CS364A: Algorithmic Game Theory
Lecture #12: More on Selfish Routing*

Tim Roughgarden†

October 30, 2013

1 Case Study: Network Over-Provisioning

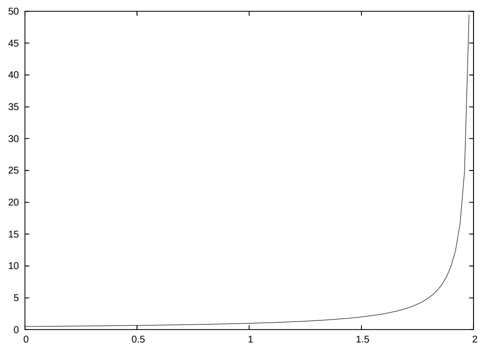
1.1 Motivation

The selfish routing model introduced last lecture can provide insight into many different kinds of networks, including transportation, communication, and electrical networks. One big advantage in communication networks is that it's often relatively cheap to add additional capacity to a network. Because of this, a popular strategy to communication network management is to install more capacity than is needed, meaning that the network will generally not be close to fully utilized (see e.g. [4]).

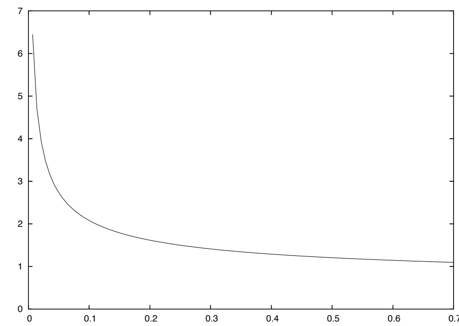
There are several reasons why network over-provisioning is common in communication networks. One reason is to anticipate future growth in demand. Beyond this, it has been observed empirically that networks tend to perform better — for example, suffering fewer packet drops and delays — when they have extra capacity. Network over-provisioning has been used as an alternative to directly enforcing "quality-of-service (QoS)" guarantees (e.g., delay bounds), for example via an admission control protocol that refuses entry to new traffic when too much congestion would result [4].

The goal of this section is develop theory to corroborate the empirical observation that network over-provisioning leads to good performance. Section 1.2 shows how to apply directly the theory developed last lecture to over-provisioned networks. Section 1.3 offers a second approach to proving the same point, that selfish routing with extra capacity is competitive with optimal routing.

†Department of Computer Science, Stanford University, 462 Gates Building, 353 Serra Mall, Stanford, CA 94305. Email: `tim@cs.stanford.edu`.



(a) M/M/1 delay function



(b) Extra capacity vs. POA curve

Figure 1: Modest overprovisioning guarantees near-optimal routing. The left-hand figure displays the per-unit cost $c(x) = 1/(u - x)$ as a function of the load x for an edge with capacity $u = 2$. The right-hand figure shows the worst-case price of anarchy as a function of the fraction of unused network capacity.

1.2 POA Bounds for Over-Provisioned Networks

The optimal price of anarchy (POA) bounds for selfish routing developed last lecture are parameterized by the class of permissible network cost functions. In this section, we consider a network in which every cost function $c_e(x)$ has the form

$$c_e(x) = \begin{cases} \frac{1}{u_e - x} & \text{if } x < u_e \\ +\infty & \text{if } x \geq u_e. \end{cases} \tag{1}$$

The parameter u_e should be thought of as the capacity of edge e. A cost function of the form (1) is the expected delay in an M/M/1 queue, meaning a queue where jobs arrive according to a Poisson process with rate x and have independent and exponentially distributed services times with mean $1/u_e$. This is generally the first and simplest cost function used to model delays in communication networks (e.g. [2]). Figure 1(a) displays such a function; it stays very flat until the traffic load nears the capacity, at which point the cost rapidly tends to $+\infty$.

For a parameter $\beta \in (0, 1)$, call a selfish routing network with M/M/1 delay functions *β-over-provisioned* if $f_e \leq (1 - \beta)u_e$ for every edge e, where f is an equilibrium flow. That is, at equilibrium, the maximum link utilization in the network is at most $(1 - \beta) \cdot 100\%$.

Figure 1(a) suggests the following intuition: when β is not too close to 0, the equilibrium flow is not too close to the capacity on any edge, and in this range the edges' cost functions behave like low-degree polynomials with nonnegative coefficients. Last lecture we saw that the POA is small in networks with such cost functions.

More formally, the main theorem from last lecture reduces computing the worst-case POA in arbitrary β-over-provisioned selfish routing networks to computing the worst-case POA merely in β-over-provisioned Pigou-like examples. A computation, which the reader

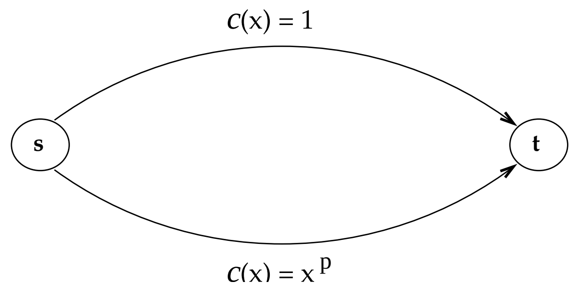


Figure 2: Nonlinear variant of Pigou's example. The POA of selfish routing can be arbitrarily large.

is encouraged to do in the privacy of their own home, shows that the worst-case POA in β-over-provisioned networks is at most

$$\frac{1}{2}\left(1+\sqrt{\frac{1}{\beta}}\right); \tag{2}$$

see Figure 1(b). As we've come to expect, very simple networks show that the bound in (2) is tight for all values of $\beta \in (0, 1)$.

Unsurprisingly, the bound in (2) tends to 1 as β tends to 1 and to $+\infty$ as β tends to 0; these are the cases where the cost functions effectively act like constant functions and like very high-degree polynomials, respectively. What's interesting to investigate is intermediate values of β. For example, if $\beta = .1$ — meaning the maximum edge utilization is at most 90% — then the POA is guaranteed to be at most 2.1. In this sense, a little over-provisioning is sufficient for near-optimal selfish routing, corroborating what has been empirically observed by Internet Service Providers.

1.3 A Resource Augmentation Bound

This section proves a guarantee for selfish routing in arbitrary networks, with no extra assumptions on the cost function. What could such a guarantee look like? Recall that the nonlinear variant of Pigou's example (Figure 2) shows that the POA in selfish routing networks with arbitrary cost functions is unbounded.

In this section, we compare the performance of selfish routing to a "weaker" optimal solution that is forced to send extra traffic.[1] For example, in Figure 2, with one unit of traffic, the equilibrium flow has cost 1 while the optimal flow has near-zero cost. If the optimal flow has to route *two* units of traffic through the network, then there is nowhere to hide: the best solution continues to route $(1-\epsilon)$ units of traffic on the lower edge, with the remaining $(1+\epsilon)$ units of traffic routed on the upper edge, for a total cost exceeding that of the equilibrium flow (with one unit of traffic).

[1]Another approach, explored in the systems community [5], is to instead make extra assumptions about the network structure and the traffic rate.

This "unfair" comparison between two flows at different traffic rates has an equivalent and easier to interpret formulation as a comparison between two flows with the same traffic rate but in networks with with different cost functions. Intuitively, instead of forcing the optimal flow to route additional traffic, we allow the equilibrium flow to use a "faster" network, with each original cost function $c_e(x)$ replaced by the "faster" function $c_e(\frac{x}{2})/2$. (See the Exercises for details.) This transformation is particularly easy to interpret for M/M/1 delay functions, since if $c_e(x) = 1/(u_e - x)$, then the "faster" function is $1/(2u_e - x)$ — an edge with double the capacity. The next theorem, after this reformulation, gives a second justification for network over-provisioning: a modest technology upgrade improves performance more than implementing dictatorial control.[2]

Theorem 1.1 ([6]) *For every selfish routing network and traffic rate r, the cost of an equilibrium flow with rate r is at most the cost of an optimal flow with rate $2r$.*

Proof: Fix a network G with nonnegative, nondecreasing, and continuous cost functions, and a traffic rate r. Let f and f^* denote equilibrium and optimal (minimum-cost) flows at the traffic rates r and $2r$, respectively.

The first part of the proof reuses the trick from last lecture of using fictitious cost functions, frozen at the equilibrium costs, to get a grip on the cost of the optimal flow f^*. Recall that since f is an equilibrium flow, all paths P used by f have a common cost $c_P(f)$, call it L. Moreover, $c_P(f) \geq L$ for every path $P \in \mathcal{P}$. Thus,

$$\sum_{P \in \mathcal{P}} \underbrace{f_P}_{\text{sums to } r} \cdot \underbrace{c_P(f)}_{= L \text{ if } f_P > 0} = r \cdot L \tag{3}$$

while

$$\sum_{P \in \mathcal{P}} \underbrace{f^*_P}_{\text{sums to } 2r} \cdot \underbrace{c_P(f)}_{\geq L} \geq 2r \cdot L. \tag{4}$$

That is, with respect to the fictitious costs $\{c_e(f_e)\}$, we get a great lower bound on the cost of f^* — at least twice the cost of the equilibrium flow f — much better than what we're actually trying to prove.

The second step of the proof shows that using the fictitious costs instead of the accurate ones overestimates the cost of f^* by at most the cost of f. Specifically, we complete the proof by showing that

$$\underbrace{\sum_{e \in E} f^*_e \cdot c_e(f^*_e)}_{\text{cost of } f^*} \geq \underbrace{\sum_{e \in E} f^*_e \cdot c_e(f_e)}_{\geq 2rL} - \underbrace{\sum_{e \in E} f_e \cdot c_e(f_e)}_{= rL}. \tag{5}$$

We prove that (5) holds term-by-term, that is, we show that

$$f^*_e \cdot [c_e(f_e) - c_e(f^*_e)] \leq f_e \cdot c_e(f_e) \tag{6}$$

[2]Like last lecture, we prove the result for networks with a single source and single sink. The same proof extends, with minor extra notation, to networks with multiple sources and sinks (see the Exercises).

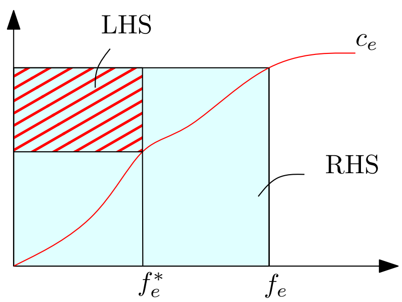


Figure 3: Proof by picture of part of Theorem 1.1.

for every edge $e \in E$. When $f_e^* \geq f_e$, the left-hand side of (6) is nonpositive and there is nothing to show. When $f_e^* < f_e$, we give a proof by picture; see Figure 3. The left-hand side of (6) is the area of the shaded region, with width f_e^* and height $c_e(f_e) - c_e(f_e^*)$. The right-hand side of (6) is the area of the solid region, with width f_e and height $c_e(f_e)$. Since $f_e^* < f_e$ and c_e is nondecreasing, the former region is a subset of the latter. This verifies (6) and completes the proof. ■

In the sense of Theorem 1.1 and its reformulation given in the Exercises, speeding up (i.e., overprovisioning) a selfish routing network by a modest amount is better than routing traffic optimally.

2 Atomic Selfish Routing

So far we've studied a *nonatomic* model of selfish routing, meaning that all players were assumed to have negligible size. This is a good model for cars on a highway or small users of a communication network, but not if a single strategic player represents, for example, all of the traffic controlled by a single Internet Service Provider. This section studies *atomic* selfish routing networks, where each player controls a non-negligible amount of traffic. While most aspects of the model will be familiar, it presents a couple of new technical complications. These complications will also be present in other classes of games that we study later.

An atomic selfish routing network has a finite number k of players. Player i has a source vertex s_i and a destination vertex t_i; these can be shared across players, or not. Each player routes 1 unit of traffic on a single s_i-t_i path, and seeks to minimize its cost.[3] Flows, equilibrium flows, and the cost of a flow are defined analogously to last lecture.

To get a feel for the atomic model, consider the variant of Pigou's example shown in Figure 4. Suppose there are two players, and recall that each controls 1 unit of flow. The optimal solution routes one player on each link, for a total cost of $1 + 2 = 3$. This is also an equilibrium flow, in the sense that neither player can decrease its cost via a unilateral

[3]Two obvious variants of the model allow players to have different sizes and/or to split traffic fractionally over multiple paths. Both variants have been extensively studied using methods similar to the ones covered in these lectures.

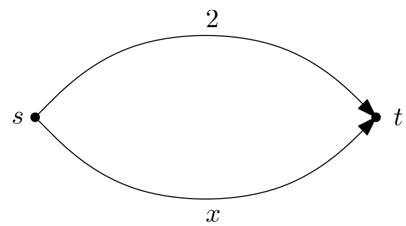


Figure 4: A pigou-like network for atomic selfish routing.

deviation. The player on the lower edge does not want to switch, since its cost would jump from 1 to 2. More interestingly, the player on the upper edge (with cost 2) has no incentive to switch to the bottom edge, where its sudden appearance would drive the cost up to 2.

There is also a *second* equilibrium in the network: if both players take the lower edge, both have a cost of 2 and neither can decrease its cost by switching to the upper edge. This equilibrium has cost 4. This illustrates an importance difference between the nonatomic and atomic models: different equilibria are guaranteed to have the same cost in the nonatomic model, but not in the atomic model.

Our current working definition of the POA — the ratio between the objective function value of an equilibrium and that of an optimal outcome — is not well defined when different equilibria have different objective function values. We extend the definition by taking a worst-case approach: the *price of anarchy (POA)* of an atomic selfish routing network is

$$\frac{\text{cost of worst equilibrium}}{\text{cost of optimal outcome}}.$$

For example, in the network in Figure 4, the POA is $\frac{4}{3}$.

A second difference between the two models is that the POA in atomic selfish routing networks can be larger than in their nonatomic counterparts. To see this, consider the four-player bidirected triangle network shown in Figure 5. Each player has two strategies, a one-hop path and a two-hop path. In the optimal flow, all players route on their one-hop paths, and the cost of this flow is 4 — these one-hop paths are precisely the four edges with the cost function $c(x) = x$. This flow is also an equilibrium flow. On the other hand, if all players route on their two-hop paths, then we obtain a second equilibrium flow. Since the first two players each incur three units of cost and the last two players each incur two units of cost, this flow has a cost of 10. As the reader should check, it is also an equilibrium. The price of anarchy of this instance is therefore $10/4 = 2.5$.

There are no worse examples with affine cost functions.[4]

Theorem 2.1 (POA Bound for Atomic Selfish Routing, Affine Cost Functions [1, 3]) *In every atomic selfish routing network with affine cost functions, the POA is at most $\frac{5}{2}$.*

[4]There are also very general and tight POA bounds known for arbitrary sets of cost functions. For example, in atomic selfish routing networks with cost functions that are polynomials with nonnegative coefficients, the POA is at most a constant that depends on the maximum polynomial degree d. The dependence on d is exponential, however, unlike the $\approx \frac{d}{\ln d}$ dependence in nonatomic selfish routing networks.

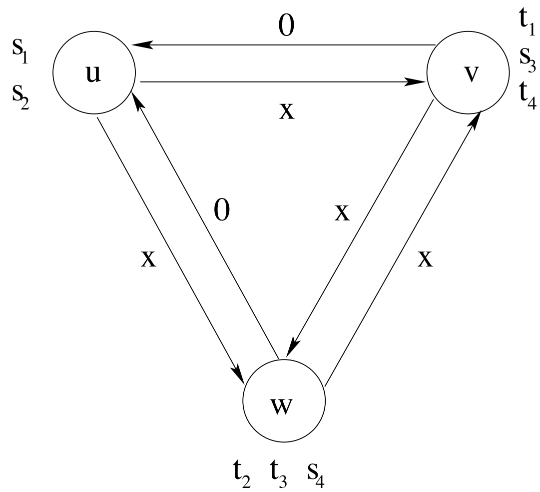


Figure 5: In atomic instances with affine cost functions, the POA can be as large as 5/2.

Proof: The following proof is a "canonical POA proof," in a sense that we'll make precise in Lecture 14. Let's just follow our nose. We need to prove a bound for every equilibrium flow; fix one f arbitrarily. Let f^* denote an optimal (minimum-cost) flow. Write f_e and f_e^* for the number of players in f and f^*, respectively, that pick a path that includes the edge e. Write each affine cost function as $c_e(x) = a_e x + b_e$ for $a_e, b_e \geq 0$.

The first step of the proof is to figure out a good way of applying our hypothesis that f is an equilibrium flow — that no player can decrease its cost through a unilateral deviation. After all, the bound of 2.5 does not generally apply to non-equilibrium flows. If we consider any player i, using path P_i in f, and any unilateral deviation to a different path $\hat{P}_i$, then we can conclude that i's equilibrium cost using P_i is at most what its cost would be if it switched to $\hat{P}_i$. This looks promising: we want an upper bound on the total cost of players in the equilibrium f, and hypothetical deviations give us upper bounds on the equilibrium costs of individual players. Which hypothetical deviations should we single out for the proof? A natural idea is to let the optimal flow f^* suggest deviations.

Formally, suppose player i uses path P_i in f and path P_i^* in f^*. Since f is an equilibrium, i's cost only increases if it switches to P_i^* (holding all other players fixed):

$$\sum_{e \in P_i} c_e(f_e) \leq \sum_{e \in P_i^* \cap P_i} c_e(f_e) + \sum_{e \in P_i^* \setminus P_i} c_e(f_e + 1),$$

where in the final term we account for the additional unit of load that i contributes to edges that it newly uses (in P_i^* but not in P_i). For all we know P_i and P_i^* are disjoint; since cost functions are nondecreasing, we have

$$\sum_{e \in P_i} c_e(f_e) \leq \sum_{e \in P_i^*} c_e(f_e + 1). \tag{7}$$

This completes the first step, in which we apply the equilibrium hypothesis to generate an upper bound (7) on the equilibrium cost of each player.

The second step of the proof sums the upper bound (7) on individual equilibrium costs over all players to get a bound on the total equilibrium cost:

$$
\begin{aligned}
\sum_{i=1}^{k} \sum_{e \in P_i} c_e(f_e) &\leq \sum_{i=1}^{k} \sum_{e \in P_i^*} c_e(f_e + 1) \\
&= \sum_{e \in E} f_e^* \cdot c_e(f_e + 1) && (8) \\
&= \sum_{e \in E} \left[a_e f_e^* (f_e + 1) + b_e f_e^* \right], && (9)
\end{aligned}
$$

where in (8) we use that the term $c_e(f_e + 1)$ is contributed exactly once by each player i that contemplates switching to a path P_i^* that includes the edge e — f_e^* times in all. This complete the second step of the proof.

The previous step gave an upper bound on a quantity that we care about — the cost of the equilibrium flow f — in terms of a quantity that we don't care about, the "entangled" version of f and f^* on the right-hand side of (9). The third and most technically challenging step of the proof is to "disentangle" the right-hand side of (9) and relate it to the only quantities that we do care about for a POA bound, the costs of f and f^*.

We next claim that, for every $y, z \in \{0, 1, 2, \ldots, \}$,

$$
y(z+1) \leq \frac{5}{3} y^2 + \frac{1}{3} z^2. \tag{10}
$$

This inequality is easy to check once guessed, and we leave the verification of it as an exercise. One can check all cases where y and z are both small, and then observe that it continues to hold when either one grows large. Note that the inequality holds with equality when $y = z = 1$ and when $y = 1$ and $z = 2$. We'll demystify how inequalities like (10) arise next week.

We now apply inequality (10) once per edge in the right-hand side of (9), with $y = f_e^*$ and $z = f_e$. We then have

$$
\begin{aligned}
C(f) &\leq \sum_{e \in E} \left[a_e \left(\frac{5}{3} (f_e^*)^2 + \frac{1}{3} f_e^2 \right) + b_e f_e^* \right] \\
&\leq \frac{5}{3} \left[\sum_{e \in E} f_e^* (a_e f_e^* + b_e) \right] + \frac{1}{3} \sum_{e \in E} a_e f_e^2 \\
&\leq \frac{5}{3} \cdot C(f^*) + \frac{1}{3} \cdot C(f).
\end{aligned}
$$

Subtracting $\frac{1}{3} C(f)$ from both sides and multiplying through by $\frac{3}{2}$ gives

$$
C(f) \leq \frac{5}{3} \cdot \frac{3}{2} \cdot C(f^*) = \frac{5}{2} \cdot C(f^*),
$$

completing the proof. ■

References

[1] B. Awerbuch, Y. Azar, and L. Epstein. The price of routing unsplittable flow. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC)*, pages 57–66, 2005.

[2] D. P. Bertsekas and R. G. Gallager. *Data Networks*. Prentice-Hall, 1987. Second Edition, 1991.

[3] G. Christodoulou and E. Koutsoupias. The price of anarchy of finite congestion games. In *Proceedings of the 37th Annual ACM Symposium on Theory of Computing (STOC)*, pages 67–73, 2005.

[4] N. Olifer and V. Olifer. *Computer Networks: Principles, Technologies and Protocols for Network Design*. Wiley, 2005.

[5] L. Qiu, Y. R. Yang, Y. Zhang, and S. Shenker. On selfish routing in Internet-like environments. *IEEE/ACM Transactions on Networking*, 14(4):725–738, 2006.

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