Impact of fairness on Internet performance — **Source link**

Thomas Bonald, Laurent Massoulié

Institutions: Orange S.A., Microsoft

Related papers:

- Fair end-to-end window-based congestion control

- Rate control for communication networks: shadow prices, proportional fairness and stability

- Bandwidth sharing and admission control for elastic traffic

- Stability and performance analysis of networks supporting elastic services

- Fluid model for a network operating under a fair bandwidth-sharing policy

Impact of Fairness on Internet Performance

Thomas Bonald, Laurent Massoulié

HAL Id: hal-01272522
https://hal.archives-ouvertes.fr/hal-01272522

Impact of Fairness on Internet Performance

Thomas Bonald
France Telecom R&D
thomas.bonald@francetelecom.com

Laurent Massoulié
Microsoft Research
lmassoul@microsoft.com

ABSTRACT

We discuss the relevance of fairness as a design objective for congestion control mechanisms in the Internet. Specifically, we consider a backbone network shared by a dynamic number of short-lived flows, and study the impact of bandwidth sharing on network performance. In particular, we prove that for a broad class of fair bandwidth allocations, the total number of flows in progress remains finite if the load of every link is less than one. We also show that provided the bandwidth allocation is "sufficiently" fair, performance is optimal in the sense that the throughput of the flows is mainly determined by their access rate. Neither property is guaranteed with unfair bandwidth allocations, when priority is given to one class of flow with respect to another. This suggests current proposals for a differentiated services Internet may lead to suboptimal utilization of network resources.

1. INTRODUCTION

Until recently, a major design objective of congestion controls mechanisms in the Internet has been the fairness of the resulting bandwidth shares made available to the users. Roughly speaking, fairness means that no user is penalized compared to others that share the same bottleneck links. This objective is approximately achieved in the current Internet as it is known that, with FIFO packet scheduling at the network links, TCP shares bottleneck bandwidth in inverse proportion to the round trip time between the source and destination. In contrast, recent proposals for the provision of differentiated services [3] no longer aim at fairness, and suggest to discriminate between traffic from different classes by introducing per-class scheduling policies at the network links.

Fairness is most frequently considered in a static regime where a fixed set of source-destination pairs share network resources for the transfer of infinite sized documents [11, 13, 14, 16, 17]. In reality, the number of flows in progress is highly dynamic, increasing at the instants of some arrival process and decreasing as the transfer of finite sized documents is completed. There is a very strong interaction between the stochastic process describing the numbers of flows in progress on different network routes and the way in which these flows share bandwidth. Very little is still known about the way fairness affects the performance perceived by users. Notable exceptions are the results derived by De Veciana et al. [7] and by Massoulié and Roberts [15]. In this paper we generalize these results and deduce some key characteristics from the analysis of a number of toy network topologies.

Earlier work [11, 13, 14, 16, 17] has analysed which equilibrium bandwidth allocations result from a number of scheduling and congestion control mechanisms, assuming a fixed population of long-lived transfers. Based on these results, we consider a dynamic population of short-lived transfers, assuming the underlying packet level mechanisms realize certain bandwidth sharing objectives perfectly. Specifically, each transfer, which in practice consists of a succession of erratically spaced packets following a certain route, is assimilated here to a fluid flow whose rate adjusts instantaneously as the number of other flows interfering directly or indirectly changes. This modelling approach relies on a "time scale separation" assumption: the time scale of the *flow level* dynamics we are interested in (that is, the duration of a flow) is assumed to be much longer than the time scale of the *packet level* dynamics considered in [11, 13, 14, 16, 17] (that is, the time of convergence to equilibrium assuming a fixed population of long-lived flows). We thus expect the approach to better capture the performance of so-called "elephants" rather than "mice".

We consider the performance of different fairness notions as measured by the average response time of transfers. Of particular interest are the stability conditions under which the response times remain finite. These conditions critically depend on the load offered to each link, defined as the ratio between the traffic intensity (flow arrival rate × average flow document size) and the link capacity. We show that, for arbitrary network topologies and a broad class of fair bandwidth allocations, stability holds if and only if the load offered to each link is less than one. We also show on simple examples that, for unfair allocations resulting from class-based scheduling, the number of flows in progress (and the associate response times) can grow unbounded while the load offered to each link is strictly less than one.

We then turn to the evaluation of average response times

when these remain finite. An equivalent performance measure is the flow throughput, defined as the average document size divided by the average time required to transfer a document. We show by analysis and simulation of a number of toy network topologies that for a broad range of fair bandwidth allocations and offered load "not too close" to one, performance is optimal in the sense that the flow throughput is mainly determined by fixed constraints (due to modem speed or maximum size of TCP window, for instance) and not by the capacity of the backbone links. In constrast, unfair bandwidth allocations can result in arbitrarily bad response times for the penalized flows, at loads considerably less than one.

The paper is organized as follows. In Section 2, we review various fairness notions and illustrate these on several examples. Section 3 is devoted to stability issues. The main result there is Theorem 1, which gives the stability conditions for arbitrary network topologies and a broad class of fair bandwidth allocations. Throughput performance under stable conditions is analysed in Section 4. Section 5 concludes the paper.

2. BANDWIDTH SHARING
In this section, we give a formal definition of the notion of Pareto-efficiency and introduce the bandwidth allocation strategies to be considered in the sequel. In order to compare different allocations we introduce the following orders. We denote by $\leq$ the natural (partial) order on $\mathbb{R}^N$, i.e. $x \leq y$ if and only if (iff) $x_i \leq y_i$ for all i. We denote by $\leq_{lg}$ the lexicographical (total) order, i.e. $x \leq_{lg} y$ iff $x = y$ or there exists some i such that $x_i < y_i$ and $x_j = y_j$ for all $1 \leq j < i$. Finally, we note $x \preceq y$ if $\sigma(x) \leq_{lg} \sigma(y)$ where $\sigma(x)$ and $\sigma(y)$ are the ordered versions of x and y, i.e., $\sigma(x)_i$ and $\sigma(y)_i$ are the i-th smallest elements of x and y, respectively.

2.1 Pareto-efficiency
Consider a backbone network as a set of links $\mathcal{L}$ where each link $l \in \mathcal{L}$ has a capacity $C_l > 0$. A number of flows compete for access to these links, each flow being associated with a route consisting of a subset of $\mathcal{L}$. We write $l \in r$ when route r goes through link l. Let $\mathcal{R}$ denote the set of routes and let x_r denote the number of flows on route $r \in \mathcal{R}$. In this section, we assume that x_r is fixed, that is there are no flow arrivals and each flow consists of the transfer of a file of unlimited size. Without loss of generality, each flow on route r is assumed to have the same bandwidth allocation. This allocation is denoted by λ_r.

DEFINITION 1. *A bandwidth allocation $\lambda = \{\lambda_r\}_{r \in \mathcal{R}}$ is called feasible if it satisfies the capacity constraints:*

$$\sum_{r \ni l} x_r \lambda_r \leq C_l, \quad \forall l \in \mathcal{L}. \tag{1}$$

The set of feasible bandwidth allocations is a convex set of $|\mathcal{R}|$-dimensional positive vectors. In the rest of the paper, we implicitly consider feasible bandwidth allocations only.

DEFINITION 2 (PARETO-EFFICIENCY). *A bandwidth allocation λ is said to be Pareto-efficient if for any bandwidth allocation μ such that $\mu \geq \lambda$, it then holds that $\lambda = \mu$.*

This condition ensures that no bandwidth is "wasted": for each route r, there exists some link $l \in r$ where the capacity constraint is attained. Equivalently, the bandwidth allocated to a given flow cannot be increased without decreasing the bandwidth allocated to some other flow.

2.2 Max-min fairness
A classical notion of fairness in data networks is *max-min fairness* [2]. This gives absolute priority to flows of lower throughput. More formally, max-min fairness is the fairest bandwidth allocation according to the following definition.

DEFINITION 3. *A bandwidth allocation λ is said to be fairer than another bandwidth allocation μ if and only if $\lambda \succeq \mu$.*

It is worth noting that max-min fairness is the only Pareto-efficient allocation such that no route is "penalised": for each route r, there exists some link $l \in r$ where the capacity constraint is reached and λ_r is the maximum bandwidth allocated to the flows that go through link l. Equivalently, the bandwidth allocated to a given flow cannot be increased without decreasing the bandwidth allocated to a flow of equal or lower throughput.

2.3 Mean throughput criteria
An intuitively appealing bandwidth allocation optimization criterion is to maximize the mean throughput of the flows. It turns out that familiar allocations correspond to this objective when the mean is arithmetic, geometric and harmonic, respectively:

- **Arithmetic mean.** Maximizing $\sum_r x_r \lambda_r / \sum_r x_r$ is equivalent to maximizing the overall throughput of the flows. This strategy does not lead to a unique bandwidth allocation.

- **Geometric mean.** Maximizing $[\prod_r \lambda_r^{x_r}]^{1/\sum_r x_r}$ is equivalent to maximizing $\sum_r x_r \log(\lambda_r)$. This strategy leads to a unique bandwidth allocation known as *proportional fairness* [11].

- **Harmonic mean.** Maximizing $\sum_r x_r / \sum_r (x_r/\lambda_r)$ is equivalent to minimizing $\sum_r x_r/\lambda_r$. This strategy leads to a unique bandwidth allocation satisfying the *minimal potential delay* criterion [14].

All three bandwidth allocations are Pareto-efficient.

2.4 A general criterion
A more general bandwidth sharing criterion was introduced by Mo and Walrand [16]. For a given positive constant $\alpha \neq 1$, consider the optimization problem:

$$\text{Maximize} \sum_r x_r \frac{\lambda_r^{1-\alpha}}{1-\alpha} \tag{2}$$

subject to capacity constraints (1). Due to the strict concavity of the function to be maximized, this defines a unique allocation referred to as α-bandwidth allocation. This allocation corresponds to the *maximum throughput* criterion

when $\alpha \to 0$, to *proportional fairness* when $\alpha \to 1$, to the *potential delay* criterion when $\alpha \to 2$, and to *max-min fairness* when $\alpha \to \infty$ [16].

A weight $w_r > 0$ can be introduced for each route r leading to a more general criterion that can be written:

$$\text{Maximize } \sum_r w_r x_r^\alpha \frac{(\lambda_r x_r)^{1-\alpha}}{1-\alpha} \quad (3)$$

subject to capacity constraints (1). Again, this defines a unique allocation referred to as weighted α-bandwidth allocation. It turns out that the bandwidth allocation achieved by the congestion avoidance algorithm of TCP is well approximated by (3) with $\alpha = 2$ and w_r equal to the reciprocal of the squared round-trip time on route r [1, 13, 17].

We now illustrate the possible allocation strategies on simple network topologies where closed form expressions can be derived.

Example 1 (Linear network).
Consider the linear network depicted in Figure 1. In this and other figures of this section, the squares represent the links and the lines represent the routes. The linear network consists of L unit capacity links with $x_0 > 0$ flows on route 0, which crosses every link, and x_l flows on route l, which uses link l alone, for $1 \le l \le L$. The weighted α-bandwidth allocation is such that $x_0 \lambda_0 + x_l \lambda_l = 1$ for all l such that $x_l > 0$, and:

$$x_0 \lambda_0 = \frac{(w_0 x_0^\alpha)^{1/\alpha}}{(w_0 x_0^\alpha)^{1/\alpha} + \left(\sum_{l=1}^L w_l x_l^\alpha\right)^{1/\alpha}}.$$

Proof. Since the allocation is Pareto-efficient, all links l such that $x_l > 0$ are saturated. The proof follows then from the fact that the rate λ_0 allocated to long flows is the maximum over $[0,1]$ of the function:

$$w_0 x_0^\alpha \frac{(x_0 \lambda_0)^{1-\alpha}}{1-\alpha} + \sum_{l=1}^L w_l x_l^\alpha \frac{(1-x_0 \lambda_0)^{1-\alpha}}{1-\alpha}.$$
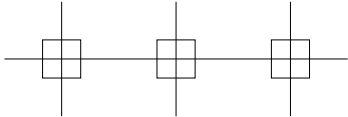
$\square$

Figure 1: Linear network

Example 2 (Cyclic network).
We consider the cyclic network depicted in Figure 2. The network consists of $2L$ unit capacity links with $2L$ routes of length L. Route l crosses links $l+1, \ldots, l+L$ (we let $2L + l \equiv l$ by convention). Assume $x_l > 0$ for all l. The weighted α-bandwidth allocation is such that all links are saturated and:

$$x_l \lambda_l = \frac{(w_l x_l^\alpha + w_{L+l} x_{L+l}^\alpha)^{1/\alpha}}{\sum_{l=1}^L \left(w_l x_l^\alpha + w_{L+l} x_{L+l}^\alpha\right)^{1/\alpha}}, \quad 1 \le l \le 2L.$$

Proof. It can be easily verified that $x_l \lambda_l = x_{L+l} \lambda_{L+l}$ for all l, and that all links are saturated. The weighted α-bandwidth allocation is then the unique solution of the optimization problem:

$$\text{Maximize } \sum_{l=1}^L (w_l x_l^\alpha + w_{L+l} x_{L+l}^\alpha) \frac{(x_l \lambda_l)^{1-\alpha}}{1-\alpha},$$

subject to the capacity constraint $\sum_{l=1}^L x_l \lambda_l = 1$. The corresponding Lagrangian is given by:

$$\sum_{l=1}^L (w_l x_l^\alpha + w_{L+l} x_{L+l}^\alpha) \frac{(x_l \lambda_l)^{1-\alpha}}{1-\alpha} + \mu \left(1 - \sum_{l=1}^L x_l \lambda_l\right),$$

where $\mu > 0$ is the Lagrange multiplier. Taking the derivative of this expression with respect to $\lambda_1, \ldots, \lambda_L$ yields:

$$(w_l x_l^\alpha + w_{L+l} x_{L+l}^\alpha)(x_l \lambda_l)^{-\alpha} = \mu, \quad \text{for } l = 1, \ldots, L.$$

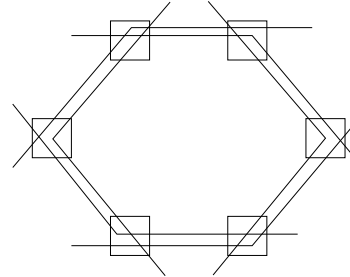The proof then follows from the fact that $\sum_{l=1}^L x_l \lambda_l = 1$. $\square$

Figure 2: Cyclic network

Example 3 (Grid network).
A natural generalization of the linear network considered above is the grid depicted in Figure 3. The network consists of $K \times L$ unit capacity links with K horizontal routes and L vertical routes. Horizontal route k crosses links (k,l), $1 \le l \le L$ and vertical route l crosses links (k,l), $1 \le k \le K$. Let w_k denote the weight of horizontal route k and v_l that of vertical route l. Let x_k be the number of flows on horizontal route k and y_l the number of flows on vertical route l. For any horizontal route k such that $x_k > 0$, the weighted α-bandwidth allocation satisfies:

$$x_k \lambda_k = \frac{\left(\sum_{k=1}^K w_k x_k^\alpha\right)^{1/\alpha}}{\left(\sum_{k=1}^K w_k x_k^\alpha\right)^{1/\alpha} + \left(\sum_{l=1}^L v_l y_l^\alpha\right)^{1/\alpha}}.$$

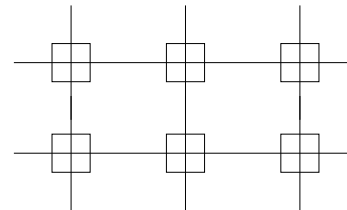The proof is similar to that of the linear network.

Figure 3: Grid network

REMARK 1. *The network looks like a "grid" when the nodes of the associated graph represent the links. This does not imply that the underlying network is a "grid". Figure 4 gives an example of the well-known "fish" network that corresponds to a "grid" network with two horizontal routes AB, CA, and three vertical routes DE, ED, CB.*
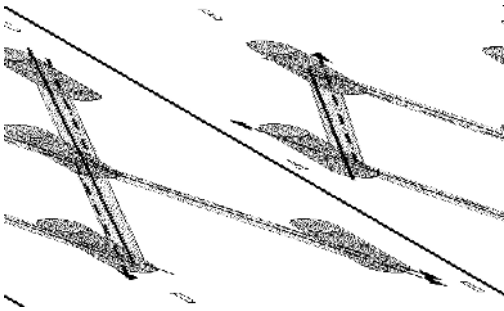

Figure 4: A fish network with five routes

In the rest of the paper, we consider a dynamical system with flow arrivals and departures. Specifically, we assume that new flows are established on route r at the instants of a stationary point process of intensity ν_r. The document size to be transferred by each flow on route r is drawn independently from a general distribution of mean $1/\mu_r$. We denote by $\rho_r = \nu_r/\mu_r$ the traffic intensity offered to route r. Clearly, the departure times of the flows depend on the bandwidth allocation considered (the "service discipline", using queuing theory terminology). Each bandwidth allocation described in this section leads to a different dynamical system.

3. STABILITY CONDITIONS

In this section, we consider the stability of the stochastic process $x = (x_r)_{r \in \mathcal{R}}$ (for notational convenience, we do not make explicit the dependency of x on time t). That is, we derive conditions on the traffic intensities ρ_r for which, starting from any initial state, the number of flows in progress on each route converges to a finite stationary regime. Clearly, the following conditions are necessary:

$$\sum_{r \ni l} \rho_r < C_l, \quad \text{for all } l \in \mathcal{L}. \tag{4}$$

One might also expect these conditions to be sufficient provided the bandwidth allocation is Pareto-efficient (which can be thought of as the network analogue of a work conserving service discipline in queuing theory). This is not the case in general as we shall see now.

3.1 Unfairness and instability

We present several examples of Pareto-efficient bandwidth allocations for which the process x is only stable under traffic conditions more strict than (4). These bandwidth allocations are deliberately unfair in the sense that some types of flow have preferential allocations compared to others. This may result in network resources being under-utilized in the sense that the total number of flows in progress grow to infinity whereas the offered load at each link is less than one.

Example 1 (Maximum throughput).
One might expect the maximum throughput allocation introduced in Section 2 to ensure a complete utilization of network resources, since the overall throughput of the flows at any instant is maximized. This is not the case in general. Consider for instance the linear network of Figure 1. For this specific network topology, the overall throughput of the flows is maximized when a preemptive priority is given to the flows on routes $1, \ldots, L$ over flows on route 0. Since flows on route 0 are "frozen out" as soon as there is at least one active flow on any route $1, \ldots, L$, assuming arrival processes on distinct routes are mutually independent, the network stability condition is given by $\rho_l < 1$ for $l = 1, \ldots, L$, and:

$$\rho_0 < \prod_{l=1}^{L} (1 - \rho_l).$$

For instance, a linear network of $L = 3$ links with the same traffic intensity ρ on each route is unstable when $\rho = 1/3$, although the total traffic load on each link is equal to $2/3$.

Example 2 (Class-based queuing).
Instability under normal traffic conditions may also occur when some milder form of priority is given to some flows. In the linear network considered above, assume that when flows on routes 0 and l are in progress at some link l, a fraction φ of the link capacity is reserved for the flows on route l, so that when $x_0 > 0$:

$$\lambda_0 x_0 = 1 - \varphi + \varphi \, \mathbb{I}_{\{x_1 = 0, \ldots, x_L = 0\}}.$$

This may be realized using *class-based queuing*, see e.g. [9]. It is not suggested that in practice one would wish to allocate bandwidth as such. This example simply shows that instability may occur under traffic conditions (4) though the discrimination between flows is not based on strict priority.

Provided $\rho_l < \varphi$ for $l = 1, \ldots, L$, assuming arrival processes on distinct routes are mutually independent, it can be shown that the network stability condition is given by:

$$\rho_0 < 1 - \varphi + \varphi \prod_{l=1}^{L} \left(1 - \frac{\rho_l}{\varphi} \right).$$

For instance, if $\rho_l = 1/2$ for all $l = 1, \ldots, L$ and $\varphi = 3/4$, this condition reads:

$$\rho_0 < \frac{1}{4} \left(1 + \frac{1}{3^{L-1}} \right).$$

This condition is more restrictive than the traffic conditions (4) as soon as $L > 1$. In the limiting case where L tends to infinity, the network is unstable whenever the total offered load at each link is larger than $3/4$.

Example 3 (Priority queuing).
A more realistic example is the "tree" network depicted in Figure 5. The network consists of one backbone link of unit capacity, and L access links of capacity $c_1, \ldots, c_L$, with $\sum c_l < 1$. Route 0 crosses the backbone link only whereas, for $l = 1, \ldots, L$, route l crosses the backbone link and access link l. Flows on route 0 might represent corporate traffic without any access constraint, for instance, whereas flows on routes l represent residential traffic typically limited by an access link.

Assume that flows on route 0 have preemptive priority over flows on routes 1 to L. This may be realized using head-of-line *priority queuing* for corporate traffic, for instance. The network stability condition is then given by:

$$\rho_l < c_l(1 - \rho_0), \quad \text{for } l = 1, \ldots, L.$$

In the particular case where $\rho_0 = 1/2$ and $c_1 = \ldots = c_L = 1/2L$ for instance, only half of the access link capacity can be used, whereas the total offered load at the backbone link does not exceed $3/4$.
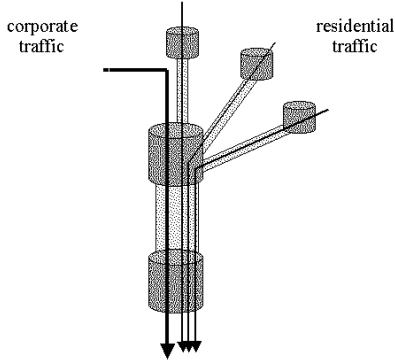

Figure 5: Tree network

REMARK 2. *In a real network, instability would not necessarily manifest itself inside the network. One would rather expect congestion control mechanisms such as TCP to hold the congestion at the network edge, which would eventually be reflected there by timeouts, and then disconnections.*

These simple examples show that the Pareto-efficiency of a bandwidth allocation is not sufficient to guarantee a complete utilization of network resources since the network may be unstable under usual traffic conditions (4). Both Pareto-efficiency *and* fairness are needed as shown by the following key result of the paper.

3.2 Fairness and stability

Consider (weighted) α-bandwidth sharing as defined by (3) for a fixed $\alpha > 0$. For the sake of simplicity, we assume that $x = (x_r)_{r \in \mathcal{R}}$ is a Markov process, i.e. the arrival process of the flows on each route is Poisson and the size of the document to be transferred is exponentially distributed. Note that this assumption is not essential for stability, and the following result can easily be extended to renewal arrival processes and generally distributed document sizes by including the residual interarrival and service times in the Markov process [6]. The transition rates of x are given by:

$$x_r \to x_r + 1 \quad \text{at rate } \nu_r,$$
$$x_r \to x_r - 1 \quad \text{at rate } \mu_r x_r \lambda_r,$$

where the rates $\{\lambda_r\}_{r \in \mathcal{R}}$ are uniquely defined by (3). The following result generalizes the known result for weighted max-min fairness [7].

THEOREM 1. *The Markov process x is ergodic if and only if traffic conditions (4) are satisfied.*

Proof. The proof consists in studying the "fluid system" obtained when the initial number of flows grows to infinity. More precisely, we consider the set of "fluid limits" defined by:

$$X(t) = \lim_{\omega \to \infty} \frac{x(\omega t)}{\omega} \quad \text{with} \quad \sum_{r \in \mathcal{R}} x_r(0) = \omega.$$

Note that if the limit exists, we have $\sum_r X_r(0) = 1$. Given an initial distribution of the fluid $X(0)$, it follows from the strong law of large number that the evolution of the fluid system $X(t)$ is uniquely defined by the differential equations:

$$\frac{d}{dt} X_r = \nu_r - \mu_r \Lambda_r(t), \quad \text{for all } r, t \text{ such that } X_r(t) > 0, \tag{5}$$

where $\Lambda_r(t)$ is the overall bandwidth allocated to the flows on route r at time t. In view of (3), the vector $\Lambda = (\Lambda_r)_{r \in \mathcal{R}}$ is the unique solution of the optimization problem:

$$\text{Maximize} \sum_r w_r X_r^{\alpha} \frac{\Lambda_r^{1-\alpha}}{1 - \alpha},$$

subject to the capacity constraints

$$\sum_{r \ni l} \Lambda_r \le C_l, \quad \forall l \in \mathcal{L}. \tag{6}$$

In the following, we show that provided traffic conditions (4) are satisfied, the fluid system empties in finite time, i.e. there exists a constant $T > 0$ independent of the considered fluid limit such that $X(t) = 0$ for all $t \ge T$. This will complete the proof in view of the general results of [6].

Consider the following function defined on the set of $|\mathcal{R}|$-dimensional positive vectors:

$$F(u) = \sum_r w_r \mu_r^{-1} \rho_r^{-\alpha} \frac{u_r^{\alpha+1}}{\alpha + 1}.$$

We get from (5):

$$\frac{d}{dt} F(X) = \sum_r w_r \rho_r^{-\alpha} X_r^{\alpha} (\rho_r - \Lambda_r). \tag{7}$$

Consider now the function

$$G(u) = \sum_r w_r X_r^{\alpha} \frac{u_r^{1-\alpha}}{1 - \alpha}.$$

The vector Λ attains the maximum of this function over the domain specified by the capacity constraints (6). Thus for any vector u in this convex domain, the gradient of G satisfies $G'(\Lambda).(u - \Lambda) \le 0$. By concavity of G, we conclude that:

$$G'(u).(u - \Lambda) \le 0.$$

Under the stability conditions (4), there exists $\epsilon > 0$ such that the vector $u = (\rho_r(1 + \epsilon))_{r \in \mathcal{R}}$ satisfies the capacity constraints. Applying the previous inequality, we get:

$$\sum_r w_r \rho_r^{-\alpha} X_r^{\alpha} (\rho_r(1 + \epsilon) - \Lambda_r) \le 0.$$

Equivalently, in view of (7), this reads

$$\frac{d}{dt} F(X) \le -\epsilon \sum_r w_r \rho_r^{-\alpha+1} X_r^{\alpha}.$$

Using straightforward bounds, we conclude that there exists a positive constant β such that:

$$\frac{d}{dt}F(X) \leq -\beta F(X)^{\frac{\alpha}{\alpha+1}}.$$

This implies that if $F(X(T)) = 0$ for some $T > 0$, $F(X(t)) = 0$ for all $t \geq T$. In addition, integrating this equation yields for all $t \geq 0$ such that $F(X(t)) > 0$,

$$F(X(t)) \leq \left(F(X(0))^{\frac{1}{\alpha+1}} - \frac{\beta}{\alpha+1}t \right)^{\alpha+1}.$$

Recalling that $\sum_r X_r(0) = 1$, this implies that $F(X(t))$ and thus $X(t)$ are identically equal to zero for all $t \geq T$, with

$$T = \frac{\alpha+1}{\beta} \left(\frac{1}{\alpha+1} \sum_r \mu_r^{-1} \rho_r^{-\alpha} \right)^{\frac{1}{\alpha+1}}.$$

$\square$

Assume now that the bandwidth allocated to any flow on route r is limited to a fixed constant a_r corresponding to the access rate of the flow. That is, the bandwidth allocation is now determined by the optimization problem (3) under the capacity constraints (1) as well as the additional access constraints $\lambda_r \leq a_r$ for all $r \in \mathcal{R}$. We have the following result.

COROLLARY 1. *Theorem 1 still holds under additional access constraints.*

Proof. The proof is completely similar to that of Theorem 1 since the access constraints disappear in the fluid limits. $\square$

4. THROUGHPUT PERFORMANCE

In this section, we consider the impact of fairness on the average transfer time of a document. We consider the broad class of weighted α-bandwidth sharing. Under the traffic conditions (4), we know from Theorem 1 that the network is stable.

Define the *flow throughput* on route r as the average document size $1/\mu_r$ divided by the average time required to transfer a document on route r. By Little's law, the flow throughput on route r is given by:

$$\gamma_r = \frac{\rho_r}{E[x_r]}. \tag{8}$$

We also define the *average flow throughput* as the average document size divided by the average time required to transfer this document, where averages are taken over all transfers, irrespective of their routes. Again, it follows from Little's law that the average flow throughput is given by:

$$\gamma = \frac{\sum_r \rho_r}{\sum_r E[x_r]} = \frac{\sum_r \rho_r}{\sum_r \rho_r/\gamma_r}. \tag{9}$$

It is worth noting that in practice, performance strongly depends on fixed throughput constraints due to modem speed or maximum size of TCP window, for instance. We know

from Corollary 1 that the network remains stable in this case. For the sake of simplicity, we will assume that the bandwidth allocated to a given flow can never exceed a fixed threshold a referred to as the *access rate*. We show in the following that provided bandwidth sharing is "sufficiently" fair and traffic intensity is "not too close" to capacity, flow throughput on each route is mainly determined by this access rate and not by the capacity of the backbone links.

4.1 Single bottleneck

Consider first the impact of fairness on performance in the case of a single bottleneck link of unit capacity.

Fair sharing.

For α-bandwidth sharing, the associated model is a processor sharing queue, independently of the value of α. In the absence of throughput limitation at the access, the number of flows in progress is then geometrically distributed with parameter ρ, provided the flow arrival process is Poisson [12]. We conclude that the flow throughput is given by:

$$\gamma = \frac{\rho}{E[x]} = 1 - \rho.$$

It is worth noting that this result is *insensitive* to the document size distribution (see e.g. [10]).

Explicit expressions of the flow throughput can also be derived in the presence of an access rate $a < 1$. For the sake of simplicity, we assume that $1/a$ is an integer. Denoting by $f(\cdot, \rho)$ the Erlang-C formula [12], we obtain from [4]:

$$\gamma = \frac{1-\rho}{1-\rho+f(1/a, \rho)}\, a,$$

The result is still insensitive to the document size distribution. Figure 6 gives the flow throughput as a fraction of the link capacity with and without an access rate $a = 0.1$. We verify that for small access rates a, the flow throughput is practically insensitive to the offered load, for a wide range of load values $\rho < 1$. In other words, the link is virtually transparent to the users, whose perceived QoS depends much more on their access rate.
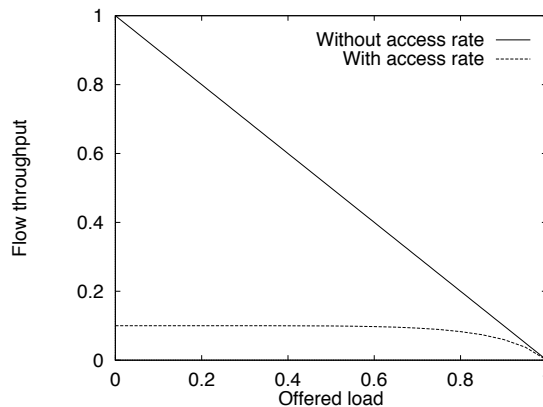
Figure 6: Flow throughput against offered load in case of fair sharing

Unequal sharing.

Now consider weighted α-bandwidth sharing. The associated model is the discriminatory processor sharing queue, with weights w_r. This discrimination may be desired and implemented by per-flow weighted fair queuing say, or unwanted when due to the impact of different round-trip times on TCP throughput, for instance. The flow throughput now depends on the document size distribution. Consider the simple case of two types of flow, with $w_1 + w_2 = 1$ and $w_1 > w_2$. That is, when flows of type 1 and 2 are in progress, flows of type 1 receive w_1/w_2 times more bandwidth than flows of type 2. Assuming both types of flow have the same exponential document size distribution for simplicity, the flow throughput of type 1 and 2 is respectively given by [8]:

$$\gamma_1 = \frac{1 - w_1\rho_1 - w_2\rho_2}{1 - w_1\rho}(1 - \rho),$$

and

$$\gamma_2 = \frac{1 - w_1\rho_1 - w_2\rho_2}{1 - w_2\rho}(1 - \rho).$$

Note that in view of (9), the average flow throughput γ is still given by $1 - \rho$, independently of the weights w_1 and w_2. We have verified by simulation that these results are approximately insensitive to the document size distribution provided the ratio w_1/w_2 is not too high (say, $w_1/w_2 < 100$). It is worth noting that γ_1/γ_2 is always much smaller than w_1/w_2, except when the total offered load ρ is close to one. This is illustrated in Figure 7 where $w_1/w_2 = 10$, $\rho_1 = \rho_2 = \rho/2$, with and without an access rate $a = 0.1$ (in this and subsequent figures, results with an access rate were obtained by simulation).
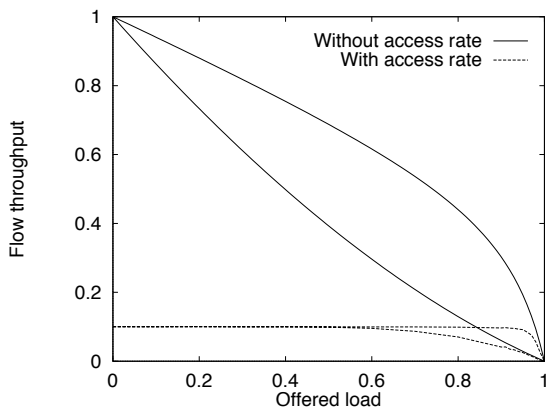

Figure 7: Flow throughput against offered load when flows of type 1 (top) receive 10 times more bandwidth than flows of type 2 (bottom)

Unfair sharing.

Finally, we consider an "unfair" bandwidth allocation where flows of type 1 have preemptive priority over flows of type 2. The duration of flows of type 2 is then extremely sensitive to the document size distribution of flows of type 1. In particular, we shall see that the flow throughput of type 2 may be equal to zero in the practically interesting case where this distribution is heavy-tailed, which again shows the potential suboptimality of "unfair" bandwidth allocations. Let $F_1(u)$

denote the complementary cumulative distribution function of the document size for flows of type 1. We have:

$$\int_0^\infty F_1(u)du = \frac{1}{\mu_1}.$$

At the arrival of a flow of type 2, the number of flows of type 1 in progress is larger than $1/a$ with a positive probability. For each of the flows of type 1 in progress, the probability that the document size remaining to be transferred is larger than u is given by [10]:

$$\mu_1 \int_u^\infty F_1(v)dv.$$

The probability that the bandwidth allocated to the considered flow of type 2 is zero during a time larger than u is greater than the probability that the duration of $1/a$ flows of type 1 in progress is larger than u, namely

$$\left(\mu_1 \int_u^\infty F_1(v)dv\right)^{1/a}.$$

We conclude that the average duration of flows of type 2 is infinite as soon as

$$\int_0^\infty \left(\mu_1 \int_u^\infty F_1(v)dv\right)^{1/a} du = \infty. \qquad (10)$$

Consider the case of the Pareto distribution that is observed in practice [5]:

$$\forall u > c, \quad F_1(u) = \left(\frac{c}{u}\right)^b,$$

where $b > 1$ is a fixed parameter, and $c = b/\mu_1(b-1)$ represents the minimum document size. The distribution is heavy-tailed when the associated variance is infinite, that is when $b \leq 2$. It can be easily verified that condition (10) simply reads:

$$\frac{b-1}{a} \leq 1.$$

If $b = 1.1$ for instance, the flow throughput for flows of type 2 is equal to zero whenever the access rate is larger than 0.1.

4.2 Multiple bottlenecks

We now consider the case where the throughput of some flows is limited by multiple bottlenecks. We are primarily interested in the (worst) case where these bottlenecks have similar capacity and offered load. This might be seen as one extreme situation, the other extreme being when bottlenecks have widely differing characteristics. In the latter case one would expect the performance to be driven solely by the bottleneck which imposes the tightest capacity constraint, and if so the results of the previous section can be applied by ignoring the less stringent bottlenecks. We limit the analysis to α-bandwidth sharing, and evaluate the impact of the parameter α on performance.

Consider the grid network introduced in Section 2 with unit link capacities. Let ρ_k be the traffic load on horizontal route k, $1 \leq k \leq K$, and ϱ_l the traffic load on vertical route l, $1 \leq l \leq L$. Assume that the arrival process of flows at each route is Poisson, and the size of the document to be transferred is exponentially distributed.

From Theorem 1, the number of active flows on each route $(x, y) = (x_k, y_l)_{k,l}$ is an ergodic Markov process, provided $\rho_k + \varrho_l < 1$ for all k, l. It turns out that when bandwidth sharing is proportional fairness, the equilibrium distribution of (x, y) can be evaluated explicitely. The following result generalizes the analysis of the linear network done in [15].

THEOREM 2. *For $\alpha = 1$, the Markov process (x, y) is reversible with equilibrium distribution*

$$\pi(x, y) = C^{-1} \left(\frac{\sum_1^K x_k + \sum_1^L y_l}{\sum_1^K x_k} \right) \prod_{k=1}^K \rho_k^{x_k} \prod_{l=1}^L \varrho_l^{y_l},$$

where C denotes the normalization constant.

Proof. Let ν_k be the arrival rate of flows on horizontal route k, $1/\mu_k$ their average document size. From §2.4, the transitions rates of the Markov process (x, y) are given by:

$$(x, y) \to (x + e_k, y) \quad \text{at rate } \nu_k,$$
$$(x, y) \to (x - e_k, y) \quad \text{at rate } \mu_k \times \frac{\sum_{k=1}^K x_k}{\sum_{k=1}^K x_k + \sum_{l=1}^L y_l}.$$

where e_k denotes the kth unit vector of $\mathbb{R}^K$. The distribution π defined above satisfies for any $x \geq e_k$:

$$\nu_k \, \pi(x - e_k, y) = \mu_k \times \frac{\sum_{k=1}^K x_k}{\sum_{k=1}^K x_k + \sum_{l=1}^L y_l} \pi(x, y).$$

This implies that (x, y) is reversible, and that π is the equilibrium distribution of (x, y), satisfying the usual global balance equations [10]. □

A property worth mentioning is that due to the reversibility of the process (x, y), the equilibrium distribution of the number of flows on each route is *insensitive* to the document size distribution, as in the case of a single bottleneck. An explicit expression of the normalization constant C as a function of $\rho_1, \ldots, \rho_K, \varrho_1, \ldots, \varrho_L$ is given in the Appendix. The following result can then be used to evaluate the average number of flows in progress on each route, thus the flow throughput on each route in view of (8).

COROLLARY 2 (GENERATING FUNCTION). *The generating function ψ defined by:*

$$\psi(u, v) = \sum_{x, y} \pi(x, y) \prod_{k=1}^K u_k^{x_k} \prod_{l=1}^L v_l^{y_l}.$$

is related to the normalization constant C by the expression:

$$\psi(u, v) = \frac{C(\rho_1 u_1, \ldots, \rho_K u_K, \varrho_1 v_1, \ldots, \varrho_L v_L)}{C(\rho_1, \ldots, \rho_K, \varrho_1, \ldots, \varrho_L)}.$$

Proof. The proof follows directly from the expression of the equilibrium distribution π. □

We now give explicit expressions of the flow throughput on each route for the simple network topologies introduced in §2.4. These expressions are obtained from Corollary 2 and calculation of the corresponding normalization constants, based on the formulas in the Appendix; these derivations are not reported here due to lack of space.

Example 1 (Linear network).
The linear network is a particular example of a grid network. The flow throughput on each route is given by:

$$\gamma_l = 1 - \rho_0 - \rho_l, \quad \text{for all } l = 1, \ldots, L,$$

and

$$\gamma_0 = \frac{1 - \rho_0}{1 + \sum_{l=1}^L \frac{\rho_l}{1 - \rho_0 - \rho_l}}.$$

Note that the flow throughput on short routes is that one would obtain if the corresponding link were in isolation, that is if the long route crossed this link only (refer to §4.1). For the long route, the flow throughput decreases as the inverse of the number of links crossed. Figure 8 illustrates this result when the traffic intensity is the same on each route. Note that in the presence of an access rate, the flow throughput becomes largely independent of the number of links crossed.
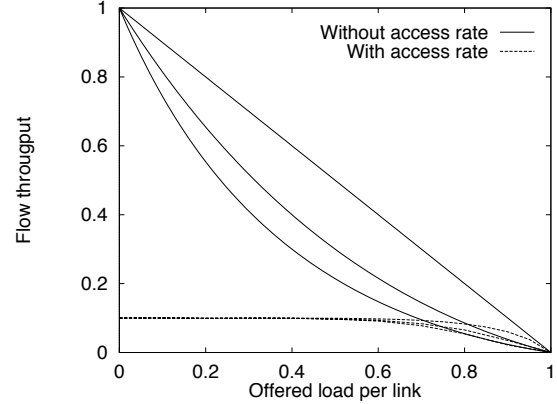

Figure 8: Flow throughput on the long route against offered load for the linear network with $L = 1, 3, 5$ links (from top to bottom)

An analytical evaluation of the impact of the allocation parameter α on flow throughput appears very difficult since models are not tractable whenever $\alpha \neq 1$. However, the simulations we present next suggest that performance is roughly insensitive to α, as long as α is not too close to the critical value $\alpha = 0$. Figures 9 and 10 compare the analytical results obtained with proportional fairness to simulation results obtained with the maximum throughput criterion and with max-min fairness, respectively, for a linear network with $L = 3$ links and the same traffic intensity offered to each route. While performance is significantly worse with the maximum throughput criterion, the results obtained with proportional fairness and with max-min fairness are practically equivalent.

Example 2 (Cyclic network).
The cyclic network with $L = 4$ links considered in §2.4 is also a particular case of a grid network. We have:

$$\gamma_1 = \left(\frac{1}{1 - \rho_1 - \rho_2} + \frac{1}{1 - \rho_1 - \rho_4} \right.$$
$$\left. - \frac{(1 - \rho_2)(1 - \rho_3)(1 - \rho_4) + \rho_2 \rho_3 \rho_4}{(1 - \rho_1)(1 - \rho_2)(1 - \rho_3)(1 - \rho_4) - \rho_1 \rho_2 \rho_3 \rho_4} \right)^{-1}.$$
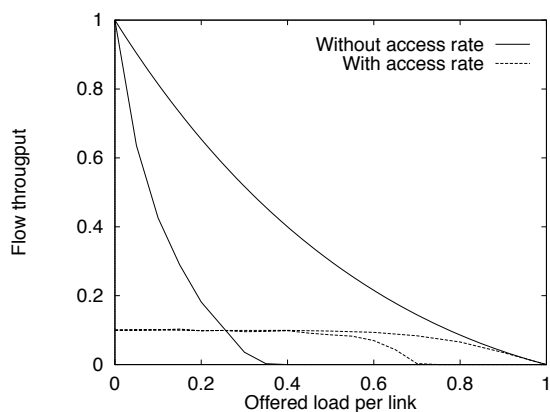
Figure 9: Flow throughput on the long route against offered load for the linear network with proportional fairness (top) and the maximum throughput criterion (bottom)
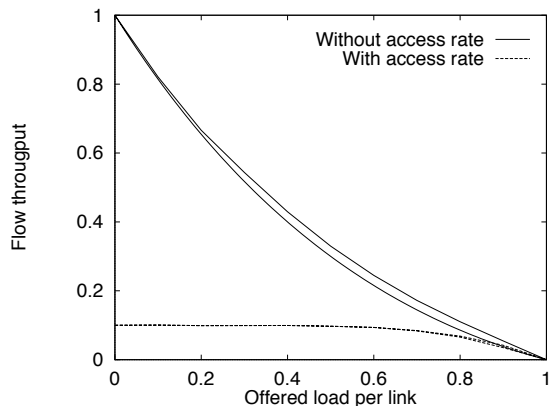

Figure 10: Flow throughput on the long route against offered load for the linear network with max-min fairness (top) and proportional fairness (bottom)

It can be easily verified that γ_1 is an increasing function of ρ_3. In particular, γ_1 is minimum when ρ_3 is equal to zero. We conclude that for each route of the cyclic network, the flow throughput is better than that one would obtain if the links on this route were in isolation, which corresponds to the linear network considered above. As for the linear network, we have verified by simulation that the flow throughput is largely independent of α, provided α is not too close to zero.

Example 3 (Grid network).
Finally, we consider the grid network with the same load ρ (resp. ϱ) on each horizontal (resp. vertical) route. Then, in heavy traffic, that is when $\rho \approx 1 - \varrho$, the flow throughput on each horizontal route k is given by:

$$\gamma_k \approx (1 - \rho - \varrho) \times \frac{K}{K + L - 1}.$$

When K grows from 1 to infinity, the flow throughput on each horizontal route grows from the flow throughput on the long route in a linear network of L links to that obtained with a single link in isolation. As for the cyclic network considered above, the linear network constitutes a worst case. Again, this suggests that for realistic values of the access rate, the flow throughput is largely independent of the number of links crossed (refer to Figure 8) and of the sharing parameter α as long as α is greater than the critical value $\alpha = 0$ (refer to Figures 9-10).

5. CONCLUSION

In this paper we have discussed the relevance of fairness as a bandwidth sharing objective. We have stressed the importance in this context of taking into account the dynamic nature of Internet traffic. To evaluate the expected flow throughput we have proposed fluid models where the impact of packet level dynamics is taken into account in an idealized way via the assumed bandwidth sharing objective. Flows arrive according to an assumed stochastic arrival process and persist for as long as it takes to transfer a randomly sized document. This modeling approach allows a clearer formalization of different notions of fairness and provides insight into their impact on realized performance.

Analytical results establish that a broad class of fair bandwidth allocations, including that achieved by the congestion avoidance algorithm of TCP, guarantees full utilization of network resources in the sense that the total number of competing flows in the network remains finite as long as demand is less than capacity on every link. This is in contrast to a number of examples we cite where deliberately biased sharing, favoring one traffic class over another, leads to instability at loads considerably less than one. This is a somewhat counter intuitive result which suggests current proposals for service differentiation should be critically examined with respect to their performance in realistic dynamic traffic conditions.

We have introduced a number of network architectures where it is possible to derive exact performance results under stable traffic conditions when the sharing objective is proportional fairness. Comparison with simulation results with max-min fairness suggests that performance is not highly dependent on the realized allocation, provided this is reasonably fair. This observation is particulary true when considering sharing in a backbone network where all flows have a relatively small upper limit on throughput, due to a modem or the maximum size of the TCP window, for example. In this case, the network is virtually transparent with respect to realized throughput performance as long as link loads are not too close to or greater than one.

While many of the analytical results derived in this paper are necessarily for toy network configurations, they considerably advance our understanding of the importance of fairness as a design objective of congestion control mechanisms in the Internet. They clearly demonstrate how throughput performance is extremely robust as long as bandwidth sharing is reasonably fair. They explain why "best effort" continues to provide excellent service in any adequately provisioned network.

6. APPENDIX

In this Appendix, we evaluate the normalization constant of the equilibrium distribution π given in Theorem 2.

We first consider the particular case where $\rho_k \neq \rho_{k'}$ for all $k \neq k'$ and $\varrho_l \neq \varrho_{l'}$ for all $l \neq l'$. We then have:

$$\sum_{x_1+\dots+x_K=n} \rho_1^{x_1}\dots\rho_K^{x_K} = \sum_{k=1}^{K} \frac{\rho_k^{n+K-1}}{\prod_{k'\neq k}(\rho_k - \rho_{k'})}.$$

and

$$\sum_{y_1+\dots+y_L=m} \varrho_1^{y_1}\dots\varrho_L^{y_L} = \sum_{l=1}^{L} \frac{\varrho_l^{m+L-1}}{\prod_{l'\neq l}(\varrho_l - \varrho_{l'})}.$$

Thus,

$$C = \sum_{n\geq 0, m\geq 0} \binom{n+m}{n} \sum_{k=1}^{K}\sum_{l=1}^{L} \frac{\rho_k^{n+K-1}}{\prod_{k'\neq k}(\rho_k - \rho_{k'})} \frac{\varrho_l^{m+L-1}}{\prod_{l'\neq l}(\varrho_l - \varrho_{l'})}.$$

Using the negative binomial formula

$$\sum_{n\geq 0}\binom{n+m}{n} z^n = (1-z)^{-m-1},$$

we obtain

$$C = \sum_{m\geq 0}\sum_{k=1}^{K}\sum_{l=1}^{L} \frac{\rho_k^{K-1}}{\prod_{k'\neq k}(\rho_k - \rho_{k'})} \frac{\varrho_l^{L-1}}{\prod_{l'\neq l}(\varrho_l - \varrho_{l'})}$$
$$\times \frac{1}{1-\rho_k}\left(\frac{\varrho_l}{1-\rho_k}\right)^m.$$

We conclude that

$$C = \sum_{k=1}^{K}\sum_{l=1}^{L} \frac{\rho_k^{K-1}\varrho_l^{L-1}}{\prod_{k'\neq k}(\rho_k - \rho_{k'})\prod_{l'\neq l}(\varrho_l - \varrho_{l'})} \times \frac{1}{1-\rho_k-\varrho_l}.$$

In the general case with $\tilde{K} \leq K$ different values of ρ_k, $\tilde{L} \leq L$ different values of ϱ_l, denote by $i_k + 1$ the number of routes with traffic load ρ_k, $j_l + 1$ the number of routes with traffic load ϱ_l. Noting that

$$\sum_{x_1+\dots+x_m=n} \rho^{x_1}\dots\rho^{x_m} = \binom{n+m-1}{n}\rho^n,$$

we have in this case:

$$C = \sum_{x_1,\dots,x_{\tilde{K}}}\sum_{y_1,\dots,y_{\tilde{L}}} \binom{\sum_1^{\tilde{K}} x_k + \sum_1^{\tilde{L}} y_l}{\sum_1^{\tilde{K}} x_k}$$
$$\times \prod_{k=1}^{\tilde{K}}\binom{x_k+i_k}{x_k}\rho_k^{x_k} \prod_{l=1}^{\tilde{L}}\binom{y_l+j_l}{y_l}\varrho_l^{y_l}.$$

Let $\tilde{C}$ denote the normalization constant of the grid network with $\tilde{K} \times \tilde{L}$ links and different traffic loads on each horizontal (resp. vertical) route for which we have an explicit expression. In view of previous equality, the normalization constant C is given by:

$$C = \frac{\partial^{i_1}\dots\partial^{i_{\tilde{K}}}\partial^{j_1}\dots\partial^{j_{\tilde{L}}}}{\partial\rho_1\dots\partial\rho_{\tilde{K}}\partial\varrho_1\dots\partial\varrho_{\tilde{L}}}$$
$$\left(\frac{\rho_1^{i_1}}{i_1!}\dots\frac{\rho_{\tilde{K}}^{i_{\tilde{K}}}}{i_{\tilde{K}}!}\frac{\varrho_1^{j_1}}{j_1!}\dots\frac{\varrho_{\tilde{L}}^{j_{\tilde{L}}}}{j_{\tilde{L}}!} \times \tilde{C}(\rho_1,\dots,\rho_{\tilde{K}},\varrho_1,\dots,\varrho_{\tilde{L}})\right).$$

7. REFERENCES

[1] S. Ben Fredj, T. Bonald, G. Regnié, J.W. Roberts, "Statistical bandwidth sharing: a study of congestion at flow level", submitted, 2001.

[2] D. Bertsekas and R. Gallager, *Data Networks*, Prentice Hall, 1987.

[3] S. Blake et. al., An Architecture for Differentiated Services, RFC 2475, December 1998.

[4] J. W. Cohen, "The multiple phase service network with generalized processor sharing", Acta Informatica 12 (1979) 245–284.

[5] M. Crovella and A. Bestravos, "Self-similarity in World Wide Web traffic Evidence and possible causes", in: *Proc. of ACM SIGMETRICS'96*, 1996.

[6] J.G. Dai "On positive Harris recurrence of multiclass queuing networks: a unified approach via fluid limit models", Ann. of Applied Probability 5 (1995) 49–77.

[7] G. De Veciana, T.J. Lee and T. Konstantopoulos, "Stability and performance analysis of networks supporting services with rate control - Could the Internet be unstable?", in: *Proc. of IEEE INFOCOM'99*, 1999.

[8] G. Fayolle, I. Mitrani and R. Iasnogorodski, "Sharing a processor among many classes", Journal of the ACM 27 (1980) 519–532.

[9] S. Floyd and V. Jacobson, "Link-sharing and resource management models for packet networks", IEEE/ACM Trans. on Networking 3-4 (1995) 365–386.

[10] F.P. Kelly, *Reversibility and Stochastic Networks*, Wiley, 1979.

[11] 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).

[12] L. Kleinrock, *Queueing Systems, Volume 2*, J. Wiley & Sons, 1975.

[13] S. Kunniyur and R. Srikant, End-to-end congestion control schemes: Utility functions, random losses and ECN marks, in:*Proc. of IEEE INFOCOM'00*, 2000.

[14] L. Massoulié and J.W. Roberts, "Bandwidth sharing: objectives and algorithms", in: *Proc. of IEEE INFOCOM'99*, 1999.

[15] L. Massoulié and J.W. Roberts, "Bandwidth sharing and admission control for elastic traffic", Telecommunication Systems 15 (2000) 185–201.

[16] J. Mo and J. Walrand, "Fair end-to-end window-based congestion control", IEEE/ACM Trans. on Networking 8-5 (2000) 556–567.

[17] M. Vojnovic, J.-Y. Le Boudec and C. Boutremans "Global fairness of additive-increase and multiplicative-decrease with heterogeneous round-trip times", in: *Proc. of IEEE INFOCOM'00*, 2000.