A review of How Bad is Selfish Routing, By Tim Roughgarden and Eva Tardos

Siddhartha Saha, CSE202, Fall 2003

December 8, 2003

1 Introduction

The problem of optimal routing in a data network is a well known and well studied problem. The vast amount of networks and nodes in today world make this problem a very practical one. But in a distributed network, there is no centralized authority that has the complete overview of the network. It is very difficult and expensive, if not nearly impossible to obtain an optimal routing allocation is a congested network.

We now describe the model and the performance criteria that we want to maximize. We are given a network, a collection of nodes and a rate of traffic between each pair of nodes and a latency function for each edge specifying the time needed to traverse the edge as a function of its congestion. The objective is to route traffic in such a way such that the sum of all travel times is minimized.

Since there is no centralized authority, each node is free to act according to its own interests. The assumption made here is that, each network user routes its own traffic on the minimum latency path available to it, given the network congestion caused by the ongoing traffics of other users.

The paper we review here [RT00] tries to quantify the inefficiency inherent in such a selfishly chosen solution and tries to compare with the theoretically computed optimal solution. In other words, this paper tries to how much the network performance degrades from the optimal solution due to this uncoordinated behavior of the network nodes.

2 Previous Work

Since this is a very much practical problem, a lot of work has already been done on this topic. This kind of problem has been modeled as network flow problem with all flow paths between a given source destination pair having equal latency. For this problem, the existence of optimal traffic equilibrium has been proven. Later, more and more efficient algorithms have been designed to compute this equilibrium.

3 Approach to Solution

The assumption made here is that the network users behave in a purely selfish manner but not in a malicious manner. Under these conditions, network users can be viewed as independent agents participating in a non-cooperative game. And in the light of the classical game theory, the solutions adopted by each network user form a Nash Equilibrium.

The primary model that the paper follows assumes that each agent controls a negligible fraction of the overall traffic. And the second assumption is that each agent can determine the latency of a particular link with arbitrary precision. Under these assumptions, equilibrium in a system is then nothing but a steady state in the system. In this case, a feasible assignment of traffic to paths in the network can be modeled as a network flow, with the amount of flow between a pair of nodes in the network being equal to the rate of traffic between the two nodes. Under this model, the Nash Equilibrium mentioned above corresponds to a flow where all flow paths from a source and a destination have equal and smallest possible latency. Since it has been shown, that if the latency each network link is a continuous non-decreasing function of the edge congestion, then a flow corresponding to Nash Equilibrium always exists. Thus, this paper tries to find the worst case ration between the total latency of a flow at Nash Equilibrium and that of an optimal flow minimizing the total latency in any such network

The results obtained by the authors are very interesting. They prove that in the case where latency of each edge is a liner function of the edge congestion, a flow at Nash Equilibrium has at most a total latency of 4/3 times that of the optimal flow. On the other hand, if the link latency functions are assumed to be continuous and non-decreasing only, then the ratio between the total latency of a flow at Nash Equilibrium and that of an optimal flow may be unbounded. The authors then propose different criteria on the optimal solution, comparing the total latency of a flow at Nash Equilibrium with that of an optimal flow in the same network but routing additional traffic between each pair of nodes. This analysis finally shows, that for any network with continuous non-decreasing latency functions, the total latency incurred by a flow at Nash Equilibrium is at most the latency incurred by an optimal flow that routes twice as much traffic through the same network

3.1 Mathematical formulation of the model

The network is modeled as a directed graph $G = (V, E)$ with vertex set V and edge set E and k source destination vertex pairs $\{s_1, t_1\}, \{s_2, t_2\}, \ldots, \{s_k, t_k\}$. The set of simple $s_i - t_i$ paths are denoted by P_i and we define $P = \cup_i P_i$. A *flow* in this case is a function $f : P \to R^+$. For a fixed flow f, define $f_e = \Sigma_{P:e\in P} f_P$. For each pair $\{s_i, t_i\}$, we associate a rate r_i, which is the amount of flow with source s_i and destination r_i. Thus, a flow is feasible if for all i, $\Sigma_{P\in P_i} f_P = r_i$. Now we define a function on each edge, which gives the latency of the edge depending on the load on the edge. The latency function is denoted by $l(.)$. For each edge $e \in E$, we assume that the latency function l_e is non negative, increasing and differentiable. An instance of the problem is thus the tripe (G, r, l). The latency of a path P with respect to a flow f is defined as the sum of all the latencies of the edges in the path, denoted by $l_P(f) = \Sigma_{e\in P} l_e(f_e)$. We define the cost $C(f)$ of a flow f in G as the total latency incurred by the flow f, i.e., $C(f) = \Sigma_{p\in P} l_p(f) f_p$.

3.2 Key Ideas

Each user of the network tries to route its own traffic along the minimal latency path available to it, and we call the system to be in Nash Equilibrium when each node in the network achieves this state. This gives rise to the following lemma:

Lemma 3.1 *A flow f feasible for problem instance (G, r, l) is at Nash Equilibrium if and only if for every $i \in 1, \ldots, k$ and $P_1, P_2 \in P_i$ with $f_{p_i} > 0, l_{p_1}(f) \leq l_{p_2}(f)$*

Since, when a flow is in Nash Equilibrium, then all $s_i - t_i$ flow paths to which f assigns a positive amount of flow have equal latency.

To compare with the flow at Nash Equilibrium with that of the optimal flow, we first need to find out the latency of an optimal flow. Finding out the minimum latency feasible flow in a network can be reduced to the problem of finding a local minimum of a convex non linear problem.

3.3 A Simple Example

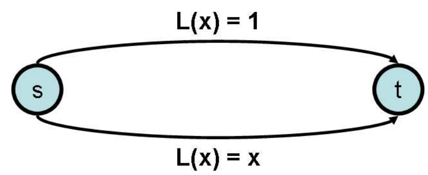

Figure 1: Picture without equations

In the network shown in figure 1 there is a single source destination pair. The rate for this pair is 1. The flow at Nash Equilibrium puts the entire unit of flow in the lower link (where $l(x) = x$). The optimal minimal latency flow spreads the flow evenly across the two links. For the case of Nash Equilibrium, the total latency is 1 whereas the total latency incurred by the optimal flow is $\frac{3}{4}$. Thus the ratio of the total latencies is $\rho = \frac{4}{3}$.

But when we relax the condition that the edge latencies be linear function of edge congestion, we see that the performance degradation ratio can be unbounded. This can be simply demonstrated from the above example. Modify the latency of the lower edge as $l(x) = x^p$, everything else being the same. The Nash Equilibrium flow puts all the traffic in the lower edge, incurring a cost of 1. But the optimal solution assigns $(p+1)^{-1/p}$ units to lower link and the remaining to the upper link. This solution has a total latency of $1 - p.(p+1)^{-(p+1)/p}$, which tends to zero as $p \rightarrow \inf$. Thus, ρ cannot be bounded above when the edge latencies cease to be a linear function of the edge congestion.

The main result of this paper is to generalize this result to any network with continuous, non decreasing edge latencies.

3.4 Worst Case with Linear Latency Functions

In this sub section, we provide brief outline of the proof. Complete reproduction of the proof with all the details is rather unnecessary. The main idea goes like this: The optimal solution is built up is a two step process. Suppose, we have a problem instance (G, r, l). In the first step, a flow optimal for the instance $(G, r/2, l)$ is sent through G. In the second step, this flow is augmented to obtain an optimal flow for (G, r, l). Then it is shown that the first flow has cost at least $\frac{1}{4}C(f)$ and the augmentation has cost at least $\frac{1}{2}C(f)$, where f is some flow which is at Nash Equilibrium.

3.5 Bicriteria Result

In addition to the above results, one more important result of this paper is the Bicriteria Result. The authors compare the cost of a flow at Nash Equilibrium to an optimal flow feasible for increased rates. In the example given in previous sections, for any p, an optimal flow feasible for twice the rate has total latency at least that of the flow at Nash Equilibrium feasible for the original flow. That is, if f is a flow at Nash Equilibrium for (G, r, l), and f^* is feasible for $(G, 2r, l)$, then $C(f) \leq C(f^*)$. This claim has also been proven rigorously in this paper.

4 Extensions to the Basic Model

The basic model assumed in the beginning of the paper suffers from some deficiencies. Firstly, agents can only evaluate path latencies approximately. To tackle this problem, a notion of Approximate Nash Equilibrium is used and it is shown that the modified model also follows the previous bicriteria claims.

Second flaw in the basic model was that it assumes a infinite number of agents, each controlling an infinitesimal amount of flow, while is real system there are only finitely many agents each controlling a strictly positive amount of flow. The paper also proves the bicriteria claims in this case also, provided that each agent can route its flow fractionally over any number of paths. But, in the case, where this relaxations of dividing flows into multiple paths is not granted, the bicriteria result does not hold.

5 Follow up Work

In [RT], the author has shown that many of the positive results about selfish routing do not require the combinatorial structure of a network, and hold more generally for a wide class of games (called nonatomic congestion games) and this is an interesting addition to the game theory literature.

In order to reduce the erraticness of behavior, some schemes to price the network edges has been thought of, with the hope that this pricing of network edges will reduce the cost incurred due to selfish routing. [CDR03b], [CDR03a] is some of the papers which have persuaded research in this direction.

Some work has also been done in studying the performance of selfish routing in Internet like environment. [QYZS03] have done one such work, and has come to conclusion that in contrast to theoretical worst cases, selfish routing achieves close to optimal average latency in such environments. However, such performance benefit comes at the expense of significantly increased congestion on certain links.

[CSSM] has tried to extend the result of [RT00] to a network model without capacities.

References

[CDR03a] R. Cole, Y. Dodis, and T. Roughgarden. Pricing network edges for heterogeneous selfish users, 2003.

[CDR03b] Richard Cole, Yevgeniy Dodis, and Tim Roughgarden. How much can taxes help selfish routing?, 2003.

[CSSM] J. R. Correa, A. S. Schulz, and N. Stier Moses. Selfish routing in capacitated networks.

[QYZS03] Lili Qiu, Yang Richard Yang, Yin Zhang, and Scott Shenker. On selfish routing in internet-like environments. In *Proceedings of the ACM SIGCOMM*, All ACM Conferences, pages 151–162, Karlsruhe, Germany, august 2003.

[RT] Tim Roughgarden and Eva Tardos. Bounding the inefficiency of equilibria in nonatomic congestion games. In *Cornell Technical Report TR2002-1866*.

[RT00] Tim Roughgarden and Eva Tardos. How bad is selfish routing? In *IEEE Symposium on Foundations of Computer Science*, pages 93–102, 2000.