

Insensitive, maximum stable allocations converge to proportional fairness

N.S. Walton

Received: 19 July 2010 / Revised: 16 March 2011 / Published online: 9 April 2011
© The Author(s) 2011. This article is published with open access at Springerlink.com

Abstract We describe a queuing model where service is allocated as a function of queue sizes. We consider allocations policies that are insensitive to service requirements and have a maximal stability region. We take a limit where the queuing model becomes congested. We study how service is allocated under this limit. We demonstrate that the only possible limit allocation is one that maximizes a proportionally fair optimization problem.

Keywords Proportional fairness · Insensitivity · Maximum stability

Mathematics Subject Classification (2000) Primary 90B20 · Secondary 60K25 · 60K30 · 90B18

1 Introduction

Consider a communication network. Documents arrive and are transferred across the different routes of the network. In a communication network, each document transfer receives a rate which may vary through time, depending on the number of transfers present on each route. What sort of behavior might we want such a network to have?

Perhaps, we would like our communication network to be stable, to be able to cope with the rate that work arrives? If a communication network does not allocate its capacity well then instability can arise. So we want to consider allocation policies that avoid instability whenever it is possible. We call such policies *maximum stable*.

We may, also, like our communication network to be *insensitive* to different document size distributions? Famous examples, such as the Erlang B formula and the

N.S. Walton (✉)
Korteweg-de Vries Institute for Mathematics, University of Amsterdam, P.O. Box 94248, 1090 GE,
Amsterdam, The Netherlands
e-mail: n.s.walton@uva.nl

processor sharing queue, are known to be insensitive to different job size distributions. As with the Erlang B formula, an advantage of insensitivity is that stationary statistics can be described without an explicit knowledge all traffic parameters. In addition, as with the processor sharing queue, an advantage of insensitivity is that the expected service time of a document is proportional to its size. Thus, each unit of a document's work can be expected to be treated equally.

A further advantage of a processor sharing discipline is that it allocates capacity proportional to the number of customers in each class. In this way, as the queue gets congested, each customer class is served at a consistent rate. Perhaps, we might want to extend this property to our communication network? In our communication network, as the number of documents on each route gets large, we might want to allocate service relative to number of document transfers on each route. This will result in a *limiting allocation policy* that assigns service according to the fraction of documents on each route.

We have described three desirable properties for our communication network to satisfy: to be maximum stable, to be insensitive and to have a limiting allocation policy. In this paper, we reason that the only limiting allocation policy that can arise from a maximum stable, insensitive network is the policy that solves the optimization problem:

$$\text{maximize } \sum_{r \in \mathcal{R}} n_r \log \Lambda_r \quad \text{over } \Lambda \in \mathcal{C}. \quad (1.1)$$

In this optimization problem, n_r gives the proportion of documents on route $r \in \mathcal{R}$, \mathcal{C} gives the set of feasible allocations and $\Lambda = (\Lambda_r : r \in \mathcal{R})$ gives a feasible allocation. The allocation policy $\Lambda^{PF}(n)$ that solves the optimization problem (1.1) for each $n = (n_r : r \in \mathcal{R})$ is called the *proportionally fair* policy.

We note that by assuming these three initial properties, we have gained an additional property: our network is attempting to *optimize* the service rate allocated to each route. We did not initially prescribe this optimization structure, instead this behavior is implicitly implied by our networking assumptions.

1.1 Literature review

The most important reference for this paper is [7]. Massoulié observes that two insensitive allocations, balanced fairness and modified proportional fairness, have a large deviations rate function given by the optimization problem (1.1). Remark 2 of Massoulié [7] notes that if balanced fairness has a limiting allocation policy then it must be proportionally fair. Massoulié proves that modified proportional fairness is maximum stable, insensitive and that proportional fairness is its limiting allocation policy.

Walton [9] proves another allocation policy arising from multi-class queuing networks is maximum stable, insensitive and has proportional fairness as its limiting allocation policy. This insensitive allocation policy was first published by Bonald and Proutière [2].

Early work on insensitivity in our modeling context is due to Whittle [10]. A good paper for results, properties and historical review on the insensitivity of allocation

policies is [1]. Procedures for calculating stationary statistics of insensitive networks are discussed in [3, 6]. Discussions on the expected service times of insensitive allocations are given in [1, 4].

1.2 Paper organization

In Sect. 2 we present the main allocation policies of interest: the proportionally fair policy and the insensitive policies. In addition, we define what it means for an allocation policy to have a limiting allocation policy. In Sect. 3 we define our stochastic model of a communication network. We call this model a stochastic flow level model. We also define what we mean by maximum stable. In Sect. 4 we define insensitivity for a stochastic flow level model, and we state a known insensitivity result. In Sect. 5 we prove the main result of this paper that the only limit allocation policy of a maximum stable, insensitive policy is the proportionally fair policy, see Theorem 1. In Sect. 6 we ask if all maximum stable, insensitive allocations converge to proportional fairness without assuming that a limiting allocation policy exists? We show this is not possible with a counterexample: Proposition 2.

2 Allocation policies

An *allocation policy* allocates the rate of document transfer across a network. Such policies allocate given the number of documents in transfer on the routes of the network and subject to the capacity constraints of the network. More explicitly, we let the set \mathcal{R} index the set of routes of the network, and we let the components of the vector $n = (n_r : r \in \mathcal{R}) \in \mathbb{Z}_+^{\mathcal{R}}$ give the number of documents in transfer on each route. An allocation policy is a function $\Lambda : \mathbb{Z}_+^{\mathcal{R}} \rightarrow \mathcal{C}$ where $\mathcal{C} \subset \mathbb{R}_+^{\mathcal{R}}$ is a closed, bounded, convex set with a non-empty interior. The set \mathcal{C} is the set of allocations that can be scheduled. We call \mathcal{C} the *schedulable region*.

An allocation policy of central interest to this paper is defined as follows.

Definition 1 (Proportionally fair policy) The allocation policy $\Lambda^{PF}(n) = (\Lambda_r^{PF}(n) : r \in \mathcal{R}) \in \mathcal{C}$, defined for $n \in \mathbb{Z}_+^{\mathcal{R}}$, is called *proportionally fair* if for all $n \in \mathbb{R}_+^{\mathcal{R}}$

$$n_r = 0 \text{ implies } \Lambda_r^{PF}(n) = 0,$$

and if $\Lambda^{PF}(n)$ solves the optimization problem¹

$$\text{maximize } \sum_{r \in \mathcal{R}} n_r \log \Lambda_r \text{ over } \Lambda \in \mathcal{C}. \tag{2.1}$$

Proportional fairness was first defined as policy for sharing bandwidth by Kelly [5].

¹We assume throughout this paper that $x \log x = 0$ for $x = 0$, and we adopt the convention that $0^0 = 1$.

We will shortly introduce a stochastic model of document transfer and allocation. For this model, we will be interested in allocation policies that show insensitivity to document sizes. Before we define this model and this insensitivity property, we will define the set of allocation policies which will exhibit insensitivity.

Definition 2 (Insensitive allocation policies) An allocation policy $\Lambda(\cdot)$ is *insensitive* if there exists a function $\Phi : \mathbb{Z}^{\mathcal{R}} \rightarrow \mathbb{R}_+$ such that $\Phi(0) = 1$, $\Phi(n) = 0 \forall n \notin \mathbb{Z}_+^{\mathcal{R}}$ and

$$\Lambda_r(n) = \frac{\Phi(n - e_r)}{\Phi(n)}, \quad \forall n \in \mathbb{Z}_+^{\mathcal{R}}. \quad (2.2)$$

Here and subsequently, e_r will denote the r th unit vector in $\mathbb{R}_+^{\mathcal{R}}$.

We want our allocation policy to have a limit when we allow the routes of the network to become congested. For this reason, we consider limiting allocation policies.

Definition 3 (Limiting allocation policy) An allocation policy $\Lambda(\cdot)$ is a *limiting allocation policy* if there exists $\tilde{\Lambda}(\cdot)$ such that for all $n \in \mathbb{R}_+^{\mathcal{R}}$,

$$\Lambda(cn + \mathbf{O}(1)) \xrightarrow{c \rightarrow \infty} \tilde{\Lambda}(n). \quad (2.3)$$

Given the conventional ‘Big-O notation’, statement (2.3) is a shorthand for the convergence statement: for all bounded sequences $\{d_c\}_c$ such that $cn + d_c \in \mathbb{Z}_+^{\mathcal{R}}$

$$\Lambda_r(cn + d_c) \xrightarrow{c \rightarrow \infty} \tilde{\Lambda}_r(n), \quad r \in \mathcal{R}.$$

3 A stochastic model

We, now, define a stochastic model that allocates service with an allocation policy. This model can be thought of as a model of document transfer across the Internet. We introduce our stochastic model for an allocation policy $\Lambda(\cdot)$.

Documents to be transferred arrive as a Poisson process. Route $r \in \mathcal{R}$ documents arrive as an independent Poisson process of rate $\nu_r > 0$. Each document has a size that is divided into stages. Each stage has a size that is independent exponentially distributed with mean δ . Each route $r \in \mathcal{R}$ document has a number of stages that is independent and equal in distribution to a random variable L_r . We assume L_r has a finite mean and we define $\mu_r^{-1} := \delta \mathbb{E}L_r$, the mean size of a route r document. When there are $n = (n_r : r \in \mathcal{R})$ documents in transfer on each route, each route r documents is served at rate $\frac{\Lambda_r(n)}{n_r}$. Documents are then processed at this rate until the number of documents in transfer changes either by a document departure or arrival.

To be explicit, given there are $n_{r,s}$ route r documents with s stages remaining, the state of this Markov chain is $x = (n_{r,s} : r \in \mathcal{R}, s \in \mathbb{N})$. Letting $e_{r,s}$ be the r st unit vector, the non-zero transition rates of this Markov chain are

$$q(x, x + e_{r,s}) = \nu_r \mathbb{P}(L_r = s),$$

$$\begin{aligned}
 q(x, x - e_{r1}) &= \delta \Lambda_r(n) \frac{n_{r1}}{n_r} \quad \text{if } n_{r1} > 0, \\
 q(x, x - e_{rs} + e_{rs-1}) &= \delta \Lambda_r(n) \frac{n_{rs}}{n_r} \quad \text{if } n_{rs} > 0 \text{ and } s \geq 1.
 \end{aligned}$$

We call this Markov chain a *stochastic flow level model operating under allocation* $\Lambda(\cdot)$. This model was introduced by Massoulié and Roberts [8].

We define the vector of *traffic intensities* $\rho \in \mathbb{R}_+^{\mathcal{R}}$ by $\rho_r := \frac{\nu_r}{\mu_r}$, $r \in \mathcal{R}$. Also, for the schedulable region \mathcal{C} , we let \mathcal{C}° be the interior of \mathcal{C} .

We consider stochastic flow level models that are stable when possible. For this purpose, we will later use the following assumption.

Assumption 1 (Maximum stable) *For $\Lambda(\cdot)$, an allocation policy on schedulable region \mathcal{C} , we say that the stochastic flow level model operating under $\Lambda(\cdot)$ is maximum stable if it is positive recurrent for all*

$$\rho \in \mathcal{C}^\circ. \tag{3.1}$$

A stochastic flow level model is transient for all ρ outside the region \mathcal{C} . Also, there exist allocation policies that are positive recurrent for all $\rho \in \mathcal{C}^\circ$. Notably, [7] proves that the proportionally fair policy is positive recurrent for all $\rho \in \mathcal{C}^\circ$. Thus, in words, Assumption 1 states that a stochastic flow level model under $\Lambda(\cdot)$ is stable in the largest possible region.

4 Insensitivity

We are interested in stochastic flow level models that show robustness to the distribution of document sizes. For this reason, we consider insensitive stochastic flow level models.

Take a stochastic flow level model operating under allocation $\Lambda(\cdot)$. For such a model, let $\pi = (\pi(n) : n \in \mathbb{Z}_+^{\mathcal{R}})$ be the stationary distribution of the number of documents in transfer on each route. In general, π is a function of the parameters of our model: $(\nu_r : r \in \mathcal{R})$, $(L_r : r \in \mathcal{R})$ and δ .

Definition 4 (Insensitive stochastic flow level model) We define a stochastic flow level model to be *insensitive* if $\pi(n)$ can be expressed as a function of parameters $(\nu_r : r \in \mathcal{R})$ and $(\mu_r : r \in \mathcal{R})$.

In words, this definition says that we do not need to know the precise distribution of document sizes to know the stationary behavior of our model; instead, we only need to know the mean size of the documents.

Derived from the work of Whittle [10], a key observation made by Bonald and Proutière [1] was that Definitions 2 and 4 are equivalent in the following sense.

Proposition 1

- (a) *A stochastic flow level model is insensitive iff it is operating under an allocation that is insensitive.*

(b) For an insensitive stochastic flow level model,

$$\pi(n) = \frac{\Phi(n)}{B(\rho)} \prod_{r \in \mathcal{R}} \rho_r^{n_r}, \quad \text{for } n \in \mathbb{Z}_+^{\mathcal{R}}, \rho \in \mathcal{C}^\circ, \tag{4.1}$$

where

$$B(\rho) := \sum_{n \in \mathbb{Z}_+^{\mathcal{R}}} \Phi(n) \prod_{r \in \mathcal{R}} \rho_r^{n_r}. \tag{4.2}$$

(c) When document sizes are exponentially distributed, an insensitive stochastic flow level model is reversible.

The proof of this proposition involves checking that reversibility is equivalent to insensitivity for these queuing networks. The stationary distribution (4.1) then arises from the detail balance equations. See [1] for a proof.

5 The main result

We can now prove the main result of this paper.

Theorem 1 For $\Lambda(\cdot)$ an insensitive, maximum stable, limiting allocation policy

$$\Lambda(cn + O(1)) \xrightarrow{c \rightarrow \infty} \Lambda^{PF}(n), \quad \text{for all } n \in \mathbb{R}_+^{\mathcal{R}}. \tag{5.1}$$

Moreover, for $\rho \in \mathcal{C}^\circ$

$$\lim_{c \rightarrow \infty} \frac{1}{c} \log \pi(\lfloor cn \rfloor) = - \max_{\Gamma \in \mathcal{C}} \sum_{r \in \mathcal{R}} n_r \log \frac{\Gamma_r}{\rho_r} \quad \text{over } \Gamma \in \mathcal{C}. \tag{5.2}$$

Here and hereafter, we use the convention that $\lfloor x \rfloor := (\lfloor x_r \rfloor : r \in \mathcal{R})$ is the lower integer part of each component of $x \in \mathbb{R}_+^{\mathcal{R}}$.

Proof To prove (5.2), we will construct an upper bound and a lower bound. To prove the upper bound, we apply a Chernoff bound. To prove the lower bound, we bound our reversible process along a sequence of transitions from $\lfloor \epsilon cn \rfloor$ to $\lfloor cn \rfloor$. The result (5.1) is a consequence of (5.2).

We start with the upper bound. We let π be the stationary distribution (4.1) and we let N be a random variable with that stationary distribution. Note from Proposition 1(b), the moment generating function of π can be expressed compactly by the expression

$$\mathbb{E} e^{\sum_r \theta_r N_r} = \frac{B(\rho e^\theta)}{B(\rho)}$$

for $\theta \in \mathbb{R}^{\mathcal{R}}$ and provided $\rho e^\theta := (\rho_r e^{\theta_r} : r \in \mathcal{R}) \in \mathcal{C}^\circ$. Thus applying Chebyshev’s inequality to the above expression:

$$\pi(\lfloor cn \rfloor) \leq \mathbb{E} e^{\sum_r \theta_r (N_r - \lfloor cn_r \rfloor)} = \begin{cases} \frac{B(\rho e^\theta)}{B(\rho)} e^{-\sum_r \theta_r \lfloor cn_r \rfloor} & \text{if } \rho e^\theta \in \mathcal{C}^\circ, \\ \infty & \text{if } \rho e^\theta \notin \mathcal{C}. \end{cases}$$

Thus, for $\rho e^\theta \in \mathcal{C}^\circ$

$$\limsup_{c \rightarrow \infty} \frac{1}{c} \log \pi(\lfloor cn \rfloor) \leq \lim_{c \rightarrow \infty} \frac{1}{c} \log \mathbb{E} e^{\sum_r \theta_r (N_r - \lfloor cn_r \rfloor)} = - \sum_{r \in \mathcal{R}} n_r \theta_r.$$

Minimizing over $\theta \in \mathbb{R}^{\mathcal{R}}$ with $\rho e^\theta \in \mathcal{C}^\circ$ gives the required upper bound:

$$\begin{aligned} \limsup_{c \rightarrow \infty} \frac{1}{c} \log \pi(\lfloor cn \rfloor) &\leq - \max_{r \in \mathcal{R}} \sum n_r \theta_r \quad \text{over } \rho e^\theta \in \mathcal{C} \\ &= - \max_{r \in \mathcal{R}} \sum n_r \log \frac{\Gamma_r}{\rho_r} \quad \text{over } \Gamma \in \mathcal{C}. \end{aligned}$$

Above, we make the substitution $\Gamma_r = \rho_r e^{\theta_r}$ for each $r \in \mathcal{R}$.

Now we prove the lower bound. We assume, for simplicity, that document sizes are exponentially distributed. Thus, by Proposition 1(c) our Markov process is reversible. We want to consider a sequence of transitions of this processes close to the line tn , $t > 0$. For $t > 0$, let m^0, m^1, m^2, \dots be the distinct values of $\lfloor tn \rfloor$ as t increases to infinity. This sequence is increasing but a transition from state m^k to m^{k+1} might take more than one document arrival. If so, split the transition from m^k to m^{k+1} into increasing transitions each involving a single arrival. Let n^0, n^1, n^2, \dots be the resulting sequence. Note that the sequence n^0, n^1, n^2, \dots is an increasing sequence of transitions and close tn . This is because for all $k \geq 0$, $n^{k+1} = n^k + e_{r_k}$ for some route r_k and because each point n^k is within distance $|\mathcal{R}|$ of the line tn , $t > 0$.²

Define $k(t)$ to be the term in our sequence which equals $\lfloor tn \rfloor$ and also take $\epsilon > 0$. By the reversibility of our process

$$\pi(\lfloor cn \rfloor) = \left[\prod_{k=k(c\epsilon)+1}^{k(c)} \frac{\rho_{r_k}}{\Lambda_{r_k}(n^k)} \right] \times \pi(\lfloor c n \epsilon \rfloor). \tag{5.3}$$

In expression (5.3), by taking $c\epsilon$ suitably large, we can make all the $\Lambda(n^k)$ terms close to $\tilde{\Lambda}(n)$. In particular given n , by our limit allocation assumption, for all $\delta > 0$ there exists c such that for all $k > k(c\epsilon)$

$$\Lambda_r(n^k) e^{-\delta} \leq \tilde{\Lambda}_r(n) \leq \Lambda_r(n^k) e^\delta, \quad \forall r \in \mathcal{R} \text{ with } n_r > 0. \tag{5.4}$$

²In this paper, we judge distances by the L^1 norm $|n| := \sum_r |n_r|$.

³We note that $\tilde{\Lambda}_r(n) > 0$ if $n_r > 0$ as otherwise $\sum_k \pi(n^k)$ would diverge.

Noting that exactly $\lfloor cn_r \rfloor - \lfloor cn_r \epsilon \rfloor$ route r transitions are used in (5.3), and applying (5.4) to (5.3) gives

$$\begin{aligned} & \liminf_{c \rightarrow \infty} \frac{1}{c} \log \pi(\lfloor cn \rfloor) \\ & \geq \liminf_{c \rightarrow \infty} - \sum_{r \in \mathcal{R}} \sum_{\substack{k: r_k=r \\ k(c\epsilon) < k \leq k(c)}} \frac{1}{c} \log \left(\frac{\Lambda_r(n^k)}{\rho_r} \right) + \liminf_{c \rightarrow \infty} \frac{1}{c} \log \pi(\lfloor cn \epsilon \rfloor) \\ & \geq - \sum_{r \in \mathcal{R}} (1 - \epsilon) n_r \log \left(\frac{\tilde{\Lambda}_r(n) e^\delta}{\rho_r} \right) + \epsilon \liminf_{c \rightarrow \infty} \frac{1}{c} \log \pi(\lfloor cn \rfloor). \end{aligned}$$

Letting $\delta \rightarrow 0$, canceling terms and dividing by $1 - \epsilon$ gives

$$\liminf_{c \rightarrow \infty} \frac{1}{c} \log \pi(\lfloor cn \rfloor) \geq - \sum_{r \in \mathcal{R}} n_r \log \frac{\tilde{\Lambda}_r(n)}{\rho_r}. \tag{5.5}$$

Combining our upper bound (5) with our new lower bound (5.5) implies

$$\sum_{r \in \mathcal{R}} n_r \log \frac{\tilde{\Lambda}_r(n)}{\rho_r} \geq \max_{\Gamma \in \mathcal{C}} \sum_r n_r \log \frac{\Gamma_r}{\rho_r} \quad \text{over } \Gamma \in \mathcal{C}.$$

But $\tilde{\Lambda}(n) \in \mathcal{C}$, as it is the limit of elements in \mathcal{C} . Thus $\tilde{\Lambda}_r(n) = \Lambda_r^{PF}(n)$. This completes part (b) of the proof and, given of our upper bound (5.3) and lower bound (5.5), we have also proven part (a). □

6 A counter example

Theorem 1 required the assumption that a limiting allocation policy existed. Can this assumption be completely removed? The answer to this is no, as the following proposition demonstrates.

Proposition 2 *From any insensitive, maximum stable, limiting allocation policy $\Lambda(\cdot)$, we can construct $\hat{\Lambda}(\cdot)$, an insensitive, maximum stable allocation policy that does not have a limit and, thus, does not converge to a proportionally fair policy.*

Proof Let $\Phi(n)$ be the potential function associated with $\Lambda(n)$ in (2.2). By Theorem 1, $\Lambda(n)$ limits to a proportionally fair allocation. Define

$$\hat{\Phi}(n) := \alpha^k \Phi(n), \quad \text{for the } k \in \mathbb{N} \text{ such that } 2^k \leq |n| < 2^{k+1}.$$

We assume $\alpha > 1$ and we take $|n| := \sum_r |n_r|$. For each $r \in \mathcal{R}$, observe that

$$\hat{\Lambda}_r(n) = \begin{cases} \Lambda_r(n), & \text{if } 2^k < |n| < 2^{k+1} \text{ for some } k \in \mathbb{N}, \\ \alpha^{-1} \Lambda_r(n), & \text{if } |n| = 2^k \text{ for some } k \in \mathbb{N}. \end{cases}$$

Since $\alpha > 1$ and $0 \in \mathcal{C}$, $\hat{\Lambda}(n) \in \mathcal{C}$ for all $n \in \mathbb{Z}_+^{\mathcal{R}}$. We can let the sequence cn hit $|cn| = 2^k$ for all k suitably large. For example, we can do this by letting $|n| = 2^m$ for some $m \in \mathbb{N}$. In this case, for $c_k = 2^k$

$$\tilde{\Lambda}(c_k n) = \frac{\Lambda(c_k n)}{\alpha} \xrightarrow{k \rightarrow \infty} \frac{\Lambda^{PF}(n)}{\alpha} \neq \Lambda^{PF}(n).$$

So along this subsequence, the allocation policy $\tilde{\Lambda}(c_k n)$ does not tend to a proportionally fair allocation. Along another subsequence, e.g. $c'_k = 2^k - 1$, the allocation policy $\tilde{\Lambda}(c'_k n)$ would tend to a proportionally fair allocation. Thus no limit allocation policy exists for $\tilde{\Lambda}(\cdot)$.

It remains to show that $\tilde{\Lambda}(\cdot)$ is a maximum stable allocation. Take $\rho \in \mathcal{C}^\circ$, there exists $\epsilon > 0$ such that $\rho e^\epsilon = (\rho_r e^\epsilon : r \in \mathcal{R}) \in \mathcal{C}^\circ$. Note also there exists n' such that $\forall |n| > n', \alpha^{\log_2 |n|} < e^{\epsilon |n|}$. Let $\hat{B}(\rho)$ and $B(\rho)$ be the normalizing constants defined by (4.2) for $\hat{\Phi}(\cdot)$ and $\Phi(\cdot)$, respectively. We know that $B(\rho e^\epsilon) < \infty$. We now show that $\hat{B}(\rho) < \infty$.

$$\begin{aligned} \hat{B}(\rho) &= \sum_{k=0}^{\infty} \sum_{n: 2^k < |n| \leq 2^{k+1}} \alpha^k \Phi(n) \prod_{r \in \mathcal{R}} \rho_r^{n_r} \\ &\leq \sum_{n: |n| > N} e^{\epsilon |n|} \Phi(n) \prod_{r \in \mathcal{R}} \rho_r^{n_r} + \sum_{n: |n| \leq N} \alpha^{\log_2 |n|} \Phi(n) \prod_{r \in \mathcal{R}} \rho_r^{n_r} \\ &= B(\rho e^\epsilon) + \sum_{n: |n| \leq N} \alpha^{\log_2 |n|} \Phi(n) \prod_{r \in \mathcal{R}} \rho_r^{n_r} \\ &< \infty. \end{aligned}$$

Thus, as the normalizing constant $\hat{B}(\rho)$ is finite for all $\rho \in \mathcal{C}^\circ$, our allocation policy $\hat{\Lambda}(\cdot)$ is maximum stable. □

Open Access This article is distributed under the terms of the Creative Commons Attribution Noncommercial License which permits any noncommercial use, distribution, and reproduction in any medium, provided the original author(s) and source are credited.

References

1. Bonald, T., Proutière, A.: Insensitivity in processor-sharing networks. *Perform. Eval.* **49**, 193–203 (2002)
2. Bonald, T., Proutière, A.: On performance bounds for balanced fairness. *Perform. Eval.* **55**, 25–50 (2004)
3. Bonald, T., Proutière, A., Roberts, J., Virtamo, J.: Computational aspects of balanced fairness. In: *Proceedings of the 18th International Teletraffic Congress (ITC-18)*, pp. 801–810 (2003)
4. Kelly, F., Massoulié, L., Walton, N.: Resource pooling in congested networks: proportional fairness and product form. *Queueing Syst.* **63**(1–4), 165–194 (2009). doi:10.1007/s11134-009-9143-8
5. Kelly, F.P.: Charging and rate control for elastic traffic. *Eur. Trans. Telecommun.* **8**, 33–37 (1997)
6. Mairesse, J., Jonckheere, M.: Towards an Erlang formula for multiclass networks. *Queueing Syst.* **6**(1), 53–78 (2010)

7. Massoulié, L.: Structural properties of proportional fairness: Stability and insensitivity. *Ann. Appl. Probab.* **17**, 809–839 (2007)
8. Massoulié, L., Roberts, J.: Bandwidth sharing and admission control for elastic traffic. *Telecommun. Syst.* **15**, 185–201 (1998)
9. Walton, N.S.: Proportional fairness and its relationship with multi-class queueing networks. *Ann. Appl. Probab.* **22**(6), 2301–2333 (2009)
10. Whittle, P.: Partial balance and insensitivity. *J. Appl. Probab.* **22**, 168–176 (1985)