



HAL
open science

Worst-case Analysis of Non-Cooperative Load Balancing

Olivier Brun, Balakrishna Prabhu

► **To cite this version:**

Olivier Brun, Balakrishna Prabhu. Worst-case Analysis of Non-Cooperative Load Balancing. *Annals of Operations Research*, 2016, 239 (2), pp.471-495. 10.1007/s10479-014-1747-7 . hal-00747244

HAL Id: hal-00747244

<https://hal.science/hal-00747244>

Submitted on 30 Oct 2012

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

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

Worst-case Analysis of Non-Cooperative Load Balancing

O. Brun, B.J. Prabhu
LAAS-CNRS
Université de Toulouse
7 Avenue du Colonel Roche
31077 Toulouse, France.
email: {brun, bjprabhu}@laas.fr

October 30, 2012

Abstract

We investigate the impact of heterogeneity in the amount of incoming traffic routed by dispatchers in a non-cooperative load balancing game. For a fixed amount of total incoming traffic, we give sufficient conditions on the cost function under which the worst-case social cost occurs when each dispatcher routes the same amount of traffic, that is, the game is symmetric. We then analyze the symmetric game and derive an explicit expression for the equilibrium flows. We also obtain lower bounds on the Price of Anarchy for *(i)* cost functions that are polynomial on server loads; and *(ii)* cost functions representing the mean delay of the Shortest Remaining Processing Time (SRPT) service discipline.

1 Introduction

We consider a distributed load-balancing system comprising of multiple dispatchers and multiple servers. The task of a dispatcher is to receive requests and decide which server to route them to. The routing decisions are taken by the dispatchers so as to minimize the cost (for example, the mean waiting time) incurred by the requests it routes. We shall assume that the dispatchers do not exchange information on their strategies and take decisions independently of the others. Thus, the dispatchers are involved in a non-cooperative game. Since the number of players (dispatchers) is finite, we shall be interested in the performance of the system at the Nash Equilibrium (NE).

Although the distributed architecture described above is attractive due to its scalability and robustness, it can be inefficient from the system performance

point of view. Indeed, since each dispatcher minimizes its own cost function, the system performance is possibly worse than that in a centralized architecture. Thus, before going in for a distributed architecture, a system designer may be interested in knowing how worse its performance can be compared to that of a centralized architecture.

Our objective is to find out under what circumstances we can expect to find the worst-case scenario for the distributed load-balancing system, and then to characterize how worse this performance can be. We first show that, for a fixed total incoming traffic to the system, the worst-case performance at the NE occurs when all the dispatchers have the same amount of traffic to be routed, that is, they are involved in a symmetric game. We give sufficient conditions on the objective function under which the above statement is true. We then analyze the symmetric game and derive an explicit expression for the equilibrium flows. Using this result, we give a lower bound on the Price of Anarchy (PoA, [8]) for cost functions of type: (i) $1 + x^m$; and (ii) $1/(1 - x)^m$.

1.1 Related work

The distributed load balancing game has been analysed previously by several authors. In [2], similar results were obtained for the $M/M/1$ type delay functions, and the present work is a generalization of that to a larger class of cost functions. In terms of methodology, our work is closely related to that of Orda *et al* [7, 9] and the arguments we use are inspired from their work. However, there are a couple of differences with their work. First, in [7], the authors restrict themselves to the $M/M/1$ delay function, whereas we consider a larger class of cost functions. In particular, we allow for the association of a heterogeneous cost per unit time (holding cost) with each server. The introduction of heterogeneous holding costs can significantly change the PoA for certain cost functions. For example, for the $M/M/1$ delay function and equal holding costs, it was shown in [6] that the PoA is upper bounded by the number of servers whereas for the same model and unequal holding costs, it was shown in [2] that the PoA is of the order of the square root of the number of dispatchers as long as there is at least one server. Second, in [7], it is shown that, for a fixed incoming traffic vector, transferring capacity from a server to another one that has a higher service rate improves the performance at the NE. We look at the complementary problem in which we fix the service capacities, and look at what happens when we transfer traffic from one dispatcher to another.

The computation of PoA for disciplines other than Processor Sharing (or the $M/M/1$ delay function) has also been previously investigated. In [1], the authors compute the PoA of routing policies with memory in server farms with First-Come-First-Served or Processor Sharing service discipline using a lower bound on the cost of the centralized architecture. The computation of PoA for server farms with SRPT discipline but without memory was given in [4]. The authors compute the PoA on assuming that the system is in heavy-traffic. Our

approach and our model differ from these papers in that we give the worst-case scenario for any fixed total traffic vector whether the system is in heavy traffic or not. Using this scenario we construct a lower bound on the PoA. Moreover, in our model, we have a finite number of dispatchers whereas in the model of both these papers each request takes a decision by itself, thereby leading to an infinite number of dispatchers.

1.2 Organization of the paper

The rest of the paper is organized as follows. In Section 2, we describe the model and state the problem. In Section 3, we explore the structure of the underlying Nash equilibria and prove their existence and uniqueness. We also establish several properties of these equilibria that form the foundation of the subsequent analysis. In Section 4, we analyze the social cost at the Nash equilibria, and show that the maximum of this cost is achieved in the symmetric case. We then analyze the symmetric game and derive an explicit expression for the equilibrium flows in Section 5. With this result at hand, in Section 6, we derive lower bounds on the PoA for the two different cost functions mentioned previously. Finally, we draw some conclusions in Section 7.

2 Problem Formulation and Main Results

Denote $\mathcal{C} = \{1, \dots, K\}$ to be the set of dispatchers and $\mathcal{S} = \{1, \dots, S\}$ to be the set of servers. Jobs received by dispatcher i are said to be jobs of class i .

Server $j \in \mathcal{S}$ has capacity r_j and a holding cost c_j per unit time is incurred for each job sent to this server. We let $u_j = \frac{c_j}{r_j}$ denote the ratio of the holding cost to the capacity for server j . By convention, $u_{S+1} = \infty$. It is assumed that servers are numbered in the order of increasing cost per unit capacity, i.e., if $m \leq n$, then $u_m \leq u_n$, so that the first servers are more attractive than the last ones. Let $\mathbf{r} = (r_j)_{j \in \mathcal{S}}$ and $\mathbf{c} = (c_j)_{j \in \mathcal{S}}$ denote the vectors of server capacities and server costs, respectively, and let $\bar{r} = \sum_{n \in \mathcal{S}} r_n$ denote the total capacity of the system.

Jobs of class $i \in \mathcal{C}$ arrive to the system according to a Poisson process and have generally distributed service-times. Let λ_i be the traffic intensity of class i . It is assumed that $\lambda_i \leq \lambda_j$ for $i \leq j$. Moreover, it will also be assumed that the vector $\boldsymbol{\lambda}$ of traffic intensities belongs to the following set:

$$\Lambda = \left\{ \boldsymbol{\lambda} \in \mathbb{R}^K : \sum_{i \in \mathcal{C}} \lambda_i = \bar{\lambda} \right\},$$

where $\bar{\lambda}$ denotes the total incoming traffic intensity.

Let $\mathbf{x}_i = (x_{i,j})_{j \in \mathcal{S}}$ denote the routing strategy of dispatcher i , with $x_{i,j}$ being the amount of traffic it sends towards server j . Let

$$\mathcal{X}_i = \left\{ \mathbf{x}_i \in \mathbb{R}^{\mathcal{S}} : 0 \leq x_{i,j} \leq r_j, \forall j \in \mathcal{S}; \sum_{j \in \mathcal{S}} x_{i,j} = \lambda_i \right\}$$

denote the set of feasible routing strategies for dispatcher i . The vector $\mathbf{x} = (\mathbf{x}_i)_{i \in \mathcal{C}}$ will be called a multi-strategy. The multi-strategies belong to the product strategy space $\mathcal{X} = \bigotimes_{i \in \mathcal{C}} \mathcal{X}_i$.

Dispatcher i seeks to minimize its cost function which is denoted by $T_i(\mathbf{x})$. This optimization problem can be formulated as follows:

$$\underset{\mathbf{x}_i \in \mathcal{X}_i}{\text{minimize}} T_i(\mathbf{x})$$

The model is illustrated in the figure below.

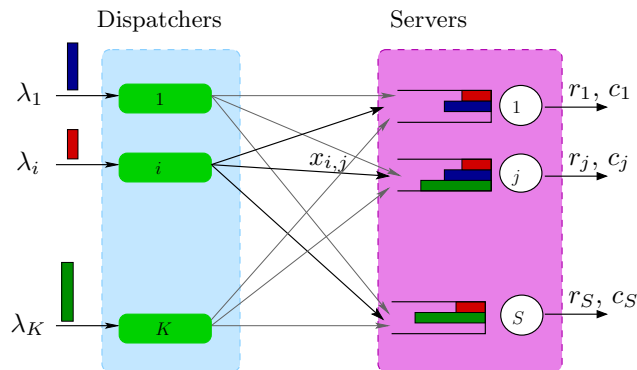


Figure 1: Distributed load balancing.

A Nash equilibrium of the routing game is a multi-strategy from which no class finds it beneficial to deviate unilaterally. Hence, $\mathbf{x} \in \mathcal{X}$ is a Nash Equilibrium Point (NEP) if

$$\mathbf{x}_i = \arg \min_{\mathbf{z} \in \mathcal{X}_i} T_i(\mathbf{x}_1, \dots, \mathbf{x}_{i-1}, \mathbf{z}, \mathbf{x}_{i+1}, \dots, \mathbf{x}_K), \quad \forall i \in \mathcal{C}.$$

Turning our attention to a server $j \in \mathcal{S}$, we let y_j be the total flow on that server, i.e., $y_j = \sum_{i \in \mathcal{C}} x_{i,j}$. We let also $\rho_j = y_j/r_j$ denote the utilization rate of server $j \in \mathcal{S}$. In the following, it will be assumed that the cost function of each user $i \in \mathcal{C}$ is the sum of server cost function, i.e.,

$$T_i(\mathbf{x}) = \sum_{j \in \mathcal{S}} \psi_j(x_{i,j}, y_j),$$

where,

$$\psi_j(x_{i,j}, y_j) = u_j x_{i,j} \phi(\rho_j).$$

Note that the cost incurred by class i on server j depends both on its amount of flow on that server, which measures its “investment” on that server, and on the total amount of flow through the server, which determines the server performance.

We shall consider two different types of congestion functions ϕ . Cost functions of the first type operate over the domain $\mathcal{D} = [0, +\infty)$ and satisfy the following assumptions:

- I-1 $\phi : [0, \infty) \rightarrow [1, \infty)$,
- I-2 $\phi(0) = 1$ and $\lim_{\rho \rightarrow \infty} \phi(\rho) = +\infty$,
- I-3 ϕ is a continuous, strictly increasing and convex function on $[0, \infty)$,
- I-4 ϕ is continuously differentiable and its second derivative ϕ'' exists.

Cost functions satisfying the above assumptions will be referred as type-I functions in the following. For type-I congestion function, it will be assumed throughout the paper that $r_j = 1$, $j \in \mathcal{S}$. In this case, $\psi_j(x_{i,j}, y_j) = c_j x_{i,j} \phi(y_j)$ for all $j \in \mathcal{S}$. Examples of such functions include polynomial functions $\phi(\rho) = 1 + \alpha_k \rho^k$ with $\alpha_k > 0$ and $k \geq 1$, as well as functions of the type $\phi(\rho) = \exp(\rho)$.

Congestion functions of the second type are defined over the domain $\mathcal{D} = [0, 1)$. They satisfy the following assumptions:

- II-1 $\phi : [0, 1) \rightarrow [1, \infty)$,
- II-2 $\phi(0) = 1$ and $\lim_{\rho \rightarrow 1^-} \phi(\rho) = +\infty$,
- II-3 ϕ is a continuous, strictly increasing and convex function on $[0, 1)$,
- II-4 ϕ is continuously differentiable and its second derivative ϕ'' exists.

Typical type-II functions are those from queueing theory, e.g. the M/G/1/PS congestion function $\phi(\rho) = \frac{1}{1-\rho}$ or the M/G/1/FCFS congestion function $\phi(\rho) = 1 + \rho \frac{1+c_b^2}{2(1-\rho)}$, where c_b^2 denotes the squared coefficient of variation of the job sizes. This class of functions also contains the delay function of the *M/Pareto/1/SRPT* in heavy-traffic, which is given by $\frac{1}{(1-\rho)^m}$, where m depends on the shape parameter of the Pareto distribution [4]. For type-II functions, it will be assumed throughout the paper that $\bar{\lambda} < \bar{\tau}$, which is the necessary and sufficient condition

to guarantee the stability of the system.

Cost functions ψ_j that comply with type-I or type-II assumptions are a special case of type-B functions, as defined in [9]. As proved in Theorem 2.1 of this reference, these assumptions are sufficient to assert the existence of a unique NEP.

Let ψ'_j be the partial derivative of ψ_j with respect to its first argument. We therefore have

$$\psi'_j(x_{i,j}, y_j) = u_j \left[\phi(\rho_j) + \frac{x_{i,j}}{r_j} \phi'(\rho_j) \right], \quad (1)$$

for all $j \in \mathcal{S}$. According to the Karush-Kuhn-Tucker (KKT) optimality conditions, $\mathbf{x} \in \mathcal{X}$ is a NEP if and only if there exist multipliers μ_i such that

$$\psi'_j(x_{i,j}, y_j) = \mu_i \quad \text{if } x_{i,j} > 0, \quad (2)$$

$$\psi'_j(x_{i,j}, y_j) \geq \mu_i \quad \text{if } x_{i,j} = 0. \quad (3)$$

Let $\mathcal{C}_j = \{i \in \mathcal{C} : x_{i,j} > 0\}$ be the set of classes which route traffic to server j . Similarly, let $\mathcal{S}_i = \{j \in \mathcal{S} : x_{i,j} > 0\}$ be the set of servers to which class i routes traffic. Note that $i \in \mathcal{C}_j \iff j \in \mathcal{S}_i$. Since ϕ is strictly increasing, we can now rewrite equations (2) and (3) as

$$\psi'_j(0, y_j) < \mu_i \iff i \in \mathcal{C}_j \iff j \in \mathcal{S}_i. \quad (4)$$

Let \mathbf{x} be the NEP for the system with K dispatchers. The global performance of the system can be assessed using the global cost

$$D_K(\boldsymbol{\lambda}, \mathbf{r}, \mathbf{c}) = \sum_{i \in \mathcal{C}} T_i(\mathbf{x}) = \sum_{j \in \mathcal{S}} c_j \rho_j \phi(\rho_j)$$

where the utilization rates ρ_j are those at the NEP. Note that when there is a single dispatcher, we have a single class whose traffic intensity is $\lambda_1 = \bar{\lambda}$. The global cost can therefore be written as $D_1(\bar{\lambda}, \mathbf{r}, \mathbf{c})$ in this case.

For our model, the PoA is defined as

$$PoA(K) = \sup_{\boldsymbol{\lambda}, \mathbf{r}, \mathbf{c}} \frac{D_K(\boldsymbol{\lambda}, \mathbf{r}, \mathbf{c})}{D_1(\bar{\lambda}, \mathbf{r}, \mathbf{c})}.$$

2.1 Main Results

Our main result states that the global cost $D_K(\boldsymbol{\lambda}, \mathbf{r}, \mathbf{c})$ achieves its maximum when $\boldsymbol{\lambda}$ is the symmetric vector $\boldsymbol{\lambda}^{\text{sym}} = \left(\frac{\bar{\lambda}}{K}, \dots, \frac{\bar{\lambda}}{K}\right)$.

Theorem 1

$$\sup_{\boldsymbol{\lambda}, \mathbf{r}, \mathbf{c}} D_K(\boldsymbol{\lambda}, \mathbf{r}, \mathbf{c}) = \sup_{\mathbf{r}, \mathbf{c}} D_K(\boldsymbol{\lambda}^{\text{sym}}, \mathbf{r}, \mathbf{c}).$$

This result implies that, for the calculation of the PoA, we can restrict ourselves to the symmetric game. This, coupled with the fact that in our setting the symmetric game is also a potential game, makes it more tractable for the analytic computation of the NEP and the global cost, thereby greatly simplifying the derivation of the lower and upper bounds on the PoA.

3 Properties of the Nash Equilibrium

In this section, we investigate properties of the traffic flow at the NEP. We first establish some properties of the cost function ϕ that will be useful in the sequel.

3.1 Properties of the cost function ϕ

We prove below some lemmata stating properties of the cost function ϕ . These properties are used in the sequel for the characterisation of the NEP.

Lemma 1

$$\forall w \in \mathcal{D}, \phi'(w) < +\infty$$

Proof. If $w \in \mathcal{D}$, then $\exists \epsilon > 0$ such that $w + \epsilon \in \mathcal{D}$. Hence, $\phi(w + \epsilon) < +\infty$. However, since ϕ is convex and continuously differentiable on \mathcal{D} , it lies above all of its tangents: $\phi(b) \geq \phi(a) + (b - a)\phi'(a)$ for all $a, b \in \mathcal{D}$. With $a = w$ and $b = w + \epsilon$, it yields $\phi'(w) \leq [\phi(w + \epsilon) - \phi(w)] / \epsilon < +\infty$. ■

Lemma 2

$$\forall a, b \in \mathcal{D} - \{0\}, \quad b \frac{\phi'(b)}{\phi(b)} \geq a \frac{\phi'(a)}{\phi(a)} \iff b \geq a$$

Proof. It is sufficient to show that the function $g : w \rightarrow g(w) = w \frac{\phi'(w)}{\phi(w)}$ is increasing on $\mathcal{D} - \{0\}$. Let $f(w) = \frac{\phi'(w)}{\phi(w)}$. Then $g(w) = w f(w)$. Note that since ϕ is twice differentiable on \mathcal{D} , the derivatives $f'(w)$ and $g'(w)$ exist for all $w \in \mathcal{D}$, and $g'(w) = f(w) + w f'(w)$. Note also that $f(w) > 0$ since ϕ is strictly increasing on \mathcal{D} .

Let us consider $w \in \mathcal{D}$. If f is increasing at point w , then $g'(w) > 0$ and g is strictly increasing at point w . Assume now that f is strictly decreasing at point w , i.e. $f'(w) < 0$. From $\phi'(w) = f(w)\phi(w)$, we obtain $\phi''(w) = f'(w)\phi(w) + f(w)\phi'(w)$. Since ϕ is convex on \mathcal{D} , we have $\phi''(w) \geq 0$, i.e.

$$f'(w)\phi(w) + f(w)\phi'(w) \geq 0 \quad (5)$$

The convexity of ϕ implies that $\phi(b) \geq \phi(a) + (b-a)\phi'(a)$ for all $a, b \in \mathcal{D}$. With $a = w$ and $b = 0$, it yields $\phi(w) \leq \phi(0) + w\phi'(w)$. From 5, we thus get

$$f'(w) [\phi(0) + w\phi'(w)] + f(w)\phi'(w) \geq 0, \quad (6)$$

and hence,

$$\phi'(w)g'(w) = \phi'(w) [f(w) + w f'(w)] \geq -\phi(0)f'(w) \geq 0, \quad (7)$$

from which we deduce that $g'(w) \geq 0$, i.e. that g is increasing at point w . ■

3.2 Properties related to traffic intensities

We prove below that there is a monotonicity among classes in their use of servers: a class with a higher demand uses more of each and every server. We first prove a series of technical lemmata before stating our main results in Proposition 1 and Corollary 1.

Lemma 3 $\mathcal{S}_i \cap \mathcal{S}_k \neq \emptyset$.

Proof. Assume the contrary, i.e., if $m \in \mathcal{S}_i$ then $m \notin \mathcal{S}_k$, and if $n \in \mathcal{S}_k$ then $n \notin \mathcal{S}_i$. For one such pair m and n , from (4), we can conclude that $\mu_i > \psi'_m(0, y_m) \geq \mu_k$ and $\mu_k > \psi'_n(0, y_n) \geq \mu_i$, which is a contradiction. ■

Since $\mathcal{S}_i \cap \mathcal{S}_k \neq \emptyset$, from (2), we have

$$\mu_i - \mu_k = \psi'_j(x_{i,j}, y_j) - \psi'_j(x_{k,j}, y_j), \quad \forall j \in \mathcal{S}_i \cap \mathcal{S}_k. \quad (8)$$

Lemma 4 $\mu_i < \mu_k \iff \exists j \in \mathcal{S}_k : x_{i,j} < x_{k,j}$.

Proof. *Straight part:* From Lemma 3, $\mathcal{S}_i \cap \mathcal{S}_k \neq \emptyset$. If $\mu_i < \mu_k$, then, from (8), $\exists j \in \mathcal{S}_k : \psi'_j(x_{i,j}, y_j) < \psi'_j(x_{k,j}, y_j)$ which implies that $\exists j \in \mathcal{S}_k : x_{i,j} < x_{k,j}$.

Converse part: $\exists j \in \mathcal{S}_k : x_{i,j} < x_{k,j}$. Either $j \in \mathcal{S}_i$ or $j \notin \mathcal{S}_i$. If $j \in \mathcal{S}_i$ then, from (8), $\psi'_j(x_{i,j}, y_j) < \psi'_j(x_{k,j}, y_j)$ implies $\mu_i < \mu_k$. If $j \notin \mathcal{S}_i$, then, from (4), $\mu_i \leq \psi'_j(0, y_j) < \psi'_j(x_{k,j}, y_j) < \mu_k$. ■

Lemma 5 *If $\mu_i < \mu_k$, then $\mathcal{S}_i \subset \mathcal{S}_k$.*

Proof. If $j \in \mathcal{S}_i$, then, from (4), $\psi'_j(0, y_j) < \mu_i$. If $\mu_i < \mu_k$ then $\psi'_j(0, y_j) < \mu_k$. Hence, from (4) we can conclude that $j \in \mathcal{S}_k$. Therefore, $\mathcal{S}_i \subset \mathcal{S}_k$. ■

Lemma 6 $\exists m \in \mathcal{S}_k : x_{i,m} < x_{k,m} \iff x_{i,j} < x_{k,j}, \quad \forall j \in \mathcal{S}_k$.

Proof. *Straight part:* If $\exists m \in \mathcal{S}_k : x_{i,m} < x_{k,m}$, then, from Lemmata 4 and 5, we have $\mu_i < \mu_k$ and $\mathcal{S}_i \subset \mathcal{S}_k$. For $j \in \mathcal{S}_i$, from (8), we have $\psi'_j(x_{i,j}, y_j) < \psi'_j(x_{k,j}, y_j)$, which implies $x_{i,j} < x_{k,j}$. For $j \in \mathcal{S}_k \setminus \mathcal{S}_i$, $x_{i,j} = 0$ and $0 < x_{k,j}$. Hence, $x_{i,j} < x_{k,j}, \forall j \in \mathcal{S}_k$.

Converse part: It is true from the statement. ■

Proposition 1 *The following statements are equivalent:*

1. $\mu_i < \mu_k$.
2. $\exists j \in \mathcal{S}_k : x_{i,j} < x_{k,j}$.
3. $x_{i,j} < x_{k,j}, \forall j \in \mathcal{S}_k$.
4. $\lambda_i < \lambda_k$.

Proof. $1 \iff 2 \iff 3$ follows from Lemmata 4 and 6. Now, we show $3 \iff 4$.

Straight part: If $x_{i,j} < x_{k,j}, \forall j \in \mathcal{S}_k$, then, from the fact that $3 \iff 1$ and Lemma 5, we can conclude that $\lambda_i = \sum_{j \in \mathcal{S}_i} x_{i,j} = \sum_{j \in \mathcal{S}_k} x_{i,j} < \sum_{j \in \mathcal{S}_k} x_{k,j} = \lambda_k$.

Converse part: Since $\lambda_k = \sum_{j \in \mathcal{S}_k} x_{k,j}$, if $\lambda_i < \lambda_k$, then $\exists j \in \mathcal{S}_k : x_{i,j} < x_{k,j}$. Since $2 \iff 3$, if $\lambda_i < \lambda_k$, then $x_{i,j} < x_{k,j}, \forall j \in \mathcal{S}_k$. ■

The above proposition shows that a class with a higher demand uses more of each and every server. The following corollary shows that if two classes have the same traffic intensity, then they send the same amount of flow on each server.

Corollary 1 *From Proposition 1, it follows that*

1. $\mu_i = \mu_k \iff \exists j \in \mathcal{S}_k : x_{i,j} = x_{k,j} \iff x_{i,j} = x_{k,j}, \forall j \in \mathcal{S}_k \iff \lambda_i = \lambda_k.$
2. *If $\lambda_i < \lambda_k$, then $\mathcal{S}_i \subset \mathcal{S}_k$.*
3. *If $\lambda_i = \lambda_k$, then $\mathcal{S}_i = \mathcal{S}_k$.*

In particular, if all classes have the same demand, i.e. $\boldsymbol{\lambda} = \boldsymbol{\lambda}^=$, then, for all server $j \in \mathcal{S}$ and for all $i \in \mathcal{C}$, we have $x_{i,j} = y_j/K$.

Recall that we have assumed that $\lambda_i \leq \lambda_k$ for $i \leq k$. Therefore, according to the above results, if we consider two classes i and $k > i$, then we have $\mathcal{S}_i \subseteq \mathcal{S}_k$, $\mu_i \leq \mu_k$ and $x_{i,j} \leq x_{k,j}$ for all servers $j \in \mathcal{S}_k$, with the equalities holding if and only if $\lambda_i = \lambda_k$.

Let $N_j = |\mathcal{C}_j|$. Proposition 1 implies that if $k \in \mathcal{C}_j$, then $i \in \mathcal{C}_j, \forall i > k$. As a consequence, the set \mathcal{C}_j has the following structure: $\mathcal{C}_j = \{K - N_j + 1, \dots, K\}$.

3.3 Properties related to server cost functions

The above results tell how an order on λ_i translates to an order on $x_{i,j}$, μ_i and \mathcal{S}_i , i.e., quantities of class i . We now give the analogous results for similar quantities of server j .

Lemma 7 $\mathcal{C}_m \cap \mathcal{C}_n \neq \emptyset$.

Proof. Assume the contrary, i.e., if $i \in \mathcal{C}_m$, then $i \notin \mathcal{C}_n$, and if $k \in \mathcal{C}_n$, then $k \notin \mathcal{C}_m$. For one such pair i and k , from (4), we can conclude that $\psi'_m(0, y_m) < \mu_i \leq \psi'_n(0, y_n)$ and $\psi'_n(0, y_n) < \mu_k \leq \psi'_m(0, y_m)$, which is a contradiction. ■

Lemma 8 *If $\psi'_n(0, y_n) \leq \psi'_m(0, y_m)$, then $\mathcal{C}_m \subseteq \mathcal{C}_n$.*

Proof. If $i \in \mathcal{C}_m$, then, from (4), $\mu_i > \psi'_m(0, y_m)$. If $\psi'_n(0, y_n) \leq \psi'_m(0, y_m)$, then $\mu_i > \psi'_n(0, y_n)$. Hence, from (4) we can conclude that $i \in \mathcal{C}_n$. Therefore, $\mathcal{C}_m \subseteq \mathcal{C}_n$. ■

Since $\mathcal{C}_m \cap \mathcal{C}_n \neq \emptyset$, from (2), we have

$$\mu_i = \psi'_m(x_{i,m}, y_m) = \psi'_n(x_{i,n}, y_n), \quad \forall i \in \mathcal{C}_m \cap \mathcal{C}_n. \quad (9)$$

Lemma 9 *The following hold:*

$$\psi'_n(0, y_n) < \psi'_m(0, y_m) \iff \exists i \in \mathcal{C}_n : \frac{u_m x_{i,m}}{r_m} \phi'(\rho_m) < \frac{u_n x_{i,n}}{r_n} \phi'(\rho_n).$$

Proof. *Straight part:* From Lemma 7, $\mathcal{C}_m \cap \mathcal{C}_n \neq \emptyset$. If $\psi'_n(0, y_n) < \psi'_m(0, y_m)$, then, from (9), $\exists i \in \mathcal{C}_n : \psi'_m(x_{i,m}, y_m) - \psi'_m(0, y_m) < \psi'_n(x_{i,n}, y_n) - \psi'_n(0, y_n)$, i.e. $\frac{u_m x_{i,m}}{r_m} \phi'(\rho_m) < \frac{u_n x_{i,n}}{r_n} \phi'(\rho_n)$.

Converse part: Assume $\exists i \in \mathcal{C}_n : \frac{u_m x_{i,m}}{r_m} \phi'(\rho_m) < \frac{u_n x_{i,n}}{r_n} \phi'(\rho_n)$. Then $\psi'_m(x_{i,m}, y_m) - \psi'_m(0, y_m) < \psi'_n(x_{i,n}, y_n) - \psi'_n(0, y_n)$. Either $i \in \mathcal{C}_m$ or $i \notin \mathcal{C}_m$. If $i \in \mathcal{C}_m$, then, from (9), $\psi'_n(0, y_n) < \psi'_m(0, y_m)$. If $i \notin \mathcal{C}_m$, then, from (3), $\psi'_m(0, y_m) \geq \mu_i = \psi'_n(x_{i,n}, y_n) > \psi'_n(0, y_n)$. Hence, $\psi'_n(0, y_n) < \psi'_m(0, y_m)$. ■

Lemma 10 *The following hold:*

$$\psi'_n(0, y_n) < \psi'_m(0, y_m) \iff \frac{u_m x_{i,m}}{r_m} \phi'(\rho_m) < \frac{u_n x_{i,n}}{r_n} \phi'(\rho_n), \quad \forall i \in \mathcal{C}_n.$$

Proof. *Straight part:* If $\psi'_n(0, y_n) < \psi'_m(0, y_m)$, then from Lemma 8, $\mathcal{C}_m \subset \mathcal{C}_n$. For $i \in \mathcal{C}_m$, from (9), $\psi'_m(x_{i,m}, y_m) - \psi'_m(0, y_m) < \psi'_n(x_{i,n}, y_n) - \psi'_n(0, y_n)$, i.e., $\frac{u_m x_{i,m}}{r_m} \phi'(\rho_m) < \frac{u_n x_{i,n}}{r_n} \phi'(\rho_n)$. For $i \in \mathcal{C}_n \setminus \mathcal{C}_m$, $x_{i,m} = 0$ and $0 < x_{i,n}$. Hence, since ϕ is strictly increasing, $\frac{u_m x_{i,m}}{r_m} \phi'(\rho_m) = 0 < \frac{u_n x_{i,n}}{r_n} \phi'(\rho_n)$.

Converse part: the proof is a direct consequence of Lemma 9. ■

The following proposition proves a monotonic property regarding the order of preference of servers as seen by each class.

Proposition 2 *The following statements are equivalent:*

1. $\psi'_n(0, y_n) < \psi'_m(0, y_m)$.
2. $\exists i \in \mathcal{C}_n : \frac{u_m x_{i,m}}{r_m} \phi'(\rho_m) < \frac{u_n x_{i,n}}{r_n} \phi'(\rho_n)$.
3. $\frac{u_m x_{i,m}}{r_m} \phi'(\rho_m) < \frac{u_n x_{i,n}}{r_n} \phi'(\rho_n), \quad \forall i \in \mathcal{C}_n$.
4. $u_n < u_m$.

Proof. $1 \iff 2 \iff 3$ follows from Lemmata 9 and 10. Next, we show $3 \iff 4$.

Straight part: If $\frac{u_m x_{i,m}}{r_m} \phi'(\rho_m) < \frac{u_n x_{i,n}}{r_n} \phi'(\rho_n), \quad \forall i \in \mathcal{C}_n$, then from the fact that $3 \iff 1$ and Lemma 8, we can conclude that

$$u_m \rho_m \phi'(\rho_m) = \sum_{i \in \mathcal{C}_m} \frac{u_m x_{i,m}}{r_m} \phi'(\rho_m) < \sum_{i \in \mathcal{C}_n} \frac{u_n x_{i,n}}{r_n} \phi'(\rho_n) = u_n \rho_n \phi'(\rho_n) \quad (10)$$

Observe that it implies $\rho_n > 0$. Assume first that $\rho_m > 0$. Note that $\psi'_n(0, y_n) < \psi'_m(0, y_m)$ can also be written as $u_n \phi(\rho_n) < u_m \phi(\rho_m)$. Together

with inequality 10, it implies that $\rho_n \frac{\phi'(\rho_n)}{\phi(\rho_n)} > \rho_m \frac{\phi'(\rho_m)}{\phi(\rho_m)}$, which according to Lemma 2 is equivalent to $\rho_n > \rho_m$. Clearly, $\rho_n > \rho_m$ also holds if $\rho_m = 0$. We thus get $u_n \phi(\rho_m) < u_n \phi(\rho_n) < u_m \phi(\rho_m)$ in both cases, which implies that $u_n < u_m$, as claimed.

Converse part: To prove $4 \implies 1$, we prove that $\neg 1 \implies \neg 4$. Assume that $u_n \phi(\rho_n) \geq u_m \phi(\rho_m)$. Since $\neg 1 \iff \neg 2$, we get $\frac{u_m x_{i,m}}{r_m} \phi'(\rho_m) \geq \frac{u_n x_{i,n}}{r_n} \phi'(\rho_n)$, $\forall i \in \mathcal{C}_n$. According to Lemma 8, $u_n \phi(\rho_n) \geq u_m \phi(\rho_m)$ implies $\mathcal{C}_n \subseteq \mathcal{C}_m$, and thus

$$u_m \rho_m \phi'(\rho_m) \geq \sum_{i \in \mathcal{C}_m} \frac{u_m x_{i,m}}{r_m} \phi'(\rho_m) \geq \sum_{i \in \mathcal{C}_n} \frac{u_n x_{i,n}}{r_n} \phi'(\rho_n) = u_n \rho_n \phi'(\rho_n).$$

Let us first assume that $\rho_n > 0$ and $\rho_m > 0$. Together with $u_n \phi(\rho_n) \geq u_m \phi(\rho_m)$, it implies that $\rho_n \frac{\phi'(\rho_n)}{\phi(\rho_n)} \leq \rho_m \frac{\phi'(\rho_m)}{\phi(\rho_m)}$, which is equivalent to $\rho_n \leq \rho_m$ according to Lemma 2. Clearly, $\rho_n \leq \rho_m$ still holds if $\rho_n = 0$. The case $\rho_m = 0 < \rho_n$ is impossible because $\mathcal{C}_n \subseteq \mathcal{C}_m$. We thus obtain that $\rho_n \leq \rho_m$ in all cases. But from $u_n \phi(\rho_n) \geq u_m \phi(\rho_m)$ and $\phi(\rho_m) \geq \phi(\rho_n)$, we deduce that $u_n \geq u_m$. We thus conclude that if $u_n < u_m$, then $u_n \phi(\rho_n) < u_m \phi(\rho_m)$, i.e., $\psi'_n(0, y_n) < \psi'_m(0, y_m)$. ■

Corollary 2 *From Proposition 2 it follows that*

1. $u_n < u_m \iff \frac{u_m x_{i,m}}{r_m} \phi'(\rho_m) < \frac{u_n x_{i,n}}{r_n} \phi'(\rho_n), \forall i \in \mathcal{C}_n \iff u_n \phi(\rho_n) < u_m \phi(\rho_m)$
2. *If $u_n < u_m$ then $\mathcal{C}_m \subset \mathcal{C}_n$.*
3. *If $u_n = u_m$ then $\mathcal{C}_m = \mathcal{C}_n$.*

The above corollary shows that we get a partition of classes among servers at the NEP: starting with a server m of maximal cost per unit capacity u_m and moving towards servers n with lower cost per unit capacity $u_n < u_m$, we observe more and more classes joining the servers, i.e. $\mathcal{C}_m \subset \mathcal{C}_n$.

Recall that it is assumed that the servers are numbered in the following order: $c_1/r_1 \leq c_2/r_2 \leq \dots \leq c_S/r_S$. According to the above properties, it implies that if we consider two servers n and $m > n$, then we have $\mathcal{C}_m \subseteq \mathcal{C}_n$, $u_n \phi(\rho_n) < u_m \phi(\rho_m)$ and $\frac{u_m x_{i,m}}{r_m} \phi'(\rho_m) < \frac{u_n x_{i,n}}{r_n} \phi'(\rho_n)$ for each class $i \in \mathcal{C}_n$, with the equalities holding if and only if $c_n/r_n = c_m/r_m$.

Let $\mathcal{S}_i = |\mathcal{S}_i|$. Proposition 2 implies that if $m \in \mathcal{S}_i$, then $n \in \mathcal{S}_i, \forall n < m$. As a consequence, the set \mathcal{S}_i has the following structure: $\mathcal{S}_i = \{1, \dots, \mathcal{S}_i\}$.

The following lemma proves that the above order on the servers implies exactly the same order on their utilization rates at the NEP.

Lemma 11

$$\forall j \in \mathcal{S}, \rho_j \geq \rho_{j+1}$$

Proof. The assertion is obviously true if $\rho_{j+1} = 0$. Let us therefore assume that $\rho_{j+1} > 0$. This implies that $\rho_j > 0$ because $\mathcal{C}_{j+1} \subseteq \mathcal{C}_j$. Since $\mathcal{C}_{j+1} \subseteq \mathcal{C}_j$ and $\frac{u_{j+1} x_{i,j+1}}{r_{j+1}} \phi'(\rho_{j+1}) < \frac{u_j x_{i,j}}{r_j} \phi'(\rho_j)$ for all $i \in \mathcal{C}_j$, we have

$$u_{j+1} \rho_{j+1} \phi'(\rho_{j+1}) = \sum_{i \in \mathcal{C}_{j+1}} \frac{u_{j+1} x_{i,j+1}}{r_{j+1}} \phi'(\rho_{j+1}) < \sum_{i \in \mathcal{C}_j} \frac{u_j x_{i,j}}{r_j} \phi'(\rho_j) = u_j \rho_j \phi'(\rho_j)$$

Together with $u_j \phi(\rho_j) \leq u_{j+1} \phi(\rho_{j+1})$, it implies that $\rho_{j+1} \frac{\phi'(\rho_{j+1})}{\phi(\rho_{j+1})} \leq \rho_j \frac{\phi'(\rho_j)}{\phi(\rho_j)}$, from which we deduce that $\rho_j \geq \rho_{j+1}$ using Lemma 2. ■

Before moving to the analysis of the set of servers used by each class at the equilibrium, we conclude this section with a last property related to the server costs per unit capacity. This technical result will play a key role when comparing the costs of two different equilibria.

Lemma 12

$$\psi'_j(y_j, y_j) \geq \psi'_{j+1}(y_{j+1}, y_{j+1}), \forall j,$$

with strict inequality if $\mathcal{C}_j \setminus \mathcal{C}_{j+1} \neq \emptyset$.

Proof. From (2), if $x_{i,j} > 0$, then

$$\mu_i = \psi'_j(x_{i,j}, y_j) = u_j \left[\phi(\rho_j) + \frac{x_{i,j}}{r_j} \phi'(\rho_j) \right],$$

from which we conclude that

$$\sum_{i \in \mathcal{C}_j} \mu_i = u_j [N_j \phi(\rho_j) + \rho_j \phi'(\rho_j)] = (N_j - 1) u_j \phi(\rho_j) + \psi'_j(y_j, y_j).$$

Similarly, we have

$$\sum_{i \in \mathcal{C}_{j+1}} \mu_i = (N_{j+1} - 1) u_{j+1} \phi(\rho_{j+1}) + \psi'_{j+1}(y_{j+1}, y_{j+1}).$$

Now,

$$\begin{aligned}
\sum_{i \in \mathcal{C}_{j+1}} \mu_i &= \sum_{i \in \mathcal{C}_j} \mu_i - \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} \mu_i \\
&= (N_j - 1) u_j \phi(\rho_j) + \psi'_j(y_j, y_j) - \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} u_j \left(\phi(\rho_j) + \frac{x_{i,j}}{r_j} \phi'(\rho_j) \right) \\
&= (N_j - 1) u_j \phi(\rho_j) + \psi'_j(y_j, y_j) - (N_j - N_{j+1}) u_j \phi(\rho_j) - u_j \phi'(\rho_j) \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} \frac{x_{i,j}}{r_j} \\
&= (N_{j+1} - 1) u_j \phi(\rho_j) + \psi'_j(y_j, y_j) - u_j \phi'(\rho_j) \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} \frac{x_{i,j}}{r_j}.
\end{aligned}$$

Thus,

$$\begin{aligned}
\psi'_j(y_j, y_j) - \psi'_{j+1}(y_{j+1}, y_{j+1}) &= (u_{j+1} \phi(\rho_{j+1}) - u_j \phi(\rho_j)) (N_{j+1} - 1) \\
&\quad + u_j \phi'(\rho_j) \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} \frac{x_{i,j}}{r_j}.
\end{aligned}$$

From Proposition 2, we have $u_{j+1} \phi(\rho_{j+1}) \geq u_j \phi(\rho_j)$. Moreover $\phi'(\rho_j) > 0$ because ϕ is strictly increasing. Since the second term on the RHS is strictly positive if $\mathcal{C}_j \setminus \mathcal{C}_{j+1} \neq \emptyset$, we can conclude that $\psi'_j(y_j, y_j) - \psi'_{j+1}(y_{j+1}, y_{j+1}) \geq 0$, with strict inequality if $\mathcal{C}_j \setminus \mathcal{C}_{j+1} \neq \emptyset$. ■

3.4 Characterization of the set of servers used

The following proposition shows that the set of servers used by each class has the so-called water-filling structure.

Proposition 3 *For each class $i \in \mathcal{C}$, the threshold S_i is such that*

$$G_{S_i} < \lambda_i \leq G_{S_i+1}, \quad (11)$$

where

$$G_s = \sum_{j=1}^s \frac{r_j}{u_j \phi'(\rho_j)} [u_s \phi(\rho_s) - u_j \phi(\rho_j)] \quad s = 1, \dots, S. \quad (12)$$

Proof. Let A be a subset of \mathcal{S}_i . Equation (2) can be rewritten as $\frac{r_j}{u_j \phi'(\rho_j)} \mu_i = \frac{r_j}{\phi'(\rho_j)} \phi(\rho_j) + x_{i,j}$ for all $j \in A$. Summing over $j \in A$, we get

$$\mu_i = \frac{\sum_{j \in A} \frac{r_j}{\phi'(\rho_j)} \phi(\rho_j) + x_{i,j}}{\sum_{j \in A} \frac{r_j}{u_j \phi'(\rho_j)}} \quad A \subseteq \mathcal{S}_i, \quad (13)$$

which in the case $A = \mathcal{S}_i$ can be written as

$$\mu_i = \frac{\lambda_i + \sum_{j \in \mathcal{S}_i} \frac{r_j}{\phi'(\rho_j)} \phi(\rho_j)}{\sum_{j \in \mathcal{S}_i} \frac{r_j}{u_j \phi'(\rho_j)}}. \quad (14)$$

Since $S_{i+1} \notin \mathcal{S}_i$, we have from (3)

$$\mu_i = \frac{\lambda_i + \sum_{j \in \mathcal{S}_i} \frac{r_j}{\phi'(\rho_j)} \phi(\rho_j)}{\sum_{j \in \mathcal{S}_i} \frac{r_j}{u_j \phi'(\rho_j)}} \leq u_{S_{i+1}} \phi(\rho_{S_{i+1}}),$$

which yields $\lambda_i \leq G_{S_{i+1}}$ after some algebra. For the lower bound, observe that, since $S_i \in \mathcal{S}_i$, (2) holds for $j = S_i$. Therefore, using (13) with $A = \{1, \dots, S_i - 1\}$, we can write

$$\mu_i = \frac{\sum_{j=1}^{S_i-1} \frac{r_j}{\phi'(\rho_j)} \phi(\rho_j) + x_{i,j}}{\sum_{j=1}^{S_i-1} \frac{r_j}{u_j \phi'(\rho_j)}} = u_{S_i} \left[\phi(\rho_{S_i}) + \frac{x_{i,S_i}}{r_{S_i}} \phi'(\rho_{S_i}) \right].$$

With $\lambda_i > \sum_{j=1}^{S_i-1} x_{i,j}$, it yields

$$\lambda_i > u_{S_i} \left[\phi(\rho_{S_i}) + \frac{x_{i,S_i}}{r_{S_i}} \phi'(\rho_{S_i}) \right] - \sum_{j=1}^{S_i-1} \frac{r_j}{u_j \phi'(\rho_j)} - \sum_{j=1}^{S_i-1} \frac{r_j}{\phi'(\rho_j)} \phi(\rho_j) + x_{i,j},$$

from which we get $\lambda_i > G_{S_i}$ since $\frac{x_{i,S_i}}{r_{S_i}} \phi'(\rho_{S_i}) > 0$. ■

Remark 1 In the special case $\lambda_i = G_{i,S_{i+1}}$, inequality (3) holds tight for $j = S_i + 1$. Therefore, in this case, we can define the set of servers used by class i as $\mathcal{S}_i = \{1, \dots, S_i, S_i + 1\}$, where server $S_i + 1$ is “marginally” used, with $x_{i,S_{i+1}} = 0$.

From Corollary 1, we can conclude that the thresholds S_1, \dots, S_K satisfy the order $S_1 \leq S_2 \leq \dots \leq S_K$.

4 Analysis of the Social Cost

In this section, it will be assumed that the cost functions vector ψ is fixed. Our goal is to prove that the social cost $D_K(\lambda, \psi)$ achieves its maximum in the symmetric case, i.e. when $\lambda = \lambda^\equiv$.

For each rate vector $\lambda \in \Lambda$, we already know that there exists a unique NEP $\mathbf{x} \in \mathcal{X}$. Let us define the function $\mathcal{N} : \Lambda \rightarrow \mathcal{X}$ such that for each vector $\lambda \in \Lambda$, $\mathcal{N}(\lambda) \in \mathcal{X}$ is this unique NEP. In the sequel, the function \mathcal{N} will be called the Nash mapping. We have the following result.

Theorem 2 *The Nash mapping \mathcal{N} is a continuous function from Λ into \mathcal{X} .*

Proof. Note that for each vector $\lambda \in \Lambda$, $\mathcal{N}(\lambda) \in \mathcal{X} \subset \bigotimes_{i \in \mathcal{C}} \bigotimes_{j \in \mathcal{S}} [0, r_j]$ and the latter set is a compact set. As a consequence, in order to apply the closed graph theorem (see Appendix A), we only need to show that the graph $\mathcal{G}_{\mathcal{N}}$ of \mathcal{N} is closed. Let us therefore consider a convergent sequence $(\lambda^n, x^n)_{n \in \mathbb{N}}$ of points in $\mathcal{G}_{\mathcal{N}}$, where $x^n = \mathcal{N}(\lambda^n)$. Let (λ, x) denote the limit of this sequence. Note that $\lambda \in \Lambda$ since Λ is closed as a topological space. We need to show that $x = \mathcal{N}(\lambda)$.

We first show that $\mathbf{x} \in \mathcal{X}$. For $i \in \mathcal{C}$ and $j \in \mathcal{S}$ fixed, the sequence $(x_{i,j}^n)_{n \in \mathbb{N}}$ takes values in the closed set $[0, r_j]$ and converges to $x_{i,j}$, from which we deduce that $0 \leq x_{i,j} \leq r_j$. Moreover, for each $n \in \mathbb{N}$ and each $i \in \mathcal{C}$, we have $\sum_{j \in \mathcal{S}} x_{i,j}^n = \lambda_i^n$. Since $x^n \rightarrow x$, we have for each class $i \in \mathcal{C}$,

$$\lambda_i = \lim_{n \rightarrow \infty} \lambda_i^n = \lim_{n \rightarrow \infty} \sum_{j \in \mathcal{S}} x_{i,j}^n = \sum_{j \in \mathcal{S}} x_{i,j}.$$

We therefore conclude that the limit point \mathbf{x} is such that $\mathbf{x} \in \mathcal{X}$. Let us now show that \mathbf{x} is the NEP associated to the rate vector λ . For each $n \in \mathbb{N}$, since \mathbf{x}^n is a NEP, there exist multipliers μ_i^n , $i \in \mathcal{C}$, such that

$$\mu_i^n = u_j \left[\phi(\rho_j^n) + \frac{x_{i,j}^n}{r_j} \phi'(\rho_j^n) \right] \quad \text{if } x_{i,j}^n > 0 \quad (15)$$

$$\mu_i^n \leq u_j \phi(\rho_j^n) \quad \text{if } x_{i,j}^n = 0. \quad (16)$$

Now, let us show that there exist $\mu_i < +\infty$ such that $\mu_i^n \rightarrow \mu_i$, $i \in \mathcal{C}$. Let $y_j^n = \sum_{i \in \mathcal{C}} x_{i,j}^n$, $\rho_j^n = y_j^n / r_j$ and y_j and ρ_j be the respective limits when $n \rightarrow \infty$. The proof is an immediate consequence of (15) and (16) for type-I congestion functions since $\phi(\rho_j^n) \rightarrow \phi(\rho_j) < +\infty$ and $\phi'(\rho_j^n) \rightarrow \phi'(\rho_j) < +\infty$ as $n \rightarrow \infty$ for all $j \in \mathcal{S}$ (Lemma 1).

Let us therefore assume that ϕ is a type-II congestion function and prove that $\mu_i^n \rightarrow \mu_i < +\infty$ as $n \rightarrow \infty$. To this end, we first prove that $\rho_1^n \rightarrow \rho_1 < 1$. Assume on the contrary that $\rho_1 = 1$. In this case, the relation $\phi(\rho_j^n) \geq \frac{u_j}{c_j} \phi(\rho_1^n)$ implies that $\phi(\rho_j^n) \rightarrow \infty$ for all $j \in \mathcal{S}$, and hence that $\rho_j^n \rightarrow 1, \forall j \in \mathcal{S}$. We thus get $y_j^n \rightarrow r_j, \forall j \in \mathcal{S}$, which implies that $\bar{\lambda} = \sum_{j \in \mathcal{S}} y_j^n \rightarrow \bar{r} = \sum_{j \in \mathcal{S}} r_j$. This is clearly a contradiction with our assumption $\bar{\lambda} < \bar{r}$.

As a consequence, we have $\rho < 1$, which implies that $\phi(\rho_1^n) \rightarrow \phi(\rho_1) < +\infty$ and $\phi'(\rho_1^n) \rightarrow \phi'(\rho_1) < +\infty$ (Lemma 1) as $n \rightarrow \infty$. Since all classes use server 1, we have

$$\mu_i^n = u_1 \left[\phi(\rho_1^n) + \frac{x_{i,1}^n}{r_1} \phi'(\rho_1^n) \right] \quad i \in \mathcal{C}.$$

Taking the limit as $n \rightarrow \infty$, we obtain that $\mu_i^n \rightarrow \mu_i < +\infty$, where

$$\mu_i = u_1 \left[\phi(\rho_1) + \frac{x_{i,1}}{r_1} \phi'(\rho_1) \right] \quad i \in \mathcal{C}.$$

Observe that it implies that $\rho_j < 1, \forall j \in \mathcal{S}$. Indeed, if $\rho_j > 0$, then there exists $k \in \mathbb{N}$ such that $\rho_j^n > 0, \forall n \geq k$. Therefore, there exists $i \in \mathcal{C}$ such that $x_{i,j}^n > 0$ for all $n \geq k$. With (4), it yields $\phi(\rho_j^n) < \frac{r_j}{c_j} \mu_i^n$ for all $n \geq k$. As a consequence, we have $\lim_{n \rightarrow \infty} \phi(\rho_j^n) \leq \frac{r_j}{c_j} \mu_i$, which proves that $\rho_j^n \rightarrow \rho_j < 1$ as $n \rightarrow \infty$.

We have proved above that $\mu_i^n \rightarrow \mu_i < +\infty$ as $n \rightarrow \infty$ for both type-I and type-II congestion functions. Now, let us consider the complementary slackness conditions for each NEP \mathbf{x}^n , that is,

$$x_{i,j}^n \left(u_j \left[\phi(\rho_j^n) + \frac{x_{i,j}^n}{r_j} \phi'(\rho_j^n) \right] - \mu_i^n \right) = 0, \quad j \in \mathcal{S}, i \in \mathcal{C}.$$

Taking the limit as $n \rightarrow \infty$, we obtain

$$x_{i,j} \left(u_j \left[\phi(\rho_j) + \frac{x_{i,j}}{r_j} \phi'(\rho_j) \right] - \mu_i \right) = 0, \quad j \in \mathcal{S}, i \in \mathcal{C}.$$

Since the above are the necessary and sufficient optimality conditions for a point to be a NEP, we conclude that $x = \mathcal{N}(\lambda)$ and thus that the graph $\mathcal{G}_{\mathcal{N}}$ is closed. Applying the closed graph theorem yields the proof. ■

In order to prove that the global cost achieves its maximum in the symmetric case, we need to compare the equilibria $\mathcal{N}(\boldsymbol{\lambda})$ and $\mathcal{N}(\hat{\boldsymbol{\lambda}})$ that are induced by two different rate vectors $\boldsymbol{\lambda}$ and $\hat{\boldsymbol{\lambda}}$ in Λ . If the resulting equilibria are such that the set of servers over which each class sends its flow do not coincide at both equilibria, then the comparisons become extremely complex, if possible at all.

To avoid this difficulty, we proceed as follows. In section 4.1, we first prove some preliminary results concerning the comparison of the equilibria induced

by two different rate vectors λ and $\hat{\lambda}$, assuming that these equilibria are such that each class sends its flow to the same servers under both equilibria. In Section 4.2, we compare the equilibria induced by two different rate vectors λ and $\hat{\lambda}$, assuming that (i) these equilibria are such that each class sends its flow to the same servers under both equilibria, and (ii) $\hat{\lambda}$ is obtained from λ using a certain transformation. In Section 4.3 we exploit the continuity of the Nash mapping to show that the global cost increases under this transformation even when the set of servers is different at the two equilibria. Finally, in Section 4.4, we show that the symmetric rate vector λ^π can be obtained from any rate vector λ with a finite number of such transformations.

4.1 Preliminary Results

In this section, we prove some lemmata that will be used in order to compare the Nash equilibria induced by two vectors $\lambda \in \Lambda$ and $\hat{\lambda} \in \Lambda$. In the sequel, if z is a certain quantity related to the Nash equilibrium induced by the vector λ then we shall denote the corresponding quantity for vector $\hat{\lambda}$ by \hat{z} .

Lemma 13 *For $i \in \mathcal{C}_j$,*

1. *if $\hat{y}_j < y_j$ and $\hat{x}_{i,j} \leq x_{i,j}$, then $\hat{\mu}_i < \mu_i$.*
2. *if $\hat{y}_j \leq y_j$ and $\hat{x}_{i,j} \leq x_{i,j}$, then $\hat{\mu}_i \leq \mu_i$.*
3. *if $\hat{y}_j \leq y_j$ and $\hat{x}_{i,j} < x_{i,j}$, then $\hat{\mu}_i < \mu_i$.*
4. *if $\hat{y}_j = y_j$ and $\hat{\mu}_i < \mu_i$, then $\hat{x}_{i,j} < x_{i,j}$.*

Proof. Proof of part 1 : for $i \in \mathcal{C}_j$, we have $\mu_i = \psi'_j(x_{i,j}, y_j)$ and $\hat{\mu}_i \leq \psi'_j(\hat{x}_{i,j}, \hat{y}_j)$. Since $\psi'_j(x, y)$ is strictly increasing in each of its two arguments, $\hat{y}_j < y_j$ and $\hat{x}_{i,j} \leq x_{i,j}$ implies that $\hat{\mu}_i \leq \psi'_j(\hat{x}_{i,j}, \hat{y}_j) < \psi'_j(x_{i,j}, y_j) = \mu_i$. The proofs of parts 2, 3, and 4 follow similarly. ■

Lemma 14 *For m and n in \mathcal{S} , and $i \in \mathcal{C}_m \cap \mathcal{C}_n$,*

if $\hat{y}_m > y_m$, $\hat{x}_{i,m} \geq x_{i,m}$, and $\hat{y}_n \leq y_n$, then $\hat{x}_{i,n} > x_{i,n}$.

Proof. Assume the contrary, that is, $\exists n, m \in \mathcal{S}$ and $i \in \mathcal{C}_m \cap \mathcal{C}_n$ such that $\hat{y}_m > y_m$, $\hat{x}_{i,m} \geq x_{i,m}$, $\hat{y}_n \leq y_n$ and $\hat{x}_{i,n} \leq x_{i,n}$. From Lemma 13.1, $\hat{y}_m > y_m$ and $\hat{x}_{i,m} \geq x_{i,m}$ implies $\hat{\mu}_i > \mu_i$. However, from Lemma 13.2, $\hat{y}_n \leq y_n$ and $\hat{x}_{i,n} \leq x_{i,n}$ implies $\hat{\mu}_i \leq \mu_i$, which is a contradiction. ■

In the rest of the section, we shall make the following assumption on the vectors λ and $\hat{\lambda}$.

Assumption 1 The vectors λ and $\hat{\lambda}$ are such that $\mathcal{C}_j = \hat{\mathcal{C}}_j, \forall j \in \mathcal{S}$.

From the above assumption, it follows that $\mathcal{S}_i = \hat{\mathcal{S}}_i, \forall i \in \mathcal{C}$.

Lemma 15 For any $j \in \mathcal{S}$:

1. $\hat{y}_j \geq y_j \iff \sum_{i \in \mathcal{C}_j} \hat{\mu}_i \geq \sum_{i \in \mathcal{C}_j} \mu_i$.
2. $\hat{y}_j > y_j \iff \sum_{i \in \mathcal{C}_j} \hat{\mu}_i > \sum_{i \in \mathcal{C}_j} \mu_i$.

Proof. Proof of part 1: from (2), if $i \in \mathcal{C}_j$, then

$$\mu_i = \psi'_j(x_{i,j}, y_j) = u_j \left[\phi(\rho_j) + \frac{x_{i,j}}{r_j} \phi'(\rho_j) \right]$$

Thus,

$$\sum_{i \in \mathcal{C}_j} \mu_i = (N_j - 1) u_j \phi(\rho_j) + \psi'_j(y_j, y_j)$$

Since $N_j = \hat{N}_j$ (from Assumption 1) and $\psi'_j(x, y)$ is strictly increasing in each of its two arguments, we can conclude that $\sum_{i \in \mathcal{C}_j} \mu_i$ is an increasing function of y_j .

The proof of part 2 follows similarly. ■

Lemma 16 If $\mathcal{C}_m = \mathcal{C}_n$ then :

1. $\hat{y}_m \geq y_m \iff \hat{y}_n \geq y_n$.
2. $\hat{y}_m > y_m \iff \hat{y}_n > y_n$.

Proof. Proof of part 1: from Lemma 15, $\hat{y}_m \geq y_m$ is equivalent to $\sum_{i \in \mathcal{C}_m} \hat{\mu}_i \geq \sum_{i \in \mathcal{C}_m} \mu_i$, which, since $\mathcal{C}_m = \mathcal{C}_n$, is equivalent to $\sum_{i \in \mathcal{C}_n} \hat{\mu}_i \geq \sum_{i \in \mathcal{C}_n} \mu_i$. Again, from Lemma 15, we can conclude that $\hat{y}_n \geq y_n$.

The proof of part 2 follows similarly. ■

Corollary 3 For $m, n \in \mathcal{S}_1, \hat{y}_m > y_m \iff \hat{y}_n > y_n$.

Proof. Recall from corollary 1 that $\mathcal{S}_1 \subseteq \mathcal{S}_k$ for all $k \in \mathcal{C}$. Since, for $m, n \in \mathcal{S}_1, \mathcal{C}_m = \mathcal{C}_n = \{1, 2, \dots, K\}$, the above statement follows from Lemma 16.2. ■

4.2 Basic Transformation of a Rate Vector

For each rate vector $\lambda \in \Lambda$, recall that by convention $\lambda_1 = \min_{i \in \mathcal{C}} \lambda_i$ and $\lambda_K = \max_{i \in \mathcal{C}} \lambda_i$. Define the sets \mathcal{C}_{min} and \mathcal{C}_{max} as follows:

$$\begin{aligned}\mathcal{C}_{min} &= \{i \in \mathcal{C} : \lambda_i = \lambda_1\}, \\ \mathcal{C}_{max} &= \{i \in \mathcal{C} : \lambda_i = \lambda_K\},\end{aligned}$$

and let $n_{min} = |\mathcal{C}_{min}|$ and $n_{max} = |\mathcal{C}_{max}|$.

Definition 1 For each rate vector $\lambda \in \Lambda$, define the function $h_\lambda : [0, n_{max} \lambda_K] \rightarrow \Lambda$ as follows:

$$h_\lambda(\epsilon) = \lambda + \epsilon \left(\frac{1}{n_{min}} \sum_{i \in \mathcal{C}_{min}} \mathbf{e}_i - \frac{1}{n_{max}} \sum_{i \in \mathcal{C}_{max}} \mathbf{e}_i \right), \quad (17)$$

where \mathbf{e}_i denotes the vector in \mathbb{R}^K with the i -th component equals to 1 and all other components are equal to 0. A rate vector $\hat{\lambda} \in \Lambda$ is said to be obtained from λ under a basic transformation if and only if there exists $\epsilon \in [0, n_{max} \lambda_K]$ such that $\hat{\lambda} = h_\lambda(\epsilon)$. In this case, ϵ is called the step of the transformation.

Note that the above transformation increases the traffic of classes $i \in \mathcal{C}_{min}$ (the classes with the smallest amount of traffic) and decrease correspondingly the traffic of classes $i \in \mathcal{C}_{max}$ (the classes with the largest amount of traffic), i.e., it preserves the total amount of traffic: $\sum_{i \in \mathcal{C}} \lambda_i = \sum_{i \in \mathcal{C}} \hat{\lambda}_i = \bar{\lambda}$. There are several other properties of the basic transformation which are worthwhile noticing. They are stated in the following lemma.

Lemma 17 Let $\hat{\lambda}$ be obtained from λ under a basic transformation, i.e., $\hat{\lambda} = h_\lambda(\epsilon)$. Then,

1. $\hat{\lambda}_i \geq \lambda_i, \forall i \in \mathcal{C}_{min}, \hat{\lambda}_i \leq \lambda_i, \forall i \in \mathcal{C}_{max}$, and $\hat{\lambda}_i = \lambda_i, \forall i \notin \mathcal{C}_{min} \cup \mathcal{C}_{max}$, where the inequalities are strict if and only if $\mathcal{C}_{min} \neq \mathcal{C}_{max}$ and $\epsilon > 0$.
2. $\hat{\lambda}_i = \hat{\lambda}_1 \forall i \in \mathcal{C}_{min}$ and $\hat{\lambda}_i = \hat{\lambda}_K \forall i \in \mathcal{C}_{max}$,
3. $\hat{\lambda}_i \leq \lambda_K \forall i \in \mathcal{C}_{min}$ and $\hat{\lambda}_i \geq \lambda_1 \forall i \in \mathcal{C}_{max}$ for $\epsilon \leq \min(n_{min}, n_{max})(\lambda_K - \lambda_1)$.
4. $\sum_{i \in \mathcal{C}_{min}} \hat{\lambda}_i - \lambda_i = - \sum_{i \in \mathcal{C}_{max}} \hat{\lambda}_i - \lambda_i$,
5. $\sum_{j \in \mathcal{S}} y_j = \sum_{j \in \mathcal{S}} \hat{y}_j$,

Proof. In part 1, if either $\mathcal{C}_{min} = \mathcal{C}_{max}$ or $\epsilon = 0$, then $\hat{\lambda} = \lambda$. Hence, the equalities are satisfied. So, we consider the case when $\mathcal{C}_{min} \neq \mathcal{C}_{max}$ and $\epsilon > 0$, and show that the inequalities are strict. From (17), it is immediate that $\hat{\lambda}_i = \lambda_i$ for $i \notin \mathcal{C}_{min} \cup \mathcal{C}_{max}$. Moreover $\hat{\lambda}_i = \lambda_i + \frac{\epsilon}{n_{min}}$ for all $i \in \mathcal{C}_{min}$ and $\hat{\lambda}_i = \lambda_i - \frac{\epsilon}{n_{max}}$ for all $i \in \mathcal{C}_{max}$. Since $\epsilon > 0$, we thus obtain $\hat{\lambda}_i > \lambda_i$ for $i \in \mathcal{C}_{min}$ and $\hat{\lambda}_i < \lambda_i$ for $i \in \mathcal{C}_{max}$, and 1 is proved.

We note that if $\mathcal{C}_{min} = \mathcal{C}_{max}$, then $\hat{\lambda} = \lambda$. Hence, the parts 2 to 5. So, for the rest of the proof, we assume that $\mathcal{C}_{min} \neq \mathcal{C}_{max}$. From $\lambda_i = \lambda_1$ for $i \in \mathcal{C}_{min}$ and $\lambda_i = \lambda_K$ for $i \in \mathcal{C}_{max}$, we obtain that $\hat{\lambda}_i = \lambda_1 + \frac{\epsilon}{n_{min}} = \hat{\lambda}_1$ for $i \in \mathcal{C}_{min}$ and $\hat{\lambda}_i = \lambda_K + \frac{\epsilon}{n_{max}} = \hat{\lambda}_K$ for $i \in \mathcal{C}_{max}$, which proves 2.

To prove 3, let us consider $i \in \mathcal{C}_{min}$. We observe that $\epsilon \leq \min(n_{min}, n_{max})(\lambda_K - \lambda_1)$ implies that $\hat{\lambda}_i = \lambda_i + \frac{\epsilon}{n_{min}} \leq \lambda_i + \frac{\min(n_{min}, n_{max})}{n_{min}}(\lambda_K - \lambda_1) \leq \lambda_i + (\lambda_K - \lambda_1) = \lambda_K$. The proof of $\hat{\lambda}_i \geq \lambda_1$ for $i \in \mathcal{C}_{max}$ is symmetric.

To prove 4, we observe that,

$$\sum_{i \in \mathcal{C}_{min}} \hat{\lambda}_i - \lambda_i = \sum_{i \in \mathcal{C}_{min}} \frac{\epsilon}{n_{min}} = \epsilon = \sum_{i \in \mathcal{C}_{max}} \frac{\epsilon}{n_{max}} = - \sum_{i \in \mathcal{C}_{max}} \hat{\lambda}_i - \lambda_i$$

Finally, the proof of property 5 is immediate since it is equivalent to $\sum_{i \in \mathcal{C}} \lambda_i = \sum_{i \in \mathcal{C}} \hat{\lambda}_i$. ■

In the following, we will compare two rate vectors λ and $\hat{\lambda}$. If z is a certain quantity related to the Nash equilibrium induced by the vector λ then we shall denote the corresponding quantity for vector $\hat{\lambda}$ by \hat{z} . The comparison is done under the following assumption.

Assumption 2 *The rate vectors $\lambda \in \Lambda$ and $\hat{\lambda} \in \Lambda$ are such that:*

1. $\hat{\lambda}$ is obtained from λ under a basic transformation,
2. $\mathcal{C}_j = \hat{\mathcal{C}}_j, \forall j \in \mathcal{S}$.

In other words, we assume that the transformation λ into $\hat{\lambda}$ leaves unaffected the set of servers used by each class.

The key point here is that in order to determine the impact of a basic transformation of the rate vector λ on the global cost, we need to compare the server loads under the equilibria $\mathbf{x} = \mathcal{N}(\lambda)$ and $\hat{\mathbf{x}} = \mathcal{N}(\hat{\lambda})$. To this end, let us define the sets \mathcal{S}^+ and \mathcal{S}^- as follows:

$$\mathcal{S}^+ = \{j \in \mathcal{S} : \hat{y}_j > y_j\} \quad \text{and} \quad \mathcal{S}^- = \mathcal{S} \setminus \mathcal{S}^+,$$

i.e., \mathcal{S}^+ is the set of servers whose load increases under the transformation while \mathcal{S}^- is the set of servers whose load is non-increasing under the transformation.

We first prove two lemmata concerning the sets \mathcal{S}^+ and \mathcal{S}^- . The first one shows that \mathcal{S}^+ is empty if and only if the load of each and every server is constant under the transformation.

Lemma 18 $y_j = \hat{y}_j, \forall j \in \mathcal{S} \iff \mathcal{S}^+ = \emptyset$.

Proof. If $\mathcal{S}^+ = \emptyset$ then $\mathcal{S}^- = \mathcal{S}$. That is, $\hat{y}_j \leq y_j, \forall j \in \mathcal{S}$. We also have $\sum_{j \in \mathcal{S}} \hat{y}_j = \sum_{j \in \mathcal{S}} y_j$. This is possible only if $\hat{y}_j = y_j, \forall j \in \mathcal{S}$.

The converse is true by definition of \mathcal{S}^+ . ■

The second lemma shows that \mathcal{S}^- cannot be empty, i.e. that there is at least one server whose load is non-increasing under the transformation.

Lemma 19 $\mathcal{S}^- \neq \emptyset$.

Proof. Assume $\mathcal{S}^- = \emptyset$, then $y_j < \hat{y}_j, \forall j \in \mathcal{S}$. Therefore, $\sum_{j \in \mathcal{S}} y_j < \sum_{j \in \mathcal{S}} \hat{y}_j$. This is in contradiction with Assumption 2 which says $\sum_{j \in \mathcal{S}} y_j = \sum_{j \in \mathcal{S}} \hat{y}_j$. ■

We now prove three fundamental propositions regarding the impact of the transformation on server loads. We first show in proposition 4 that if there exists at least one server whose load increases under the transformation, then the load of each and every server used by class 1 increases. We then prove in proposition 5 that the load of all servers is non-increasing under the transformation if and only if all traffic classes use the same set of servers. Finally, proposition 6 proves that the transformation induces a monotonic partition of servers: there exists a threshold $J < S$ such that for all servers $j > J$ the load is non-increasing under the transformation.

Proposition 4 *If $\mathcal{S}^+ \neq \emptyset$ then $\mathcal{S}_1 \subset \mathcal{S}^+$.*

Proof. Assume by contradiction that we can find a server $s \in \mathcal{S}_1$ such that $s \in \mathcal{S}^-$. Then, according to Corollary 3, $\mathcal{S}_1 \subset \mathcal{S}^-$. Since $\mathcal{S}^+ \neq \emptyset$ and $\hat{y}_j > y_j$ for all $j \in \mathcal{S}^+$, we have $\sum_{j \in \mathcal{S}^+} \hat{y}_j > \sum_{j \in \mathcal{S}^+} y_j$, i.e.,

$$\sum_{i \in \mathcal{C}} \left(\sum_{j \in \mathcal{S}^+} \hat{x}_{i,j} \right) > \sum_{i \in \mathcal{C}} \left(\sum_{j \in \mathcal{S}^+} x_{i,j} \right),$$

from which we conclude that there exists i such that $\sum_{j \in \mathcal{S}^+} \hat{x}_{i,j} > \sum_{j \in \mathcal{S}^+} x_{i,j}$. Since $\mathcal{S}_k = \mathcal{S}_1 \subset \mathcal{S}^-$ for all $k \in \mathcal{C}_{min}$, we necessarily have $i \notin \mathcal{C}_{min}$ and thus $\hat{\lambda}_i \leq \lambda_i$. Therefore,

$$\hat{\lambda}_i = \sum_{j \in \mathcal{S}^-} \hat{x}_{i,j} + \sum_{j \in \mathcal{S}^+} \hat{x}_{i,j} \leq \sum_{j \in \mathcal{S}^-} x_{i,j} + \sum_{j \in \mathcal{S}^+} x_{i,j} = \lambda_i.$$

Thus,

$$\sum_{j \in \mathcal{S}^-} \hat{x}_{i,j} \leq \sum_{j \in \mathcal{S}^-} x_{i,j} + \left(\sum_{j \in \mathcal{S}^+} x_{i,j} - \sum_{j \in \mathcal{S}^+} \hat{x}_{i,j} \right) < \sum_{j \in \mathcal{S}^-} x_{i,j}.$$

We therefore conclude that class i is such that $\sum_{j \in \mathcal{S}^+} \hat{x}_{i,j} > \sum_{j \in \mathcal{S}^+} x_{i,j}$ and $\sum_{j \in \mathcal{S}^-} \hat{x}_{i,j} < \sum_{j \in \mathcal{S}^-} x_{i,j}$. Therefore, we can find a server $m \in \mathcal{S}^+$ and a server $n \in \mathcal{S}^-$ such that $\hat{x}_{i,m} > x_{i,m}$ and $\hat{x}_{i,n} < x_{i,n}$. But according to Lemma 14, this is impossible. We therefore conclude that $\mathcal{S}_1 \subset \mathcal{S}^+$. ■

Proposition 5 $\mathcal{S}^+ = \emptyset \iff \mathcal{S}_1 = \mathcal{S}_K$.

Proof. We first prove that if $\mathcal{S}^+ = \emptyset$ then $\mathcal{S}_1 = \mathcal{S}_K$. From Lemma 18, this is equivalent to proving that if $y_j = \hat{y}_j, \forall j \in \mathcal{S}$ then $\mathcal{S}_1 = \mathcal{S}_K$. Assume the contrary, that is $\mathcal{S}_1 \subsetneq \mathcal{S}_K$. Then, $\exists m : m \in \mathcal{S}_K, m \notin \mathcal{S}_1$.

Since $y_m = \hat{y}_m$, from Lemma 15, we get $\sum_{i \in \mathcal{C}_m} \mu_i = \sum_{i \in \mathcal{C}_m} \hat{\mu}_i$, which we can rewrite as

$$\sum_{i \in \mathcal{C}_{max}} \mu_i + \sum_{i \in \mathcal{C}_m \setminus \mathcal{C}_{max}} \mu_i = \sum_{i \in \mathcal{C}_{max}} \hat{\mu}_i + \sum_{i \in \mathcal{C}_m \setminus \mathcal{C}_{max}} \hat{\mu}_i. \quad (18)$$

We shall show that the above equality is not possible, which then proves the claim.

For $i \in \mathcal{C}_{max}$, since $\lambda_i > \hat{\lambda}_i$, $\sum_{j \in \mathcal{S}_i} x_{i,j} > \sum_{j \in \mathcal{S}_i} \hat{x}_{i,j}$. Thus, there exists an $n \in \mathcal{S}_i$ such that $x_{i,n} > \hat{x}_{i,n}$. Since $y_n = \hat{y}_n$, from Lemma 13.3, we can conclude that $\mu_i > \hat{\mu}_i$, and that $\sum_{i \in \mathcal{C}_{max}} \mu_i > \sum_{i \in \mathcal{C}_{min}} \hat{\mu}_i$, which, upon substitution in (18), leads to

$$\sum_{i \in \mathcal{C}_m \setminus \mathcal{C}_{max}} \mu_i < \sum_{i \in \mathcal{C}_m \setminus \mathcal{C}_{max}} \hat{\mu}_i.$$

If $\mathcal{C}_m \setminus \mathcal{C}_{max} = \emptyset$, then the above inequality cannot be possible which then proves the claim. So, assume $\mathcal{C}_m \setminus \mathcal{C}_{max} \neq \emptyset$. Then the above inequality implies that $\exists i \notin \mathcal{C}_{min} \cup \mathcal{C}_{max} : \mu_i < \hat{\mu}_i$. Since $y_j = \hat{y}_j, \forall j \in \mathcal{S}_i$, application of Lemma 13.4 leads to $x_{i,j} < \hat{x}_{i,j}, \forall j \in \mathcal{S}_i$, and consequently to $\lambda_i = \sum_{j \in \mathcal{S}_i} x_{i,j} < \sum_{j \in \mathcal{S}_i} \hat{x}_{i,j} = \hat{\lambda}_i$. However, for $i \notin \mathcal{C}_{min} \cup \mathcal{C}_{max}$, from Lemma 17.1, $\lambda_i = \hat{\lambda}_i$. Hence, there is a contradiction, and we can conclude that $\mathcal{S}_1 = \mathcal{S}_K$. ■

As a direct consequence of the above proposition, we get the following corollary that tells us that if at equilibria \mathbf{x} and $\hat{\mathbf{x}}$ all classes use the same set of servers, then the server loads are constant under the transformation.

Corollary 4 $y_j = \hat{y}_j, \forall j \in \mathcal{S} \iff \mathcal{S}_1 = \mathcal{S}_K$.

We now turn our attention to the set \mathcal{S}^- and prove the following result.

Proposition 6 *For all $j \in \mathcal{S}$, if $j \in \mathcal{S}^-$ then $j + 1 \in \mathcal{S}^-$.*

Proof. If $\mathcal{S}^+ = \emptyset$ then the proposition is true. So, assume $\mathcal{S}^+ \neq \emptyset$. Then, from Proposition 4, $\mathcal{S}_1 \subset \mathcal{S}^+$. In order to prove the proposition, assume by contradiction that there exists a server $j \in \{S_1 + 1, \dots, S_K - 1\}$ such that $j \in \mathcal{S}^-$ and $j + 1 \in \mathcal{S}^+$. Again, if $S_1 + 1 = S_K$ then the proposition is true. So, assume that $S_1 + 1 < S_K$.

Since $j \in \mathcal{S}^-$ and $j + 1 \in \mathcal{S}^+$, from Lemma 16,

$$\sum_{i \in \mathcal{C}_j} \hat{\mu}_i \leq \sum_{i \in \mathcal{C}_j} \mu_i, \quad (19)$$

and

$$\sum_{i \in \mathcal{C}_{j+1}} \hat{\mu}_i > \sum_{i \in \mathcal{C}_{j+1}} \mu_i, \quad (20)$$

Moreover, from the contrapositive of Lemma 16, we can conclude that $\mathcal{C}_j \setminus \mathcal{C}_{j+1} \neq \emptyset$. Note that since $j < S_K$, classes $i \in \mathcal{C}_{max}$ do not belong to $\mathcal{C}_j \setminus \mathcal{C}_{j+1}$. Similarly, since $j > S_1$, classes $i \in \mathcal{C}_{min}$ do not belong to $\mathcal{C}_j \setminus \mathcal{C}_{j+1}$.

Since $\mathcal{C}_{j+1} \subset \mathcal{C}_j$, we have, for all $i \in \mathcal{C}_{j+1}$,

$$\mu_i = \psi'_j(x_{i,j}, y_j) = u_j \left[\phi(\rho_j) + \frac{x_{i,j}}{r_j} \phi'(\rho_j) \right].$$

Therefore, $\sum_{i \in \mathcal{C}_{j+1}} \hat{\mu}_i > \sum_{i \in \mathcal{C}_{j+1}} \mu_i$ is equivalent to

$$(N_{j+1} - 1) \phi(\hat{\rho}_j) + \psi'_j \left(\sum_{i \in \mathcal{C}_{j+1}} \hat{x}_{i,j}, \hat{y}_j \right) > (N_{j+1} - 1) \phi(\rho_j) + \psi'_j \left(\sum_{i \in \mathcal{C}_{j+1}} x_{i,j}, y_j \right)$$

and since $\hat{y}_j \leq y_j$, this implies that $\sum_{i \in \mathcal{C}_{j+1}} \hat{x}_{i,j} > \sum_{i \in \mathcal{C}_{j+1}} x_{i,j}$. Since $\hat{y}_j \leq y_j$, necessarily $\sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} \hat{x}_{i,j} < \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} x_{i,j}$. However, since all classes $k \in \mathcal{C}_{min} \cup \mathcal{C}_{max}$ do not belong to $\mathcal{C}_j \setminus \mathcal{C}_{j+1}$, we know that $\hat{\lambda}_i = \lambda_i$ for all $i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}$, and thus

$$\sum_{l=1}^j \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} x_{i,l} = \sum_{l=1}^j \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} \hat{x}_{i,l},$$

from which we obtain

$$\sum_{l < j} \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} x_{i,l} = \sum_{l < j} \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} \hat{x}_{i,l} + \left(\sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} \hat{x}_{i,j} - \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} x_{i,j} \right),$$

and therefore

$$\sum_{l < j} \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} x_{i,l} < \sum_{l < j} \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} \hat{x}_{i,l}. \quad (21)$$

Subtracting (20) from (19), we obtain

$$\sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} \hat{\mu}_i < \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} \mu_i.$$

Hence, for each server $l < j$,

$$(N_j - N_{j+1}) \phi(\hat{\rho}_l) + \psi'_l \left(\sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} \hat{x}_{i,l}, \hat{y}_l \right) < (N_j - N_{j+1}) \phi(\rho_l) + \psi'_l \left(\sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} x_{i,l}, y_l \right).$$

But, for $l < j$ and $l \in \mathcal{S}^+$, it implies that

$$\sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} \hat{x}_{i,l} < \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} x_{i,l},$$

and thus

$$\sum_{l < j, l \in \mathcal{S}^+} \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} \hat{x}_{i,l} < \sum_{l < j, l \in \mathcal{S}^+} \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} x_{i,l}. \quad (22)$$

From (21), we have

$$\sum_{l < j, l \in \mathcal{S}^-} \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} \hat{x}_{i,l} > \sum_{l < j, l \in \mathcal{S}^-} \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} x_{i,l} + \left(\sum_{l < j, l \in \mathcal{S}^+} \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} x_{i,l} - \sum_{l < j, l \in \mathcal{S}^+} \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} \hat{x}_{i,l} \right),$$

and using (22) it leads to

$$\sum_{l < j, l \in \mathcal{S}^-} \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} \hat{x}_{i,l} > \sum_{l < j, l \in \mathcal{S}^-} \sum_{i \in \mathcal{C}_j \setminus \mathcal{C}_{j+1}} x_{i,l}. \quad (23)$$

According to (20), for each server $l < j$,

$$(N_{j+1} - 1) \phi(\hat{\rho}_l) + \psi'_l \left(\sum_{i \in \mathcal{C}_{j+1}} \hat{x}_{i,l}, \hat{y}_l \right) > (N_{j+1} - 1) \phi(\rho_l) + \psi'_l \left(\sum_{i \in \mathcal{C}_{j+1}} x_{i,l}, y_l \right).$$

But, for $l < j$, $l \in \mathcal{S}^-$, it implies that

$$\sum_{i \in \mathcal{C}_{j+1}} \hat{x}_{i,l} > \sum_{i \in \mathcal{C}_{j+1}} x_{i,l},$$

and thus

$$\sum_{l < j, l \in \mathcal{S}^-} \sum_{i \in \mathcal{C}_{j+1}} \hat{x}_{i,l} > \sum_{l < j, l \in \mathcal{S}^-} \sum_{i \in \mathcal{C}_{j+1}} x_{i,l} \quad (24)$$

Now, summing (24) and (23) gives

$$\sum_{l < j, l \in \mathcal{S}^-} \sum_{i \in \mathcal{C}_j} \hat{x}_{i,l} > \sum_{l < j, l \in \mathcal{S}^-} \sum_{i \in \mathcal{C}_j} x_{i,l}. \quad (25)$$

However, for each server $l \in \mathcal{S}^-$, we have $\hat{y}_l \leq y_l$ and thus $\sum_{l < j, l \in \mathcal{S}^-} \hat{y}_l \leq \sum_{l < j, l \in \mathcal{S}^-} y_l$. Since, for $l < j$, y_l can also be written as $y_l = \sum_{i \in \mathcal{C}_j} x_{i,l} + \sum_{i \notin \mathcal{C}_j} x_{i,l}$, it yields

$$\sum_{l < j, l \in \mathcal{S}^-} \sum_{i \notin \mathcal{C}_j} \hat{x}_{i,l} \leq \sum_{l < j, l \in \mathcal{S}^-} \sum_{i \notin \mathcal{C}_j} x_{i,l} + \left(\sum_{l < j, l \in \mathcal{S}^-} \sum_{i \in \mathcal{C}_j} x_{i,l} - \sum_{l < j, l \in \mathcal{S}^-} \sum_{i \in \mathcal{C}_j} \hat{x}_{i,l} \right),$$

and using (25),

$$\sum_{l < j, l \in \mathcal{S}^-} \sum_{i \notin \mathcal{C}_j} \hat{x}_{i,l} < \sum_{l < j, l \in \mathcal{S}^-} \sum_{i \notin \mathcal{C}_j} x_{i,l}. \quad (26)$$

Therefore, there exists a class $i \notin \mathcal{C}_j$ such that

$$\sum_{l < j, l \in \mathcal{S}^-} \hat{x}_{i,l} < \sum_{l < j, l \in \mathcal{S}^-} x_{i,l}. \quad (27)$$

It implies that, for this class i , we can find a server $n \in \mathcal{S}_i$ and $n \in \mathcal{S}^-$ such that $\hat{x}_{i,n} < x_{i,n}$. Since $\mathcal{C}_{max} \subsetneq \mathcal{C}_j$, we know that $i \notin \mathcal{C}_{max}$. Moreover, since $\mathcal{S}_k = \mathcal{S}_1 \subset \mathcal{S}^+$ for all $k \in \mathcal{C}_{min}$, $i \notin \mathcal{C}_{min}$. We therefore have $\lambda_i = \lambda_j$. Thus,

$$\sum_{l \in \mathcal{S}^-} \hat{x}_{i,l} + \sum_{l \in \mathcal{S}^+} \hat{x}_{i,l} = \sum_{l \in \mathcal{S}^-} x_{i,l} + \sum_{l \in \mathcal{S}^+} x_{i,l},$$

which implies

$$\sum_{l \in \mathcal{S}^+} \hat{x}_{i,l} = \sum_{l \in \mathcal{S}^+} x_{i,l} + \left(\sum_{l \in \mathcal{S}^-} x_{i,l} - \sum_{l \in \mathcal{S}^-} \hat{x}_{i,l} \right),$$

and with (27), it yields

$$\sum_{l \in \mathcal{S}^+} \hat{x}_{i,l} > \sum_{l \in \mathcal{S}^+} x_{i,l}.$$

This implies that there exists a server $m < j$, $m \in \mathcal{S}^+$ such that $\hat{x}_{i,m} > x_{i,m}$. But, according to Lemma 14, there cannot be two servers $m, n \in \mathcal{S}$ such that

$\hat{y}_m > y_m$, $\hat{y}_n \leq y_n$, $\hat{x}_{i,m} > x_{i,m}$ and $\hat{x}_{i,n} < x_{i,n}$. This is a contradiction. Therefore, if $j \in \mathcal{S}^-$, then $j+1 \in \mathcal{S}^-$ for all servers $j \in \mathcal{S}$. ■

Proposition 6 proves that the transformation induces a monotonic partition of servers: there exists a threshold $J < S$ such that for all servers $j > J$ the load is non-increasing under the transformation.

Using the above results regarding the impact of the transformation on the server loads, the following two theorems compare the costs $D(\lambda)$ and $D(\hat{\lambda})$. The first theorem uses the following lemma.

Lemma 20 *If b_i , $i = 1, 2, \dots$, is such that*

1. $b_1 > 0$,
2. $b_i \leq 0 \Rightarrow b_{i+1} \leq 0$, and
3. $\sum_i b_i = 0$,

and a_i , $i = 1, 2, \dots$, is such that

1. $a_i \geq a_{i+1}$, and
2. $a_I - a_{I+1} > 0$,

where $I = \max\{i : b_i > 0\}$, then $\sum_i a_i b_i > 0$.

Proof. We have

$$\begin{aligned}
\sum_i a_i b_i &= \sum_{i \leq I} a_i b_i + \sum_{i > I} a_i b_i \\
&\geq a_I \sum_{i \leq I} b_i + \sum_{i > I} a_i b_i \\
&\geq a_I \sum_{i \leq I} b_i - a_{I+1} \sum_{i > I} |b_i| \\
&\geq (a_I - a_{I+1}) \sum_{i \leq I} b_i \\
&> 0.
\end{aligned}$$

■

We are now in position to state our main results.

Theorem 3 $D(\lambda) < D(\hat{\lambda}) \iff \mathcal{S}_1 \subsetneq \mathcal{S}_K$.

Proof. We first show that if $\mathcal{S}_1 \subsetneq \mathcal{S}_K$ then $D(\boldsymbol{\lambda}) < D(\hat{\boldsymbol{\lambda}})$. Let $\Delta y_j = \hat{y}_j - y_j$ for all $j \in \mathcal{S}$. Since the global cost D is a convex function of the offered traffics y_j , $j \in \mathcal{S}$, we have

$$D(\hat{\boldsymbol{\lambda}}) \geq D(\boldsymbol{\lambda}) + \nabla D(\boldsymbol{\lambda}) \cdot \Delta \mathbf{y}.$$

We thus have

$$D(\hat{\boldsymbol{\lambda}}) - D(\boldsymbol{\lambda}) \geq \sum_j \frac{\partial D}{\partial y_j} \Delta y_j = \sum_j \psi'_j(y_j, y_j) \Delta y_j \quad (28)$$

We now show that the RHS in the above inequality is strictly positive. Since $\mathcal{S}_1 \subsetneq \mathcal{S}_K$, from Proposition 4 and Lemma 19, we can infer that $\mathcal{S}^+ \neq \emptyset$ and $\mathcal{S}^- \neq \emptyset$. From Proposition 4, we can also infer that $\mathcal{S}_1 \subset \mathcal{S}^+$. Hence, $\Delta y_1 > 0$. From Proposition 6, if $j \in \mathcal{S}^-$ then $j+1 \in \mathcal{S}^-$. Therefore, the sequence Δy_j , $j \in \mathcal{S}$ is such that

1. $\Delta y_1 > 0$,
2. $\Delta y_j \leq 0 \Rightarrow \Delta y_{j+1} \leq 0$, and
3. $\sum_{j \in \mathcal{S}} \Delta y_j = 0$.

Let $J = \max\{j : j \in \mathcal{S}^+\}$. Then, $J+1 = \min\{j : j \in \mathcal{S}^-\}$. Note that $\mathcal{C}_J \neq \mathcal{C}_{J+1}$, otherwise from Lemma 16, either both J and $J+1$ belong to \mathcal{S}^+ or both belong to \mathcal{S}^- . From Lemma 12, we can conclude that

1. $\psi'_j(y_j, y_j) \geq \psi'_{j+1}(y_{j+1}, y_{j+1})$, $\forall j$, and
2. $\psi'_j(y_J, y_J) \geq \psi'_{J+1}(y_{J+1}, y_{J+1})$.

Since the sequences $\psi'_j(y_j, y_j)$ and Δy_j satisfy the conditions of Lemma 20, we have

$$\sum_j \psi'_j(y_j, y_j) \Delta y_j > 0,$$

and hence, from (28), we can conclude that $D(\boldsymbol{\lambda}) < D(\hat{\boldsymbol{\lambda}})$.

To show the converse, if $D(\boldsymbol{\lambda}) < D(\hat{\boldsymbol{\lambda}})$ then necessarily there exists an m such that $y_m \neq \hat{y}_m$. From Proposition 4, we obtain $\mathcal{S}_1 \neq \mathcal{S}_K$. Since $\mathcal{S}_1 \subset \mathcal{S}_K$, we can conclude that $\mathcal{S}_1 \subsetneq \mathcal{S}_K$. ■

Theorem 3 shows that if all the classes do not use the same set of servers at the equilibrium \mathbf{x} , then the transformation will strictly increase the cost. The following theorem proves that the cost is constant under the transformation if all classes use the same set of servers.

Theorem 4 $D(\boldsymbol{\lambda}) = D(\hat{\boldsymbol{\lambda}}) \iff \mathcal{S}_1 = \mathcal{S}_K$.

Proof. From Lemma 19 and Proposition 4, if $\mathcal{S}_1 = \mathcal{S}_K$ then $y_j = \hat{y}_j, \forall j \in \mathcal{S}$ and therefore, $D(\boldsymbol{\lambda}) = D(\hat{\boldsymbol{\lambda}})$. To prove the inverse, if $\mathcal{S}_1 \neq \mathcal{S}_K$ then necessarily $\mathcal{S}_1 \subsetneq \mathcal{S}_K$. From Theorem 3, we can conclude that $D(\boldsymbol{\lambda}) \neq D(\hat{\boldsymbol{\lambda}})$. ■

4.3 Maximum Step of a Basic Transformation

Theorems 3 and 4 enable the comparison of the equilibria induced by two different rate vectors $\boldsymbol{\lambda}$ and $\hat{\boldsymbol{\lambda}}$, provided that $\hat{\boldsymbol{\lambda}}$ can be obtained from $\boldsymbol{\lambda}$ under a basic transformation which leaves unaffected the set of servers used by each class. The main limitation of these results comes from the latter assumption. However, as will be shown below, the continuity of the Nash mapping can be exploited to prove that, under certain conditions, the global cost is non-decreasing under the transformation even if some classes change the set of servers they use.

Definition 2 For each rate vector $\boldsymbol{\lambda} \in \Lambda$, the maximum step of the transformation $h_{\boldsymbol{\lambda}}$ is

$$\Delta = \min(n_{min} \Delta_{min}, n_{max} \Delta_{max}), \quad (29)$$

where $\Delta_{min} = -\lambda_1 + \min\left(\frac{\bar{\lambda}}{K}, \min_{i \in \mathcal{C} \setminus \mathcal{C}_{min}} \lambda_i\right)$ and $\Delta_{max} = \lambda_K - \max\left(\frac{\bar{\lambda}}{K}, \max_{i \in \mathcal{C} \setminus \mathcal{C}_{max}} \lambda_i\right)$.

Intuitively, if the step ϵ of a basic transformation is lower than the maximum step Δ , then the sets \mathcal{C}_{min} and \mathcal{C}_{max} will be unaffected by the transformation. On the contrary, if $\epsilon = \Delta$, then, after the transformation, we will have either (i) one more class in the set \mathcal{C}_{min} or \mathcal{C}_{max} , or (ii) $\boldsymbol{\lambda} = \boldsymbol{\lambda}^{\bar{}}$.

For each rate vector $\boldsymbol{\lambda}$, let $\boldsymbol{\lambda}(\epsilon) = h_{\boldsymbol{\lambda}}(\epsilon)$ for $\epsilon \in [0, \Delta]$. All quantities of interest can be treated as functions of ϵ . Therefore, in the following, if z is a certain quantity related to the Nash equilibrium induced by the vector $\boldsymbol{\lambda}$ then we shall denote the corresponding quantity for vector $\boldsymbol{\lambda}(\epsilon)$ by $z(\epsilon)$.

We first prove the following properties of the transformation when $\epsilon \leq \Delta$.

Lemma 21 For each $\epsilon \leq \Delta$,

1. $\lambda_i(\epsilon) \leq \min\left(\frac{\bar{\lambda}}{K}, \min_{k \in \mathcal{C} \setminus \mathcal{C}_{min}} \lambda_k\right)$ for all $i \in \mathcal{C}_{min}$, and the inequality is strict if $\epsilon < n_{min} \Delta_{min}$ whereas it holds as an equality if $\epsilon = \Delta = n_{min} \Delta_{min}$,

2. $\lambda_i(\epsilon) \geq \max\left(\frac{\bar{\lambda}}{K}, \max_{k \in \mathcal{C} \setminus \mathcal{C}_{max}} \lambda_k\right)$ for all $i \in \mathcal{C}_{max}$, and the inequality is strict if $\epsilon < n_{max} \Delta_{max}$, whereas it holds as an equality if $\epsilon = n_{max} \Delta_{max}$.

Proof. For $i \in \mathcal{C}_{min}$, we have $\lambda_i(\epsilon) = \lambda_1 + \frac{\epsilon}{n_{min}}$. Since $\epsilon \leq n_{min} \Delta_{min}$, it yields $\lambda_i(\epsilon) \leq \lambda_1 + \Delta_{min}$, i.e. $\lambda_i(\epsilon) \leq \min\left(\frac{\bar{\lambda}}{K}, \min_{k \in \mathcal{C} \setminus \mathcal{C}_{min}} \lambda_k\right)$, as claimed. Note that the inequality is strict if $\epsilon < n_{min} \Delta_{min}$ and that it holds as an equality if $\epsilon = n_{min} \Delta_{min}$. The proof of property 2 is symmetric. ■

The following two lemmata detail how the sets \mathcal{C}_{min} and \mathcal{C}_{max} evolve under the transformation.

Lemma 22 For each $\epsilon \leq \Delta$,

1. $\mathcal{C}_{min} \subseteq \mathcal{C}_{min}(\epsilon)$ and $\mathcal{C}_{max} \subseteq \mathcal{C}_{max}(\epsilon)$,
2. If $\epsilon < n_{min} \Delta_{min}$, then $\mathcal{C}_{min} = \mathcal{C}_{min}(\epsilon)$,
3. If $\epsilon < n_{max} \Delta_{max}$, then $\mathcal{C}_{max} = \mathcal{C}_{max}(\epsilon)$,
4. If $\epsilon = n_{min} \Delta_{min}$ and $\mathcal{C}_{min} \neq \mathcal{C}$, then $\mathcal{C}_{min} \subsetneq \mathcal{C}_{min}(\epsilon)$,
5. If $\epsilon = n_{max} \Delta_{max}$ and $\mathcal{C}_{max} \neq \mathcal{C}$, then $\mathcal{C}_{max} \subsetneq \mathcal{C}_{max}(\epsilon)$.

Proof. We just prove the relations between \mathcal{C}_{min} and $\mathcal{C}_{min}(\epsilon)$, since the proofs of the relations between \mathcal{C}_{max} and $\mathcal{C}_{max}(\epsilon)$ are symmetric. We first prove assertion 1. Let $i \in \mathcal{C}_{min}$. From Lemma 21.1, we have $\lambda_i(\epsilon) \leq \lambda_k$ for all $k \in \mathcal{C} \setminus \mathcal{C}_{min}$. For $k \notin \mathcal{C}_{min} \cup \mathcal{C}_{max}$, Lemma 17.1 states that $\lambda_k(\epsilon) = \lambda_k$, which implies that $\lambda_i(\epsilon) \leq \lambda_k(\epsilon)$ for all $k \notin \mathcal{C}_{min} \cup \mathcal{C}_{max}$. From Lemma 21.1, we also have $\lambda_i(\epsilon) \leq \frac{\bar{\lambda}}{K} \leq \lambda_k(\epsilon)$ for all $k \in \mathcal{C}_{max}$, where the last inequality comes from Lemma 21.2. We therefore conclude that $\lambda_i(\epsilon) \leq \lambda_k(\epsilon)$ for all $k \in \mathcal{C} \setminus \mathcal{C}_{min}$. However, from Lemma 17.2, we have $\lambda_k(\epsilon) = \lambda_i(\epsilon) = \lambda_1(\epsilon)$ for all $k \in \mathcal{C}_{min}$. We conclude that if $i \in \mathcal{C}_{min}$, then $\lambda_i(\epsilon) \leq \lambda_k(\epsilon)$ for all $k \in \mathcal{C}$, and thus $i \in \mathcal{C}_{min}(\epsilon)$. This shows that $\mathcal{C}_{min} \subset \mathcal{C}_{min}(\epsilon)$.

Let us now prove assertion 2. Assume $\epsilon < n_{min} \Delta_{min}$. Since $\mathcal{C}_{min} \subset \mathcal{C}_{min}(\epsilon)$, we just need to prove that $\mathcal{C}_{min}(\epsilon) \subset \mathcal{C}_{min}$. It is sufficient to show that if $k \notin \mathcal{C}_{min}$, then $k \notin \mathcal{C}_{min}(\epsilon)$. Let $k \in \mathcal{C} \setminus \mathcal{C}_{min}$. If $k \notin \mathcal{C}_{max}$, then, according to Lemma 21.1, $\lambda_1(\epsilon) < \lambda_k = \lambda_k(\epsilon)$, whereas if $k \in \mathcal{C}_{max}$, $\lambda_1(\epsilon) < \frac{\bar{\lambda}}{K} \leq \lambda_k(\epsilon)$, also from Lemma 21. Since $\lambda_1(\epsilon) = \min_{i \in \mathcal{C}} \lambda_i(\epsilon)$, we conclude that $k \notin \mathcal{C}_{min}(\epsilon)$, and thus that $\mathcal{C}_{min} = \mathcal{C}_{min}(\epsilon)$.

We now prove assertion 4. From Lemma 21.1, we have either $\lambda_1(\epsilon) = \frac{\bar{\lambda}}{K}$ or $\lambda_1(\epsilon) = \min_{k \in \mathcal{C} \setminus \mathcal{C}_{min}} \lambda_k$.

- If $\lambda_1(\epsilon) = \frac{\bar{\lambda}}{K}$, then it clearly implies that $\lambda_i(\epsilon) = \frac{\bar{\lambda}}{K}$ for all $i \in \mathcal{C}$. However, since $\mathcal{C}_{min} \neq \mathcal{C}$, it implies that class K belongs to $\mathcal{C}_{min}(\epsilon)$ but not to \mathcal{C}_{min} .
- If $\lambda_1(\epsilon) = \min_{k \in \mathcal{C} \setminus \mathcal{C}_{min}} \lambda_k$, we can find $j \in \mathcal{C} \setminus \mathcal{C}_{min}$ such that $\lambda_j \leq \frac{\bar{\lambda}}{K}$ and $\lambda_j = \min_{k \in \mathcal{C} \setminus \mathcal{C}_{min}} \lambda_k$. From $\lambda_j \leq \frac{\bar{\lambda}}{K}$ we deduce that $j \notin \mathcal{C}_{max}$. Therefore from Lemma 17.1 we obtain $\lambda_j(\epsilon) = \lambda_j = \lambda_1(\epsilon)$. We conclude that class j belongs to $\mathcal{C}_{min}(\epsilon)$ but not to \mathcal{C}_{min} .

Since in both cases we can find a class $i \in \mathcal{C}_{min}(\epsilon)$ such that $i \notin \mathcal{C}_{min}$, we conclude that $\mathcal{C}_{min} \subsetneq \mathcal{C}_{min}(\epsilon)$. ■

Lemma 23 *The following statements hold.*

1. If $\mathcal{C}_{min} \cup \mathcal{C}_{max} \neq \mathcal{C}$, then $\mathcal{C}_{min} \cup \mathcal{C}_{max} \subsetneq \mathcal{C}_{min}(\Delta) \cup \mathcal{C}_{max}(\Delta)$.
2. If $\mathcal{C}_{min} \cup \mathcal{C}_{max} = \mathcal{C}$, then $\lambda(\Delta) = \lambda^-$.

Proof. We first prove assertion 1. Assume that $\mathcal{C}_{min} \cup \mathcal{C}_{max} \neq \mathcal{C}$. We have either $\Delta = n_{min} \Delta_{min}$, or $\Delta = n_{max} \Delta_{max}$. According to Lemmata 22.4 and 22.5, if $\Delta = n_{min} \Delta_{min}$, then $\mathcal{C}_{min} \subsetneq \mathcal{C}_{min}(\epsilon)$, while if $\Delta = n_{max} \Delta_{max}$ we have $\mathcal{C}_{max} \subsetneq \mathcal{C}_{max}(\epsilon)$. We therefore conclude that $\mathcal{C}_{min} \cup \mathcal{C}_{max} \subsetneq \mathcal{C}_{min}(\Delta) \cup \mathcal{C}_{max}(\Delta)$. ■

The following proposition states that if we consider two rate vectors obtained from λ under basic transformations of steps lower than the maximum step, then one can be obtained from the other by a basic transformation.

Proposition 7 *Let $\epsilon_1, \epsilon_2 \in [0, \Delta]$, $\epsilon_1 < \epsilon_2$. Then $\lambda(\epsilon_2)$ can be obtained from $\lambda(\epsilon_1)$ under a basic transformation.*

Proof. Since $\epsilon_1 < \epsilon_2$ implies $\epsilon_1 < \Delta$, from Lemmata 22.2 and 22.3 we have $\mathcal{C}_{min}(\epsilon_1) = \mathcal{C}_{min}$ and $\mathcal{C}_{max}(\epsilon_1) = \mathcal{C}_{max}$. Accordingly, $\lambda(\epsilon_2)$ can be written as

$$\begin{aligned} \lambda(\epsilon_2) &= \lambda + \epsilon_1 \left(\frac{\sum_{i \in \mathcal{C}_{min}} \mathbf{e}_i}{n_{min}} - \frac{\sum_{i \in \mathcal{C}_{max}} \mathbf{e}_i}{n_{max}} \right) \\ &\quad + (\epsilon_2 - \epsilon_1) \left(\frac{\sum_{i \in \mathcal{C}_{min}(\epsilon_1)} \mathbf{e}_i}{n_{min}(\epsilon_1)} - \frac{\sum_{i \in \mathcal{C}_{max}(\epsilon_1)} \mathbf{e}_i}{n_{max}(\epsilon_1)} \right), \end{aligned}$$

i.e., $\lambda(\epsilon_2) = h_{\lambda(\epsilon_1)}(\epsilon_2 - \epsilon_1)$. ■

We now show that even if some classes change the set of servers they use, the global cost is non-decreasing under the transformation $\lambda(\epsilon) = h_{\lambda}(\epsilon)$ provided that $\epsilon \leq \Delta$. The proof is based on the following theorem which is proved in [7] (Theorem 5, page 321), and closely parallels the discussion in section III.B of the above reference.

Theorem 5 (Theorem 5 in [7]) *Let $f : X \rightarrow \mathbb{R}$, where $X \subset \mathbb{R}$ is a closed interval. Consider a family $\mathcal{A} = \{A_1, \dots, A_n\}$ of closed subsets of X , such that (i) $\cup_{i=1}^n A_i = X$, and (ii) for every $A_i \in \mathcal{A}$, we have : $x, y \in A_i$ and $x < y \Rightarrow f(x) < f(y)$. Then f is non-decreasing in X .*

The following theorem extends Theorems 3 and 4 to the case when the transformation changes the set of servers used by some classes.

Theorem 6 *For $\epsilon \leq \Delta$, $D(\lambda(\epsilon)) \geq D(\lambda)$.*

Proof. Consider a rate vector λ and the transformation $\lambda(\epsilon) = h_{\lambda}(\epsilon)$ of this rate vector for $\epsilon \in [0, \Delta]$. We want to prove that $D(\lambda(\epsilon)) \geq D(\lambda)$. Since all quantities of interest, and in particular the global cost, can be treated as functions of ϵ , it suffices to show that D is a non-decreasing function of ϵ on $[0, \Delta]$.

Let $A_{i,j} = \{\epsilon \in [0, \Delta] : G_{i,j}(\epsilon) \leq \lambda_i(\epsilon) \leq G_{i,j+1}(\epsilon)\}$, denote the set of $\epsilon \in [0, \Delta]$ for which class i sends flow to servers $\{1, \dots, j\}$ under equilibrium $\mathcal{N}(\epsilon)$. From (12), one can see that $G_{i,j}$ is a continuous function of the ρ_j , which in turn are continuous function of the equilibrium strategies of all classes. The continuity of the Nash mapping then implies that $G_{i,j}$ is a continuous function of $\epsilon \in [0, \Delta]$. Continuity of the functions $G_{i,j}(\epsilon)$ and $\lambda_i(\epsilon)$ implies that $A_{i,j}$ is a closed set.

For each $\mathbf{S} \in \mathcal{S}^K$, define

$$A_{\mathbf{S}} = \cap_{i \in \mathcal{C}} A_{i, S_i},$$

which is also a closed set. If $\epsilon_1, \epsilon_2 \in A_{\mathbf{S}}$, then each class sends its flow to the same set of servers under equilibria $\mathcal{N}(\epsilon_1)$ and $\mathcal{N}(\epsilon_2)$.

Consider a vector $\mathbf{S} \in \mathcal{S}^K$ and assume that we can find $\epsilon_1, \epsilon_2 \in A_{\mathbf{S}}$ such that $\epsilon_1 < \epsilon_2$, i.e. $A_{\mathbf{S}}$ is neither empty nor reduced to an isolated point. According to Proposition 7, the vector $\lambda(\epsilon_2)$ can be obtained from $\lambda(\epsilon_1)$ under a basic transformation. Since $\epsilon_1, \epsilon_2 \in A_{\mathbf{S}}$, this transformation satisfies Assumption 1, and according to Theorems 3 and 4 we have either $D(\epsilon_2) > D(\epsilon_1)$ or $D(\epsilon_2) = D(\epsilon_1)$. We therefore conclude that if we can find $\epsilon_1, \epsilon_2 \in A_{\mathbf{S}}$ such that $\epsilon_1 < \epsilon_2$, then $D(\epsilon_2) \geq D(\epsilon_1)$.

Since $[0, \Delta] = \cup_{\mathbf{S} \in \mathcal{S}^K} A_{\mathbf{S}}$, all conditions of Theorem 5 are fulfilled, and we can conclude that D is a non-decreasing function of ϵ on $[0, \Delta]$. ■

4.4 Maximum of the Global Cost

The purpose of this section is to prove that the global cost achieves its maximum in the symmetric case, i.e., when $\lambda = \lambda^\ominus = \left(\frac{\bar{\lambda}}{K}, \dots, \frac{\bar{\lambda}}{K}\right)$. To this end, starting from a fixed rate vector λ , we build a sequence $(\lambda^k)_{k \in \mathbb{N}}$ of rate vectors such that:

- $\lambda^0 = \lambda$, and
- λ^{k+1} is obtained from λ^k under a basic transformation of maximum step, i.e., $\lambda^{k+1} = h_{\lambda^k}(\Delta^k)$.

The following proposition shows that the sequence $(\lambda^k)_{k \in \mathbb{N}}$ converges to λ^\ominus in a finite number of steps.

Proposition 8 *The sequence $(\lambda^k)_{k \in \mathbb{N}}$ converges to λ^\ominus in at most K steps.*

Proof. Let w^k be the number of classes in $\mathcal{C}_{min}^k \cup \mathcal{C}_{max}^k$. Note that $w^0 \geq 2$. According to Lemma 23.2, if $w^k = K$, then $\lambda^{k+1} = \lambda^\ominus$. Otherwise, according to Lemma 23.1, we have $\mathcal{C}_{min}^k \cup \mathcal{C}_{max}^k \subsetneq \mathcal{C}_{min}^{k+1} \cup \mathcal{C}_{max}^{k+1}$, and thus $w^k < w^{k+1} \leq K$. This structure implies that in at most K steps we have $w^k = K$, and thus $\lambda^{k+1} = \lambda^\ominus$. ■

We now prove Theorem 1.

Proof of Theorem 1. For each $\lambda \in \Lambda$, the sequence $(\lambda^k)_{k \in \mathbb{N}}$ converges to λ^\ominus in a finite number of steps. According to Theorem 6, we have $D(\lambda^{k+1}) \geq D(\lambda^k)$. This implies that $D(\lambda^\ominus) \geq D(\lambda)$. ■

5 Analysis of the Symmetric Game

In the sequel, we will need the following assumption:

Assumption 3 $\phi(0) = 1$.

Note that the above assumption is satisfied by type-II functions. The results established below are therefore valid for all type-II functions. They also hold for all type-I functions that satisfy assumption 3.

According to Theorem 1, we have

$$PoA(K) = \sup_{\lambda, \mathbf{r}, \mathbf{c}} \frac{D_K(\lambda, \mathbf{r}, \mathbf{c})}{D_1(\bar{\lambda}, \mathbf{r}, \mathbf{c})} = \sup_{\mathbf{r}, \mathbf{c}} \frac{D_K(\lambda^=, \mathbf{r}, \mathbf{c})}{D_1(\bar{\lambda}, \mathbf{r}, \mathbf{c})}. \quad (30)$$

Therefore, in order to analyze the PoA, we can focus on the symmetric case. We first introduce some preliminary results in Section 5.1 and then analyze the symmetric game in Section 5.2 where we derive an explicit expression for the equilibrium flows.

5.1 Preliminary Results

We introduce below some results that will be used to derive the NEP of the symmetric game.

Let $\overset{\circ}{\mathcal{D}}$ be the interior of the interval \mathcal{D} upon which the function ϕ is defined. For each server $j \in \mathcal{S}$ and for $K \geq 1$, let the function $f_j(K, \rho) : \overset{\circ}{\mathcal{D}} \rightarrow (u_j, +\infty)$ be defined as follows

$$f_j(K, \rho) = u_j \left(\phi(\rho) + \frac{\rho}{K} \phi'(\rho) \right) \quad \forall \rho \in \overset{\circ}{\mathcal{D}}. \quad (31)$$

From our assumptions regarding the function ϕ , it is easy to see that $f_j(K, \rho)$ is a continuous, strictly increasing and differentiable function on $\overset{\circ}{\mathcal{D}}$. It follows that the inverse function of $f_j(K, \rho)$ exists and maps $(u_j, +\infty)$ onto $\overset{\circ}{\mathcal{D}}$. In the following, it will be denoted as $\rho_j(K, z)$, i.e., for each $z \in (u_j, +\infty)$, $\rho_j(K, z)$ is the unique value of $\rho \in \overset{\circ}{\mathcal{D}}$ such that $f_j(K, \rho) = z$.

The following lemma establish some properties of the function $\rho_j(K, z)$.

Lemma 24 *The function $\rho_j(K, z)$ has the following properties:*

- a) *For a fixed K , the function $\rho_j(K, z)$ is continuous and strictly increasing in z on $(u_j, +\infty)$.*
- b) *For a fixed K , the function $\rho_j(K, z)$ is differentiable on $(u_j, +\infty)$ and*

$$\rho_j'(K, z) = \frac{1}{f_j'(K, \rho_j(K, z))} \quad (32)$$

- c) *if $f_j(K, \rho)$ is convex, then $\rho_j(K, z)$ is concave.*

- d) For a fixed z , the function $\rho_j(K, z)$ is strictly increasing in K .
- e) $\lim_{z \rightarrow u_j^+} \rho_j(K, z) = 0$
- f) For two strictly positive integers K_1 and K_2 , and for each $z \in (u_j, u_{j+1})$, the following statement holds:

$$\exists i \leq j, \rho_i(K_1, z) < \rho_i(K_2, z) \iff \rho_s(K_1, z) < \rho_s(K_2, z), \forall s \leq j$$

- g) For type-I functions, $\lim_{z \rightarrow +\infty} \rho_j(K, z) = +\infty$, while for type-II function $\lim_{z \rightarrow +\infty} \rho_j(K, z) = 1$.

Proof. Properties (a), (b) and (c) are well known properties of the inverse function of a continuous, strictly increasing and differentiable function (with a strictly positive derivative everywhere). See e.g. [3].

To prove (d), let us consider two strictly positive integers K_1 and K_2 such that $K_1 < K_2$. For $z > u_j$, assume that $\rho_j(K_1, z) \geq \rho_j(K_2, z)$. Since $f_j(K, \rho)$ is strictly decreasing in K and strictly increasing in ρ , it yields

$$f_j(K_1, \rho_j(K_1, z)) > f_j(K_2, \rho_j(K_1, z)) \geq f_j(K_2, \rho_j(K_2, z)),$$

which is clearly a contradiction with $f_j(K_1, \rho_j(K_1, z)) = f_j(K_2, \rho_j(K_2, z)) = z$. As a consequence, $\rho_j(K_1, z) < \rho_j(K_2, z)$, which shows that for a fixed z , $\rho_j(K, z)$ is increasing in K .

Property (e) follows from $\phi(0) = 1$ and the fact that ϕ and ϕ' are continuous functions.

Let us now prove property (f). Straight part: if $\exists i \leq j$ such that $\rho_i(K_1, z) < \rho_i(K_2, z)$ then $K_1 < K_2$ because $\rho_i(K, z)$ is strictly increasing in K . Since $z \in (u_s, +\infty)$ and $\rho_s(K, z)$ is strictly increasing in K , it implies that $\rho_s(K_1, z) < \rho_s(K_2, z)$ for all $s \leq j$, as claimed. Converse part: it is true from the statement.

Finally, to prove property (g), assume that $z = f_j(K, \rho) \rightarrow \infty$. In view of (31), this is only possible if we have either $\phi(\rho) \rightarrow \infty$, or $\rho \phi'(\rho) \rightarrow \infty$, or both. Since $\phi(w) < +\infty$ and $\phi'(w) < +\infty$ for all $w \in \mathcal{D}$, this implies that $\rho_j(K, z) \rightarrow \infty$ for type-I functions, while $\rho_j(K, z) \rightarrow 1$ for type-II functions. ■

Let us now define the function $W(K, z) = \sum_{j \in \mathcal{S}} W_j(K, z)$, where

$$W_j(K, z) = \mathbb{1}_{\{z \in [u_j, u_{j+1})\}} \cdot \left(\sum_{s=1}^j r_s \rho_s(K, z) - \bar{\lambda} \right).$$

The following lemma states some properties of the function $W(K, z)$.

Lemma 25 *The function $W(K, z)$ is such that:*

1. for a fixed K , the function $W : [u_1, \infty) \rightarrow \mathbb{R}$ is continuous, strictly increasing and differentiable in z ,
2. for a fixed z , $W(K, z)$ is strictly increasing in K ,
3. for a fixed K , $W(K, z) = 0$ has a unique solution in the interval (u_1, ∞) .

Proof. Let us first prove property 1. By definition $W(k, x) = W_j(K, x)$ in the interval $[u_j, u_{j+1})$. The functions $\rho_s(K, z)$ are continuous and strictly increasing in $(u_s, +\infty)$, which implies that W is continuous and strictly increasing in (u_j, u_{j+1}) . To conclude the proof, we need to verify that W is continuous at $u_j, j = 2, 3, \dots, S$. We have

$$\begin{aligned}
\lim_{x \rightarrow u_j^+} W(K, x) - \lim_{x \rightarrow u_j^-} W(K, x) &= \lim_{x \rightarrow u_j^+} W_j(K, x) - \lim_{x \rightarrow u_j^-} W_{j-1}(K, u_j) \\
&= \left(\sum_{i=1}^{j-1} r_i \rho_i(K, u_j) + r_j \lim_{x \rightarrow u_j^+} \rho_j(K, x) - \bar{\lambda} \right) \\
&\quad - \left(\sum_{i=1}^{j-1} r_i \rho_i(K, u_j) - \bar{\lambda} \right) \\
&= r_j \lim_{x \rightarrow u_j^+} \rho_j(K, x) \\
&= 0,
\end{aligned}$$

where the last equality follows from Lemma 24.(e). This shows that the function $W(K, x)$ is also continuous at the points $u_j, j = 2, 3, \dots, S$.

To prove property 2, let us consider two strictly positive integers K_1 and K_2 such that $K_1 < K_2$. For $z \in (u_j, u_{j+1})$, we have using Lemma 24.(f)

$$\begin{aligned}
W(K_1, z) < W(K_2, z) &\iff \sum_{s=1}^j r_s \rho_s(K_1, z) - \bar{\lambda} < \sum_{s=1}^j r_s \rho_s(K_2, z) - \bar{\lambda} \\
&\iff \rho_s(K_1, z) < \rho_s(K_2, z) \\
&\iff K_1 < K_2,
\end{aligned}$$

which shows that for a fixed $z \in (u_j, u_{j+1})$, $W(K, z)$ is strictly increasing in K . Continuity of $W(K, z)$ at points $u_j, j = 2, 3, \dots, S$ implies that this is true for all $z \in [u_1, +\infty)$.

Finally, let us now prove property 3. First, we note that $W(K, u_1) = r_1 \rho_1(K, u_1) - \bar{\lambda} = -\bar{\lambda}$. From Lemma 24.(g), we have $W(K, \infty) = +\infty$ for type-I functions, while for type-II functions $W(K, \infty) = \bar{r} - \bar{\lambda}$. In both cases, $W(K, \infty)$ is strictly positive. Also according to property 1, $W(K, z)$ is continuous and increasing in the interval $[u_1, \infty)$. Hence, there is a unique value of z for which $W(K, z) = 0$. ■

In the following, we let $\gamma(K)$ be the unique solution of $W(K, z) = 0$ in $[u_1, \infty)$.

5.2 Analysis of the Symmetric Game

It is well known that in this case the non-cooperative routing game is a potential game, i.e., the equilibrium flows are the global minima of a standard convex optimization problem (see e.g. Theorem 4.1 in [5]). This is formally stated in the following proposition.

Proposition 9 *If the vector ρ is global optimum of the following convex optimization problem*

$$\begin{aligned} & \underset{\rho}{\text{minimize}} && \sum_{j \in \mathcal{S}} c_j \rho_j \phi(\rho_j) + (K-1) \int_0^{\rho_j} c_j \phi(z) dz \\ & \text{s.t.} && \sum_{j \in \mathcal{S}} r_j \rho_j = \bar{\lambda}, \\ & && 0 \leq \rho_j < 1, \forall j \in \mathcal{S}, \end{aligned} \tag{33}$$

then the multi-strategy \mathbf{x} such that $x_{i,j} = r_j \frac{\rho_j}{K}$, $\forall i \in \mathcal{C}, \forall j \in \mathcal{S}$, is the NEP of the symmetric game.

Proof. The statement follows from Theorem 4.1 in [5] with $c_a(x_a) = u_a \phi(\frac{x_a}{r_a})$. ■

Note that when $K = 1$, the above problem reduces to the global optimization problem solved by the centralized scheme. When $K \rightarrow \infty$, the equivalent problem states the common function that is optimized jointly by an infinite number of players and is characteristic of the Wardrop equilibrium.

The following proposition gives the solution of the symmetric game.

Proposition 10 *The set of servers used at the NEP of the symmetric game is $S^*(K) = \{1, \dots, j^*(K)\}$, where $j^*(K)$ is the unique server such that $W(K, u_{j^*(K)}) < 0 \leq W(K, u_{j^*(K)+1})$. Moreover, the offered traffic of server $j \in \mathcal{S}$ at the NEP is 0 if $u_j \leq \gamma(K)$ and $\rho_j(K, \gamma(K))$ otherwise.*

Proof. Let ρ be an optimal solution of the equivalent problem stated in Proposition 9. According to the KKT conditions, there exists γ such that for each $j \in \mathcal{S}$,

$$f_j(K, \rho_j) = u_j \left[\phi(\rho_j) + \frac{\rho_j}{K} \phi'(\rho_j) \right] \geq \gamma, \quad (34)$$

with equality if and only if $\rho_j > 0$.

Let us now consider a server $j \in \mathcal{S}$. Let us first assume that $u_j < \gamma$. In this case, a necessary condition for (34) to hold is $f_j(K, \rho_j) > u_j$, which implies $\rho_j > 0$. We therefore obtain that (34) holds as an equality, and thus that $\gamma = f_j(K, \rho_j)$, i.e. $\rho_j = \rho_j(K, \gamma)$.

Let us now assume on the contrary that $u_j \geq \gamma$. If $\rho_j > 0$, then $f_j(K, \rho_j) > u_j \geq \gamma$, which implies that $\rho_j = 0$, i.e. a contradiction. Therefore, if $u_j \geq \gamma$, then $\rho_j = 0$.

Let $j^*(K)$ be the unique value of j such that $u_j < \gamma \leq u_{j+1}$. Then $\sum_{j=1}^{j^*(K)} r_j \rho_j = \bar{\lambda}$ can be written as $\sum_{j=1}^{j^*(K)} r_j \rho_j(K, \gamma) - \bar{\lambda} = 0$, i.e. $W(K, \gamma) = 0$, from which we deduce that $\gamma = \gamma(K)$. From Lemma 25.1 and $u_{j^*(K)} < \gamma(K) \leq u_{j^*(K)+1}$, we then obtain that $W(K, u_{j^*(K)}) < 0 \leq W(K, u_{j^*(K)+1})$. ■

We now prove that the distributed scheme with K dispatchers uses only a subset of the servers used by the centralized scheme. The proof is based on the following proposition.

Proposition 11 *The function $\gamma(K)$ is decreasing in K .*

Proof. For $K_1 < K_2$, we have $W(K_1, z) < W(K_2, z)$ for all $z \in [u_1, +\infty)$. With $z = \gamma(K_1)$, it yields $0 = W(K_1, \gamma(K_1)) < W(K_2, \gamma(K_1))$. Since $W(K_2, z)$ is strictly increasing in z , we conclude that $\gamma(K_2) < \gamma(K_1)$. Therefore, $\gamma(K)$ is decreasing in K . ■

The fact that $\gamma(K)$ is decreasing in K implies that $j^*(K)$ is non-increasing in K . We therefore have the following important corollary.

Corollary 5 *For $K \geq 1$, $\mathcal{S}^*(K+1) \subset \mathcal{S}^*(K)$.*

As an immediate consequence, we can conclude that $\mathcal{S}^*(K) \subset \mathcal{S}^*(1)$, i.e., the distributed scheme with K dispatchers uses only a subset of the servers used by the centralized scheme.

6 Lower bound on Price of Anarchy

In this section we give lower bounds on the PoA for the functions of type (i) $\phi(x) = \frac{1}{(1-x)^m}$, and (ii) $\phi(x) = 1 + x^m$.

According to Theorem 1, we can restrict the analysis to that of the symmetric game.

In order to prove the lower bounds, we shall construct an example with two servers and symmetric dispatchers. For this symmetric game, we use its analysis given in Section 5.2 to show that the ratio D_K/D_1 attains the desired value, which will then give a lower bound on the PoA.

Proposition 12 For $\phi(x) = \frac{1}{(1-x)^m}$,

$$PoA(K) \geq \frac{K}{(K^{1/(m+1)} + mK^{1/(m+1)} - m)}. \quad (35)$$

Proof. We consider the symmetric game of section 5.2 instance in which all the classes use only server 1 in the distributed case whereas they use all the servers in the centralized case. Let $\bar{\lambda} < r_1$, and $c_j = r_j = 1$ for $j \geq 2$. For simplicity, let us consider the case of two servers. The same result holds for any number of servers as well. From the KKT conditions, if

$$\begin{aligned} \frac{c_1}{r_1} \left[\phi(\bar{\lambda}/r_1) + \frac{\bar{\lambda}}{r_1} \phi'(\bar{\lambda}/r_1) \right] &\geq \frac{c_2}{r_2}, \\ \frac{c_1}{r_1} \left[\phi(\bar{\lambda}/r_1) + \frac{\bar{\lambda}}{Kr_1} \phi'(\bar{\lambda}/r_1) \right] &\leq \frac{c_2}{r_2}, \end{aligned}$$

then in the distributed case all the classes use only the first server whereas in the centralized case all the servers are used. For $\phi(x) = \left(\frac{1}{1-x}\right)^m$, the above condition become

$$\frac{c_1}{r_1} \left[\left(\frac{r_1}{r_1 - \bar{\lambda}}\right)^m + \frac{\bar{\lambda}}{r_1} m \left(\frac{r_1}{r_1 - \bar{\lambda}}\right)^{m+1} \right] \geq \frac{c_2}{r_2} \quad (36)$$

$$\frac{c_1}{r_1} \left[\left(\frac{r_1}{r_1 - \bar{\lambda}}\right)^m + \frac{\bar{\lambda}}{Kr_1} m \left(\frac{r_1}{r_1 - \bar{\lambda}}\right)^{m+1} \right] \leq \frac{c_2}{r_2} \quad (37)$$

Let us assume that c_1 is such that the above two inequalities are satisfied, and that it is closer to the upper bound. We note that when $\bar{\lambda} \uparrow r_1$, the first term within the square brackets of (37) can be neglected with respect to the second term so that

$$\frac{c_1}{r_1} \frac{m}{K} \left(\frac{r_1}{r_1 - \bar{\lambda}}\right)^{m+1} \rightarrow 1 \quad (38)$$

In the distributed case, since all the classes use only server 1, $D_K = c_1 \frac{\bar{\lambda}}{r_1} \left(\frac{r_1}{r_1 - \bar{\lambda}} \right)^m$. As $\bar{\lambda} \uparrow r_1$, we can approximate D_K as

$$D_K \approx \frac{K}{m} (r_1 - \bar{\lambda}). \quad (39)$$

In the centralized case, in order to compute the total cost, D_1 , we need to first find the traffic on each of the servers. From the KKT conditions, the loads y_j satisfy

$$\frac{c_j}{r_j} \left[\left(\frac{r_j}{r_j - y_j} \right)^m + \frac{y_j}{r_1} m \left(\frac{r_1}{r_1 - y_j} \right)^{m+1} \right] = \gamma(1), \quad (40)$$

$$\sum_j y_j = \bar{\lambda}. \quad (41)$$

From the optimality conditions we know that $\gamma(1) > c_2/r_2 = 1$, and we also know that when $\bar{\lambda} \uparrow r_1$, $\gamma(1) \rightarrow c_2/r_2 = 1$. From (40) and (38) as $\bar{\lambda} \uparrow r_1$,

$$r_1 - y_1 \approx K^{1/(m+1)} (r_1 - \bar{\lambda}). \quad (42)$$

Since $y_1 + y_2 = \bar{\lambda}$,

$$\begin{aligned} y_2 &= \bar{\lambda} - y_1 \\ &\approx \bar{\lambda} - (r_1 - K^{1/(m+1)} (r_1 - \bar{\lambda})) \\ &= (r_1 - \bar{\lambda})(K^{1/(m+1)} - 1), \end{aligned}$$

that is, y_2 tends to zero in proportion to $(r_1 - \bar{\lambda})$.

As $\bar{\lambda} \uparrow r_1$, the centralized cost can now be computed as

$$\begin{aligned} D_1 &= c_1 \frac{y_1}{r_1} \left(\frac{r_1}{r_1 - y_1} \right)^m + c_2 \frac{y_2}{r_2} \left(\frac{r_2}{r_2 - y_2} \right)^m \\ &\approx r_1 \left(\frac{m}{K} \left(\frac{r_1}{r_1 - \bar{\lambda}} \right)^{m+1} \right)^{-1} \frac{y_1}{r_1} \left(\frac{r_1}{r_1 - y_1} \right)^m + c_2 \frac{y_2}{r_2} \\ &\approx \frac{K}{m} (r_1 - \bar{\lambda}) \left(\frac{r_1 - \bar{\lambda}}{r_1 - y_1} \right)^m + (r_1 - \bar{\lambda})(K^{1/(m+1)} - 1) \\ &\approx (r_1 - \bar{\lambda}) \left(\frac{K}{m} \frac{1}{K^{m/(m+1)}} + K^{1/(m+1)} - 1 \right) \\ &= (r_1 - \bar{\lambda}) \left(\frac{K^{1/(m+1)}}{m} + K^{1/(m+1)} - 1 \right). \end{aligned} \quad (43)$$

Using (43) and (39), we obtain the lower bound (35) on the PoA as was claimed. ■

Conjecture 1 For $\phi(x) = 1 + x^m$ and $m \geq 1$,

$$PoA(K) \geq \frac{(1 + m/K)^{-1}}{(1 + m/K)^{-1} \left(\frac{1+m/K}{1+m} \right)^{\frac{m+1}{m}} + m^{-1} \log \left(\frac{1+m}{1+m/K} \right)}. \quad (44)$$

We shall give numerical evidence to support our claim that (44) is a lower bound for the PoA. We first note that for the polynomial function, the stability condition is $\bar{\lambda} < \infty$. We now construct an example in which the number of servers is $\bar{\lambda}$ and $\bar{\lambda} \rightarrow \infty$. Following the steps in the proof of Proposition 12, we can show that

$$\frac{D_K}{D_1} = \frac{(1 + m/K)^{-1}}{(1 + m/K)^{-1} (1 - \rho_2)^{m+1} + \rho_2 (1 + \rho_2^m)}, \quad (45)$$

where ρ_2 is the solution of

$$\left(\frac{1 + m/K}{1 + m} \right)^{1/m} (1 + (1 + m)\rho_2^m)^{1/m} + \rho_2 = 1.$$

We now claim that the denominator of D_K/D_1 is upper bounded by $(1 + m/K)^{-1} \left(\frac{1+m/K}{1+m} \right)^{\frac{m+1}{m}} + m^{-1} \log \left(\frac{1+m}{1+m/K} \right)$, which then supports the conjecture.

In support of the conjecture, in figures 2, 3 and 4, we plot the exact value of D_K/D_1 , and the lower bound for various values of m and K .

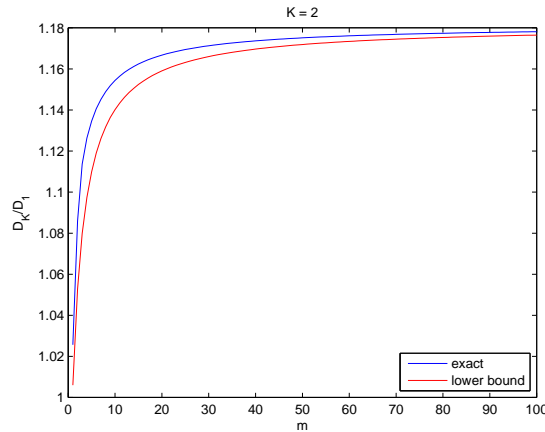


Figure 2: Comparison of D_K/D_1 and its lower bound for $m = 1, 2, \dots, 100$ and $K = 2$.

Remark 2 1. The lower bounds obtained above are independent of \mathbf{r} and \mathbf{c} .

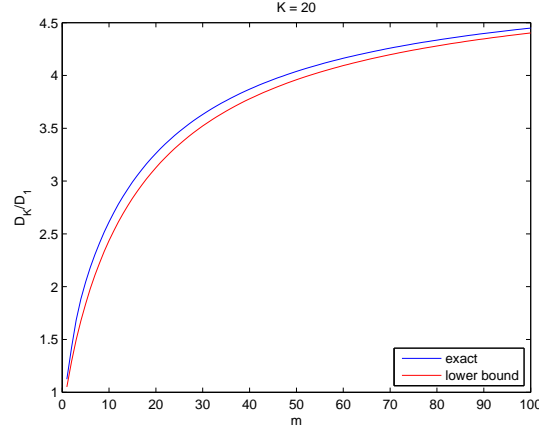


Figure 3: Comparison of D_K/D_1 and its lower bound for $m = 1, 2, \dots, 100$ and $K = 20$.

2. As $K \rightarrow \infty$, the lower bound of (44) tends to a constant, unlike the lower bound (35) which tends to infinity. Thus, for delay function of the type of $M/\text{Pareto}/1/\text{SRPT}$ queues, the PoA can be unbounded when each job minimizes its own mean delay.
3. Moreover, for $K = \infty$, the lower bound (44) is of the order of $m/\log(m)$, which matches the PoA obtained by Roughgarden [10] for polynomial functions for the Wardrop equilibrium.
4. For $m = 1$ the delay function of $M/\text{Pareto}/1/\text{SRPT}$ tends towards that of the $M/G/1/PS$. From the lower bound formula given above, we retrieve the lower bound in [2].

Although we do not have the upper bounds for PoA as in [2], we conjecture that the lower bounds constructed using the above method give the right order of the PoA, just as was proved in [2] for the case of $M/M/1$ delay functions.

7 Conclusion

We investigated the performance of non-cooperative load-balancing in processor-sharing server-farms. We have first shown that the worst global performance is obtained when all K dispatchers route exactly the same amount of traffic. This result implies that the analysis of the PoA can be done by focusing on the symmetric case, and therefore using the potential function method.

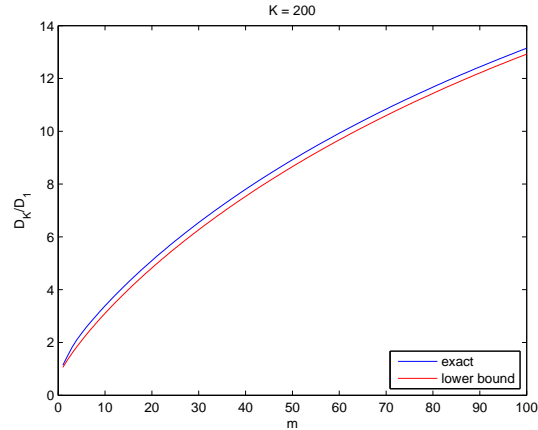


Figure 4: Comparison of D_K/D_1 and its lower bound for $m = 1, 2, \dots, 100$ and $K = 200$.

8 Acknowledgments

The authors would like to thank U. Ayesta (Basque Center for Applied Mathematics, Bilbao, Spain) for useful discussions.

References

- [1] J. Anselmi and B. Gaujal. Optimal routing in parallel, non-observable queues and the price of anarchy revisited. In *22nd International Teletraffic Congress (ITC)*, Amsterdam, 2010.
- [2] U. Ayesta, O. Brun, and B. J. Prabhu. Price of anarchy in non-cooperative load balancing games. *Performance Evaluation*, 68:1312–1332, 2011.
- [3] N. Bourbaki. *Fonctions d'une variable réelle*. Hermann, Paris, 1965.
- [4] H. L. Chen, J. Marden, and A. Wierman. The effect of local scheduling in load balancing designs. In *Proceedings of IEEE Infocom*, 2009.
- [5] R. Cominetti, J. R. Correa, and N. E. Stier-Moses. The impact of oligopolistic competition in networks. *Operations Research, Published online in Articles in Advance*, DOI: 10.1287/opre.1080.0653, June 2009.
- [6] M. Haviv and T. Roughgarden. The price of anarchy in an exponential multi-server. *Operations Research Letters*, 35:421–426, 2007.

- [7] Y. Korilis, A. Lazar, and A. Orda. Capacity allocation under noncooperative routing. *IEEE Transactions on Automatic Control*, 42(3):309–325, March 1997.
- [8] E. Koutsoupias and C. H. Papadimitriou. Worst-case equilibria. In *STACS 1999*, 1999.
- [9] A. Orda, R. Rom, and N. Shimkin. Competitive routing in multi-user communication networks. *IEEE/ACM Transactions on Networking*, 1:510–521, October 1993.
- [10] T. Roughgarden. The price of anarchy is independent of the network topology. *J. Comput. Syst. Sci.*, 67(2):341–364, 2003.