Computational Game Theory **Spring Semester, 2003/4**

Lecture 3: Coordination Ratio of Selfish Routing

Lecturer: Yishay Mansour *Scribe: Anat Axelrod, Eran Werner*

3.1 Lecture Overview

In this lecture we consider the problem of routing traffic to optimize the performance of a congested and unregulated network. We are given a network, a rate of traffic between each pair of nodes and a latency function specifying the time needed to traverse each edge given its congestion. The goal is to route traffic while minimizing the *total latency*. In many situations, network traffic cannot be regulated, thus each user minimizes his latency by choosing among the available paths with respect to the congestion caused by other users. We will see that this "selfish" behavior does not perform as well as an optimized regulated network.
We start by exploring the characteristics of *Nash equilibrium* and *minimal latency optimal flow* to investigate the *coordination ratio*. We prove that if the latency of each edge is a linear function of its congestion, then the coordination ratio of selfish routing is at most 4/3. We also show that if the latency function is only known to be continuous and nondecreasing in the congestion, then there is no bounded coordination ratio; however, we prove that the total latency in such a network is no more than the total latency incurred by optimally routing twice as much traffic on the same network.

3.2 Introduction

We shall investigate the problem of routing traffic in a network. The problem is defined as follows: Given a rate of traffic between each pair of nodes in a network find an assignment of the traffic to paths so that the total latency is minimized. Each link in the network is associated with a latency function which is typically load-dependent, i.e. the latency increases as the link becomes more congested.
In many domains (such as the internet or road networks) it is impossible to impose regulation of traffic, and therefore we are interested in those settings where each user acts according to his own selfish interests. We assume that each user will always select the minimum latency path to its destination. In other words, we assume all users are rational and non malicious. This can actually be viewed as a noncooperative game where each user plays the best response given the state of all other users, and thus we expect the routes chosen to form a Nash equilibrium.
The network contains a numerous amount of users where each user holds only a negligible portion of the total traffic. Alternatively, we can think of a model with a finite

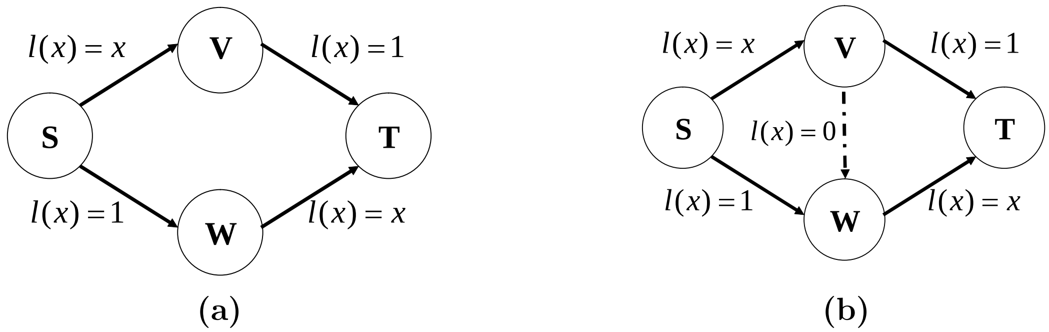

Figure 3.1:

number of users that are allowed to split their load between different paths. Our target function is to minimize the average (or total) latency suffered by all users. We will compare the overall performance under a Nash equilibrium against the theoretically optimal performance of a regulated network.

Before we continue, let's examine an example setting which has inspired much of the work in this traffic model. Consider the network in Figure 3.1(a). There are two disjoint paths from S to T. Each path follows exactly two edges. The latency functions are labelled on the edges. Suppose one unit of traffic needs to be routed from S to T. The optimal flow coincides with the Nash equilibrium such that half of the traffic takes the upper path and the other half takes the lower path. In this manner, the latency perceived by each user is $\frac{3}{2}$. In any other nonequal distribution of traffic among the two paths, there will be a difference in the total latency of the two paths and users will be motivated to reroute to the less congested path.

Note *Incidentally, we will soon realize that in any scenario in which the flow at Nash is split over more than a single path, the latency of all the chosen paths must be equal.*

Now, consider Figure 3.1(b) where a fifth edge of latency zero is added to the network. While the optimum flow has not been affected by this augmentation, Nash will only occur by routing the entire traffic on the single S $\rightarrow$ V $\rightarrow$ W $\rightarrow$ T path, hereby increasing the latency each user experiences to 2. Amazingly, adding a new zero latency link had a negative effect for all agents. This counter-intuitive impact is known as *Braess's paradox*.

Anecdote 1 *Two live and well known examples of Braess's paradox occurred when 42nd street was closed in New York City and instead of the predicted traffic gridlock, traffic flow actually improved. In the second case, traffic flow worsened when a new road was constructed in Stuttgart, Germany, and only improved after the road was torn up.*

3.2.1 The Model - Formal Definition

- We consider a directed graph $\boldsymbol{G = (V,E)}$ with $\boldsymbol{k}$ pairs $\boldsymbol{(s_i, t_i)}$ of source and destination vertices.
- r_i - The amount of flow required between s_i and t_i.
- P_i - The set of simple paths connecting the pair (s_i, t_i). $\mathcal{P} = \bigcup_i P_i$.
- Flow $\boldsymbol{f}$ - A function that maps a path to a positive real number. Each path P is associated with a flow f_P.
- f_e - The flow on edge e defined for a fixed flow function. $f_e = \sum_{P:e \in P} f_P$.
- A flow f is said to be *feasible* if $\forall i, \ \sum_{P \in P_i} f_P = r_i$.
- Each edge $e \in E$ is given a load-dependent latency function denoted $\ell_e(\cdot)$. We restrict our discussion to nonnegative, differentiable and nondecreasing latency functions.
- (G, r, ℓ) - A triple which defines an *instance* of the routing problem.
- The latency of a path ℓ_P is defined as the sum of latencies of all edges in the path. $\ell_P(f) = \sum_{e \in P} \ell_e(f_e)$.
- $\boldsymbol{C(f)}$ - The total latency, also defined as the *cost* of a flow *f*. $C(f) = \sum_{P \in \mathcal{P}} \ell_P(f) f_P$. Alternatively, we can accumulate over the edges to get $C(f) = \sum_{e \in E} \ell_e(f_e) f_e$.

3.3 Characterizations of Nash & OPT Flows

3.3.1 Flows at Nash Equilibrium

Lemma 3.3.1 *A feasible flow f for instance (G, r, ℓ) is at Nash equilibrium iff for every $i \in \{1, ..., k\}$ and $P_1, P_2 \in P_i$ with $f_{P_1} > 0$, $\ell_{P_1}(f) \leq \ell_{P_2}(f)$.*

From the lemma it follows that flow at Nash equilibrium will be routed only on best response paths. Consequently, all paths assigned with a positive flow between (s_i, t_i) have equal latency denoted by $L_i(f)$.

Corollary 3.1 *If f is a flow at a Nash equilibrium for instance (G, r, ℓ) then $C(f) = \sum_{i=1}^{k} L_i(f) r_i$.*

3.3.2 Optimal (Minimum Total Latency) Flows

Recall that a cost of a flow f is expressed by $C(f) = \sum_{e \in E} \ell_e(f_e) f_e$. We seek to minimize this function for finding an optimal solution.

Observation 3.2 *Finding the minimum latency feasible flow is merely a case of the following non-linear program:*

$$\min \sum_{e \in E} c_e(f_e)$$

subject to:

$$\text{(NLP)} \qquad \begin{array}{ll} \sum_{P \in P_i} f_P = r_i & \forall i \in \{1, ...k\} \\ f_e = \sum_{P \in \mathcal{P}: e \in P} f_P & \forall e \in E \\ f_P \geq 0 & \forall P \in \mathcal{P} \end{array}$$

where in our problem we assign $c_e(f_e) = \ell_e(f_e) f_e$.

Note *For simplicity the above formulation of (NLP) is given with an exponential number of variables (there can be an exponential number of paths). This formulation can be easily modified with decision variables only on edges giving a polynomial number of variables and constraints.*

In our case we assume that for each edge $e \in E$ the function $c_e(f_e) = \ell_e(f_e) f_e$ is a convex function and therefore, our target function *C(f)* is also convex. This is a special case of convex programming. We wish to optimize (minimize) a convex function *F(x)* where x belongs to a convex domain.
Recall the following properties of convex sets and functions:

1. If f is strictly convex then the solution is unique.
2. If f is convex then the solution set U is convex.
3. If y is not optimal ($\exists x : F(x) < F(y)$) then y is not a local minimum. Consequently, any local minimum is also the global minimum.

Lemma 3.3.2 *The flow f is optimal for the convex program of the form (NLP) iff* $\forall i \in \{1, ..., k\}$ *and* $P_1, P_2 \in P_i$ *with* $f_{P_1} > 0$, $c'_{P_1}(f) \leq c'_{P_2}(f)$.

Notice the striking similarity between the characterization of optimal solutions (Lemma 3.3.2) and Nash equilibrium (Lemma 3.3.1). In fact, an optimal flow can be interpreted as a Nash equilibrium with respect to a different edge latency functions.
Let $x\ell_e(x)$ be a convex function for all $e \in E$. Define $\ell^*_e(f_e) = (\ell_e(f_e) f_e)'$.

Corollary 3.3 *A feasible flow f is an optimal flow for* (G, r, ℓ) *iff it is at Nash equilibrium for the instance* (G, r, ℓ^*).

Proof. f is OPT for $\ell \Leftrightarrow c'_{P_1}(f) \leq c'_{P_2}(f) \Leftrightarrow \ell^*_{P_1}(f) \leq \ell^*_{P_2}(f) \Leftrightarrow f$ is Nash for ℓ^* $(\forall i \; \forall P_1, P_2 \in P_i)$. □

3.3.3 Existence of Flows at Nash Equilibrium

We exploit the similarity between the characterizations of Nash and OPT flows to establish that a Nash equilibrium indeed exists and its cost is unique.
For the outline of the proof we define an edge cost function $h_e(x) = \int_0^x \ell_e(t)dt$.
By definition $(h_e(f_e))' = \frac{d}{dx}h_e(f_e) = \ell_e(f_e)$ thus h_e is differentiable with non decreasing derivative ℓ_e and therefore convex. Next, we consider the following convex program:

$$\min \sum_{e\in E} h_e(f_e)$$

subject to:

$$\text{(NLP2)} \qquad \begin{array}{ll} \sum_{P\in P_i} f_P = r_i & \forall i \in \{1, ..., k\} \\ f_e = \sum_{P\in\mathcal{P}:e\in P} f_P & \forall e \in E \\ f_P \geq 0 & \forall P \in \mathcal{P} \end{array}$$

Observation 3.4 *The optimal solution for (NLP2) is Nash for the modified instance where $\ell_e(x) = h'_e(x)$.*

Proof. The proof follows directly from Lemma 3.3.1 and Lemma 3.3.2 □

Since Nash is an optimal solution for a different convex setting we conclude that:

- Nash equilibrium exists.
- The cost at Nash equilibrium is unique.

3.3.4 Bounding the Coordination ratio

The relationship between Nash and OPT characterizations provide a general method for bounding the coordination ratio $\rho = \frac{C(f)}{C(f^*)} = \frac{Nash}{OPT}$.

Theorem 3.5 *For an instance (G, r, ℓ), if there exists a constant $\alpha \geq 1$ s.t.*

$$\forall e \in E\ \forall x \in \mathbb{R}^+ \quad x\ell_e(x) \leq \alpha \int_0^x \ell_e(t)dt$$

then $\rho(G, r, \ell) \leq \alpha$.

Proof.

$$\begin{aligned} C(f) &= \sum_{e\in E} \ell_e(f_e)f_e \\ &\leq \alpha \sum_{e\in E} \int_0^{f_e} \ell_e(t)dt \\ &\leq \alpha \sum_{e\in E} \int_0^{f_e^*} \ell_e(t)dt \\ &\leq \alpha \sum_{e\in E} \ell_e(f_e^*)f_e^* \\ &= \alpha \cdot C(f^*) \end{aligned}$$

The first inequality follows from the hypothesis, the second follows from the fact that Nash flow f is OPT for the function $\sum_{e \in E} \int_0^x \ell_e(t)dt$ and the final inequality follows from the assumption that the latency functions ℓ_e are nondecreasing.
□

Corollary 3.6 *If every latency function ℓ_e has the form $\ell_e(x) = \sum_{i=0}^{d} a_{e,i} x^i$ (meaning latency is a polynomial function of order d) then $\rho(G, r, \ell) \leq d + 1$.*

Note *From the corollary, an immediate coordination ratio of 2 is established for linear latency functions. Later, we will show a tighter bound of $\frac{4}{3}$.*

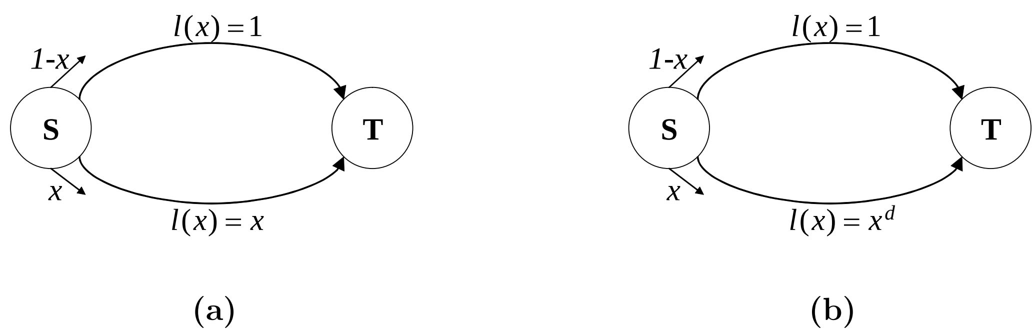

Figure 3.2:

Figure 3.2(a) shows an example for which Nash flow will only traverse in the lower path while OPT will divide the flow equally among the two paths. The target function is $1(1 - x) + x \cdot x$ and it reaches minimum with value $\frac{3}{4}$ when x = $\frac{1}{2}$, giving a coordination ratio of $\frac{4}{3}$ for this example. Combining the example with the tighter upper bound to be shown, we demonstrate a tight bound of $\frac{4}{3}$ for linear latency functions.

In Figure 3.2(b) the flow at Nash will continue to use only the lower path but OPT will reach minimum for the cost function $x \cdot x^d + (1 - x) \cdot 1$ at $x = (d + 1)^{-\frac{1}{d}}$, giving a total latency 1-$\frac{d}{d+1}(d + 1)^{-1/d}$ which approaches 0 as d $\rightarrow \infty$. So, $\lim_{d \rightarrow \infty} \rho = \infty$ meaning, ρ cannot be bounded from above in some cases when nonlinear latency functions are allowed.

3.4 A Bicriteria Bound for latency functions

We now examine an interesting *bicriteria* result. We show that the cost of a flow at Nash equilibrium can be bounded by the cost of an optimal flow feasible for twice the amount of traffic.

Theorem 3.7 *If f is a flow at Nash equilibrium for instance (G, r, ℓ) and f^* is a feasible flow for instance $(G, 2r, \ell)$ (same network but with twice the required rate), then C(f) $\leq C(f^*)$.*

Proof. Let $L_i(f)$ be the latency of a $s_i - t_i$ flow path, so that $C(f) = \sum_i L_i(f) r_i$. We define a new latency function:

$$\bar{\ell}_e(x) = \begin{cases} \ell_e(f_e) & \text{if } x \leq f_e \\ \ell_e(x) & \text{if } x \geq f_e \end{cases}$$

This latency function will allow us to approximate the original latencies as well as to lower bound the cost of any feasible flow.

- Step 1: Let's compare the cost of f^* under the new latency function $\bar{\ell}$ with respect to the original cost $C(f^*)$.
  From the construction of $\bar{\ell}_e(x)$ we get:

  $$\begin{array}{ll} \bar{\ell}_e(x) - \ell_e(x) = 0 & \text{for } x \geq f_e \\ \bar{\ell}_e(x) - \ell_e(x) \leq \ell_e(f_e) & \text{for } x \leq f_e \end{array}$$

  So, for all x we get $x[\bar{\ell}_e(x) - \ell_e(x)] \leq \ell_e(f_e) f_e$.

  The difference between the new cost under $\bar{\ell}_e$ and the original cost under ℓ is:

  $$\begin{aligned} \sum_e \bar{\ell}_e(f_e^*) f_e^* - C(f^*) &= \sum_{e \in E} f_e^* (\bar{\ell}_e(f_e^*) - \ell_e(f_e^*)) \\ &\leq \sum_{e \in E} \ell_e(f_e) f_e \\ &= C(f). \end{aligned}$$

  The cost of OPT with the latency function $\bar{\ell}$ increased by at most the cost of Nash (an additive $C(f)$ factor).

- Step 2: Denote z_0 the zero flow in G. For the pair $s_i - t_i$ we can observe that by construction, $\forall P \in P_i \;\; \bar{\ell}_P(z_0) \geq \ell_P(f) \geq L_i(f)$.
  Hence, since $\bar{\ell}_e$ is nondecreasing for each edge e, $\forall P \in P_i \;\; \bar{\ell}_P(f^*) \geq \bar{\ell}_P(z_0) \geq \ell_P(f) \geq L_i(f)$, revealing that the cost of f^* with respect to $\bar{\ell}$ can be bounded as follows:

  $$\begin{aligned} \sum_P \bar{\ell}_P(f^*) f_P^* &\geq \sum_i \sum_{P \in P_i} L_i(f) f_P^* \\ &= \sum_i 2 L_i(f) r_i = 2C(f). \end{aligned}$$

Combining the results from the previous two steps finishes the proof of the theorem:

$$\begin{aligned} C(f^*) &\geq \sum_P \bar{\ell}_P(f^*) f_P^* - C(f) \\ &\geq 2C(f) - C(f) = C(f). \end{aligned}$$

$\square$

3.5 A Tight Bound for Linear Latency Functions

Finally, we consider a scenario where all edge latency functions are linear $\ell_e(x) = a_e x + b_e$, for constants $a_e, b_e \geq 0$. A fairly natural example for such a model is a network employing a congestion control protocol such as TCP. We have already seen in Figure 3.2(a) an example where the coordination ratio was $\frac{4}{3}$. We have also established an upper bound of 2 according to Corollary 3.6. We shall now show that the $\frac{4}{3}$ ratio is also a tight upper bound. Prior to this result, we examine two simple cases:

1. $\ell_e(x) = b$
2. $\ell_e(x) = a_e x$.

For both these cases we will show that OPT=Nash.

- Case 1 is obvious since the latency on each path is constant, so both OPT and Nash will route all the flow to the paths with minimal latency.
- Case 2:
  - Using Lemma 3.3.1, a flow f is at Nash equilibrium *iff* for each source-sink pair i and $P, P' \in P_i$ with $f_P > 0$ then $\ell_P(f) = \sum_{e \in P} \ell_e(f_e) = \sum_{e \in P} a_e f_e \leq \sum_{e' \in P'} a_{e'} f_{e'} = \ell_{P'}(f)$.
  - Using Lemma 3.3.2, a flow f^* is an optimal flow *iff* for each source-sink pair i and $P, P' \in P_i$ with $f^*_P > 0$ then $C'_P(f^*) = \sum_{e \in P} C'_e(f^*_e) = \sum_{e \in P}((a_e f^*_e) f^*_e)' = \sum_{e \in P} 2a_e f^*_e \leq \sum_{e' \in P'} 2a_{e'} {f^*}'_e = C'_{P'}(f^*)$.

Corollary 3.8 *For the latency functions $\ell_e(x) = a_e(x)$ f is at Nash equilibrium iff f is an optimal flow.*

Observation 3.9 *In the example shown in Figure 3.2(a) we showed that even a simple combination of the two sets of functions is enough to demonstrate that OPT $\neq$ Nash.*

Theorem 3.10 *Let f be a flow at Nash equilibrium and f^* an optimal flow . If the latency functions are all of the form $\ell_e(x) = a_e x + b_e$ then $\rho \leq \frac{4}{3}$.*

Proof. We define a new latency function $\bar{\ell}_e$,

$$\bar{\ell}_e(x) = (\ell_e(f_e)) \cdot x = \ell_e^f \cdot x$$

Under this definition of $\bar{\ell}_e$, $OPT \equiv Nash$ (by Corollary 3.8).
Hence, f is at Nash equilibrium with respect to $\bar{\ell} \Leftrightarrow$ for every feasible flow x where $C^f(\cdot)$ is the cost with respect to $\bar{\ell}$, $C^f(f) \leq C^f(x)$.

$$\begin{aligned} C^f(x) &= \textstyle\sum_e (a_e f_e + b_e) \cdot x_e \\ &\le \textstyle\sum_e (a_e x_e + b_e) x_e + \frac{1}{4} \sum_e a_e f_e^2 \\ &\le C(x) + \tfrac{1}{4} C(f) \end{aligned}$$

The first inequality is justified by the following algebraic steps:

$$\begin{array}{crl} & (a_e f_e + b_e) x_e & \le (a_e x_e + b_e) x_e + \frac{a_e f_e^2}{4} \\ \Leftrightarrow & a_e f_e x_e & \le a_e x_e x_e + \frac{a_e f_e^2}{4} \\ \Leftrightarrow & f_e x_e & \le x_e^2 + \frac{f_e^2}{4} \\ \Leftrightarrow & 0 & \le x_e^2 - f_e x_e + \frac{f_e^2}{4} = (x_e - \frac{f_e}{2})^2 \end{array}$$

Since f brings $C^f(\cdot)$ to minimum,

$$\begin{aligned} C(f) = C^f(f) &\le C^f(x) \\ &\le C(x) + \tfrac{1}{4} C(f) \end{aligned}$$

or,

$$\frac{3}{4} C(f) \le C(x).$$

As this is true for all x, let's plug-in $x = f^*$:

$$C(f) \le \frac{4}{3} C(f^*).$$

$\square$

3.6 FIN

All good things must come to an end.