On the Inefficiency of Atomic Splittable Routing Games over Parallel Links

Olivier Brun[1,2] and Josu Doncel[3]

[1]CNRS, LAAS, Toulouse, 31400, France.
[2]University of Toulouse, INSA, LAAS Toulouse, 31400, France.
[3]Mathematics Department, University of the Basque Country, UPV/EHU, Leioa, 48940, Spain.

Contributing authors: brun@laas.fr; josu.doncel@ehu.eus;

Abstract

Several recent works on non-atomic routing games suggest that the performance degradation of selfish routing with respect to optimal routing is overall low and far from worst-case scenarios. In this work, we study the performance degradation induced by the lack of coordination in an atomic routing game over parallel links in which there are two types of links. The latency function of "cheap" links is of the form $c_1\phi(x)$, whereas the latency function of "expensive" links is of the form $c_2\phi(x)$, where $c_2 > c_1$. We obtain an explicit characterization of the optimal and equilibrium flow configurations, and establish sufficient conditions on the latency function $\phi(x)$ under which the worst traffic conditions occur when all users have the same traffic demand and the total traffic demand is such that "expensive" link are marginally used by selfish routing. We also obtain some partial results on the worst network configuration for the inefficiency of selfish routing. All in all, our results suggest that the worst-case scenario for the inefficiency of selfish routing corresponds to very specific traffic conditions and to highly asymmetric network configurations, and thus that the *Price of Anarchy* is probably an overly pessimistic performance measure for non-cooperative routing games, as advocated in the above-mentioned works.

Keywords: Parallel links, Inefficiency, Price of Anarchy

1 Introduction

1.1 Motivation

In networks based on a centralized routing scheme, a central node computes the least-cost path between source and destination nodes by using some global knowledge of the network and then distributes the resulting routes to other nodes so that they can forward user traffic. The main advantage of such an approach is that it can enforce an optimal routing policy minimizing the overall cost (e.g., the overall mean packet-delay) of all users. However, it is broadly admitted that a centralized routing scheme cannot be used in large networks due to scalability, robustness and complexity reasons. An alternative approach is to let each network user selfishly decide on which path to route its traffic demand according to its own interest. Although more robust and scalable, this decentralized scheme may lead to a loss in performance as the individual optimizations performed by many interacting self-interested users does not necessarily converge to an optimal routing policy.

Noncooperative routing games provide the natural framework to study the performance degradation in the above decentralized routing scheme. These games are mathematical models of strategic interactions between selfish, uncoordinated network users. One usually distinguishes two types of such games. Atomic routing games refer to games in which there are finitely many users. For such games, a Nash equilibrium Nash (1951) is defined as a set of routing strategies employed by network users such that no user can decrease its own routing cost by deviating from its strategy unilaterally. In contrast, non-atomic routing games refer to games in which there is a continuum of users, each one controlling a negligible amount of traffic, and an equilibrium state (which is known as a Wardrop equilibrium Wardrop (1952)) is defined as a set of routing strategies such that the traffic demand of each user is forwarded along a minimum-cost path. For both types of games, the equilibrium flow configuration does not always correspond to that of a globally optimal routing policy. In game theory terminology, such an optimal policy is usually referred to as a social optimum because it minimizes the social cost, that is the sum of all user costs.

A vast body of literature has been devoted to the study of the inefficiency of selfish routing under a variety of traffic models. The most popular measure of the inefficiency of equilibria is the Price of Anarchy (PoA) which was introduced by Koutsoupias and Papadimitriou Koutsoupias and Papadimitriou (1999) and is defined as the performance ratio between the overall cost of an optimal routing policy and that of the worst Nash/Wardrop equilibrium (that is, an equilibrium with the largest social cost). As discussed in Section 1.3, it has been shown that the PoA of some selfish routing games can be arbitrarily large. Several recent works aiming at understanding when is selfish routing bad suggest however that the PoA is an overly pessimistic measure and that non-cooperative routing achieves near-optimal performance in most realistic settings. Nevertheless, most of these works have been carried out for non-atomic routing games, which are usually much simpler to analyze thanks to the assumption of a continuum of players.

In contrast to the above mentioned works, the present paper studies the efficiency of selfish routing in atomic routing games. Owing to the complexity of the analysis, we restrict ourselves to a topology of parallel links, as introduced in the seminal paper of Orda, Rom and Shimkin Orda et al. (1993). Given a strictly increasing and convex function $\phi(x)$ and two cost parameters c_1 and c_2 such that $c_1 < c_2$, we further assume that there are two types of links: "cheap" links whose cost is a function of the flow on the link of the form $c_1\phi(x)$, and "expensive" links whose cost function is of the form $c_2\phi(x)$. We also assume a finite number of users, each one splitting its traffic demand over the parallel links so as to minimize its own routing cost, which is the sum of the costs incurred over all the links it uses. As the cost of a link is a non-decreasing function of the traffic flow on that link, the optimal strategy of a user depends on how the other users split their traffic demands. In this context, a Nash equilibrium is a flow configuration on the links in which no user can benefit from a unilateral deviation of its own routing strategy.

For the above atomic routing game, we assume that the number of links of each type as well as their cost parameters are fixed, and study the efficiency of selfish routing as a function of the traffic demands of users. Our goal is to understand under which traffic conditions the worst inefficiency of selfish routing is obtained for a fixed network configuration. As a measure of efficiency, we adopt the ratio of the social costs obtained at the Nash equilibrium and under a socially optimal solution. This ratio is at least one, when selfish routing is optimal, and is upper bounded by the PoA of the game. The worst value of this ratio (over all possible vectors of traffic demands) corresponds to the *Inefficiency* of the routing game, a concept introduced in Doncel et al. (2014) for load-balancing games. As opposed to the PoA, the *Inefficiency* depends on the network configuration. By calculating the worst possible value of the *Inefficiency* over all network configurations, one retrieves the PoA.

1.2 Contributions

The main contributions in this work are the following:

- For an arbitrary network configuration, we characterize the traffic conditions associated with the *Inefficiency*, i.e., the traffic conditions under which the ratio of social costs is maximum. Specifically, we establish sufficient conditions on the latency function $\phi(x)$ under which the worst traffic conditions occur when all users have the same traffic demand and when the total traffic in the network is such that "expensive" links are marginally used at Nash equilibrium.
- We show that these sufficient conditions are in particular satisfied by $\phi(x) = (1+x)^m$ for $m \geq 2$ and $\phi(x) = e^{\nu x}$ for $\nu > 0$. These latency functions are used throughout the paper for illustration purposes. The former belongs to the class of polynomial latency functions commonly used in transportation research to model travel times in road networks US Bureau of Public Roads (1964). The latter is reminiscent of exponential growth models used to model many physical phenomena and corresponds to situations in which the rate at which the delay over a link increases is proportional to its value, that is, $\phi'(x) = \nu\,\phi(x)$.
- We provide an explicit characterization of the optimal and equilibrium flow configurations on the links. In particular, we show that under the worst traffic conditions the ratio of the flows obtained under the decentralized and centralized schemes is

maximum for the "cheap" links, whereas it is minimum for the "expensive" links. The latter result holds true under weaker conditions on the latency function $\phi(x)$, which are satisfied in particular by $\phi(x) = 1+x^m$ and $\phi(x) = 1/(1-x)^m$ for $m \geq 2$.

- We show that the *Inefficiency* depends only on the ratio of the number of links of each type and on the ratio $\frac{c_1}{c_2}$ of their costs (but not directly on the total number of links nor on their costs). We prove that it implies that the worst value of the *Inefficiency* is obtained when there is only one "cheap" link and the rest of the links are "expensive".
- For $\phi(x) = e^{\nu x}$, we provide a lower-bound on the PoA. Besides, we conjecture from numerical experiments that the PoA is obtained when the cost parameter of the "expensive" links c_2 is infinitely larger than that of "cheap" links c_1. Assuming that this conjecture holds, we obtain an upper-bound on the PoA. This is in contrast to the situation for $\phi(x) = (1+x)^m$ for which we observe that the *Inefficiency* is not monotone as a function of the ratio $\frac{c_1}{c_2}$.

Due to the lack of space, all proofs are omitted, except that of our main result, Proposition 5, which can be found in Appendix A. Nevertheless, detailed proofs of our results can be found in a companion technical report Brun and Doncel (2023).

1.3 Related work

We first review relevant works on nonatomic routing games. The analysis of the efficiency of Wardrop equilibria has a long history which dates back to 1920 and the well-known Pigou's example which shows that the outcome of a selfish routing game can lead to a performance degradation with respect to a centrally designed outcome Pigou (1920). It was shown in 1968 by Dietrich Braess that adding resources to a transportation network can sometimes hurt performance at equilibrium, a phenomemon now known as the famous Braess's paradox Braess (1968). More recently, Roughgarden and Tardos have shown that the value of the PoA of nonatomic congestion games with affine costs is bounded above by 4/3, and that this bound is tight Roughgarden and Tardos (2002). This shows that selfish routing is always efficient for such routing games. However, it was shown in Roughgarden (2002) that the PoA for networks with latency functions that are polynomials with nonnegative coefficients and degree at most d is asymptotically $\Theta\left(\frac{d}{\log(d)}\right)$ as $d \to \infty$, indicating that selfish routing can be very inefficient in such games. Similarly, it was shown in Haviv and Roughgarden (2007) that the PoA of nonatomic load-balancing games over parallel servers corresponds to the number of servers (see also Altman et al. (2011); Bell and Stidham (1983)). Other relevant works on the PoA of nonatomic routing games are Roughgarden and Tardos (2004); Correa et al. (2008, 2004, 2007).

On the empirical side, several recent works studied the efficiency of Wardrop equilibria in real networks and observed that the performance degradation with respect to optimal routing is overall low in spite of large theoretical values of the PoA. For instance, Monnot et al. analyze data of commuting students in Singapore and conclude that routing choices are near optimal Monnot et al. (2017) (see also Youn et al. (2008) for a similar study). On a more theoretical side, the authors in Colini-Baldeschi et al. (2016, 2019, 2020) prove that Wardrop equilibria are efficient when the network is lightly or highly congested. The authors in Wu et al. (2021) extend the results of Colini-Baldeschi et al. (2020) to a more general setting and show that selfish routing is efficient when the total traffic demand gets very large. In Cominetti et al. (2021), Cominetti et al. study the efficiency of Wardrop equilibria as a function of the total traffic demand in the network. As an efficiency metric, they focus as we do on the ratio of social costs obtained under the equilibrium and optimal routing strategies. For affine link costs, they show that this ratio has a finite number of local maxima, which are achieved where the set of active links changes. In summary, all the above works suggest that the PoA is an overly pessimistic measure of the inefficiency of selfish routing and that the performance degradation is often low and far from the worst-case scenarios.

Efficiency results for atomic routing games are much scarcer, as these games are much harder to analyze. Most known results are only valid for topologies of parallel links, as introduced in Orda et al. (1993), where the existence and unicity of the Nash equilibrium are shown for a broad class of latency functions. Ayesta et al. consider in Ayesta et al. (2011) an atomic load-balancing game in which K users selfishly route their jobs to a system of S parallel M/G/1/PS servers and prove that in this case the PoA is of the order $\sqrt{K}$, independently of the number S of servers as long as $S \geq 2$. Other results on the PoA of selfish load balancing can be found in Suri et al. (2004); Anselmi and Gaujal (2010); Czumaj et al. (2002); Chen et al. (2009) (see also Ghosh and Hassin (2021) for a recent survey).

A closely related work to ours is presented in Doncel et al. (2014), where the authors consider an atomic load-balancing game with "fast" and "slow" servers, which are modeled as M/G/1/PS queues. The setting they consider is thus similar to the one considered here, but restricted to the latency function $\phi(x) = 1/(1-x)$ for the parallel links representing the servers. They study the ratio of social costs as a function of the total incoming traffic in the system and prove that this ratio attains its maximum when the "slow" servers are marginally used by the decentralized load-balancing scheme. Our work extends the results of Doncel et al. (2014). Whereas the analysis in Doncel et al. (2014) heavily relies on the properties of the function $\phi(x) = 1/(1-x)$, and in particular on the fact that $\phi'(x) = \phi(x)^2$, we establish sufficient technical conditions under which a similar result holds and show that these conditions are in particular satisfied by the latency functions $\phi(x) = e^{\nu x}$ and $\phi(x) = (1+x)^m$. We emphasize however that all our results hold for any other latency function ϕ satisfying Assumptions 1-5 stated in Section 2.3. Besides, we show in Section 4.1 that the worst traffic conditions for the ratio of social costs are precisely those that maximize the ratio of the equilibrium and optimal flows on the "cheap" links, a result which does not appear in Doncel et al. (2014) and hold under weaker conditions satisfied as well by other latency functions such as $\phi(x) = 1/(1-x)^m$ or $\phi(x) = 1 + x^m$. Our work also complements the work in Cominetti et al. (2021) which also studies the ratio of social costs as a function of the total traffic demand in general network topologies, but for a nonatomic routing game and affine cost functions on the links, whereas we consider an atomic routing game over parallel links and non-linear cost functions on the links.

As observed for non-atomic routing games, our work suggests that the PoA is probably an overly pessimistic performance measure for non-cooperative routing games.

Interestingly, some works have reached similar conclusions for other types of games. For instance, in Feldman et al. (2016) the authors focus on large auction games (that is, auction games with many players, the influence of each one on the game outcome being small), and show that the inefficiency in such games is much smaller than what the worst-case bound suggests. Similarly, the authors of Hamers et al. (2015) study the PoA of scheduling games and observe by simulations that on average the PoA is relatively small with respect to the worst-case tight upperbound they obtain. For an analysis of the relevance of the PoA as a performance measure in games with multiple equilibria, see Seaton and Brown (2023) and Balcan et al. (2013).

1.4 Organization of the paper

We present the mathematical model of the atomic routing game considered in this paper in Section 2. In Section 3, we establish some preliminary results and obtain the characterization of the equilibrium and optimal flow configurations as solution of convex optimization problems. Our results on the worst-case total traffic are established in Section 4. In Section 5, we investigate the worst network configuration for the inefficiency of the decentralized routing scheme, that is, we study how the PoA is obtained from the *Inefficiency*. In Section 6, we discuss some possible extensions of this work. Finally, the conclusions of this work are drawn in Section 7.

2 Problem formulation

2.1 Non-cooperative routing game

As illustrated in Figure 1, we consider an atomic splittable routing game in which K users have to send their traffic demands from a source node to a destination node. We let $\mathcal{C} = \{1, \ldots, K\}$ be the set of users and denote by $\lambda_u > 0$ the traffic demand of user $u \in \mathcal{C}$. We also denote by $\bar{\lambda} = \sum_{u \in \mathcal{C}} \lambda_u$ the total traffic in the system. We consider a *decentralized routing scheme* in which each user freely decides how to split its traffic demand over the N parallel links joining the source node and the destination node. We shall denote by $x_{u,j}$ the quantity of traffic sent by user u on link $j = 1, 2, \ldots, N$.

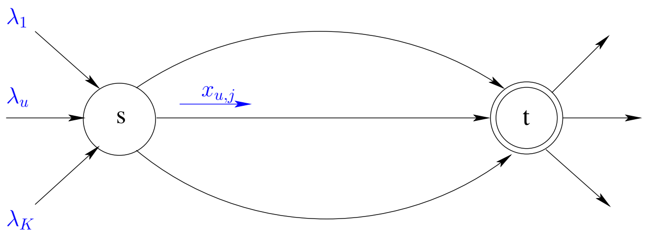


Fig. 1: K users route their traffic demands over N parallel links.

The routing strategy of user u is represented by the vector $\mathbf{x}_u = (x_{u,j})_{j=1,\ldots,N}$. We shall denote by $\mathcal{X}_u$ the set of feasible routing strategies for this user, that is, the set of vectors $\mathbf{x}_u \geq \mathbf{0}$ such that $\sum_j x_{u,j} = \lambda_u$. A strategy profile is then a vector $\mathbf{x} = (\mathbf{x}_u)_{u \in \mathcal{C}}$ belonging to the product strategy space $\mathcal{X} = \bigotimes_{u \in \mathcal{C}} \mathcal{X}_u$. It describes the routing strategy used by each player and represents in a way the state of the game. Given a state $\mathbf{x} \in \mathcal{X}$ of the game, we denote by $\mathbf{x}_{-u}$ the vector $(\mathbf{x}_1, \ldots, \mathbf{x}_{u-1}, \mathbf{x}_{u+1}, \ldots, \mathbf{x}_K)$ which describes the routing strategies of all other users than u.

We assume that the links are of two types. There are $n_1 \geq 1$ type-1 links and $n_2 = N - n_1 \geq 1$ type-2 links. In the following, we let $\mathcal{S}_1 = \{1, \ldots, n_1\}$ be the set of type-1 links and $\mathcal{S}_2 = \{n_1 + 1, \ldots, N\}$ be the set of type-2 links. A crucial assumption in our model is that the links have traffic-dependent cost functions. More precisely, it is assumed that the cost per unit flow on link $j \in \mathcal{S}_k$ is of the form $c_k \phi(y_j)$, where $y_j = \sum_{u \in \mathcal{C}} x_{u,j}$ represents the total traffic on the link, c_k is a cost parameter which depends on the type of the link, and ϕ is a given cost function. In this work, we shall assume that $c_1 < c_2$ and refer to type-1 links (resp. type-2 links) as "cheap" links (resp. "expensive" links).

In state $\mathbf{x} \in \mathcal{X}$ of the game, the routing cost $T_u(\mathbf{x})$ of user u is defined as the sum of the costs it incurs on all links, that is

$$T_u(\mathbf{x}) = \sum_{k=1}^{2} c_k \sum_{j \in \mathcal{S}_k} x_{u,j} \, \phi(y_j) \, .$$

Note that the cost incurred by the routing agent u on a link j depends both on the amount of flow $x_{u,j}$ that it routes to that link, but also on the total traffic $\sum_{i \neq u} x_{i,j}$ sent by the other users on that link.

Another key assumption is that users are self-interested agents seeking to minimize their own routing cost. More precisely, given the strategies $\mathbf{x}_{-u}$ of the other users, user u chooses its routing strategy $\mathbf{x}_u^*$ so as to solve the following optimization problem

$$\underset{\mathbf{x}_u^* \in \mathcal{X}_u}{\text{minimize}} \; T_u(\mathbf{x}_u^*, \mathbf{x}_{-u})$$

and the optimal strategy $\mathbf{x}_u^*$ is known as the best-response of user u. The best-response of user u depends on the routing strategies of the other users, which gives a non-cooperative routing game between the users. We say that the game is symmetric if all users control the same amount of traffic, that is, $\lambda_u = \bar{\lambda}/K$ for all $u \in \mathcal{C}$. Otherwise, the game is asymmetric.

A Nash equilibrium (NE) of the game is a stable state $\mathbf{x}^{ne} \in \mathcal{X}$ from which no user finds it beneficial to deviate unilaterally, that is

$$\mathbf{x}_u^{ne} \in \arg\min_{\mathbf{z} \in \mathcal{X}_u} \, T_u\left(\mathbf{z}, \mathbf{x}_{-u}^{ne}\right), \quad \forall u \in \mathcal{C}.$$

Throughout the paper, we shall only consider latency functions ϕ satisfying Assumptions 1-5 stated in Section 2.3. These assumptions are in particular satisfied by the functions $\phi(x) = e^{\nu x}$ for $\nu > 0$ and $\phi(x) = (1+x)^m$ for $m \geq 2$, which are used for illustration purposes in the present paper. Under Assumption 1, the existence and uniqueness of the NE follow from Orda et al. (1993).

2.2 Inefficiency of the decentralized routing scheme

It is well-known in game theory that the outcome of a non-cooperative game between selfish players can lead to an overall performance degradation with respect to a centrally designed outcome (see, e.g. Chapter 17 of Nisan et al. (2007)). In our setting, it means that letting users optimize their own performances without any coordination may lead to a performance degradation with respect to a centralized routing scheme which would optimize the global performance of all users.

We adopt the point of view of a network operator who has a known *network configuration*. For the model introduced above, a network configuration is defined as a vector of parameters $\mathbf{p} = (\mathbf{n}, \mathbf{c})$, where $\mathbf{n} = (n_1, n_2)$ specifies the numbers of "cheap" and "expensive" links, whereas $\mathbf{c} = (c_1, c_2)$ gives the costs of the two types of links. The network operator does not know the traffic demands $\lambda_1, \ldots, \lambda_K$ of the users and does not control how they route their traffic demands into the network. We assume a decentralized scheme in which each user minimizes its own routing cost, without coordination with the others, as explained above. The objective is to evaluate, for the fixed network configuration $\mathbf{p}$, the performance degradation resulting from the absence of a central authority under the worst-case traffic conditions.

In order to make things more formal, we introduce below some definitions. Consider a network configuration $\mathbf{p}$ and a vector $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_K)$ of traffic demands. Let $\mathbf{x}^{ne}$ be the corresponding NE of the routing game and define $\mathbf{y}^{ne} = (y_1^{ne}, y_2^{ne}, \ldots, y_N^{ne})$, where $y_j^{ne} = \sum_{u \in \mathcal{C}} x_{u,j}^{ne}$ is the total flow on link j at the NE. The global performance (or social cost) of the *decentralized routing scheme* with K players is defined as the sum of the individual player' costs at the NE

$$D_K(\boldsymbol{\lambda}, \mathbf{p}) = \sum_{u \in \mathcal{C}} T_u(\mathbf{x}^{ne}).$$

Introducing $F_k(y) = c_k \, y \, \phi(y)$ for $k = 1, 2$, the social cost can be written as follows

$$\begin{aligned} D_K(\boldsymbol{\lambda}, \mathbf{p}) &= c_1 \sum_{j \in \mathcal{S}_1} \sum_{u \in \mathcal{C}} x_{u,j}^{ne} \, \phi\left(\sum_{u \in \mathcal{C}} x_{u,j}^{ne}\right) + c_2 \sum_{j \in \mathcal{S}_2} \sum_{u \in \mathcal{C}} x_{u,j}^{ne} \, \phi\left(\sum_{u \in \mathcal{C}} x_{u,j}^{ne}\right) \\ &= \sum_{j \in \mathcal{S}_1} F_1(y_j^{ne}) + \sum_{j \in \mathcal{S}_2} F_2(y_j^{ne}) \\ &= F(\mathbf{y}^{ne}), \end{aligned}$$

where $F(\mathbf{y}) = \sum_{j \in \mathcal{S}_1} F_1(y_j) + \sum_{j \in \mathcal{S}_2} F_2(y_j)$ for all $\mathbf{y} \geq \mathbf{0}$. In this paper, our objective is to compare the performance of the decentralized scheme with the optimal performance that could be achieved

$$F(\mathbf{y}^*) = \min \left\{ F(\mathbf{y}) \ : \ \mathbf{y} \geq \mathbf{0}, \sum_j y_j = \bar{\lambda} \right\}, \tag{1}$$

where $\bar{\lambda} = \sum_{u \in \mathcal{C}} \lambda_u$ is the total traffic in the system. Note that $F(\mathbf{y}^*)$ is the performance achieved by an optimal routing strategy minimizing the social cost, and that it corresponds to the cost obtained when there is a single user controlling all the traffic in the system, i.e. such that $\lambda_1 = \bar{\lambda}$. The global cost of this *centralized routing scheme* can be written as $D_1(\bar{\lambda}, \mathbf{p})$, and we thus have $F(\mathbf{y}^*) = D_1(\bar{\lambda}, \mathbf{p})$.

In order to compare the two routing schemes, we shall use the concept of *inefficiency* introduced in Doncel et al. (2014) in the context of server farms. The *inefficiency* of the decentralized scheme is defined as the ratio between the performance obtained by the NE and the global optimal solution under the worst possible traffic conditions, that is

$$\textit{Inefficiency } I_K(\mathbf{p}) = \sup_{\boldsymbol{\lambda}} \frac{D_K(\boldsymbol{\lambda}, \mathbf{p})}{D_1(\bar{\lambda}, \mathbf{p})} = \sup_{\boldsymbol{\lambda}} \frac{F(\mathbf{y}^{ne})}{F(\mathbf{y}^*)}, \tag{2}$$

where the supremum is taken over all vectors $\boldsymbol{\lambda} \geq 0$ of traffic demands that the network may have to accommodate such that $\sum_{u=1}^{K} \lambda_u = \bar{\lambda}$. We emphasize that the inefficiency depends on the network configuration $\mathbf{p}$ but not on the traffic demands of the users. Its values are between 1 and ∞, a higher value indicating a worse performance of decentralized routing compared to centralized routing.

Another widely used measure of how the efficiency of a system degrades due to selfish behavior of its users is the so-called *Price of Anarchy* (PoA) Koutsoupias and Papadimitriou (1999), which, in our case, is related to the *inefficiency* as follows

$$PoA(K, N) = \sup_{\mathbf{p}} I_K(\mathbf{p}), \tag{3}$$

where the supremum is taken over all parameter vectors $\mathbf{p} = (\mathbf{n}, \mathbf{c})$ such that $n_1 + n_2 = N$ and $0 < c_1 < c_2$.

2.3 Assumptions on latency functions

We regroup in this section all the technical assumptions on the latency function ϕ that are required for our results to hold, and systematically indicate in which proof an assumption is used. We emphasize that, with the exceptions of Lemmas 3 and 4, all the results presented in the present paper hold for any latency function ϕ satisfying Assumptions 1-5 below. Some of our results hold under weaker assumptions and this will be indicated in the text.

Our basic assumption on the latency function ϕ is formally stated in Assumption 1 below.

Assumption 1. *The latency function ϕ is a continuously differentiable, strictly increasing and convex function over the interval $[0, +\infty)$ which verifies that $\phi(0) = 1$ and $\lim_{x \to +\infty} \phi(x) = +\infty$. In addition, its second derivative $\phi''()$ exists at all points in the interval $[0, +\infty)$.*

As discussed in Brun and Prabhu (2016), any latency function ϕ for which Assumption 1 is satisfied is such that the conditions given in Orda et al. (1993) for the existence of a unique NE for the routing game are satisfied. Our second assumption is as follows.

Assumption 2. *The function $g : [0, +\infty) \to [0, +\infty)$ defined by*

$$g(y) = \frac{y\,\phi(y)}{\phi(y) + y\phi'(y)} \tag{4}$$

is strictly increasing and strictly concave on $[0, +\infty)$.

The third assumption that we shall require is stated below.

Assumption 3. *The function* $f : [0, +\infty) \to [0, +\infty)$ *defined by*

$$f(y) = \frac{2\phi'(y) + y\phi''(y)}{\phi(y) + y\phi'(y)}\, y \tag{5}$$

is strictly increasing over $[0, +\infty)$.

We shall also need Assumption 4 below.

Assumption 4. *Let*

$$H(x, y) = x\phi'(x) - \phi(x)\left(1 + y\frac{\phi''(y)}{\phi'(y)}\right), \tag{6}$$

for $x, y \in [0, +\infty)$. *Then, it is assumed that*

- *the function* $A(x, y) = [c_1 H(x, y) - c_2 H(y, y)]\,/\phi'(x)$ *is decreasing in* x *over* $\left[y, \phi^{-1}(\frac{c_2}{c_1}\phi(y))\right)$ *for* y *fixed, and increasing in* y *over* $\left[\phi^{-1}(\frac{c_1}{c_2}\phi(x)), x\right)$ *for* x *fixed, and at least one of the monotonicities is strict;*
- *the function* $B(x, y) = [c_1 H(x, x) - c_2 H(y, x)]\,/\phi'(y)$ *is increasing in* x *over* $\left[y, \phi^{-1}(\frac{c_2}{c_1}\phi(y))\right)$ *for* y *fixed, and decreasing in* y *over* $\left[\phi^{-1}(\frac{c_1}{c_2}\phi(x)), x\right)$ *for* x *fixed, and at least one of the monotonicities is strict.*

Finally, in order to prove Proposition 5, we will need one last assumption. Before introducing this assumption, we define some additional notations. Given a fixed value of $\bar{\lambda} \geq \bar{\lambda}^{ne}$, let y_1^* (resp. y_N^*) be the flow on an arbitrary "cheap" (resp."expensive") link under the centralized routing strategy. Let us define the vector-valued function $\mathbf{y}(\Delta) = \left(y_1^* + \Delta, y_N^* - \frac{n_1}{n_2}\Delta\right)$ for $\Delta \in \left[0, \frac{n_2}{n_1} y_N^*\right]$, and

$$Q(\mathbf{y}) = \frac{n_1 F_1(y_1) + n_2 F_2(y_N)}{\delta F_1'(y_1) + (1 - \delta) F_2'(y_N)}, \tag{7}$$

where $\delta = n_1 \frac{dy_1^*}{d\bar{\lambda}}$. With a slight abuse of notation, we write $Q(\Delta)$ for $Q(\mathbf{y}(\Delta))$. Note that $\mathbf{y}(0) = \mathbf{y}^*$ and that there exists $\Delta^{ne} \in \left(0, \frac{n_2}{n_1} y_N^*\right)$ such that $\mathbf{y}(\Delta^{ne}) = \mathbf{y}^{ne}$. Our last assumption is then as follows.

Assumption 5. *The latency function* ϕ *is such that* $Q(\Delta) > Q(0)$ *for all* $\Delta \in \left(0, \frac{n_2}{n_1} y_N^*\right]$.

Table 1 summarizes which assumption is required for which proof and shows whether or not an assumption is satisfied by one of the latency functions considered in the present paper. Note that all assumptions hold in particular for the latency functions $\phi(x) = e^{\nu x}$ and $\phi(x) = (1 + x)^m$ (for proofs, see Appendix A of Brun and Doncel (2023)), which are used for illustration purposes in the present paper.

Assumption	Required for	$\phi(x)$			
		$e^{\nu x}$	$(1+x)^m$	$(1-x)^{-m}$	$1+x^m$
Assumption 1	Existence of a unique NE and Theorem 1	✓	✓	✓	✓
Assumption 2	Assertion (b) of Proposition 5	✓	✓	✓	✗
Assumption 3	Proposition 4	✓	✓	✓	✓
Assumption 4	Proposition 3	✓	✓	✓	✓
Assumption 5	Proposition 5	✓	✓	✗	✓

Table 1: This table shows which assumptions are satisfied or not by some classic latency functions and indicates where they are used.

3 Preliminary results

We first recall in Section 3.1 some results regarding the worst-case traffic conditions for the performance of the decentralized routing scheme. We then provide simple relations for the characterization of the centralized and decentralized routing solutions in Sections 3.2 and 3.3, respectively.

3.1 Worst traffic conditions for a fixed total traffic

If the total incoming traffic $\bar{\lambda}$ is fixed, it is proven in Brun and Prabhu (2016) that the global cost $D_K(\boldsymbol{\lambda}, \mathbf{p})$ achieves its maximum for the symmetric game, that is

$$\sup_{\boldsymbol{\lambda}} D_K(\boldsymbol{\lambda}, \mathbf{p}) = D_K\left(\frac{\bar{\lambda}}{K}\mathbf{1}, \mathbf{p}\right), \tag{8}$$

where $\mathbf{1} = (1, 1, \ldots, 1)$, implying that

$$I_K(\mathbf{p}) = \sup_{\bar{\lambda}>0} \frac{D_K\left(\frac{\bar{\lambda}}{K}\mathbf{1}, \mathbf{p}\right)}{D_1(\bar{\lambda}, \mathbf{p})}. \tag{9}$$

As a consequence, for the calculation of the *inefficiency*, 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 NE routing solution. More precisely, it is shown in Brun and Prabhu (2016) (see also Theorem 4.1 in Cominetti et al. (2009)) that the decentralized routing solution is the solution of a convex optimization problem, as stated in Theorem 1 below.

Theorem 1. *Let the vector* $\mathbf{y}$ *be the global optimum of the following convex optimization problem*

$$\begin{aligned} &\underset{\mathbf{y}}{\textit{minimize}} \sum_{k=1}^{2}\sum_{j\in\mathcal{S}_k} F_k(y_j) + (K-1)\int_0^{y_j} c_k\,\phi(z)dz \\ &\textit{s.t.} \\ &\qquad \textstyle\sum_{j=1}^{N} y_j = \bar{\lambda}, \\ &\qquad y_j \geq 0,\ j = 1, \ldots, N. \end{aligned} \tag{P}$$

The strategy profile $\mathbf{x}$ *such that* $x_{u,j} = \frac{y_j}{K}$ *for all* $u \in \mathcal{C}$ *and* $j = 1, \ldots, N$ *is the NE of the symmetric game.*

In words, Theorem 1 states that in a symmetric game, although each user u updates selfishly its routing strategy $\mathbf{x}_u$ according to its own objective function $T_u(\mathbf{x}_u, \mathbf{x}_{-u})$, users collectively solve the convex optimization problem (P). The objective function of this problem is a potential function for the symmetric game, so that each best-response of a player implies a decrease in the objective function value of problem (P). The equilibrium flows on the links at the NE correspond to the optimal solution $\mathbf{y}$ of this problem. As all players control the same amount of traffic in a symmetric game, the NE routing strategies are such that $x_{u,j} = y_j/K$ for all $u \in \mathcal{C}$ and $j = 1, \ldots, N$.

3.2 Characterization of the centralized routing strategy

Under the centralized routing strategy, the vector $\mathbf{y}^*$ of link flows is defined as the optimal solution of problem (1). For a type-k link $j \in \mathcal{S}_k$, the KKT conditions then imply that $y_j^* > 0$ if and only if the marginal cost

$$F_k'(y_j^*) = c_k \left[\phi\left(y_j^*\right) + y_j^*\, \phi'\left(y_j^*\right)\right], \tag{10}$$

is minimal. Defining $\bar{\lambda}^*$ as the unique solution of $F_1'(\frac{\bar{\lambda}^*}{n_1}) = F_2'(0)$, that is,

$$c_1 \left[\phi\left(\frac{\bar{\lambda}^*}{n_1}\right) + \frac{\bar{\lambda}^*}{n_1}\, \phi'\left(\frac{\bar{\lambda}^*}{n_1}\right)\right] = c_2, \tag{11}$$

it follows from the assumption $c_1 < c_2$ that

- for $\bar{\lambda} \leq \bar{\lambda}^*$, the centralized routing strategy forwards all the traffic to the cheap links and they all receive the same amount of traffic, that is, $y_k^* = \frac{\bar{\lambda}}{n_1}$ for all $k \in \mathcal{S}_1$ and $y_j^* = 0$ for all $j \in \mathcal{S}_2$,
- for $\bar{\lambda} > \bar{\lambda}^*$, the centralized routing strategy is such that all links receive a positive amount of traffic (that is, $y_l^* > 0$ for all $l = 1, \ldots, N$), and

$$c_1 \left[\phi\left(y_k^*\right) + y_k^*\, \phi'\left(y_k^*\right)\right] = c_2 \left[\phi\left(y_j^*\right) + y_j^*\, \phi'\left(y_j^*\right)\right], \forall k \in \mathcal{S}_1,\ \forall j \in \mathcal{S}_2. \tag{12}$$

Note that in both cases two links of the same type receive exactly the same amount of traffic, that is, $y_l^* = y_m^*$ if $l, m \in \mathcal{S}_k$ for $k = 1, 2$. This implies that for $\bar{\lambda} \leq \bar{\lambda}^*$ the optimal social cost is simply $F(\mathbf{y}^*) = n_1 F_1\left(\frac{\bar{\lambda}}{n_1}\right)$, whereas for $\bar{\lambda} > \bar{\lambda}^*$, it can be written as $F(\mathbf{y}^*) = n_1 F_1\left(y_1^*\right) + n_2 F_2\left(y_N^*\right)$.

3.3 Characterization of the decentralized routing strategy

Under the decentralized routing strategy, the vector $\mathbf{y}^{ne}$ of link flows is the optimal solution of the optimization problem stated in Theorem 1. Defining $\bar{\lambda}^{ne}$ as the unique solution of

$$c_1 \left[K\phi\left(\frac{\bar{\lambda}^{ne}}{n_1}\right) + \frac{\bar{\lambda}^{ne}}{n_1}\, \phi'\left(\frac{\bar{\lambda}^{ne}}{n_1}\right)\right] = c_2 K, \tag{13}$$

it follows from the KKT conditions and the assumption $c_1 < c_2$ that

- for $\bar{\lambda} \leq \bar{\lambda}^{ne}$, the decentralized routing strategy forwards all the traffic to the cheap links only and they all receive the same amount of traffic, that is, $y_k^{ne} = \frac{\bar{\lambda}}{n_1}$ for all $k \in \mathcal{S}_1$ and $y_j^{ne} = 0$ for all $j \in \mathcal{S}_2$,
- for $\bar{\lambda} > \bar{\lambda}^{ne}$, in the decentralized routing strategy all links receive a positive amount of traffic (that is $y_l^{ne} > 0$ for all l), and

$$c_1 \left[K\phi\left(y_k^{ne}\right) + y_k^{ne}\,\phi'\left(y_k^{ne}\right)\right] = c_2\left[K\phi\left(y_j^{ne}\right) + y_j^{ne}\,\phi'\left(y_j^{ne}\right)\right], \forall k \in \mathcal{S}_1,\ \forall j \in \mathcal{S}_2. \quad (14)$$

Note that, as in the centralized setting, links of the same type always receive the same amount of traffic. As a direct consequence, the social cost at the NE is $F(\mathbf{y}^{ne}) = n_1 F_1\left(\frac{\bar{\lambda}}{n_1}\right)$ for $\bar{\lambda} \leq \bar{\lambda}^{ne}$, whereas it can be written as $F(\mathbf{y}^{ne}) = n_1 F_1\left(y_1^{ne}\right) + n_2 F_2\left(y_N^{ne}\right)$ for $\bar{\lambda} > \bar{\lambda}^{ne}$.

It directly follows from (11) and (13) that $\bar{\lambda}^* < \bar{\lambda}^{ne}$ for $K > 1$. It means that when there are more than one user, the decentralized routing strategy uses only the "cheap" links longer than what would be optimal. Of course, when there is only one user, that is for $K = 1$, we have $\bar{\lambda}^{ne} = \bar{\lambda}^*$ and conditions (14) and (12) are equivalent in this case. In other words, the centralized routing strategy and the decentralized one coincide when there is only one user.

4 Worst-case total traffic

As discussed in Section 3.1, for a fixed total traffic $\bar{\lambda} > 0$, the worst inefficiency is obtained when all users control the same amount of traffic $\frac{\bar{\lambda}}{K}$, that is, for the symmetric game. We now study the worst-case total traffic $\bar{\lambda}$ for the ratio of social costs $D_K\left(\frac{\bar{\lambda}}{K}\mathbf{1}, \mathbf{p}\right)/D_1(\bar{\lambda}, \mathbf{p})$. As $D_K\left(\frac{\bar{\lambda}}{K}\mathbf{1}, \mathbf{p}\right) = F(\mathbf{y}^{ne})$ and $D_1(\bar{\lambda}, \mathbf{p}) = F(\mathbf{y}^*)$, we first establish some results pertaining to the comparison of equilibrium and optimal flow configurations in Section 4.1. In particular, we show that for any "cheap" link k the ratio y_k^{ne}/y_k^* reaches its maximum for $\bar{\lambda} = \bar{\lambda}^{ne}$. We then study in Section 4.2 the ratio of social costs as a function of $\bar{\lambda}$ and prove that it also achieves its maximum for $\bar{\lambda} = \bar{\lambda}^{ne}$.

4.1 Link flows under the centralized and decentralized routing strategies

Note that for $\bar{\lambda} > \bar{\lambda}^*$ (resp. $\bar{\lambda} > \bar{\lambda}^{ne}$) the link flows y_l^* (resp. y_l^{ne}) obtained under the centralized (resp. decentralized) routing scheme are implicitly defined by equation (12) (resp. (14)). We first prove in Lemma 1 below that under both routing schemes these link flows are continuous functions of $\bar{\lambda}$.

Lemma 1. *The vectors $\mathbf{y}^*$ and $\mathbf{y}^{ne}$ are continuous in $\bar{\lambda}$ over $[0, \infty)$.*

Proposition 2 below proves some inequalities satisfied by the flows on cheap and expensive links, which are valid under both strategies. It is worthwhile noticing that the proof of this proposition exploits only the strict monotonicity and the convexity of the latency function ϕ.

Proposition 2. *For $K \geq 1$, it holds that* *(a) The flow on the expensive links is strictly lower than that on the cheap links, that is, $y_j^{ne} < y_k^{ne}$ for all $j \in \mathcal{S}_2$ and $k \in \mathcal{S}_1$,*

(b) For any cheap link $k \in \mathcal{S}_1$ and any expensive link $j \in \mathcal{S}_2$, it holds that $c_1\phi\left(y_k^{ne}\right) < c_2\phi\left(y_j^{ne}\right)$ and $c_1 y_k^{ne}\phi'\left(y_k^{ne}\right) > c_2 y_j^{ne}\phi'\left(y_j^{ne}\right)$.

(c) For $\bar{\lambda} \geq \bar{\lambda}^{ne}$, the marginal cost of the cheap links is greater than or equal to that of the expensive ones, that is

$$F_1'(y_k^{ne}) \geq F_2'(y_j^{ne}), \quad \text{for all } k \in \mathcal{S}_1 \text{ and } j \in \mathcal{S}_2, \tag{15}$$

and the inequality is strict for $K > 1$.

We emphasize that the properties stated in Proposition 2 also hold for $K = 1$, that is, for the centralized routing strategy. Hence, the centralized routing strategy forwards more traffic on the cheap links than on the expensive links, exactly as does the decentralized one. However, for $\bar{\lambda} \geq \bar{\lambda}^{ne}$, the marginal costs of cheap and expensive links are always equal under the centralized routing strategy, whereas the marginal costs of cheap links are strictly greater than those of expensive links under the decentralized routing strategy.

We now turn our attention to the comparison of the link flows obtained under both routing strategies when $\bar{\lambda}$ varies. When $\bar{\lambda}$ is in the interval $(0, \bar{\lambda}^*]$, both strategies coincide: they both forward all the traffic only on the cheap links. In the interval $(\bar{\lambda}^*, \bar{\lambda}^{ne}]$, the centralized strategy deviates a fraction of the traffic onto the expensive links, whereas the decentralized one keeps using only the cheap links. Finally, in the interval $(\bar{\lambda}^{ne}, \infty)$, both the centralized and decentralized strategies use both type of links. Proposition 3 below states our main result regarding the comparison of the equilibrium and optimal link flows when $\bar{\lambda}$ is in the latter interval. Interestingly, we note that Proposition 3 as well as all other results in the present section are valid for any latency function satisfying Assumptions 1, 3 and 4 in Section 2.3, including of course $\phi(x) = e^{\nu x}$ and $\phi(x) = (1+x)^m$, but also other functions such as $\phi(x) = 1/(1-x)^m$ and $\phi(x) = 1 + x^m$ (see Table 1).

Proposition 3. *For $\bar{\lambda} > \bar{\lambda}^{ne}$, it holds that*

(a) The decentralized routing strategy forwards more (resp. less) traffic on cheap (resp. expensive) links than the centralized one does, that is

$$y_k^{ne} > y_k^* \text{ for all } k \in \mathcal{S}_1 \text{ and } y_j^{ne} < y_j^* \text{ for all } j \in \mathcal{S}_2. \tag{16}$$

(b) For any cheap link $k \in \mathcal{S}_1$, the difference $y_k^{ne} - y_k^*$ between the amount of flow forwarded on this link by the decentralized routing strategy and the centralized one decreases as $\bar{\lambda}$ increases, that is,

$$\frac{dy_k^{ne}}{d\bar{\lambda}} < \frac{dy_k^*}{d\bar{\lambda}}. \tag{17}$$

(c) For any expensive link $j \in \mathcal{S}_2$, the difference $y_j^{ne} - y_j^*$ between the amount of flow forwarded on this link by the decentralized routing strategy and the centralized one increases as $\bar{\lambda}$ increases, that is,

$$\frac{dy_j^{ne}}{d\bar{\lambda}} > \frac{dy_j^*}{d\bar{\lambda}}. \tag{18}$$

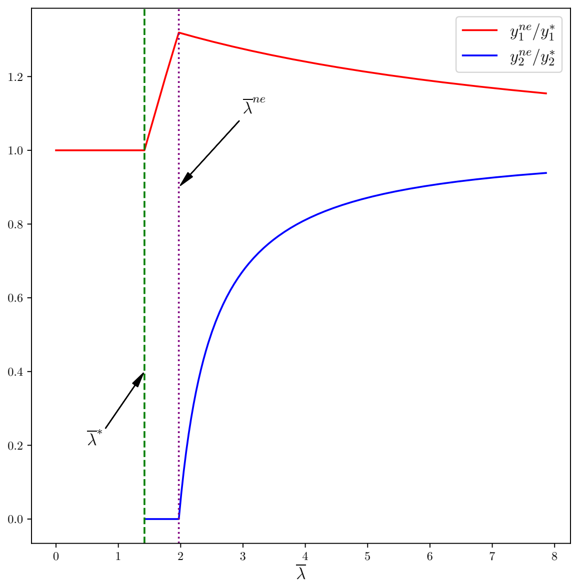


Fig. 2: Evolution of the link flow ratios y_1^{ne}/y_1^* (for the cheap link) and y_2^{ne}/y_2^* (for the expensive links) as a function of $\bar{\lambda}$ for the latency function $\phi(x) = e^x$. In this example, there are $K = 5$ users, one cheap link with $c_1 = 1$ and nine expensive links with $c_2 = 10$.

Using Proposition 2 and Proposition 3, Proposition 4 below characterizes the behaviour of the ratios y_l^{ne}/y_l^* of link flows obtained under both settings when the total traffic $\bar{\lambda}$ varies.

Proposition 4. *Let $k \in \mathcal{S}_1$ be an arbitrary cheap link and $j \in \mathcal{S}_2$ be an arbitrary expensive link and consider the ratios y_k^{ne}/y_k^* and y_j^{ne}/y_j^* as functions of the total traffic $\bar{\lambda}$. It holds that*

- *the ratio y_k^{ne}/y_k^* is strictly increasing in $\bar{\lambda}$ over the interval $\left(\bar{\lambda}^*, \bar{\lambda}^{ne}\right]$, and strictly decreasing over the interval $\left(\bar{\lambda}^{ne}, \infty\right)$,*
- *The ratio y_j^{ne}/y_j^* is 0 over the interval $\left(\bar{\lambda}^*, \bar{\lambda}^{ne}\right]$, and strictly increasing in $\bar{\lambda}$ over the interval $\left(\bar{\lambda}^{ne}, \infty\right)$.*

It directly follows from Proposition 4 and Lemma 1 that for any cheap link k the maximum value of the ratio y_k^{ne}/y_k^* is obtained when $\bar{\lambda} = \bar{\lambda}^{ne}$, as formally stated in Corollary 1.

Corollary 1. *For all cheap links $k \in \mathcal{S}_1$, the ratio y_k^{ne}/y_k^* achieves its maximum for $\bar{\lambda} = \bar{\lambda}^{ne}$. At this point, the ratio y_j^{ne}/y_j^* is minimum for all expensive links $j \in \mathcal{S}_2$.*

To summarize, in the interval $(0, \bar{\lambda}^*]$, the ratio y_k^{ne}/y_k^* is constant and equal to 1 for any cheap link k. In the interval $(\bar{\lambda}^*, \bar{\lambda}^{ne}]$, the ratio y_k^{ne}/y_k^* increases as $\bar{\lambda}$ increases and it reaches its maximum for $\bar{\lambda} = \bar{\lambda}^{ne}$. From this point onwards, the ratio y_k^{ne}/y_k^* decreases with $\bar{\lambda}$. Similarly, for any expensive link j, the ratio y_j^{ne}/y_j^* is 0 over the interval $(\bar{\lambda}^*, \bar{\lambda}^{ne}]$ (it is undefined for $\bar{\lambda} \leq \bar{\lambda}^*$) and it increases with $\bar{\lambda}$ over $(\bar{\lambda}^{ne}, \infty)$. Figure 2 illustrates this behaviour of the link flow ratios y_k^{ne}/y_k^* and y_j^{ne}/y_j^* for $\phi(x) = e^x$, $K = 5$ users and $N = 10$ parallel links. In this example, there is only one cheap link of cost $c_1 = 1$ and there are nine expensive links of cost $c_2 = 10$.

4.2 Worst-case total traffic for the ratio of social costs

We now study the ratio of social costs. It directly follows from Lemma 1 that this ratio is continuous in $\bar{\lambda}$.

Lemma 2. *As a function of $\bar{\lambda}$, the ratio $D_K\left(\frac{\bar{\lambda}}{K}\mathbf{1}, \mathbf{p}\right)/D_1(\bar{\lambda}, \mathbf{p})$ is continuous over $(0, \infty)$.*

Proposition 5 below characterizes the behaviour of the ratio of social costs as $\bar{\lambda}$ varies over $(0, \infty)$.

Proposition 5. *For $K > 1$, as a function of $\bar{\lambda}$, the ratio $D_K\left(\frac{\bar{\lambda}}{K}\mathbf{1}, \mathbf{p}\right)/D_1(\bar{\lambda}, \mathbf{p})$ of the social costs obtained under the decentralized routing strategy and the centralized one is*

(a) constant and equal to 1 in the interval $(0, \bar{\lambda}^*]$,
(b) strictly increasing with $\bar{\lambda}$ in the interval $\left(\bar{\lambda}^*, \bar{\lambda}^{ne}\right]$, and,
(c) strictly decreasing with $\bar{\lambda}$ in the interval $\left(\bar{\lambda}^{ne}, \infty\right)$.

Proof. See Appendix A. □

We have shown in Section 4.1 that for any cheap link $k \in \mathcal{S}_1$, the ratio y_k^{ne}/y_k^* of the flows on this link obtained under the decentralized routing strategy and the optimal one achieves its maximum for $\bar{\lambda} = \bar{\lambda}^{ne}$. As stated in Corollary 2, the same is true for the ratio of social costs.

Corollary 2. *The ratio $D_K\left(\frac{\bar{\lambda}}{K}\mathbf{1}, \mathbf{p}\right)/D_1(\bar{\lambda}, \mathbf{p})$ achieves its maximum when $\bar{\lambda} = \bar{\lambda}^{ne}$, implying that*

$$I_K(\mathbf{p}) = \frac{n_1 F_1\left(\frac{\bar{\lambda}^{ne}}{n_1}\right)}{n_1 F_1\left(y_1^*\right) + n_2 F_2\left(y_N^*\right)}, \tag{19}$$

where y_1^* and y_N^* are the links flows over cheap and expensive links, respectively, obtained under the centralized routing strategy for $\bar{\lambda} = \bar{\lambda}^{ne}$.

Proof. The proof directly follows from Lemma 2 and Proposition 5. □

Figure 3 illustrates the evolution of the ratio of social costs as the total traffic $\bar{\lambda}$ in the system varies for the latency functions $\phi(x) = e^x$, $\phi(x) = (1+x)^3$ and $\phi(x) = (1+x)^4$. The setting is the same as in Figure 2, that is, there are $K = 5$ users, one cheap link with $c_1 = 1$ and nine expensive links with $c_2 = 10$. For $\phi(x) = e^x$, the ratio is constant before $\bar{\lambda}^* = 1.42$, it increases from $\bar{\lambda}^*$ up to $\bar{\lambda}^{ne} = 1.97$ where it reaches a maximum value of 1.21, and then decreases with $\bar{\lambda}$. Similarly, for $\phi(x) = (1+x)^4$, the ratio of social costs is constant before $\bar{\lambda}^* = 0.45$. From this point onwards it increases up to $\bar{\lambda}^{ne} = 0.66$ where it reaches a maximum value of 1.18, and then decreases with $\bar{\lambda}$.

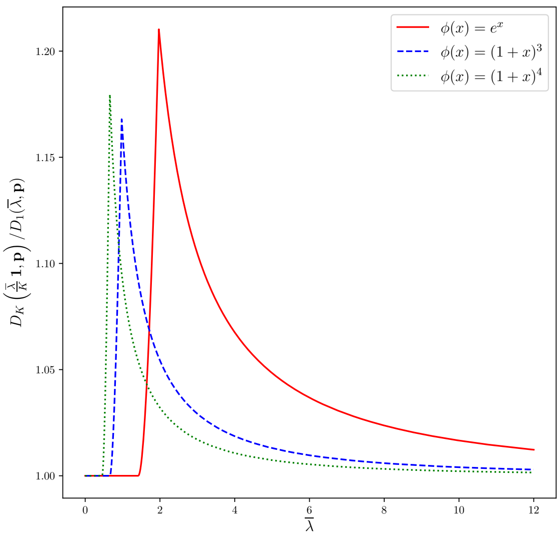


Fig. 3: Evolution of the ratio $D_K\left(\frac{\bar{\lambda}}{K}\mathbf{1},\mathbf{p}\right)/D_1(\bar{\lambda},\mathbf{p})$ as $\bar{\lambda}$ increases for the latency functions $\phi(x) = e^x$, $\phi(x) = (1+x)^3$ and $\phi(x) = (1+x)^4$. The setting is the same as in Figure 2.

In summary, we have shown that, given a fixed network configuration $\mathbf{p}$, the worst inefficiency of the decentralized routing scheme is obtained when all users control the same amount of traffic and when the total traffic in the system is $\bar{\lambda} = \bar{\lambda}^{ne}$. This corresponds to the value of the total traffic for which the decentralized routing scheme starts using the expensive links. When $K > 1$, this value is strictly greater than $\bar{\lambda}^*$, which means that selfish users send all their traffic demands on the cheap links longer than what would be globally optimal. Although this result is proven here only for latency functions satisfying Assumptions 1-5, we note that a similar result was proven in Doncel et al. (2014) for the M/M/1 latency function, that is, for $\phi(x) = 1/(1-x)$. As discussed in Section 6, numerical experiments suggest that Corollary 2 seems to hold for a much broader class of latency functions.

We would like to remark that the result of Corollary 2 is consistent with the work of Cominetti et al. (2021), in which the authors show that, for nonatomic routing games and affine costs, the local maxima of the ratio of social costs are obtained when the total traffic is such that a new set of links is used. Another direct consequence of Corollary 2 is that, given a fixed network configuration, the worst inefficiency of the decentralized routing scheme is obtained neither for $\bar{\lambda} \to 0$ nor for $\bar{\lambda} \to \infty$, as could be expected. Interestingly, this result is consistent with the work of Colini-Baldeschi et al. (2020) where the authors show that the PoA of nonatomic routing games is not achieved when the traffic is very small or very large. Nevertheless, for $\phi(x) = (1+x)^m$, the value of $\bar{\lambda}^{ne}$ can be made arbitrarily small, as shown in Lemma 3 below.

Lemma 3. *For $\phi(x) = (1+x)^m$, $\bar{\lambda}^{ne}$ decreases with m and $\bar{\lambda}^{ne} \to 0$ as $m \to \infty$.*

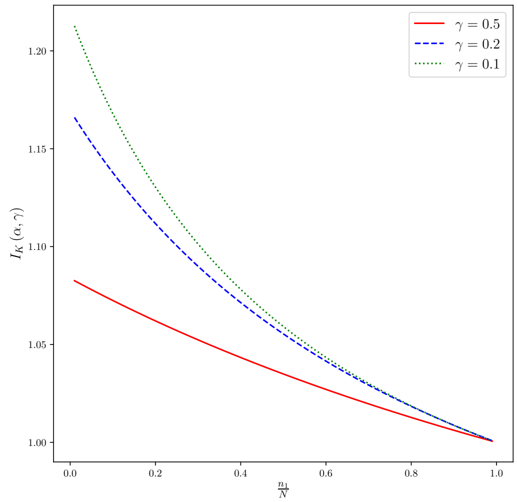


Fig. 4: Inefficiency $I_K(\alpha, \gamma)$ as a function of $\frac{n_1}{N}$ for different values of γ and for $\phi(x) = (1 + x)^3$.

5 From Inefficiency to Price of Anarchy

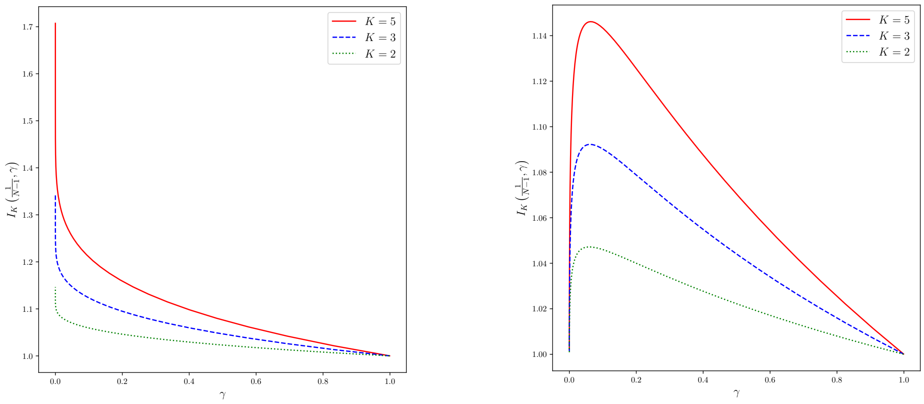
In Section 4, we have characterized the worst traffic conditions for the inefficiency $I_K(\mathbf{p})$ of the decentralized routing scheme, assuming a fixed network configuration $\mathbf{p}$. In this section, we assume that the worst traffic conditions are met and investigate the worst network configuration for the inefficiency of the decentralized routing scheme. In other words, we study the PoA for the routing game over parallel links, where in (3), the PoA was defined as the supremum over all the network configurations $\mathbf{p}$ of $I_K(\mathbf{p})$. We first show that the *Inefficiency* depends only on the ratios n_1/n_2 and c_1/c_2.

Proposition 6. *The Inefficiency $I_K(\mathbf{p})$ depends on the parameters $\mathbf{p} = (\mathbf{n}, \mathbf{c})$ only through the ratios $\alpha = \frac{n_1}{n_2}$ and $\gamma = \frac{c_1}{c_2}$.*

Note in particular that the *Inefficiency* depends on the total number of links only through the values that the ratio $\alpha = \frac{n_1}{n_2}$ can take. As a consequence of Proposition 6, we shall write $I_K(\alpha, \gamma)$ instead of $I_K(\mathbf{n}, \mathbf{c})$ to denote the *Inefficiency* in the following. We study below how the *Inefficiency* varies with α and γ.

We first study how $I_K(\alpha, \gamma)$ varies with α for a fixed value of γ. Assuming that $K = 5$ and $N = 100$, Figure 4 shows $I_K(\alpha, \gamma)$ as a function of n_1/N for $\gamma = 0.5$, $\gamma = 0.2$ and $\gamma = 0.1$ when $\phi(x) = (1+x)^3$. We observe that the inefficiency of the decentralized routing scheme seems to decrease as the proportion of cheap links increases. A similar behavior was observed for $\phi(x) = e^x$. As $n_1/N = \alpha/(1 + \alpha)$ is an increasing function of α, this suggests that the inefficiency decreases with the ratio α of the numbers of cheap and expensive links. This is formally proven in Proposition 7 below.

Proposition 7. *The Inefficiency $I_K(\alpha, \gamma)$ is strictly decreasing with α.*



(a) $\phi(x) = e^x$. (b) $\phi(x) = (1+x)^2$.

Fig. 5: Inefficiency $I_K(\frac{1}{N-1}, \gamma)$ as a function of γ for different values of K and for (a) $\phi(x) = e^x$ and (b) $\phi(x) = (1+x)^2$.

An immediate consequence is the following corollary.

Corollary 3. *The PoA is obtained when there is only one "cheap" link and $N-1$ " expensive" links, that is,*

$$PoA(K, N) = \sup_{\alpha,\gamma} I_K(\alpha, \gamma) = \sup_{\gamma} I_K\left(\frac{1}{N-1}, \gamma\right). \tag{20}$$

In the following, we shall therefore assume that $\alpha = 1/(N-1)$ and study how the *Inefficiency* varies as a function of γ. Assuming that $\phi(x) = e^x$, Figure 5a shows the values obtained for $I_K\left(\frac{1}{N-1}, \gamma\right)$ as γ varies from 0 to 1 in scenarios with $K = 2$, $K = 3$ and $K = 5$ users and $N = 10$ parallel links. We observe that for all values of K the *Inefficiency* is strictly decreasing with γ, which, according to (20), implies that the PoA is obtained when γ tends to zero. As a result, we conjecture that for $\phi(x) = e^{\nu x}$,

$$PoA(K, N) = \lim_{\gamma \to 0} I_K\left(\frac{1}{N-1}, \gamma\right).$$

Besides, we get the following bounds on the performance degradation for $\phi(x) = e^{\nu x}$.

Lemma 4. *For $\phi(x) = e^{\nu x}$ and $N \geq 2$, it holds that*

$$K^{\frac{N-1-\nu}{N-1}} \frac{N-1}{N(1+\log K)-1} \leq \lim_{\gamma \to 0} I_K\left(\frac{1}{N-1}, \gamma\right) \leq K,$$

from which it follows that $K/(1+\log(K)) \leq \lim_{N\to\infty} PoA(K, N)$.

In words, we conjecture that for $\phi(x) = e^{\nu x}$ the worst inefficiency is achieved when the cost of the "expensive" links is infinitely larger than the cost of the "cheap" link. If this conjecture holds, this implies that the PoA is upper bounded by K. Furthermore, regardless this conjecture is true or not, the above result implies that the PoA for $\phi(x) = e^{\nu x}$ is, at least, $K^{\frac{N-1-\nu}{N-1}}(N-1)/[N(1+\log K)-1]$. For $\nu = 1$, $N = 10$ and $K = 5$, it yields 1.499, which is to be compared to the value 1.7 obtained in Figure 5a. Moreover, we can conclude that the PoA for $\phi(x) = e^{\nu x}$ is unbounded in nonatomic routing games as $K/(1+\log(K))$ tends to ∞ when $K \to \infty$. Surprisingly, the monotonicity property shown in Figure 5a for $\phi(x) = e^{\nu x}$ does not seem to hold for $\phi(x) = (1+x)^m$. As illustrated for $\phi(x) = (1+x)^2$ in Figure 5b, in which we also assume that $N = 10$, the inefficiency $I_K(\frac{1}{N-1}, \gamma)$ obtained for different values of K is not monotone as a function of γ. A similar behaviour was observed for other values of m. Unfortunately, we were not able to characterize the value of γ yielding the worst inefficiency. Therefore, the precise value of γ for which the PoA is achieved when $\phi(x) = (1+x)^m$ remains an open question.

6 Extensions of this work

Our main result is that the worst inefficiency of the decentralized routing scheme is obtained when the traffic demands of all users are $\bar{\lambda}^{ne}/K$. The key ingredient to prove this result is Proposition 5, which characterizes how the ratio of equilibrium and optimal social costs varies with the total traffic demand $\bar{\lambda}$. Proposition 5 has been established under sufficient conditions on the latency function $\phi(x)$, assuming that there are only two types of links with the same latency function $\phi(x)$. We discuss below several interesting extensions of this work.

6.1 Generalization to other latency functions

As already mentioned, it was proven in Doncel et al. (2014) that for $\phi(x) = 1/(1-x)$ the ratio of social costs varies with $\bar{\lambda}$ exactly as stated in Proposition 5 for latency functions satisfying Assumptions 1-5 such as $\phi(x) = e^{\nu x}$ and $\phi(x) = (1+x)^m$. Our numeric experiments suggest however that Proposition 5 holds for a much broader class of latency functions.

For instance, in Figure 6 we plot the evolution of the ratio of social costs as $\bar{\lambda}$ increases when $\phi(x) = 1 + x^m$ for different values of m. A similar behavior of the ratio of social costs is obtained for $\phi(x) = 1/(1-x)^m$. For both types of latency function, we observe that the ratio of social costs varies with $\bar{\lambda}$ as stated in Proposition 5, and that the worst inefficiency is obtained when $\bar{\lambda} = \bar{\lambda}^{ne}$.

Unfortunately, the arguments used to prove Proposition 5 do not readily apply to these latency functions, as we now briefly explain:

- For $\phi(x) = 1 + x^m$, with $m \geq 1$, it is straightforward to show that Assumption 5 holds, and thus that the ratio of social costs is strictly decreasing with $\bar{\lambda}$ over the interval $(\bar{\lambda}^{ne}, +\infty)$. However, the argument used to prove that this ratio is strictly increasing with $\bar{\lambda}$ over $(\bar{\lambda}^*, \bar{\lambda}^{ne})$ does not apply because the latency function $\phi(x) =$

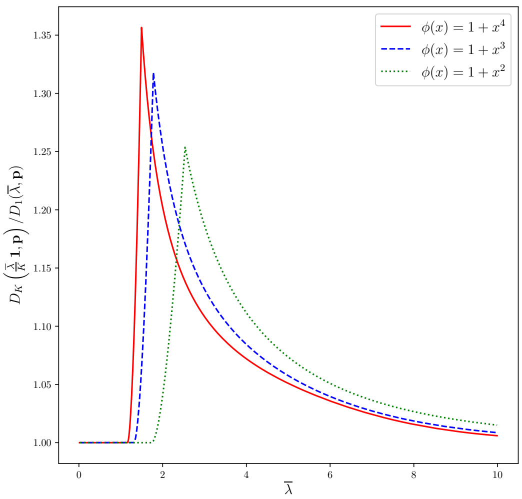


Fig. 6: Evolution of the ratio of social costs as $\bar{\lambda}$ increases for the latency function $\phi(x) = 1 + x^m$. In this example, there are $K = 5$ users, one cheap link with $c_1 = 1$ and nine expensive links with $c_2 = 10$.

$1+x^m$ does not satisfy Assumption 2 in Section 2.3. Therefore, a different argument should be used to prove that the ratio of social costs increases over $\left(\bar{\lambda}^*, \bar{\lambda}^{ne}\right)$.

- For $\phi = \frac{1}{(1-x)^m}$ with $m \geq 2$, the same approach as for $\phi(x) = e^{\nu x}$ and $\phi(x) = (1+x)^m$ can be used to show that the ratio of social costs is strictly increasing with $\bar{\lambda}$ over $\left(\bar{\lambda}^*, \bar{\lambda}^{ne}\right)$. Unfortunately, our numerical experiments suggest that Assumption 5 in Section 2.3 is not met by $\phi(x) = 1/(1-x)^m$, implying that a different approach should be used to prove that the ratio of social costs is strictly decreasing with $\bar{\lambda}$ for $\bar{\lambda} > \bar{\lambda}^{ne}$.

6.2 Extension to more than two types of links

Our results on the inefficiency of selfish routing have been established assuming that there are only two types of links which differ by their cost parameters c_1 and $c_2 > c_1$. Numeric experiments suggest however that similar results hold for more than two types of links.

In Figure 7 we plot the ratio of social costs obtained in the symmetric game for $\phi(x) = e^x$ and $\phi(x) = (1+x)^2$ as a function of the total traffic demand $\bar{\lambda}$ when there are 4 links, each of a different type. It can be observed that as $\bar{\lambda}$ increases the ratio goes through peaks and valleys, and finally converges towards 1 as $\bar{\lambda} \to \infty$. The peaks correspond to values of the total traffic demand $\bar{\lambda}$ at which the decentralized routing scheme starts using one new link (these values are shown with dotted vertical lines in Figure 7). A similar behavior of the ratio of social costs was observed in Doncel et al.

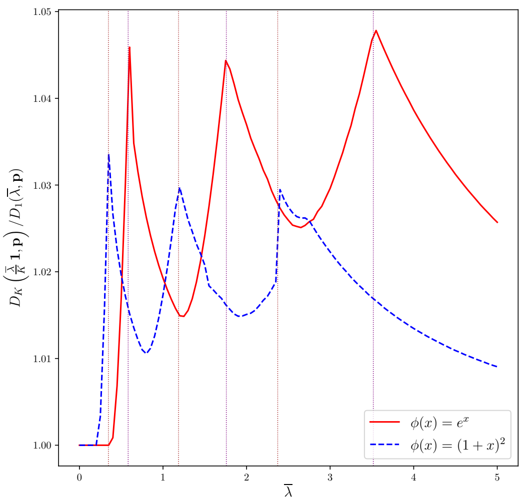


Fig. 7: Evolution of the ratio of social costs as a function of the total traffic demand for $\phi(x) = e^x$ and $\phi(x) = (1 + x)^2$ when there are $K = 5$ users and $N = 4$ links, each of a different type. The cost parameters of the links are as follows: $c_1 = 1$, $c_2 = 2$, $c_3 = 4$ and $c_4 = 8$.

(2014) for the latency function $\phi(x) = 1/(1 - x)$, and in nonatomic routing games with affine costs Cominetti et al. (2021).

The analysis is nevertheless much more complex for more than two types of links as one needs to compare multiple local maxima to determine the *Inefficiency*. The extension of our results to more than two types of links is therefore left for future work.

6.3 Extension to heterogeneous latency functions

It was assumed throughout this paper that the two types of links share the same latency function $\phi(x)$ and differ only through their cost parameters c_1 and c_2. A natural extension of this work would be to investigate the inefficiency of selfish routing for heterogeneous latency functions, that is, when the cost function of type-i links is $c_i\,\phi_i(x)$ (e.g., c_1e^x for type-1 links and $c_2(1 + x)^m$ for type-2 links). It is known that even in this case the Nash Equilibrium exists and is unique under mild assumptions on the latency functions $\phi_i(x)$ Orda et al. (1993). It is not clear however whether the global cost $D_K(\boldsymbol{\lambda}, \mathbf{p})$ achieves its maximum for the symmetric game $\boldsymbol{\lambda} = \left(\frac{\bar\lambda}{K}, \ldots, \frac{\bar\lambda}{K}\right)$ in this case. The proof of this result in Brun and Prabhu (2016) relies on a monotonic property regarding the order of preference of links as seen by each user (see Proposition 2 and Lemma 1 in Brun and Prabhu (2016)). Proving that this monotonic property is still valid in the case of heterogeneous latency functions is highly non-trivial. As a consequence, it is not clear whether the worst inefficiency of the decentralized routing scheme is obtained for a symmetric game, which is crucial to characterize the decentralized routing strategy as the solution of a convex optimization problem (see Theorem 1). This extension is therefore also left for future work.

7 Conclusions

For the specific atomic routing game considered in this paper, it was shown that the worst traffic conditions occur when all users have the same traffic demand and the total traffic demand is such that "expensive links" are marginally used by selfish routing. Moreover, if these worst traffic conditions are met, the worst inefficiency of the selfish routing scheme is obtained when the number of "expensive" links is infinitely larger than the number of "cheap" links and under a very specific condition on the ratio c_1/c_2 (which we conjecture to be $c_1/c_2 \to 0$ for $\phi(x) = e^{\nu x}$). The worst-case scenarios for the inefficiency of selfish routing therefore corresponds to very specific traffic conditions and to highly asymmetric network configurations, which explain why the PoA is probably an overly pessimistic performance measure, as advocated in many recent works on non-atomic routing games.

Compliance with Ethical Standards:

Funding: This work has been partially funded by the Consolidated Research Group MATHMODE (IT1456-22) given by the Department of Education of the Basque Government.

Ethical approval: This article does not contain any studies with human participants or animals performed by any of the authors.

Conflict of Interest: Olivier Brun declares that he has no conflict of interest. Josu Doncel declares that he has no conflict of interest.

References

Altman, E., Ayesta, U., Prabhu, B.J.: Load balancing in processor sharing systems. Telecommunication Systems **47**, 35–48 (2011)

Ayesta, U., Brun, O., Prabhu, B.J.: Price of anarchy in non-cooperative load balancing games. Performance Evaluation **68**(12), 1312–1332 (2011)

Anselmi, J., Gaujal, B.: Optimal routing in parallel, non-observable queues and the price of anarchy revisited. In: 22nd International Teletraffic Congress (ITC), Amsterdam (2010)

Balcan, M.-F., Blum, A., Mansour, Y.: Circumventing the price of anarchy: Leading dynamics to good behavior. SIAM Journal on Computing **42**(1), 230–264 (2013) https://doi.org/10.1137/110821317 https://doi.org/10.1137/110821317

Brun, O., Doncel, J.: On the Inefficiency of Atomic Routing Games over Parallel Links. working paper or preprint (2023). https://laas.hal.science/hal-03863930

Brun, O., Prabhu, B.: Worst-case analysis of non-cooperative load balancing. Annals of Operations Research **239**(2), 471–495 (2016) https://doi.org/10.1007/s10479-014-1747-7

Braess, D.: Über ein paradoxon aus der verkehrsplanung. Unternehmensforschung **12**(1), 258–268 (1968)

Bell, C.H., Stidham, S.: Individual versus social optimization in the allocation of customers to alternative servers. Management Science **29**, 831–839 (1983)

Colini-Baldeschi, R., Cominetti, R., Mertikopoulos, P., Scarsini, M.: When is selfish routing bad? the price of anarchy in light and heavy traffic. Operations Research **68**(2), 411–434 (2020)

Colini-Baldeschi, R., Cominetti, R., Scarsini, M.: On the price of anarchy of highly congested nonatomic network games. In: International Symposium on Algorithmic Game Theory, pp. 117–128 (2016). Springer

Colini-Baldeschi, R., Cominetti, R., Scarsini, M.: Price of anarchy for highly congested routing games in parallel networks. Theory of Computing Systems **63**(1), 90–113 (2019)

Cominetti, R., Correa, J.R., Stier-Moses, N.E.: The impact of oligopolistic competition in networks. Operations Research **57**(6), 1421–1437 (2009) https://doi.org/10.1287/opre.1080.0653

Cominetti, R., Dose, V., Scarsini, M.: The price of anarchy in routing games as a function of the demand. Mathematical Programming, 1–28 (2021)

Czumaj, A., Krysta, P., Vocking, B.: Selfish traffic allocation for server farms. In: Proceedings of STOC (2002)

Chen, H.L., Marden, J., Wierman, A.: The effect of local scheduling in load balancing designs. In: Proceedings of IEEE Infocom (2009)

Correa, J.R., Schulz, A.S., Stier-Moses, N.E.: Selfish routing in capacitated networks. Mathematics of Operations Research **29**(4), 961–976 (2004)

Correa, J.R., Schulz, A.S., Stier-Moses, N.E.: Fast, fair, and efficient flows in networks. Operations Research **55**(2), 215–225 (2007)

Correa, J.R., Schulz, A.S., Stier-Moses, N.E.: A geometric approach to the price of anarchy in nonatomic congestion games. Games and Economic Behavior **64**(2), 457–469 (2008)

Doncel, J., Ayesta, U., Brun, O., Prabhu, B.: Is price of anarchy the right measure for load-balancing games? ACM Transactions on Internet Technology **14**(2–3) (2014)

Feldman, M., Immorlica, N., Lucier, B., Roughgarden, T., Syrgkanis, V.: The price of anarchy in large games. In: Proceedings of the Forty-Eighth Annual ACM Symposium on Theory of Computing. STOC '16, pp. 963–976. Association for Computing Machinery, New York, NY, USA (2016). https://doi.org/10.1145/2897518.2897580 . https://doi.org/10.1145/2897518.2897580

Ghosh, S., Hassin, R.: Inefficiency in stochastic queueing systems with strategic customers. European Journal of Operational Research **295**(1), 1–11 (2021)

Haviv, M., Roughgarden, T.: The price of anarchy in an exponential multi-server. Operations Research Letters **35**, 421–426 (2007)

Hamers, H., Slikker, M., Klijn, F.: Price of Anarchy in Sequencing Situations and the Impossibility to Coordinate. Working Papers 709, Barcelona School of Economics (September 2015). https://ideas.repec.org/p/bge/wpaper/709.html

Koutsoupias, E., Papadimitriou, C.H.: Worst-case equilibria. In: STACS 1999 (1999)

Monnot, B., Benita, F., Piliouras, G.: Routing games in the wild: Efficiency, equilibration and regret. In: International Conference on Web and Internet Economics, pp. 340–353 (2017). Springer

Nash, J.: Non-cooperative games. Annals of mathematics, 286–295 (1951)

Nisan, N., Roughgarden, T., Tardos, E., Vazirani, V.V. (eds.): Algorithmic Game Theory. Cambridge University Press, ??? (2007). https://doi.org/10.1017/CBO9780511800481

Orda, A., Rom, R., Shimkin, N.: Competitive routing in multi-user communication networks. IEEE/ACM Transactions on Networking **1**, 510–521 (1993)

Pigou, A.: The Economics of Welfare. Macmillan, London, ??? (1920)

Roughgarden, T.: The price of anarchy is independent of the network topology. In: Proceedings of the Thiry-fourth Annual ACM Symposium on Theory of Computing, pp. 428–437 (2002)

Roughgarden, T., Tardos, É.: How bad is selfish routing? Journal of the ACM (JACM) **49**(2), 236–259 (2002)

Roughgarden, T., Tardos, É.: Bounding the inefficiency of equilibria in nonatomic congestion games. Games and economic behavior **47**(2), 389–403 (2004)

Seaton, J.H., Brown, P.N.: On the intrinsic fragility of the price of anarchy. IEEE Control Systems Letters **7**, 3573–3578 (2023) https://doi.org/10.1109/LCSYS.2023.

3335315

Suri, S., Tóth, C.D., Zhou, Y.: Selfish load balancing and atomic congestion games. In: Proceedings of the Sixteenth Annual ACM Symposium on Parallelism in Algorithms and Architectures. SPAA '04, pp. 188–195. Association for Computing Machinery, New York, NY, USA (2004). https://doi.org/10.1145/1007912.1007941 . https://doi.org/10.1145/1007912.1007941

US Bureau of Public Roads: Traffic Assignment Manual for Application with a Large, High Speed Computer vol. 2. U.S. Department of Commerce, Bureau of Public Roads, Office of Planning, Urban Planning Division, ??? (1964). https://books.google.fr/books?id=gkNZAAAAMAAJ

Wardrop, J.G.: Road paper. some theoretical aspects of road traffic research. Proceedings of the institution of civil engineers **1**(3), 325–362 (1952)

Wu, Z., Möhring, R.H., Chen, Y., Xu, D.: Selfishness need not be bad. Operations Research **69**(2), 410–435 (2021)

Youn, H., Gastner, M.T., Jeong, H.: Price of anarchy in transportation networks: efficiency and optimality control. Phys Review Letter **101**(12) (2008)

Appendix A Proof of Proposition 5

In this appendix, we prove Proposition 5.

Proof of Proposition 5. We first prove assertion (a). It readily follows from $\bar{\lambda}^* < \bar{\lambda}^{ne}$ that for $\bar{\lambda} \leq \bar{\lambda}^*$

$$\frac{D_K\left(\frac{\bar{\lambda}}{K}\mathbf{1}, \mathbf{p}\right)}{D_1(\bar{\lambda}, \mathbf{p})} = \frac{F(\mathbf{y}^{ne})}{F(\mathbf{y}^*)} = \frac{n_1 F_1(\frac{\bar{\lambda}}{n_1})}{n_1 F_1(\frac{\bar{\lambda}}{n_1})} = 1.$$

We now consider assertion (b) and assume that $\bar{\lambda} \in \left(\bar{\lambda}^*, \bar{\lambda}^{ne}\right]$. Since for $\bar{\lambda} \leq \bar{\lambda}^{ne}$ we have $\mathbf{y}^{ne} = \left(\frac{\bar{\lambda}}{n_1}, \ldots, \frac{\bar{\lambda}}{n_1}, 0, \ldots, 0\right)$, the ratio $D_K\left(\frac{\bar{\lambda}}{K}\mathbf{1}, \mathbf{p}\right)/D_1(\bar{\lambda}, \mathbf{p})$ is strictly increasing in $\bar{\lambda}$ if and only if

$$F_1'\left(\frac{\bar{\lambda}}{n_1}\right) F(\mathbf{y}^*) > \left(n_1 F_1'(y_1^*)\frac{dy_1^*}{d\bar{\lambda}} + n_2 F_2'(y_N^*)\frac{dy_N^*}{d\bar{\lambda}}\right) n_1 F_1(\frac{\bar{\lambda}}{n_1}). \tag{A1}$$

The constraint $n_1 y_1^* + n_2 y_N^* = \bar{\lambda}$ implies that $n_1 \frac{dy_1^*}{d\bar{\lambda}} + n_2 \frac{dy_N^*}{d\bar{\lambda}} = 1$. Moreover, $\bar{\lambda} > \bar{\lambda}^*$ implies that $F_1'(y_1^*) = F_2'(y_N^*)$. It yields $n_1 F_1'(y_1^*)\frac{dy_1^*}{d\bar{\lambda}} + n_2 F_2'(y_N^*)\frac{dy_N^*}{d\bar{\lambda}} = F_1'(y_1^*)$. Therefore, the ratio $D_K\left(\frac{\bar{\lambda}}{K}\mathbf{1}, \mathbf{p}\right)/D_1(\bar{\lambda}, \mathbf{p})$ is strictly increasing in $\bar{\lambda}$ if and only if

$$F_1'\left(\frac{\bar{\lambda}}{n_1}\right)[n_1 F_1(y_1^*) + n_2 F_2(y_N^*)] > n_1 F_1'(y_1^*) F_1\left(\frac{\bar{\lambda}}{n_1}\right),$$

which can equivalently be written as

$$n_1 \frac{F_1(y_1^*)}{F_1'(y_1^*)} + n_2 \frac{F_2(y_N^*)}{F_2'(y_N^*)} > n_1 \frac{F_1(\frac{\bar{\lambda}}{n_1})}{F_1'(\frac{\bar{\lambda}}{n_1})},$$

where we have used the equality $F_1'(y_1^*) = F_2'(y_N^*)$. Observing that $\frac{F_1(y_1^*)}{F_1'(y_1^*)} = g(y_1^*)$ and $\frac{F_2(y_N^*)}{F_2'(y_N^*)} = g(y_N^*)$, where the function $g()$ is defined by (4) in Section 2.3, it is therefore enough to show that

$$n_1 g(y_1^*) + n_2 g(y_N^*) > n_1 g\left(\frac{\bar{\lambda}}{n_1}\right). \tag{A2}$$

To this end, let us consider the function $h(t) = n_1 g\left(\frac{t\bar{\lambda}}{n_1}\right) + n_2 g\left(\frac{(1-t)\bar{\lambda}}{n_2}\right)$. We have

$$h'(t) = \bar{\lambda} \times \left(g'\left(\frac{t\bar{\lambda}}{n_1}\right) - g'\left(\frac{(1-t)\bar{\lambda}}{n_2}\right)\right),$$

which implies that $h'(t) < 0$ for $t \in \left[\frac{n_1}{n_1+n_2}, 1\right]$ because g is strictly concave (see Assumption 2 in Section 2.3). We know from Proposition 2 that $y_1^* > y_N^*$. Together with $n_1 y_1^* + n_2 y_N^* = \bar{\lambda}$, it implies that $\frac{y_1^*}{\bar{\lambda}} > \frac{1}{n_1+n_2}$. Taking $t = \frac{n_1 y_1^*}{\bar{\lambda}} < 1$, we hence obtain that $h(1) < h(t)$, which proves (A2). We thus conclude that the ratio $D_K\left(\frac{\bar{\lambda}}{K}\mathbf{1}, \mathbf{p}\right)/D_1(\bar{\lambda}, \mathbf{p})$ is strictly increasing in $\bar{\lambda}$ over the interval $\left(\bar{\lambda}^*, \bar{\lambda}^{ne}\right]$.

Finally, we focus on assertion (c) assuming that $\bar{\lambda} > \bar{\lambda}^{ne}$. The ratio $D_K\left(\frac{\bar{\lambda}}{K}\mathbf{1}, \mathbf{p}\right)/D_1(\bar{\lambda}, \mathbf{p})$ is strictly decreasing in $\bar{\lambda}$ if and only if

$$\left(n_1 F_1'(y_1^{ne})\frac{dy_1^{ne}}{d\bar{\lambda}} + n_2 F_2'(y_N^{ne})\frac{dy_N^{ne}}{d\bar{\lambda}}\right) F(\mathbf{y}^*) < \left(n_1 F_1'(y_1^*)\frac{dy_1^*}{d\bar{\lambda}} + n_2 F_2'(y_N*)\frac{dy_N^*}{d\bar{\lambda}}\right) F(\mathbf{y}^{ne}). \tag{A3}$$

From Proposition 3, we know that $\frac{dy_1^{ne}}{d\bar{\lambda}} < \frac{dy_1^*}{d\bar{\lambda}}$. Let $\delta = n_1 \frac{dy_1^*}{d\bar{\lambda}}$. We have

$$\begin{aligned} n_1 F_1'(y_1^{ne})\frac{dy_1^{ne}}{d\bar{\lambda}} + n_2 F_2'(y_N^{ne})\frac{dy_N^{ne}}{d\bar{\lambda}} &= n_1 F_1'(y_1^{ne})\frac{dy_1^{ne}}{d\bar{\lambda}} + n_2 F_2'(y_N^{ne})\frac{1}{n_2}\left(1 - n_1\frac{dy_1^{ne}}{d\bar{\lambda}}\right) \\ &= n_1\left(F_1'(y_1^{ne}) - F_2'(y_N^{ne})\right)\frac{dy_1^{ne}}{d\bar{\lambda}} + F_2'(y_N^{ne}) \\ &< \delta\left(F_1'(y_1^{ne}) - F_2'(y_N^{ne})\right) + F_2'(y_N^{ne}) \\ &= \delta F_1'(y_1^{ne}) + (1-\delta)F_2'(y_N^{ne}), \end{aligned}$$

where we have used the inequality $F_1'(y_1^{ne}) > F_2'(y_N^{ne})$ which is proven in Proposition 2. As a consequence, a sufficient condition for (A3) to hold is that

$$\left(\delta F_1'(y_1^{ne}) + (1-\delta)F_2'(y_N^{ne})\right) F(\mathbf{y}^*) < \left(\delta F_1'(y_1^*) + (1-\delta)F_2'(y_N^*)\right) F(\mathbf{y}^{ne}), \tag{A4}$$

which can equivalently be written as follows

$$\frac{F(\mathbf{y}^*)}{\delta F_1'(y_1^*) + (1-\delta)F_2'(y_N^*)} < \frac{F(\mathbf{y}^{ne})}{\delta F_1'(y_1^{ne}) + (1-\delta)F_2'(y_N^{ne})}, \tag{A5}$$

that is $Q(0) < Q(\Delta^{ne})$. Assumption 5 completes the proof. □