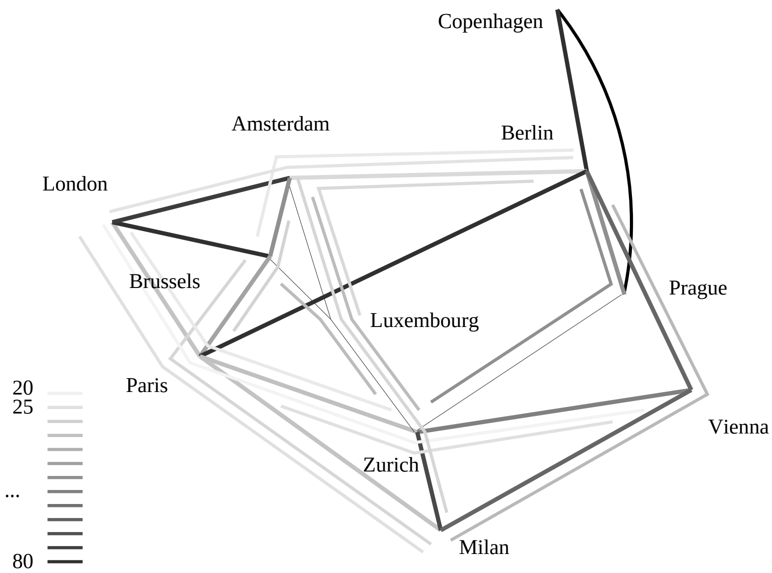


Fig. 12. COST network: NBS allocation.

are represented with darker colors. We can again note that the connections using the smaller number of (saturated) links received more bandwidth. We can even point out that four connections (Copenhagen-Prague, Copenhagen-Berlin, London-Brussels and Paris-Berlin) are alone in their respective link, and are therefore allocated their maximum demand, that is to say 80% of the link capacity. Those links are thus not saturated. The allocations of connections Berlin-Prague and Berlin-Prague-Zurich are equal and are 50% of the link capacity because they compete for the same bottleneck link (the reason is therefore similar that for connections 1 and 4 of the previous network).

Network-aware optimization We then considered the case of network-aware optimization and fractional routing. In this example we added 3 links: London-Copenhagen, Amsterdam-Copenhagen and Luxembourg-Prague to increase the possible routing strategies. Moreover we got interested in comparing different values of α. We present here the results obtained for $\alpha = 0.5$ and $\alpha = 5$.

The link capacity is given in arbitrary units and we considered the 110 possible demands (corresponding to all the possible pairs source-destination). The demands are different from one to another, as illustrated in Table 3. Here, although the problem size is much higher than the previous example (due to the increase of the number of links and connections), the computation time remained lower than one minute. The link usages (in percentage) for $\alpha = 0.5$ and $\alpha = 5$ are given in the Tables 4 and 5 respectively and graphically represented in Figures 13 and 14.

Pa - Mi	5	Pa - Zu	6	Pa - Lu	1	Am - Lu	1	Lo - Pr	1	Pr - Mi	1	Pr - Be	2	Zu - Lo	3
Pa - Lo	10	Pa - Vi	2	Pa - Am	5	Pa - Co	1	Lo - Br	4	Lo - Am	5	Pr - Zu	1	Zu - Br	6
Pa - Pr	1	Pa - Be	11	Mi - Am	2	Mi - Lu	1	Lo - Be	8	Lo - Co	1	Pr - Br	1	Pr - Lu	1
Pa - Br	6	Mi - Be	9	Mi - Co	1	Zu - Lu	1	Am - Br	4	Br - Co	1	Lu - Lo	1	Pr - Am	1
Mi - Vi	3	Mi - Br	2	Am - Co	1	Lu - Br	1	Be - Br	6	Co - Be	3	Vi - Co	1	Vi - Lo	2
Mi - Zu	6	Mi - Lo	3	Zu - Am	3	Zu - Co	1	Vi - Br	1	Be - Lu	2	Pr - Vi	1		
Zu - Vi	3	Zu - Be	11	Pr - Co	1	Be - Am	8	Vi - Be	9	Vi - Lu	1	Vi - Am	1		

Table 3
COST network: demands

Link	BP	Link	BP	Link	BP	Link	BP	Link	BP	Link	BP	Link	BP
Pa - Mi	57	Zu - Pr	87	Vi - Be	96	Am - Lo	61	Mi - Zu	61	Pa - Br	68	Mi - Br	73
Pa - Zu	68	Pr - Vi	55	Be - Am	92	Am - Co	29	Mi - Vi	81	Pr - Lu	32	Am - Lu	62
Pa - Be	97	Pr - Be	93	Be - Co	81	Am - Br	90	Zu - Lu	73	Lu - Br	61	Br - Lo	62
Pa - Lo	91	Zu - Vi	41	Pr - Co	61	Lo - Co	87						

Table 4
Results for $\alpha = 0.5$

Link	BP	Link	BP	Link	BP	Link	BP	Link	BP	Link	BP	Link	BP
Pa - Mi	52	Mi - Zu	73	Zu - Lu	76	Am - Lo	73	Mi - Br	76	Pa - Lo	78	Pr - Co	74
Pa - Zu	73	Pr - Vi	76	Be - Am	92	Am - Co	47	Vi - Be	92	Am - Br	77	Lo - Co	73
Pa - Be	92	Pr - Be	92	Be - Co	92	Lu - Br	73	Br - Lo	74	Mi - Vi	73	Pa - Br	76
Zu - Pr	76	Pr - Lu	47	Am - Lu	74	Zu - Vi	47						

Table 5
Results for $\alpha = 5$

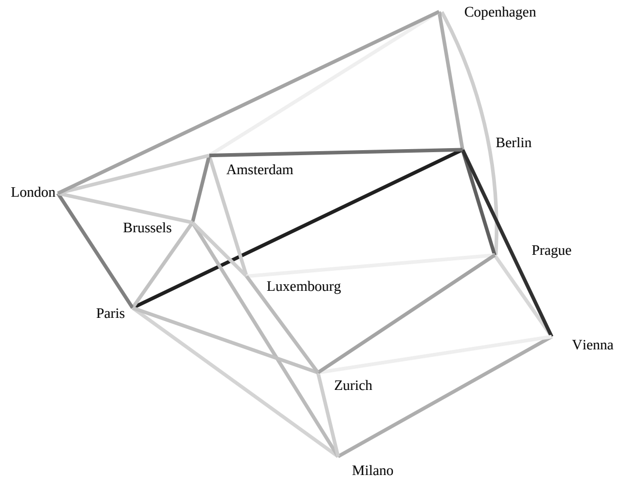


Fig. 13. Bandwidth allocation for COST network, $\alpha = 0.5$.

We observe that, when the parameter α grows, we tend to a more homogenous coloration of the links, which agrees with theoretical results. When α grows, the allocations are more "equaled" and we thus observe a more homogeneous use of the links. Accordingly, the use of the more saturated link decrease from 97.36% to 92%.

When α grows from 0.5 to 5, the average link usage increases (from 70.38%

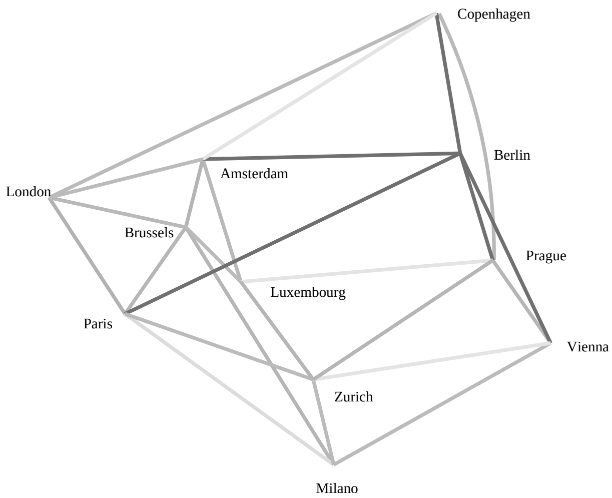


Fig. 14. Bandwidth allocation for COST network, $\alpha = 5$.

to 73.83%). This can seem surprising, since the allocated bandwidth remain constant, as well as the link capacity. In fact, when α grows, the traffic becomes more split between the different possible routes in order make the link usage more homogeneous. Therefore the average number of links used by a connection increase as well and so does the network usage.

It is also interesting to note that the usage of some links, like the Berlin-Copenhagen, have increased almost to their maximum usage for the solution $\alpha = 5$, supposedly more "fair" ! This shows that the search for fairness in a global scale sometimes lead some users to critical situations while their remain in their bound in cases that are less fair.

10 Conclusion

We have applied in this paper the NBS approach for bandwidth allocation that is sensitive to the utilities of connections. We have studied some of the characteristics of these concepts, and showed that they are indeed more suitable for applications that have concave utility. We proposed a simple parametrization of the concavity of the utility function using quadratic functions. We finally proposed some computational approaches that allows us to handle large networks: a Lagrangian approach and a novel approach based on SDP.

The SDP approach has the advantage that it is simple to implement using any general SDP software. Furthermore additional conditions (for instance those linked to integer programming, or other telecommunication requests) can be introduced without requiring any study on the stability or the convergence of the algorithm. This is a clear improvement compared to many other specific methods, and in particular to the iterative ones that have been proposed for solving fairness problems.

Also, we considered a fairness approach that takes into account not just the assigned throughput of each connection but instead the utility that the assigned throughput represents. This type of fairness concept agrees with the game theoretic definition of fairness given by NBS and is an interesting generalization of the fairness criteria that have been used so far in the telecommunication context.

References

[1] Cost 239. Ultra high capacity optical transmission networks, final report of Cost project 239. *ISBN 953-184-013-X*, 1999.

[2] Traffic management specification, version 4.0. Technical Report AF-TM-0056.000, ATM Forum Traffic Management Working Group, April 1996.

[3] V. Bharghavan, S. Lu, and T. Nandagopal. Fair queuing in wireless networks, issues and applications. *IEEE Personal Communications*, 6(1), Feb 1999.

[4] J.Y Le Boudec. *Rate Adaptation, Congestion Control and Fairness: A Tutorial.* Ecole Polytechnique Federale de Lausanne, Dec 2000.

[5] S. Boyd, L. El Ghaoui, E. Feron, and V. Balakrishnan. *Linear matrix inequalities in system and control theory.* Springer, 1994.

[6] X. Cao, H. Shen, R. Milito, and P. Wirth. Internet pricing with a game theoretical approach : Concepts and examples. *IEEE/ACM Trans. on Networking*, 10(2):208–216, Apr 2002.

[7] X. R. Cao. Preference functions and bargaining solutions. In *IEEE Conf. Decision and Control*, pages 164–171, 1982.

[8] X.R Cao. Preference functions and bargaining solutions. In *Proc. of IEEE CDC*, pages 164–171, Florida, Dec. 1982.

[9] G. W. Cermak. Subjective quality of speech over packet networks as a function of packet loss, delay and delay varitation. *International Journal of Speech Technology*, 5(1):65–84, Jan 2002. ISSN 1381-2416.

[10] A. Ştefănescu and M. V. Ştefănescu. The arbitrated solution for multi-objective convex programming. *Rev. Roum. Math. Pure Applicat.*, 29(7):593–598, 1984.

[11] Z. Dziong. *ATM Network Resource Management.* McGraw-Hill, 1997.

[12] M. Haviv. The aumann-shappley price mechanism for allocating costs in congested systems. *Operation research letters*, 29:221–229, 2001.

[13] C. Helmberg. *Semidefinite Programming for Combinatorial Optimization.* ZIB-Report ZR-00-34, Konrad-Zuse-Zentrum fr Informationstechnik Berlin, Takustrae 7, D-14195 Berlin, Germany, October 2000.

[14] J. Janssen, D. De Vieeschauwer, M. Bchli, and G. H. Petit. Assessing voice

quality in packet based telephony. *IEEE Internet Computing*, 6(3):48–56, 2002.

[15] K. Kar, S. Sarkar, and L. Tassiulas. Optimization based rate control for multirate multicast sessions. In *Proc. of IEEE INFOCOM*, April 2001.

[16] K. Kar, S. Sarkar, and L. Tassiulas. A simple rate control algorithm for maximizing total user utility. In *Proc. of IEEE INFOCOM*, April 2001.

[17] F. P. Kelly. Charging and rate control for elastic traffic. *European Trans. on Telecom.*, 8:33–37, 1998.

[18] F. P. Kelly, A. Maulloo, and D. Tan. Rate control for communication networks: shadow prices, proportional fairness and stability. *Journal of the Operational Research Society*, 49, 1998.

[19] S. Kunniyur and R. Srikant. A time scale decomposition approach to adaptive ECN marking. In *Infocom*, pages 1330–1339, 2001.

[20] S. Low, L. Peterson, and L. Wang. Understanding TCP Vegas: A Duality Model. In *Proc of ACM SIGMETRICS*, pages 226–235, June 2001.

[21] L. Massoulié and J. Roberts. Bandwidth sharing: objectives and algorithms. In *Proc. of IEEE INFOCOM*, volume 3, pages 1395–1403, 1999.

[22] R. Mazumdar, Lorne G. Mason, and C. Douligeris. Fairness in network optimal flow control: optimality of product forms. *IEEE Trans. on Comm.*, 39:775–782, 1991.

[23] J. Mo and J. Walrand. Fair end-to-end window-based congestion control. In *SPIE '98, International Symposium on Voice, Video and Data Communications*, 1998.

[24] J. Nash. The bargaining problem. *Econometrics*, 18:155–162, 1950.

[25] A. Nemirovski. Five lectures on modern convex optimization. CORE Summer School on Modern Convex Optimization, August 2002.

[26] S. Shenker. Fundammental design issues for the future internet. *IEEE J. on Selected Areas in Comm.*, 1995.

[27] C. Touati, E. Altman, and J. Galtier. Utility based fair bandwidth allocation. In *Proc. of the IASTED International Conference on Networks, Parallel and Distributed Processing and Applications*, pages 126–131, Japan, Oct. 2002.

[28] C. Touati, E. Altman, and J. Galtier. Semi-definite programming approach for bandwidth allocation and routing in networks. *Game Theory and Applications*, 9:169–179, Dec 2003. Nova Publisher.

[29] H. Yaiche, R. R. Mazumdar, and C. Rosenberg. A game theoretic framework for bandwidth allocation and pricing in broadband networks. *IEEE/ACM Transactions on Networking*, 8(5):667–677, 2000.

A Proof of Proposition 17

We first prove the following lemma:

Lemma 38 *As $|\mathcal{L}|$ grows to infinity, the only possible limits x_{lim} of the bandwidth assigned to connection* 0 *are* MR_0 *and* $C_0 - b_1$.

PROOF. Since x is bounded (from $MR_0 \leq x \leq PR_0$), the left part of (8) is bounded too. Therefore, the limit x_{lim}, if any, is such that $a_1(C_0 - x - b_1) \cdot f_0(x) = 0$.

We want to determine which of the three solutions are possible, that is to say belonging to the set $[MR_0, PR_0]$. From (7) we have $PR_0 + PR_1 > C$. As $b_1 > PR_1$ (from (5)), then $C - b_1 < PR_0$. Therefore the solution $x_{lim} = C - b_1$ is acceptable if $C - b_1 \geq MR_0$. The two other solutions are the zeros of f_0. By definition the larger one is strictly greater than PR_0 and therefore not acceptable. $\diamond$

The second lemma compares the possible limits.

Lemma 39 *Let f and g be defined as in (8). Then $g(C - b_1) \geq g(MR_0)$ and $h(C - b_1) \geq h(MR_0)$.*

PROOF. We have $g(MR_0) = \dfrac{|\mathcal{L}|}{1-\alpha} f_1(C - MR_0)^{1-\alpha}$ and
$g(C - b_1) = \dfrac{1}{1-\alpha}\left[f_0(C - b_1)^{1-\alpha} + |\mathcal{L}|(f_1(b_1))^{1-\alpha}\right]$.

From the definition of b_1 we have $f_1(b_1) \geq f_1(C - MR_0)$. Moreover, if $C - b_1 \geq MR_0$ then $f_0(C - b_1) \geq 0$ and therefore $g(C - b_1) \geq g(MR_0)$.

If $\alpha = 1$, then $h(MR_0) = 0$ and $h(C - b_1) = f_0(C - b_1)(f_1(b_1))^{\mathcal{N}} > 0$. $\diamond$

We can finally prove Proposition 17. We have seen that for any value of α, MR_0 and $C - b_1$ are the only two possible limits. Moreover, if $C - b_1 \geq MR_0$ then $g(C - b_1) \geq g(MR_0)$ and $h(C - b_1) \geq h(MR_0)$. Therefore, the limit of the allocation, if any, is $\max(C - b_1, MR_0)$.

We can show that this value is indeed the limit of the allocation. We detail here the case of $\alpha = 1$. The case of $\alpha \neq 1$ being similar. If $C - b_1 > MR_0$, then $f_1(C - Id)$ is maximal for $C - b_1$ (by definition of b_1) and $f_0(C - b_1) > 0$ (since $C - b_1 > MR_0$), then, $lim_{|\mathcal{L}| \to \infty} f_0(x)(f_1(C - x))^{|\mathcal{L}|} = C - b_1$.

Suppose now that $C - b_1 \leq MR_0$. $f_1(C - \mathrm{Id})$ is a parabola those maximum is for $C - b_1 \leq MR_0$ and that is null in MR_0. Therefore the function is decreasing in $[b_1, +\infty]$, and in particular in $[C - MR_0, C - PR_0]$ (since $C - b_1 > MR_0$). Therefore its maximum is achieved in $C - MR_0$. As f_0 is positive on the set $[MR_0, PR_0]$ then $x_{lim} = MR_0$.