51st IEEE Conference on Decision and Control
December 10-13, 2012. Maui, Hawaii, USA

On the Characterization and Computation of Nash Equilibria on Parallel Networks with Horizontal Queues

Walid Krichene   Jack Reilly   Saurabh Amin   Alexandre Bayen

Abstract— **We study inefficiencies in parallel networks with horizontal queues due to the selfish behavior of players, by comparing social optima to Nash equilibria. The article expands studies on routing games which traditionally model congestion with latency functions that increase with the flow on a particular link. This type of latency function cannot capture congestion effects on horizontal queues. Latencies on horizontal queues increase as a function of density, and flow can decrease with increasing latencies. This class of latency functions arises in transportation networks. For static analysis of horizontal queues on parallel-link networks, we show that there may exist multiple Nash equilibria with different total costs, which contrasts with results for increasing latency functions. We present a novel algorithm, quadratic in the number of links, for computing the Nash equilibrium that minimizes total cost (best Nash equilibrium). The relative inefficiencies of best Nash equilibria are evaluated through analysis of the price of stability, and analytical results are presented for two-link networks. Price of stability is shown to be sensitive to changes in demand when links are near capacity, and congestion mitigation strategies are discussed, motivated by our results.**

I. Introduction

A. Routing games and Nash equilibria

Routing games (or congestion games) form an important class of non-atomic games that is used to model the interaction of players who are sharing resources on a network, in which the cost on each edge depends on the fraction of players using that edge. Extensive work has been dedicated to studying Nash equilibria (or user optimal assignments) of congestion games [9], [12], in which all players are assumed to choose the routes that minimize their respective individual costs. Under some assumptions on the latency functions, Nash equilibria can be computed as a solution of a convex optimization problem [5]. Nash equilibria of congestion games are known to be inefficient compared to system optimal assignments, in which a coordinator, or a central authority, assigns flow as to minimize a cost function over all players [2], [16].

B. A new class of latency functions

The latency functions that have been studied so far in routing games literature satisfy the following assumptions: if $\ell_n(x_n)$ is the latency on a link n, where x_n is the flow, then ℓ_n is assumed to be non-decreasing, and $x_n \mapsto x_n\ell_n(x_n)$ is assumed to be convex [11]. While this class of latency functions provides a good model of congestion for a considerable range of networks, such as communication networks, it does not accurately model congestion with horizontal queues, such as congestion on transportation networks [4], [15], [7]. Intuitively, a given flow x on a road can correspond to either a large concentration of cars moving slowly (high density on a congested road), in which case the latency is large, or few cars moving quickly (low density), in which case the latency is small. We introduce a new model of latency that captures this phenomenon by adding a binary state variable m_n on each link n that specifies if the link is in free-flow or in congestion. We show that such latency functions can be derived from a macroscopic model of traffic flow developed by Lighthill and Whitham [7].

A large body of literature has applied game-theoretic concepts to horizontal queues, such as dynamic traffic assignment for user equilibria [8] and system optimal assignments [16]. Due to the complexity of modeling horizontal queues, approaches to solving the user equilibrium on general networks usually involve non-linear optimization techniques that limit the size of networks that can be considered. By restricting our analysis to parallel networks, we exploit the structure of the network to improve upon previous approaches to computing Nash equilibria.

C. Contributions of the article

We introduce a new class of latency functions, the HQSF latency class, that is expressive enough to model congestion on networks with horizontal queues, and study, for this class, the Nash equilibria of the routing game on a parallel network. This leads to novel results for characterizing and computing Nash equilibria:

- We show that there is no essential uniqueness of Nash equilibria (not all Nash equilibria have equal total costs), unlike point-queueing models usually considered in routing games [12].[1]
- We characterize the structure of the flow of the *best Nash equilibrium* (the Nash equilibrium that minimizes the total network latency) and show that this equilibrium can be computed in $O\left(N^2\right)$ time where N is the size of the network.
- We give the analytical expression of the *price of stability* on a two-link parallel network. This gives insight into

Walid Krichene is with the department of Electrical Engineering and Computer Sciences, University of California at Berkeley. walid@eecs.berkeley.edu

Jack Reilly is with the department of Civil and Environmental Engineering, University of California at Berkeley. jackdreilly@berkeley.edu

Saurabh Amin is with the department of Civil and Environmental Engineering, Massachusetts Institute of Technology. amins@mit.edu

Alexandre Bayen is with the department of Electrical Engineering and Computer Sciences, and the department of Civil and Environmental Engineering, University of California at Berkeley. bayen@berkeley.edu

[1] Under different modeling assumptions, similar non-uniqueness results exist for capacitated networks.[13]

978-1-4673-2066-5/12/$31.00 ©2012 IEEE

the qualitative behavior of congestion in Nash equilibria on networks with horizontal queues. We show in particular that when the lowest-latency link in a network is slightly above capacity, diverting a small amount of flow to a slower link can significantly decrease the total network latency.

These results provide a framework for efficient computation of Nash equilibria on parallel networks, which, in turn, give a high-level explanation of congestion patterns on such networks. While the assumption of a parallel network may seem restrictive, there are many examples of highway networks that can be accurately modeled by a parallel network connecting two highly populated areas (such as the San Francisco bay area).

D. Organization

We start by defining the model and introducing a new class of latency functions in Section II, and show as an example how such latency functions can be derived from known macroscopic models of traffic flow. In Section III, we study Nash equilibria of routing games for this new class of latency, and show that the essential uniqueness property does not hold. We then bound the number of Nash equilibria and give a tractable algorithm for computing the set of Nash equilibria. In Section IV we characterize in particular the best Nash equilibrium and give an explicit algorithm for its computation. In Section V, we study the inefficiency of best Nash equilibria using price of stability as measure of inefficiency.

II. The model

A. Routing games on on parallel edge network

We consider a non-atomic [14] routing game on a parallel network similar to the one studied in [11], shown in Figure 1. The network has a single source and a single sink. Connecting the source and sink are N parallel edges (or links) indexed by $n \in \{1, \dots, N\}$. The network is subject to a constant positive flow demand r at the source. We will denote by (N, r) an instance of the routing game on a network with N parallel links subject to demand r. A feasible flow assignment for the instance (N, r) is a vector $\boldsymbol{x} \in \mathbb{R}_+^N$ such that $\sum_{n=1}^N x_n = r$ where x_n is the flow on link n.

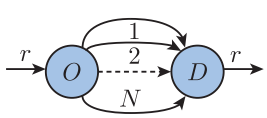


Fig. 1: Parallel network with N links, under demand r.

We then introduce a cost function, or latency function ℓ_n, on each link n. The cost on a link n can be thought of as the latency experienced by a job assigned to machine n in the case of job scheduling [11], or the travel time of a vehicle using road n in the case of traffic networks. In a routing game, every non-atomic player, represented by an infinitesimal amount of flow, chooses a route in order to minimize her/his individual latency [9], [12].

B. Modeling congestion with latency functions

To model the effects of queueing on a given link n, the latency ℓ_n on the link is typically assumed to be a non-decreasing function of the amount of flow x_n on link n [2], [3], [12]. While this class of latency accurately models congestion on a broad range of networks, such as communication networks, it fails to correctly model congestion for networks with horizontal queues. For example, consider a link (or road) n in a traffic network. A given flow x_n may correspond to two different scenarios: few vehicles on the road are moving quickly (the road is in *free-flow*), in which case the latency on the road is low, or a large number of vehicles on the road are moving slowly (the road is *congested*), in which case the latency on the road is high. This phenomenon is not captured if the latency is only a function of flow. One way to address this limitation is to introduce a binary state variable $m_n \in \{0, 1\}$ that specifies whether the link is in free-flow ($m_n = 0$) or in congestion ($m_n = 1$). The latency is then modeled to depend on flow x_n and congestion state m_n.

We next show that such latency functions can be derived from macroscopic models of flow on horizontal queuing networks.

C. Deriving latency functions for networks with horizontal queues

The relationship between the flow on a link and the *density* is usually expressed by a function called the flux function in the physical sciences and conservation law theory and fundamental diagram in traffic flow theory [4], [10]. Figure 2a shows an example of a triangular flux function that arises in traffic networks.

While such flow models have been popular for many decades in specific domains (such as traffic and fluid mechanics), less attention has been given to these models in the literature studying routing games, which focuses on modeling latency as a non-decreasing function of flow. In order to characterize Nash equilibria on horizontal queues, we develop a novel approach by introducing a new model of latency.

Consider a link n with length L_n, and assume the flow x_n on the link is given by a continuous function of density:

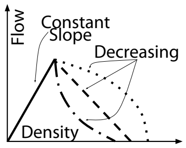
$$\begin{aligned} x_n^\rho : [0, \rho_n^{\max}] &\to \mathbb{R}_+ \\ \rho_n &\mapsto x_n = x_n^\rho(\rho_n) \end{aligned}$$

The function x_n^ρ maps density ρ_n to flow, is defined on the domain $[0, \rho_n^{\max}]$, and corresponds to the fundamental diagram of traffic. The latency is given by a function

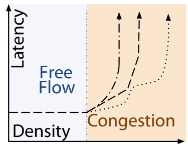
$$\begin{aligned} \ell_n^\rho : [0, \rho_n^{\max}] &\to \mathbb{R}_+ \\ \rho_n &\mapsto \ell_n^\rho(\rho_n) \end{aligned}$$

We observe that latency is related to flow and density through the relation:

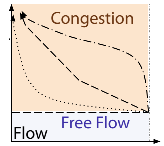
$$\ell_n^\rho(\rho_n) = \frac{L_n \rho_n}{x_n^\rho(\rho_n)}, \tag{1}$$



(a) Fundamental diagram examples.



(b) Latency as a function of density examples.



(c) Latency as a function of flow and state.

Fig. 2: Examples of flux functions and corresponding latency functions that satisfy the assumptions in Section II-C.

We make three assumptions on the flow and latency functions x_n^ρ and ℓ_n^ρ for horizontal queues, which are illustrated in Figure 2a.

1) There exists a critical density $\rho_n^{\text{crit}} > 0$, such that the latency is constant and minimal on the interval $\left[0, \rho_n^{\text{crit}}\right]$. This is equivalent to the flow function x_n^ρ being linear on that interval.
2) $\lim_{\rho_n \to \rho_n^{\max}} x_n^\rho = 0$ (the flow vanishes when the density approaches maximum density), and x_n^ρ is maximal at ρ_n^{crit}. The value of the flow at critical density is denoted $x_n^{\max} = x_n^\rho(\rho_n^{\text{crit}})$ and referred to as maximum flow or *capacity* of the link.
3) The latency function ℓ_n^ρ, is continuous, non-decreasing.

We define the *free-flow region* as $\rho_n \in \left[0, \rho_n^{\text{crit}}\right]$ and *congested region* as $\rho_n > \rho_n^{\text{crit}}$. These assumptions define a class of supported fundamental diagrams. Assumption 2 just states that $\forall \rho_n \geq 0$, $x_n^\rho\left(\rho_n\right) \in [0, x_n^{\max}]$. From Equation (1), Assumption 3 can be expressed equivalently in terms of the flow function x_n^ρ as $\frac{dx_n^\rho(\rho_n)}{d\rho_n} \leq \frac{x_n^\rho(\rho_n)}{\rho_n}$. This gives reasonable restrictions on the shape of fundamental diagrams in the congestion region, and flexible enough to include concave fundamental diagrams, and even some non-concave or non-decreasing functions x_n^ρ (as long as the conditions above are satisfied). Examples of allowable fundamental diagrams are given in Figure 2a, and corresponding examples of latency functions are given in Figure 2b. Note that from these assumptions, we can write the latency function for the horizontal queueing model as:

$$\ell_n^\rho\left(\rho_n\right) = \begin{cases} \frac{L_n \rho_n^{\text{crit}}}{x_n^{\max}} & \rho_n \in \left[0, \rho_n^{\text{crit}}\right] \\ \frac{L_n \rho_n}{x_n(\rho_n)} & \text{otherwise} \end{cases}$$

Example 1: Triangular fundamental diagrams

One particular class of fundamental diagrams x^ρ that satisfy the above assumptions are triangular fundamental diagrams [4], which are linear with positive slope v^f in the free-flow region, affine with negative slope v^c in the congestion region, and have maximum flow $x^{\max} = \rho^{\text{crit}} v^f$. Assumptions 1 and 2 are satisfied by definition, and Assumption 3 is satisfied since $\frac{dx(\rho)}{d\rho} = v^f = \frac{x(\rho)}{\rho}$ $\forall \rho \in [0, \rho^{\text{crit}}]$ and $\frac{dx(\rho)}{d\rho} = v^c \leq 0 \leq \frac{x(\rho)}{\rho}$ $\forall \rho \geq \rho^{\text{crit}}$. The dotted line in Figure 2a shows a triangular fundamental diagram. The latency function is then given by:

$$\ell_\triangle^\rho\left(\rho\right) = \begin{cases} \frac{L}{v^f} & 0 \leq \rho \leq \frac{x^{\max}}{v^f} \\ \frac{L\rho}{v^c(\rho - \rho^{\max})} & \frac{x^{\max}}{v^f} < \rho \leq \rho^{\max} \end{cases}$$

where $\rho^{\max} = x^{\max}\left(\frac{1}{v^f} - \frac{1}{v^c}\right)$.

D. A class of latency functions for horizontal queues

While expressing latency as a function of density is intuitive and succinct for horizontal queues, expressing it as a function of flow proves to be more convenient in the study of congestion games. This is largely due to the fact that total flow must be conserved in traffic assignment problems, and not density. For this reason, we introduce an equivalent formulation of latency using flow and congestion state. Let the congestion state m_n of link n be defined as:

$$m_n := \begin{cases} 0 & \text{if } n \text{ is in free-flow} \\ 1 & \text{if } n \text{ is congested} \end{cases}$$

We can now define a general class of latency functions ℓ_n as a function of both flow and congestion state:

$$\begin{aligned} D_n &\to \mathbb{R}_+ \\ (x_n, m_n) &\mapsto \ell_n\left(x_n, m_n\right), \end{aligned}$$

defined on $D_n = \ell_n : [0, x_n^{\max}] \times \{0\} \cup (0, x_n^{\max}) \times \{1\}$. Note that the latency in congestion $\ell_n(\cdot, 1)$ is defined on the open interval $(0, x_n^{\max})$. In particular, if $x_n = 0$ then $m_n = 0$ (an empty link is in free-flow) and if $x_n = x_n^{\max}$ then $m_n = 0$ (if a link is at maximum capacity, it is considered, by convention, to be in free-flow. It is in fact on the boundary of the free-flow and congestion regions, and we choose this convention to simplify the discussion). We also assume that ℓ_n satisfies the following properties, which are equivalent to the assumptions in Section II-C:

- The latency in free-flow is constant. Equivalently, $\forall x_n \in [0, x_n^{\max}]$, $\ell_n\left(x_n, 0\right) = a_n$, where a_n is the constant *free-flow latency*.
- $\lim_{x_n \to x_n^{\max}} \ell_n\left(x_n, 1\right) = \ell_n\left(x_n^{\max}, 0\right) = a_n$
- $x_n \mapsto \ell_n\left(x_n, 1\right)$ is decreasing from $(0, x_n^{\max})$ onto $(a_n, +\infty)$.

One interesting result is that the latency under congestion $\ell(x, 1)$ is a *decreasing* function offlow. Intuitively, as the link becomes more congested, agents slow down, so their latency increases, and the amount of flow on the link decreases. Some examples of latency functions in this class are illustrated in Figure 2c. Again, the latency function corresponding to a triangular fundamental diagram can be readily expressed in this form:

$$\ell_\triangle\left(x, 0\right) = \frac{L}{v^f}$$
$$\ell_\triangle\left(x, 1\right) = L\left(\frac{\rho^{\max}}{x} + \frac{1}{v^c}\right)$$

E. Total System Cost

If a non-atomic player chooses link n, the latency experienced by the player is $\ell_n(x_n, m_n)$. Therefore, the total cost experienced by all players on link n is $C_n\left(x_n, m_n\right) = \ell_n\left(x_n, m_n\right) x_n = L_n \rho_n$. Then, the total system cost is the sum of the costs of the individual links $C\left(\boldsymbol{x}, \boldsymbol{m}\right) = \sum_{n=1}^N C_n\left(x_n, m_n\right)$, where $\boldsymbol{x} = (x_1, \ldots, x_N)$ is the vector of flows, and $\boldsymbol{m} = (m_1, \ldots, m_N)$ is the vector of congestion states for the entire network.

III. Nash Equilibria

In this section, we characterize pure non-atomic Nash equilibria of the network (also called Wardrop equilibria in the transportation literature), which we simply refer to as Nash equilibria.

A. Characterization of Nash Equilibria

We first recall the fundamental notion of Nash equilibrium for the routing game instance (N, r) [12], [9].

Definition 1: Nash Equilibrium

An assignment $(\boldsymbol{x}, \boldsymbol{m}) \in \mathbb{R}_+^N \times \{0,1\}^N$ for the routing game instance (N, r) is a Nash equilibrium if $\forall n$

$$x_n > 0 \Rightarrow \forall k \in \{1, \dots, N\},\ \ell_n(x_n, m_n) \leq \ell_k(x_k, m_k)$$

In particular, every non-atomic agent cannot improve her/his latency by switching to another link. As a consequence, all links that are in the support of x have the same latency ℓ_0, and links that are not in the support have latency greater than or equal to ℓ_0. We will denote by $\mathrm{Supp}\,(\boldsymbol{x})$ the support of $\boldsymbol{x}$, i.e. the set $\{n \in \{1, \dots, N\} \,|x_n > 0\}$.

Note that to fully describe the equilibrium, one needs to specify the congestion state vector m in addition to the flow assignment x, since the latency on a link depends on whether the link is congested or not. The following lemma gives an equivalent characterization of Nash equilibria.

Lemma 1: Characterization of a Nash Equilibrium

A feasible assignment $(\boldsymbol{x}, \boldsymbol{m})$ for a routing game instance (N, r) is a Nash equilibrium if and only if $\exists\, \ell_0 > 0$ such that

$$x_n > 0 \Rightarrow \ell_n(x_n, m_n) = \ell_0$$
$$x_n = 0 \Rightarrow \ell_n(0, 0) \geq \ell_0$$

The total latency incurred by the network is $C(\boldsymbol{x}, \boldsymbol{m}) = r\ell_0$. Note that links that have zero flow are necessarily in free-flow $x_n = 0 \Rightarrow m_n = 0$.

B. Multiple Nash equilibria on networks with horizontal queues

Let $\mathrm{NE}\,(N, r)$ denote the set of Nash Equilibria for routing game instance (N, r). For our class of latency functions, the essential uniqueness property of Nash equilibrium [12] does not hold. To see this, consider for example a routing game instance (N=2, r=1) where $x_1^{\max} = x_2^{\max} = 1$, the free-flow latencies are $a_1 = 1$ and $a_2 = 2$, and the congested latency functions are given respectively by $\ell_1(x_1, 1) = \frac{1}{x_1}$ and $\ell_2(x_2, 1) = \frac{2}{x_2}$. Then it is easy to see that $(\boldsymbol{x}, \boldsymbol{m}) = ((1, 0), (0, 0))$, $(\boldsymbol{x}', \boldsymbol{m}') = ((\frac{1}{2}, \frac{1}{2}), (1, 0))$, and $(\boldsymbol{x}'', \boldsymbol{m}'') = ((\frac{1}{3}, \frac{2}{3}), (1, 1))$ are all Nash equilibria for this instance, and they have different costs: $C(\boldsymbol{x}, \boldsymbol{m}) = 1$, $C(\boldsymbol{x}', \boldsymbol{m}') = 2$ and $C(\boldsymbol{x}'', \boldsymbol{m}'') = 3$. This simple example shows that there are at least two types of Nash equilibria: equilibria for which every link in the support is congested (this is the case for $(\boldsymbol{x}'', \boldsymbol{m}'')$ in the previous example), and equilibria that have one link in free-flow in their support (this is the case for both $(\boldsymbol{x}, \boldsymbol{m})$ and $(\boldsymbol{x}', \boldsymbol{m}')$). In this section, we show that these are in fact the only possible types of equilibria. To simplify the discussion, we assume without loss of generality, that the links are ordered by increasing free-flow latencies, and that free-flow latencies are different to avoid degenerate cases where the set of Nash equilibria is infinite:

Assumption: $(a_1 < a_2 < \ldots < a_N)$.

We start by deriving properties that the congestion state vector $\boldsymbol{m}$ needs to satisfy for a Nash equilibrium $(\boldsymbol{x}, \boldsymbol{m})$.

Lemma 2: Congestion of lower links

Let $(\boldsymbol{x}, \boldsymbol{m}) \in \mathrm{NE}\,(N, r)$. Then

$$j \in \mathrm{Supp}\,(\boldsymbol{x}) \Rightarrow m_i = 1 \quad \forall i \in \{1, \dots, j-1\}$$

Proof: Let $i \in \{1, \dots, j-1\}$. Then $m_i = 0 \Rightarrow \ell_i(x_i, m_i) = a_i < a_j \leq \ell_j(x_j, m_j)$, which violates the characterization of Nash equilibrium in Lemma 1. Therefore, $m_i = 1 \quad \forall i \in \{1, \dots, j-1\}$. ■

Corollary 1: Congestion states under equilibrium

Let $(\boldsymbol{x}, \boldsymbol{m}) \in \mathrm{NE}\,(N, r)$. Assume that $\exists j \in \mathrm{Supp}\,(x)$ such that $m_j = 0$. Then $\boldsymbol{m} = (1, \dots, \overset{j-1}{1}, \overset{j}{0}, \dots, 0)$ and $\mathrm{Supp}\,(\boldsymbol{x}) = \{1, \dots, j\}$.

Proof: We have from Lemma 2 that $\forall i \in \{1, \dots, j-1\}$, $m_i = 1$. And we have $\forall i \in \{j+1, \dots, N\}$, $\ell_i(x_i, m_i) \geq a_i$ by definition of the latency function, and $a_i > a_j$ since $i > j$. Therefore the latency on link $i \in \{j+1, \dots, N\}$ is strictly greater than the latency on link $j \in \mathrm{Supp}\,(\boldsymbol{x})$, therefore $i \notin \mathrm{Supp}\,(\boldsymbol{x})$ (follows from the characterization of Nash equilibrium in Lemma 1) and $m_i = 0$. ■

The corollary states that if some link j in the support of a Nash equilibrium is in free-flow, this completely determines the congestion state vector of the equilibrium: links $\{1, \dots, j-1\}$ are in the support and are congested, and links $\{j+1, \dots, N\}$ are not in the support. We will call such Nash equilibria (where a single link in the support is in free-flow) *single-link-free-flow equilibri*a. In general a Nash equilibrium does not necessarily have a link in free-flow: this defines a second type of equilibria where all links in the support are congested, i.e. $m_{\max \mathrm{Supp}(\boldsymbol{x})} = 1$. We will call such equilibria *congested equilibria*.

Lemma 3: Enumerating Nash Equilibria

For a given congestion state $\boldsymbol{m}$, there are at most two flow assignments $\boldsymbol{x}$ such that $(\boldsymbol{x}, \boldsymbol{m})$ is a Nash equilibrium: one single-link-free-flow equilibrium and one congested equilibrium.

The proof requires a long argument and is not presented here. For a a detailed proof, we refer the reader to [6].

Lemma 3 shows that there are at most $2N$ Nash equilibria for the instance (N, r): N single-link-free-flow equilibria, corresponding to congestion states $\boldsymbol{m} = (0, \dots, 0)$, $\boldsymbol{m} = (1, 0, \dots, 0)$, …, $\boldsymbol{m} = (1, \dots, 1, 0)$, and N congested equilibria, corresponding to congestion states $\boldsymbol{m} = (1, 0, \dots, 0)$, …, $\boldsymbol{m} = (1, \dots, 1)$. Next, we characterize single-link-free-flow equilibria.

C. Single link free-flow Equilibria

Consider a Nash equilibrium $(\boldsymbol{x}, \boldsymbol{m})$ and let $k = \max\,[\mathrm{Supp}\,(\boldsymbol{x})]$. Assume $m_k = 0$ (i.e. $(\boldsymbol{x}, \boldsymbol{m})$ is a free-flow Nash equilibrium). We have from Corollary 1 that links $\{1, \dots, k-1\}$ are congested and link k is in free-flow.

Therefore we must have $\forall n \in \{1,\ldots,k-1\}$, $\ell_n(x_n,1) = \ell_k(x_k,0) = a_k$. This uniquely determines the flow on congested links $n \in \{1,\ldots,k-1\}$. We define this flow to be $\hat{x}_n(k)$. More precisely,

Definition 2: Congestion flow

For $1 \le n < k \le N$, the congestion flow $\hat{x}_n(k)$ is defined as the unique flow in $(0, x_n^{\max})$ that satisfies $\ell_n(\hat{x}_n(k), 1) = a_k$.

Proposition 1: Congestion Flows are decreasing

$$\hat{x}_n(k) = \ell_n(\cdot,1)^{-1}(a_k) \tag{2}$$

is a decreasing function of k since a_k is increasing in k and $\ell_n(\cdot,1)^{-1}$ is decreasing.

We can now characterize single-link-free-flow equilibria. All single link free-flow equilibria are of the form $\left(\bar{\boldsymbol{x}}^{k,r}, \bar{\boldsymbol{m}}^k\right)$ where

$$\bar{\boldsymbol{m}}^k := \big(1,\ldots,\overset{k-1}{1},\overset{k}{0},\ldots,0\big) \tag{3}$$

$$\bar{\boldsymbol{x}}^{k,r} := \left(\hat{x}_1(k),\ldots,\hat{x}_{k-1}(k), r - \sum_{n=1}^{k-1}\hat{x}_n(k), 0,\ldots,0\right) \tag{4}$$

Illustrations of Equations (2), (3) and (4) are shown in Figure 3.

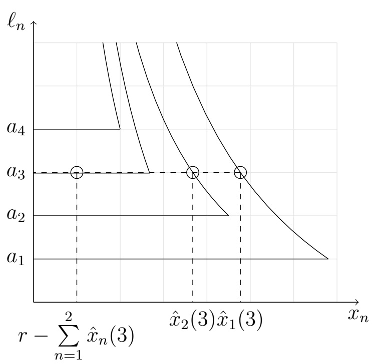


Fig. 3: Example of a single-link-free-flow equilibrium $(\bar{\boldsymbol{x}}^{3,r}, \bar{\boldsymbol{m}}^3)$. Links 1 and 2 are congested, link 3 is in free-flow, and link 4 is empty.

Proposition 2: Single link free-flow Nash Equilibria

If $\bar{\boldsymbol{x}}^{k,r}$ is a feasible assignment, i.e. $r - \sum_{n=1}^{k-1}\hat{x}_n(k) \in [0, x_k^{\max}]$, then $\left(\bar{\boldsymbol{x}}^{k,r}, \bar{\boldsymbol{m}}^k\right)$ is a Nash Equilibrium for the instance (N, r).

Proof: Follows directly from the definitions in Equatinos (2), (3) and (4). ■

D. Existence of a single-link free-flow Nash Equilibrium

From Proposition 2, we have a simple characterization of single-link-free-flow equilibria. Next, we show that if the set of Nash equilibria is non-empty, then it contains a single-link-free-flow equilibrium.

Lemma 4: Existence of a single-link-free-flow Nash equilibrium

Consider instance (N, r). If the set of Nash equilibria is non empty, $\mathrm{NE}(N,r) \neq \emptyset$, then there exists a single-link-free-flow Nash equilibrium $\left(\bar{\boldsymbol{x}}^{j,r}, \bar{\boldsymbol{m}}^j\right) \in \mathrm{NE}(N,r)$ for some $j \le N$.

Proof: We first observe that for a network of N links, $\mathrm{NE}(N,r) \neq \emptyset$ only if $r \le \max_{k\in\{1,\ldots,N\}}\left\{x_k^{\max} + \sum_{n=1}^{k-1}\hat{x}_n(k)\right\}$. We denote this quantity with $r^{\mathrm{NE}}(N)$. Therefore, from Lemma 2, it suffices to show the following property:

$\mathbf{P}_N$: $\forall r \in \left[0, r^{\mathrm{NE}}(N)\right]$, there exists a single-link-free-flow Nash equilibrium for the instance (N, r).

We show $\mathbf{P}_N$ by induction on N, the size of the network. For $N = 1$, it is clear that if $0 \le r \le x_1^{\max}$, there is a single-link free-flow equilibrium $(\boldsymbol{x}, \boldsymbol{m}) = (r, 0)$.

Now let $N \ge 1$, assume $\mathbf{P}_N$ is true and let us show $\mathbf{P}_{N+1}$. Let $0 \le r \le r^{\mathrm{NE}}(N+1)$ and consider a routing game instance $(N+1, r)$. ■

Case 1: If $r \le r^{\mathrm{NE}}(N)$, then by the induction hypothesis there exists a single-link-free-flow Nash equilibrium $(\boldsymbol{x}, \boldsymbol{m})$ for the instance (N, r). Then assignment $(\boldsymbol{x}', \boldsymbol{m}')$ defined as $\boldsymbol{x}' = (x_1,\ldots,x_N,0)$ and $\boldsymbol{m}' = (m_1,\ldots,m_N,0)$ is clearly a single-link free-flow Nash equilibrium for the instance $(N+1, r)$.

Case 2: If $r^{\mathrm{NE}}(N) < r \le r^{\mathrm{NE}}(N+1)$ then we can show that $\left(\bar{\boldsymbol{x}}^{N+1,r}, \bar{\boldsymbol{m}}^{N+1}\right) \in \mathrm{NE}(N+1,r)$. From Proposition 2, we only need to show that

$$0 \le r - \sum_{n=1}^{N}\hat{x}_n(N+1) \le x_{N+1}^{\max}. \tag{5}$$

First, we note that since $r^{\mathrm{NE}}(N) < r^{\mathrm{NE}}(N+1)$, then $r^{\mathrm{NE}}(N+1) = x_{N+1}^{\max} + \sum_{n=1}^{N}\hat{x}_n(N+1)$, thus from $r < r^{\mathrm{NE}}(N+1)$, we have $r \le x_{N+1}^{\max} + \sum_{n=1}^{N}\hat{x}_n(N+1)$ which proves the second inequality of (5). To show the first inequality, we have

$$\begin{aligned}
r &\ge x_N^{\max} + \sum_{n=1}^{N-1}\hat{x}_n(N) && \text{since } r^{\mathrm{NE}}(N) < r \\
&\ge x_N^{\max} + \sum_{n=1}^{N-1}\hat{x}_n(N+1) && \text{since } \hat{x}_n(N) \ge \hat{x}_n(N+1) \\
&\ge \hat{x}_N(N+1) + \sum_{n=1}^{N-1}\hat{x}_n(N+1) && \text{since } x_N^{\max} \ge \hat{x}_N(N+1)
\end{aligned}$$

which achieves the induction.

Corollary 2: Cost of single-link-free-flow Equilibria

If there exists a congested equilibrium $(\boldsymbol{x}, \boldsymbol{m}) \in \mathrm{NE}(N,r)$, then there exists a single-link free-flow equilibrium $(\boldsymbol{x}', \boldsymbol{m}')$ with lower cost.

Proof: Let $(\boldsymbol{x}, \boldsymbol{m}) \in \mathrm{NE}(N,r)$ be a congested equilibrium, i.e. $m_k = 1$ where $k = \max \mathrm{Supp}(\boldsymbol{x})$. Then we have $r \le x_k^{\max} + \sum_{n=1}^{k-1}\hat{x}_n(k)$ and by the property $\mathbf{P}_k$, there exists a single-link free-flow equilibrium $(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{m}}) \in \mathrm{NE}(k,r)$, and the cost of this equilibrium is $C(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{m}}) \le a_k r$. But this also provides a single-link free-flow equilibrium $(\boldsymbol{x}', \boldsymbol{m}')$ for the original instance (N, r) defined by $\boldsymbol{x}' = (\tilde{x}_1,\ldots,\tilde{x}_k,0,\ldots,0)$ and $\boldsymbol{m}' = (\tilde{m}_1,\ldots,\tilde{m}_k,0,\ldots,0)$, and $C(\boldsymbol{x}', \boldsymbol{m}') = C(\tilde{\boldsymbol{x}}, \tilde{\boldsymbol{m}}) \le a_k r$. To conclude, we simply note that the cost of the congested equilibrium is $C(\boldsymbol{x}, \boldsymbol{m}) = \ell_k(x_k,1)r > a_k r$, thus $C(\boldsymbol{x}, \boldsymbol{m}) > C(\boldsymbol{x}', \boldsymbol{m}')$. ■

IV. Best Nash Equilibria

A. Determining minimum cost Nash equilibria

In order to study the inefficiency of Nash equilibria, we focus our attention on *best Nash equilibria* and *price of stability* as a measure of their inefficiency (see for example [1]).

Definition 3: Best Nash Equilibrium
The set of Best Nash Equilibria (BNE) is the set of minimizers of the total cost function

$$\text{BNE}(N,r) = \underset{(\boldsymbol{x},\boldsymbol{m})\in\text{NE}(N,r)}{\arg\min} \; C(\boldsymbol{x},\boldsymbol{m})$$

We show that the minimizer is, in fact, unique, and that it satisfies some properties, given in the following Theorem.

Theorem 1: Characterization of Best Nash Equilibria
For a routing game instance (N,r), if the set of Nash equilibria is nonempty, then there exists a unique best Nash equilibrium, and it is the single-link free-flow equilibrium that has smallest support

$$\text{BNE}(N,r) = \underset{(\boldsymbol{x},\boldsymbol{m})\in\text{NE}_\text{f}(N,r)}{\arg\min} \; \{\max[\text{Supp}(x)]\}$$

Since the best Nash equilibrium is unique, we will, with a slight abuse of notation, identify the set $\text{BNE}(N,r)$ with its unique element. *Proof:* From Corollary 2 we have that if $(\boldsymbol{x},\boldsymbol{m}) \in \text{NE}(N,r)$ is a congested equilibrium, then these exists a single-link free-flow equilibrium with lower cost. Therefore the Best Nash Equilibrium is a single-link free-flow equilibrium. To show that the BNE has smallest support, we simply note that if $(\boldsymbol{x},\boldsymbol{m}) \in \text{NE}_\text{f}(N,r)$ is a single-link free-flow equilibrium and $k = \max \text{Supp}(x)$, then its cost is $C(\boldsymbol{x},\boldsymbol{m}) = a_k r$. Note that uniqueness is immediate since two single-link free-flow equilibria $(\boldsymbol{x},\boldsymbol{m})$ and $(\boldsymbol{x}',\boldsymbol{m}')$ that have the same support, hence the same congestion state $\boldsymbol{m} = \boldsymbol{m}'$, coincide by Lemma 3. ∎

B. Computational complexity of finding Best Nash Equilibria

In this section, we present a constructive algorithm for finding the best Nash equilibrium of a routing game instance (N,r) and then show the running time to be in $O(N^2)$.

In Algorithm 1, subroutine `freeFlowConfig` outputs a candidate *single-link-free-flow* assignment for the instance (N,r), such that link i is the last link in the support (Equation (4)). Starting with link 1 in free-flow, `bestNE` checks if the output of `freeFlowConfig` is a feasible assignment. If this is the case, the candidate assignment is the Best Nash Equilibrium, and `bestNE` terminates. If not, the free-flow link index is incremented by one, and the process is repeated until either a feasible assignment is found, or the number of links exceeds N, in which case no Nash equilibrium exists.

We first note that we can precompute $\hat{x}_i(k)\, \forall 1 \le i < k \le N)$ in $O(N^2)$. The subroutine `freeFlowConfig` runs in $O(N)$ time. Finally, for each loop of the `bestNE` outer routine (with N iterations), the running time is a constant plus the running time of `freeFlowConfig`. Therefore, the overall running time of the algorithm is $O(N^2)$.

Algorithm 1 Best Nash Equilibrium

```
procedure bestNE(N, r)
Inputs: Size of the network N, demand r
Outputs: Best Nash equilibrium (x, m)
for k ∈ {1, ..., N}:
    let (x, m) = freeFlowConfig(N, r, k)
    if x_i ∈ [0, x_i^max]:
        return (x, m)
return No-Solution

procedure freeFlowConfig(N, r, k)
Inputs: Size of the network N,
        demand r, free-flow link index k
Outputs: assignment (x, m) = (x̄^{r,k}, m̄^k)  (Eq. (6) and (7))
for i ∈ {1, ..., N}:
    if i < k:
        x_i = x̂_i(k), m_i = 1          (x̂_i(k) defined in Eq. (2))
    elseif i == k:
        x_i = r − Σ_{n=1}^{k} x_n, m_i = 0
    else:
        x_i = 0, m_i = 0
return (x, m)
```

V. Inefficiency of Best Nash Equilibria

To study the inefficiency of Nash equilibria, in particular of the best Nash equilibrium, we use price of stability as a measure of inefficiency [1]. Price of stability is defined as the ratio between the cost of the best Nash Equilibrium and the social optimal cost. First we give an overview of social optimum for our model. Then we consider a simple two link parallel network and derive the price of stability for a triangular fundamental diagram. This example illustrates in particular the dependency of the price of stability on the flow demand and the free-flow latencies.

A. Social Optima

Consider an instance (N,r) where the flow demand r does not exceed the maximum capacity of the network $r \le \sum_n x_n^{\max}$. Since the total cost function is $C(\boldsymbol{x},\boldsymbol{m}) = \sum_{n=1}^{N} x_n \ell_n(x_n,m_n)$, the social optimum of the network is a solution to the optimization problem: $\min_{\boldsymbol{x},\boldsymbol{m}} \sum_n x_n \ell_n(x_n,m_n)$ such that $\sum_n x_n = r$. It can be shown that a socially optimal flow is necessarily in free flow on all links, which leads to an equivalent linear program: $\min_{\boldsymbol{x},\boldsymbol{m}} \sum_n x_n a_n$ such that $\sum_n x_n = r$. Then, since the links are ordered by increasing free-flow latency $a_1 < \cdots < a_N$, the social optimum is simply given by the assignment that saturates most efficient links first. Formally, if $k_0 = \max\{k | r > \sum_{n=1}^{k} x_n^{\max}\}$ then the social optimal assignment $\boldsymbol{x}^\star$ is:

$$\boldsymbol{x}^\star = \left(x_1^{\max}, \dots, x_{k_0}^{\max}, r - \sum_{n=1}^{k_0} x_n^{\max}, 0, \dots, 0\right) \quad (6)$$

B. Price of Stability on a Two-Link Network

Consider a routing game instance $(2,r)$ such that $a_1 < a_2$ and $x_2^{\max} + \hat{x}_1(2) > x_1^{\max}$. Let $\text{BNE}(2,r) = (\boldsymbol{x}_{\text{BNE}}(r), \boldsymbol{m}_{\text{BNE}}(r))$ be the best Nash equilibrium and $(\boldsymbol{x}^\star(r), 0)$ be the social optimum, as defined by (6). The price of stability is then defined as

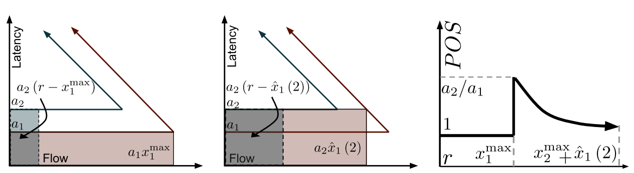
$$\text{POS}(N,r) = \frac{C(\boldsymbol{x}_{\text{BNE}}(r), \boldsymbol{m}_{\text{BNE}}(r))}{C(\boldsymbol{x}^\star(r), \boldsymbol{0})}$$

a) Case 1: $0 \leq r \leq x_1^{\max}$*:* Using (6), all the demand will be on link 1 in free-flow. Similarly, from Theorem 1 we have that since link 1 can accommodate r in free-flow and the support cannot be smaller than a single link, then $\text{BNE}\,(2, r)$ has all flow demand on link 1 in free-flow, and is coincides with the social optimum. In this case, the price of stability is equal to 1.

b) Case 2: $x_1^{\max} < r \leq x_2^{\max} + \hat{x}_1(2)$*:* We know that all flow demand cannot be accommodated by link 1. From Equation (6), the social optimum assignment is given by $\boldsymbol{x}^\star(r) = (x_1^{\max}, r - x_1^{\max})$. From Theorem 1 we have that $\text{BNE}\,(2, r)$ has a single link in free-flow. Since the total demand exceeds the capacity of link 1, then under a best Nash equilibrium, link 2 is in free-flow, and link 1 is congested. Therefore $\boldsymbol{m}_{\text{BNE}}(r) = (1, 0)$. The corresponding flow $\boldsymbol{x}_{\text{BNE}}(r)$ is $(\hat{x}_1(2), r - \hat{x}_1(2))$. The comparison of the social optimum and Nash equilibrium assignments are depicted in Figure 4. Then it can be shown that

$$\text{POS}\,(2, r > x_1^{\max}) = \left(1 - \frac{x_1^{\max}}{r}\left(1 - \frac{a_1}{a_2}\right)\right)^{-1}$$

In this simple two-link parallel network, the price of stability is maximal at $r = (x_1^{\max})^+$ and equal to a_2/a_1 (Figure 4c). This shows in particular that for the general class of horizontal queuing congestion latencies, the price of stability is unbounded, since for any demand r and any positive constant A, we can design an instance $(2, r)$ such that the price of stability is $\frac{a_2}{a_1} > A$.



(a) Social optimum (b) Nash equilibrium (c) POS as a function of demand.

Fig. 4: Visualization of POS on two-link network. Differences in flow assignments between social optimum and Nash equilibrium are shown in 4a and 4b. The area of the shaded regions in 4a,4b are the total costs attributed to each link. In 4c, the flat region corresponds to $r \leq x_1^{\max}$ (Case 1) and the decreasing region to $r > x_1^{\max}$ (Case 2).

We also observe that for a fixed flow demand $r > x_1^{\max}$, the price of stability is an increasing function of $\frac{a_2}{a_1}$. And as $a_2 \to a_1$, the price of stability goes to 1. Intuitively, the inefficiency of Nash equilibria can be directly attributed to the difference in free-flow latency between the links.

Additionally, as the demand $r > x_1^{\max}$ increases, the price of stability decreases. This occurs because the difference in total latency between social optimum and Nash equilibrium is constant for $r > x_1^{\max}$.

This also shows that selfish routing is most costly when a free-flow link is near maximum capacity (note the discontinuity in total latency for Nash equilibrium that occurs when demand exceeds the capacity of the first link $r > x_1^{\max}$). If a controller were to anticipate a scenario where demand was slightly above this capacity, they could dramatically reduce the inefficiency of the Nash equilibrium by rerouting a small fraction of the flow and keeping the link in free-flow.

VI. Conclusion

We introduced a new class of latency functions that models congestion in horizontal queuing networks, and studied the resulting Nash equilibria for non-atomic congestion games on parallel networks. We showed the essential uniqueness property does not hold for this new class, and that there may be up to $2N$ equilibria for a routing game instance (N, r). Then we focused our attention on the best Nash equilibrium $\text{BNE}\,(N, r)$, which we proved is the single-link-free-flow equilibrium with smallest support, and then presented a constructive, quadratic time algorithm for finding this equilibrium. Finally, we derived price of stability results for a two-link network, then showed that if a link is anticipated to be near capacity, congestion can be completely averted by diverting only a small fraction of the demand.

References

[1] E. Anshelevich, A. Dasgupta, J. Kleinberg, E. Tardos, T. Wexler, and T. Roughgarden. The Price of Stability for Network Design with Fair Cost Allocation. *45th Annual IEEE Symposium on Foundations of Computer Science*, pages 295–304, 2004.

[2] A. Aswani and C. Tomlin. Game-theoretic routing of GPS-assisted vehicles for energy efficiency. In *American Control Conference (ACC), 2011*, pages 3375–3380. IEEE, 2011.

[3] M. Babaioff, R. Kleinberg, and C. H. Papadimitriou. Congestion games with malicious players. *Games and Economic Behavior*, 67(1):22–35, 2009.

[4] C. F. Daganzo. The cell transmission model: A dynamic representation of highway traffic consistent with the hydrodynamic theory. *Transportation Research Part B: Methodological*, 28(4):269–287, 1994.

[5] F. P. Kelly. Mathematical modelling of the Internet. *Mathematics unlimited-2001 and beyond*, pages 685–702, 2001.

[6] W. Krichene, J. Reilly, S. Amin, and Bayen A. M. Stackelberg routing on parallel networks with horizontal queues, http://www.eecs.berkeley.edu/~walid/papers/stack.pdf. 2012.

[7] M. J. Lighthill and G. B. Whitham. On kinematic waves. II. A theory of traffic flow on long crowded roads. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 229(1178):317, 1955.

[8] H. K. Lo and W. Y. Szeto. A cell-based variational inequality formulation of the dynamic user optimal assignment problem. *Transportation Research Part B: Methodological*, 36(5):421–443, 2002.

[9] C. Papadimitriou and G. Valiant. A new look at selfish routing. *Innovations in Computer Science (ICS)*, 2010.

[10] M. Papageorgiou, J. M. Blosseville, and H. Hadj-Salem. Macroscopic modelling of traffic flow on the Boulevard Périphérique in Paris. *Transportation Research Part B: Methodological*, 23(1):29–47, 1989.

[11] T. Roughgarden. Stackelberg scheduling strategies. In *Proceedings of the thirty-third annual ACM symposium on Theory of computing*, pages 104–113. ACM, 2001.

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

[13] A. S. Schulz and N. E. Stier-Moses. *On the performance of user equilibria in traffic networks*, pages 86–87. Society for Industrial and Applied Mathematics, 2003.

[14] D. Sehmeidler. Equilibrium points of nonatomic games. *Journal of Statistical Physics*, 7(4):295–300, 1973.

[15] D. B. Work, S. Blandin, O. P. Tossavainen, B. Piccoli, and A. M. Bayen. A traffic model for velocity data assimilation. *Applied Mathematics Research eXpress*, 2010(1):1, April 2010.

[16] A. K. Ziliaskopoulos. A linear programming model for the single destination system optimum dynamic traffic assignment problem. *Transportation science*, 34(1):37, 2000.