On Selfish Routing in Internet-Like Environments

Lili Qiu, UT Austin, lili@cs.utexas.edu
Yang Richard Yang, Yale University, yry@cs.yale.edu
Yin Zhang, UT Austin, yzhang@cs.utexas.edu
Scott Shenker, UC Berkeley, shenker@icir.org

***Abstract*— A recent trend in routing research is to avoid inefficiencies in network-level routing by allowing hosts to either choose routes themselves (*e.g.*, source routing) or use overlay routing networks (*e.g.*, Detour or RON). Such approaches result in *selfish* routing, because routing decisions are no longer based on system-wide criteria but are instead designed to optimize host-based or overlay-based metrics. A series of theoretical results showing that selfish routing can result in suboptimal system behavior have cast doubts on this approach. In this paper, we use a game-theoretic approach to investigate the performance of selfish routing in Internet-like environments based on realistic topologies and traffic demands in our simulations. We show that in contrast to theoretical worst cases, selfish routing achieves close to optimal average latency in such environments. However, such performance benefits come at the expense of significantly increased congestion on certain links. Moreover, the adaptive nature of selfish overlays can significantly reduce the effectiveness of traffic engineering by making network traffic less predictable.**

***Index Terms*— Selfish routing, overlay, game theory, traffic equilibrium, traffic engineering, optimization, relaxation.**

I. Introduction

For decades, it has been the responsibility of the network to route traffic. Recent studies (*e.g.*, [36], [41]) have shown that there is inherent inefficiency in network-level routing from the user's perspective. In response to these observations, we have seen an emergent trend to allow end hosts to choose routes themselves by using either source routing (*e.g.*, Nimrod [5]) or overlay routing (*e.g.*, Detour [36] or RON [3]). These end-to-end route selection schemes are shown to be effective in addressing some deficiencies in today's IP routing. For example, measurements [36] from the Detour project show that in the Internet, a large percentage of flows can find better alternative paths by relaying among overlay nodes, thereby improving their performance. RON [3] also demonstrates the benefits of overlay routing using real implementation and deployment.

Such end-to-end route selection schemes are selfish by nature in that they allow end users to greedily select routes to optimize their own performance without considering the system-wide criteria. Recent theoretical results suggest that in the worst case selfish routing can result in serious performance degradation due to lack of cooperation. In particular, Roughgarden and Tardos prove that the *price of anarchy* (*i.e.*, the worst-case ratio between the total latency of selfish routing and that of the global optimal) for selfish routing can be unbounded for general latency functions [35].

Despite much theoretical advance, an open question is how selfish routing performs in Internet-like environments. This is a challenging question because today's Internet is unique in the following respects.

First, topologies and traffic demands of the Internet are not arbitrary but have certain structures. The worst-case results may not be applicable to realistic topologies and traffic demands. A general open question is *whether selfish routing results in poor performance in Internet-like environments (*i.e.*, under realistic network topologies and traffic demands)*.

Second, users in overlay networks do not have full flexibility in specifying their end-to-end paths. Due to limited availability of source routing support in the routers, the path between any two network nodes is dictated by the Internet routing protocols, such as OSPF, MPLS, or BGP. While overlay networks provide another mechanism to enable users to control their routes by relaying through overlay nodes, the route between two overlay nodes is still governed by the underlying routing protocol. A natural question is *how to model such selfish overlay routing and whether selfish overlay routing results in poor performance*.

Third, even if selfish overlays (*i.e.*, overlays consisting of selfish traffic) yield good performance, they can be deployed only incrementally. As a result, background traffic and overlay traffic will interact with each other. We call such interactions *horizontal interactions*. An important question is *how such selfish traffic affects the remaining traffic routed using the traditional routing protocols*. A related question is *whether multiple overlays result in poor performance*.

Fourth, the way in which selfish users choose their routes can interact with traffic engineering. We call such interactions *vertical interactions*, which can be viewed as the following iterative process. First, Internet Service Providers (ISPs) adjust network-level routing according to traffic demands, using schemes in [4], [13], [14], [43], to minimize *network cost*. Then selfish users adapt to changes in the underlying default routes by choosing different overlay paths to optimize their end-to-end performance. Such adaptation changes traffic demands and triggers traffic engineering to readjust the default routes, which in turn makes selfish users adapt to new routes. Given the mismatch between the objectives of selfish routing and traffic engineering, an interesting question is *whether selfish routing interacts poorly with traffic engineering*.

In this paper, we seek to answer the above questions through extensive simulations. We take a game-theoretic approach to compute the traffic equilibria of various routing schemes and then evaluate their performance. We focus on *intra-domain* network environments because recent advances in topology mapping [38] and traffic estimation [44] allow us to use realistic network topologies and traffic demands for such scenarios. Understanding selfish routing in inter-domain environments is also of great interest but is more challenging. First, we do not have realistic models for inter-domain traffic demands. Second, despite some recent progress towards understand-

A shorter version of this paper appears in [31].

Yang Richard Yang was supported in part by NSF grant ANI-0207399, and Scott Shenker was supported in part by NSF grants ITR-0205519, ANI-0207399, ITR-0121555, ITR-0081698, ITR-0225660 and ANI-0196514.

ing autonomous system relationships [16], [39], [28], more research efforts are needed to develop realistic models for inter-domain routing policies. Finally, the large size of inter-domain topologies makes it computationally prohibitive to derive traffic equilibria. Due to these difficulties, in this paper we conduct a preliminary investigation of selfish routing in *inter-domain* environments. We leave a more thorough study of selfish routing in inter-domain environments (*e.g.*, considering a larger-scale network with different types of routing policies and realistic traffic demands) as future work.

Our key contributions and results can be summarized as follows. First, we formulate and evaluate selfish routing in overlay networks. Selfish routing in overlay networks is different from traditional selfish source routing in that (i) the route between any two overlay nodes is dictated by network-level routing, and (ii) different overlay links may share common physical links and therefore traditional algorithms to compute traffic equilibria do not apply. We fill the gap by presenting algorithms to compute one of (the potentially multiple) traffic equilibria.

Second, we find that in contrast to theoretical worst cases, selfish routing in Internet-like environments yields close to optimal average latency, which can be much lower than that of default network-level routing. This is true for both source routing and overlay routing. Moreover, we show selfish routing achieves good performance without hurting the traffic that uses default network-level routing.

Third, we show that an important impact of selfish routing on Internet-like environments is the fundamental mismatch between the objectives of selfish routing and traffic engineering. In particular, our results show that the low latency of selfish routing is often achieved at the expense of increased congestion on certain links. Moreover, the adaptive nature of selfish routing makes traffic demands less predictable, and can significantly reduce the effectiveness of traffic engineering.

The rest of the paper is organized as follows. In Section II, we review related work. In Section III, we present our network model. In Section IV and Section V, we specify the routing schemes we evaluate and the algorithms we use to compute traffic equilibria. In Section VI, we describe our evaluation methodology. We study the performance of selfish source routing in Section VII and that of selfish overlay routing in Section VIII. In Section IX and Section X, we investigate horizontal and vertical interactions, respectively. We examine the impacts of multi-AS nature of the Internet on routing performance in Section XI. Finally we conclude in Section XII.

II. Related Work

A number of recent studies have reported that network-level routing is inefficient from the user's perspective. For example, Savage *et al.* [36] use Internet measurements to show that the default routing path is often suboptimal in terms of latency, loss rate, and TCP throughput. The suboptimal performance of network-level routing is inevitable due to routing hierarchy and policy [41], as well as different routing objectives used by network operators, whose goal is to avoid high utilization. Moreover, stability problems with routing protocols, such as BGP, could make things even worse. As a result, there has been a movement to give users more autonomy in choosing their routes by using source routing (*e.g.*, Nimrod [5]) or overlay routing networks (*e.g.*, Detour [36] and RON [3]).

Recently a series of theoretical results show that selfish routing can result in extremely suboptimal performance in worst cases. The pioneering work in this area is by Koutsoupias and Papadimitriou [24], who compare the worst-case Nash equilibrium with a global optimal solution in minimizing network congestion in a two-node network. Roughgarden and Tardos are interested in a different performance metric – latency. In [35], they prove that the price of anarchy (*i.e.*, the worst-case ratio between the average latency of a Nash equilibrium and that of the global optimal) depends on the "steepness" of the network latency functions. They show that the price of anarchy is unbounded for a general latency function such as M/M/1. In contrast to the theoretical studies, our study focuses on a practical setting by using realistic network topologies and traffic demands. Different from the measurement studies, our study considers a more general setting, and investigates networks with a large amount of selfish traffic under different network configurations (including both static and dynamic network control).

The inefficiency of selfish routing motivates researchers to design strategies to reduce the cost of uncooperation. For example, Korilis, Lazar, and Orda in [23] and Roughgarden in [34] study a network with a mixture of selfish traffic and "centrally controlled" traffic. Roughgarden shows that it is NP-hard to compute the optimal strategy for "centrally controlled" traffic (*i.e.*, a Stackelberg strategy), and gives a simple algorithm to approximate the optimal strategy in a network of parallel links with total latency no more than a constant times that of the minimum latency [33]. In [26], Kumar and Marathe give a FPTAS to the Stackelberg scheduling strategies in a parallel-link setting and also extend the results to slightly more general topologies. Another way to cope with selfish behavior is to introduce pricing and taxation. The authors in [7], [20], [11] develop pricing schemes to minimize total latency under homogeneous and heterogeneous traffic, assuming that selfish users are conscious of both latency and taxes.

Although the price of anarchy can be high in the worst-case, some theoretical studies have also shown that the degradation is less severe from some other perspectives (*e.g.*, [15]). For example, Friedman shows that for "most" traffic rate vectors in a range, the price of anarchy is lower than that of the worst cases [15]. He also analyzes the effects of TCP rate adaptation in a parallel-link network and shows that the performance loss is small. Roughgarden and Tardos [35] show (essentially) that the performance degradation due to selfish routing can be compensated by doubling the bandwidth on all links. However, this is often not a practical option for the Internet at least in the short-term.

The interaction between selfish overlay routing and traffic engineering studied in this paper has attracted other researchers to have a more in-depth investigation of the issue. For example, in [27], Yong *et al.* formulate the interaction as a non-cooperative non-zero-sum two-player game. Using this framework, they identify cases, both analytically and experimentally, where the Nash equilibrium is not Pareto efficient [8]. Keralapura *et al.* [21] study the dynamics due to the interaction between overlay routing and traffic engineering. They show that uncoordinated effort by both entities to recover from failures may cause performance degradation for both overlay and non-overlay traffic.

There are also other ways in which end users can selfishly optimize the performance of their traffic. For example, a user can greedily inject traffic into a network. A number of papers (*e.g.*, [1]) consider such a congestion game. In practice, it is possible to have a hybrid game that consists of a route selection game and a congestion game, but we defer its study to future work.

III. Network Model

In this section, we describe our network model, especially the network-level routing protocols. In the next section, we describe the schemes of how traffic demands are routed through the network. In Section VI, we describe the network topologies, traffic demands, and latency functions that we use to instantiate our network model.

Physical network: We study the performance of realistic physical networks. We model a physical network as a directed graph $G = (V, E)$, where V is the set of nodes, and E the set of directed links. We assume that the latency of each physical link is a function of its load. The exact latency functions we use will be described in Section VI-C.

Demands: We partition network traffic into demands. A demand represents a given amount of traffic from a source to a destination. In particular, we identify a special type of demands, called infinitesimal demands. A collection of infinitesimal demands models a large aggregation of independent, small transactions such as web transactions, and the generator of each transaction makes an independent decision.

Overlays: An overlay consists of overlay nodes, directed overlay links, and a set of demands originated from the overlay nodes. The overlay nodes agree to forward each other's traffic along one or more overlay links. The physical route for an overlay link is dictated by network-level routing and may involve multiple physical links. Different overlay links may share one or more physical links. The overlay nodes and overlay links form the overlay topology. To limit the parameter space, we only consider the *fully connected* overlay topology in this work. That is, we assume that there is an overlay link between every pair of overlay nodes. We plan to investigate the effects of different overlay topologies in our future work.

Users: We assume that the network consists of a collection of users. Each user decides how its traffic should be routed. The objective of a user is to minimize the average latency of its traffic. We choose to use latency as the optimization objective of selfish routing for the following reasons: 1) many applications such as short Web transfers and IP telephony require low latency; 2) most previous theoretical analyses are based on latency, and one of the major objectives of this study is to investigate whether the theoretical worst-case results apply to Internet-like environments. We plan to investigate the effects of alternative routing objectives in our future work.

Route controller: Besides users, we also have a route controller, which controls the network-level routing in the physical network. (We use network-level routing and physical routing interchangeably in this paper.) We consider the following network-level routing: (i) OSPF, which uses shortest-path with equal-weight splitting, and (ii) MPLS, which uses the more general multi-commodity flow routing. For OSPF routing, we consider three weight assignments:

- Hop-count OSPF routing, which assigns a unit weight to each physical link;
- Random-weight OSPF routing, which assigns a random weight to each physical link;
- Optimized-compliant OSPF routing, which sets OSPF link weights to minimize network cost [13] (see Section VI-D), when all traffic demands are compliant, and thus follow the routes determined by the network. The network cost is a piece-wise linear convex function over all links. This metric has been considered as a good objective for traffic engineering because it not only avoids overloading physical links, but also avoids taking very long paths [13], [14].

We represent network-level routing by a routing matrix R, where $R[p, e]$ specifies the fraction of traffic between the source-destination pair p that goes through the physical link e. The routing matrix R is computed by the routing protocol under study.

In our study, the route controller can change network routing to optimize overall network performance. In other words, it can perform traffic engineering. An MPLS-based route controller can directly adjust the routing matrix R. An OSPF-based route controller can adjust the weights of the physical links to influence network routing [13], [14].

IV. Routing Schemes

For a comprehensive study, we consider the following five routing schemes: (i) source routing, (ii) optimal routing, (iii) overlay source routing, (iv) overlay optimal routing, and (v) compliant routing. Below we describe these routing schemes in details.

A. Routing on the physical network

The first two routing schemes allow a user to route its traffic directly through any paths on the physical network.

Source routing: Source routing results in selfish routing, since the source of the traffic makes an independent decision about how the traffic should be routed. The selfish routing scheme studied in most previous theoretical work is source routing.

Optimal routing: Optimal routing refers to *latency* optimal routing. It models a scenario where a single authority makes the routing decision for all the demands to minimize the average latency.

B. Overlay routing

The next two routing schemes are the overlay versions of source routing and optimal routing.

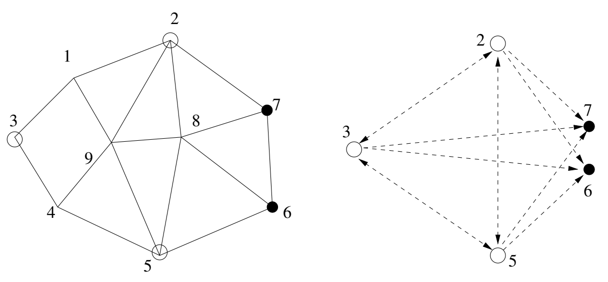
Overlay source routing: Overlay source routing is selfish routing through overlay nodes. Similar to source routing, it is the traffic source that controls the routes.

Overlay optimal routing: Overlay optimal routing refers to overlay *latency* optimal routing. It models a scenario where the demands in the overlay have complete cooperation in minimizing the average latency.

As mentioned in Section I, overlay routing is different from routing directly on the physical network. In particular, the physical route for an overlay link is dictated by network-level routing and may involve multiple physical links. Moreover, different overlay links may share common physical links, and therefore may interfere with each other. As a result, we cannot apply the traditional linear approximation algorithms to compute traffic equilibria for such schemes.

We use the following approach to compute traffic equilibria for overlay routing. For each overlay, we build a logical network from the physical network. The nodes in the logical network consist of the union of the nodes in the overlay and the nodes that are the destinations of nonzero demands in the overlay. The links in the logical network consist of all the overlay links, as well as a link from each overlay node to each node that is the destination of some traffic demands but does not belong to the overlay.

Given this model, each logical link can be mapped to a collection of physical links. More specifically, assume that the logical link p is for the source-destination pair p (we use the same symbol p to denote the logical link p and the source-destination pair p), then the logical link consists of all the physical links e such that $R[p, e] > 0$. If a demand sends f units of traffic through a logical link p, then each physical link e will carry $f \cdot R[p, e]$ amount of traffic for this demand. Fig. 1 shows an example of a physical network, and the logical network for an overlay formed by nodes 2, 3, and 5.



(a) Physical Network (b) Logical Network of an Overlay

Fig. 1. A physical network and the logical network for the overlay formed by nodes 2, 3, and 5. Nodes 6 and 7 are not overlay nodes but nodes 2, 3, and 5 have demands to them. The logical link from node 2 to 5 consists of two physical paths: 2 to 9 to 5, and 2 to 8 to 5, if hop-count OSPF routing is used.

C. Compliant routing

For comparison, we also consider the default network-level routing, which we term compliant routing.

Compliant routing: Traffic demands using compliant routing follow the routes determined by the network-level routing protocol.

V. COMPUTING TRAFFIC EQUILIBRIA

We evaluate each of the preceding selfish routing schemes by computing its performance at traffic equilibria. Using a game-theoretic approach, we define a traffic equilibrium as a state where no user can improve the latency of its traffic by unilaterally changing the amount of traffic it sends along different network paths. One possible way of computing traffic equilibria is through simulation. More specifically, one could simulate the moves of each individual user and wait until the system reaches equilibrium. However, given the size of the network we are considering (see Section VI-A), such simulation-based approach may take a prohibitively long time to converge. Instead, we compute traffic equilibria directly using the following algorithms.

Computing traffic equilibrium for non-overlay traffic: As shown in [34], the traffic equilibrium for selfish traffic is achieved when the integral of the latency function over all traffic is minimized. In comparison, the traffic equilibrium for latency optimal routing is achieved when the latency over all traffic is minimized. For both cases, we use a linear approximation algorithm (a variant of Frank-Wolfe algorithm) [12] to compute traffic equilibrium with the corresponding objective functions. The linear approximation algorithm is a gradient algorithm for solving non-linear optimization problems. Specifically, in each iteration we compute shortest paths based on the current traffic assignment, and use them to construct the gradient direction. We then move towards that direction by taking a step size that optimizes the objective function. The number of iterations is controlled by the stopping condition from [12]. When the link latency functions satisfy the monotonicity condition, which is the case for our latency functions, there is a unique equilibrium.

Computing traffic equilibrium for selfish overlay routing: Using the logical networks we described in Section IV, we can compute the traffic equilibrium of overlay routing by either a relaxation algorithm or a modified linear approximation algorithm.

$\triangleright$ Assume $l_e(x)$ is increasing and convex for any edge e.
$\triangleright$ Assume $xl_e(x)$ is convex for any edge e.
$\triangleright$ If the overlay is latency optimal, $f = \sum_e xl_e(x)$;
$\triangleright$ otherwise, $f = \oint l_e(x)$;

set other overlay's traffic as background traffic
repeat
  assume the current traffic vector on each edge is x_t
  determine link latency according to x_t
  use Dijkstra's algorithm to find all-or-nothing traffic assignment y_t
  use line search to find optimal λ so that $f(x_t + \lambda(y_t - x_t))$ is minimal.
until (best lower bound gap $<$ threshold)

Fig. 2. The linear approximation algorithm to compute the best response of source routing or overlay routing, when the network is symmetric, assuming the other overlay's traffic is background.

Specifically, for an asymmetric logical network (*i.e.*, any two logical links that share the same physical link send the same fraction of traffic through the physical link; an example of a symmetric logical network is OSPF routing without equal weight splitting), we can formulate the problem as an optimization problem by using a line integral to replace the normal summation of cost on each link. As a result, we still can use the linear approximation algorithm. Fig. 2 specifies the structure of our algorithm. Note that for overlay networks, the traffic equilibrium may not be unique [22], [2], and our algorithm identifies only one equilibrium.

For a logical network that is asymmetric (*i.e.*, there are two logical links that share the same physical link but send different fractions of traffic through the physical link), we use Jacob's relaxation algorithm on top of Sheffi's diagonalization method [37] to determine the traffic equilibrium, since in this case we cannot formulate the equilibrium problem as an optimization problem.

Computing traffic equilibrium for multiple overlays: Guaranteeing convergence poses a major challenge in computing traffic equilibrium when there are multiple overlays. To this end, we use the relaxation framework proposed in [25], [42] to ensure convergence to one equilibrium. Fig. 3 shows the algorithm. The basic structure of the algorithm is that in each round, each overlay computes its best response by considering the other's traffic as background traffic. Then the best response

```
▷ N is the number of overlays.
▷ x_t(i) is a vector of overlay i's traffic at round t.
▷ y_t(i) is the best response of overlay i at round t.
▷ Assume Σ_t α_t → ∞; α_t → 0 as t → ∞.

repeat
    assume the traffic state is x_t(i) of overlay i
    for each i
        computes its best response y_t(i),
            assuming other overlays as background.
    for each overlay i
        set x_{t+1}(i) ← (1 − α_t)x_t(i) + α_t y_t(i).
until (change between round < threshold)
```

Fig. 3. The relaxation framework to compute the traffic equilibrium of N overlays.

and the previous state are merged using the relaxation factor α_t. The conditions $\sum_t \alpha_t \rightarrow \infty$ and $\alpha_t \rightarrow 0$ as $t \rightarrow \infty$ guarantee convergence to one equilibrium.

VI. Evaluation Methodology

In this section, we first describe the network topologies, traffic demands, and link latency functions used in our evaluation. Then we discuss the performance metrics that we use as a basis for comparing the efficiency of different routing schemes.

A. Network topologies

We use both real and synthetic topologies in our evaluation.

Real topology: We use a real router-level backbone topology from an operational tier-1 ISP, referred to as $ISPTopo$, with on the order of a hundred backbone routers connected by OC48 (*i.e.*, 2.48 Gbps) and OC192 (*i.e.*, 10 Gbps) links (the exact numbers are omitted for proprietary reasons). For each link in the real topology, we use the actual link capacity in our study. The propagation delay of each link is estimated using the actual fiber length divided by the speed of light.

Rocketfuel topologies: Rocketfuel applies several effective techniques to obtain fairly complete ISP maps [38]. We use the POP-level maps published by the authors, shown in Table I, as part of our topologies. For each Rocketfuel topology, we use two bandwidth settings: all links are either OC3 (*i.e.*, 155 Mbps) or OC48 (*i.e.*, 2.48 Gbps). The propagation delay of each link is approximated using geographical distance divided by the speed of light.

ISP	Loc.	#Nodes	#Non-leaf Nodes	#Edges
ATT	US	108	30	282
Abovenet	US	22	13	160
Exodus	US	22	17	102
Level3	US	53	37	912
Sprint	US	44	21	212
Verio	US	122	82	620
EBONE	Intl.	28	25	132
Telstra	Intl.	58	8	120
Tiscali	Intl.	51	38	258

TABLE I

ISP topologies as measured by Rocketfuel.

Random topologies: In addition to real topologies, for diversity we also randomly generate power-law topologies using BRITE [29], since a number of papers (*e.g.*, [9], [40]) have shown that the power-law captures the Internet structure quite well. We generate 100-node router-level topologies with edge density (*i.e.*, the number of neighboring nodes that each new node connects to) varying from 2 to 10. In the following sections, we use PowerDn to denote a power-law topology with edge density n. For each power-law topology, we use two bandwidth settings: all links are either OC3 or OC48. The propagation delay of each link is drawn uniformly between $0 - 10$ ms.

B. Traffic demands

We use both real and synthetic traffic demands in our evaluation.

Real traffic demands: Our real traffic demands are estimated from SNMP link data using the *tomogravity* method [44], which has been shown to yield accurate estimates especially for large traffic matrix elements. We use the backbone router to backbone router traffic matrices during three randomly chosen hours in November 2002.

Synthetic traffic demands: The real traffic demands are only available for $ISPTopo$. For the other topologies, we generate synthetic traffic demands as follows. For a Rocketfuel topology, we generate synthetic traffic by randomly mapping POPs in $ISPTopo$ to non-leaf nodes in the Rocketfuel topology, using several different random seeds. Specifically, let $m(.)$ denote a random mapping from the cities in $ISPTopo$ to those in a Rocketfuel topology. Let $T(s,d)$ denote the traffic demand from city s to city d in $ISPTopo$. Then the traffic demand from city $m(s)$ to city $m(d)$ in the topology under study is set to $T(s,d)$. For synthetic power-law topologies, we perform similar mappings at the router level to derive demands.

Load scale factor: To control system load, we scale up the demands so that when all the traffic is compliant and routed based on *shortest hop-count*, the maximum link utilization is $100 \cdot F\%$, where F is a *load scale factor* (sometimes abbreviated as LSF).

C. Link latency functions

As shown in [34], link latency functions play an important role in determining the effectiveness of selfish routing. In our evaluations, we use five representative latency functions: M/M/1, M/D/1 [17], P/M/1, P/D/1 [18], and BPR [6]. We also implement piecewise-linear, increasing, convex functions to approximate any other latency functions. In all latency functions, we include a term for propagation delay (Section VI-A shows how we determine its value for each physical link).

Our first two latency functions belong to the general M/G/1 class of latency functions: M/M/1 and M/D/1. For a M/G/1 queue, the latency can be expressed as $l(x) = \frac{1}{\mu} + \frac{x\cdot(1+\sigma^2\mu^2)}{2\mu(\mu-x)} + prop$, where x is the traffic load, μ the link capacity, σ the standard deviation of the service time, and $prop$ the propagation delay. The M/M/1 latency function is M/G/1 with $\sigma = \frac{1}{\mu}$; therefore $l(x) = \frac{1}{\mu-x} + prop$. The M/D/1 latency function is M/G/1 with $\sigma = 0$; therefore $l(x) = \frac{0.5}{\mu-x} + \frac{0.5}{\mu} + prop$. To avoid the discontinuity when the load approaches capacity, we approximate the M/M/1 or M/D/1 function with a linear function beyond 99% utilization. To test sensitivity to the threshold, we also try 90% and 99.9%. The results are similar,[1]

[1]We can construct scenarios where different thresholds can yield significantly different results. However, our interest is in typical Internet-like environments.

and in the interest of brevity we present the results using 99% as the threshold.

Our next two latency functions, P/M/1 and P/D/1, have heavy-tail inter-arrival times. Here P stands for Pareto. We set the shape parameter $\beta = 1.5$ so that the resulting distribution has infinite variance. Since there is no closed-form expression for either P/M/1 or P/D/1, we approximate each of them using a piecewise-linear, increasing, convex function. We use the results in [18] to approximate P/M/1. For P/D/1, we derive a linear approximation of its shape using ns-2 [30] simulations. Specifically, we generate Pareto traffic to compete for a single bottleneck link with a large FIFO drop-tail queue and observe the latency as we vary the load.

For comparison purposes, we also run some experiments with the latency function BPR [6], which is used as a standard latency function in transportation networks. The expression for this latency function is $l(x) = prop \cdot \left[1 + 0.15\ (\frac{x}{\mu})^4\right]$. Table II summarizes the above five latency functions.

Notation	Latency function
M/M/1	$l(x) = \frac{1}{\mu - x} + prop$
M/D/1	$l(x) = \frac{0.5}{\mu - x} + \frac{0.5}{\mu} + prop$
P/M/1	approx. with Pareto $\beta = 1.5$, see [18]
P/D/1	approx. with Pareto $\beta = 1.5$
BPR	l(x) = $prop \cdot \left[1 + 0.15\ (\frac{x}{\mu})^4\right]$

TABLE II

LINK LATENCY FUNCTIONS.

D. Performance metrics

We use the following performance metrics to evaluate routing efficiency: (i) average latency, (ii) maximum link utilization, and (iii) network cost. The first metric reflects end-to-end user performance, while the next two reflect the perspective of network operators, who aim to avoid link overloads in their networks. These performance metrics are computed from traffic equilibria, as we discussed in the previous section.

The *utilization* (or *traffic intensity*) of a link is the amount of traffic on the link divided by its capacity. When a link's utilization is beyond 100%, the link is overloaded. The maximum link utilization is the maximum utilization over all links in a network.

The maximum link utilization is an intuitive metric; however, it is dominated by a single bottleneck, as pointed out in [13]. To get a more complete picture, we also adopt a metric to capture the overall network cost. According to [13], [14], the cost of a link can be modeled using a piecewise-linear, increasing, convex function with slopes specified as follows:

$$u_e(x/c) = \begin{cases} 1 & : \quad x/c \in [0, 1/3) \\ 3 & : \quad x/c \in [1/3, 2/3) \\ 10 & : \quad x/c \in [2/3, 9/10) \\ 70 & : \quad x/c \in [9/10, 1) \\ 500 & : \quad x/c \in [1, 11/10) \\ 5000 & : \quad x/c \in [11/10, \infty), \end{cases}$$

where x is the load on link e, and c its capacity. We refer to the points at which the slope changes (*e.g.*, $1/3$ and $2/3$) as the cut-points. The overall network cost is the sum of all links' costs. In [13], Fortz, Rexford, and Thorup show that OSPF weights derived from one set of cut-points and slopes also tend to give good performance for other sets of cut-points and slopes. Therefore the above cost function is a general metric to consider. For all three metrics, lower values are preferred.

VII. SELFISH SOURCE ROUTING

We first investigate the performance of selfish source routing; that is, all the demands are infinitesimal and the selfish traffic can use any routes in the physical network. This is the type of selfish routing scheme analyzed in most theoretical studies. As shown in [34], the worst-case latency degradation of selfish source routing compared with optimal routing can be unbounded due to lack of cooperation. In this section, we aim to answer the following question: how does selfish routing perform in Internet-like environments?

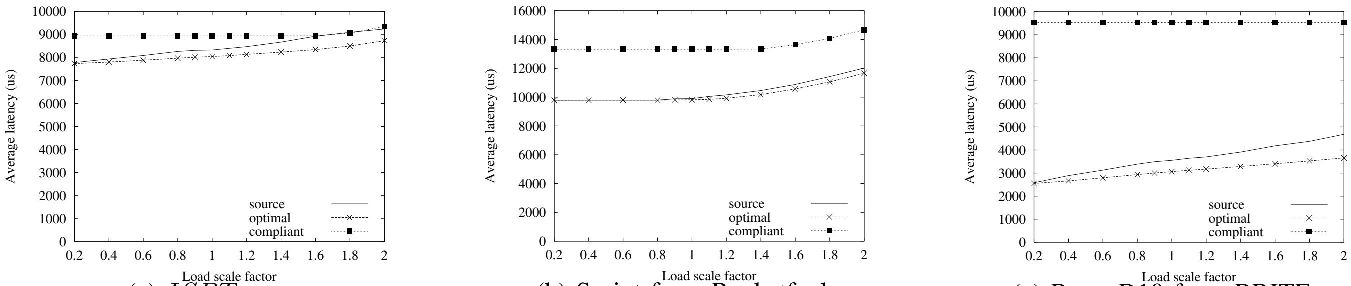
A. Are Internet-like environments among the worst cases?

Effects of network load: We begin our investigation of selfish routing by varying network load. Fig. 4 shows the latency for three representative topologies, as we vary the network load scale factor from 0.2 to 2.

We make the following observations. First, under various loads, selfish routing yields lower latency than compliant routing, which is based on optimized-compliant OSPF weights. This result complements the previous findings, such as Detour [36] and RON [3], and shows that the performance benefits of selfish routing over compliant routing exist even in a single AS network; moreover such benefits do not disappear even if all traffic is selfish (as opposed to just having a small portion of selfish traffic in RON). It is not surprising that compliant routing results in higher latency, because the OSPF weights are optimized mainly to avoid link overloads rather than minimize end-to-end user latency. As we will see later, the lower latency of selfish routing comes at the cost of increased congestion on certain links.

Second, compared with optimal routing, selfish routing yields similar average latency—the difference is close to 0 in most cases and is always within 30%. In other words, unlike the theoretical worst cases, the price of anarchy in Internet-like environments is close to 1. There are two main reasons for this. First, the worst-case result arises when there is mismatch between link bandwidth and link propagation delay (*e.g.*, when there are two parallel links between the source and destination, where one link has high propagation delay but large bandwidth whereas the other link has low propagation delay but small bandwidth). Such mismatch is not common in current Internet-like topologies. Second, under realistic network topologies and realistic traffic demands, traffic is spread across the network and only a few links get congested even with selfish routing. This is because real networks are designed so that even under common failures the network can still carry all of the traffic (often without having to reconfigure the routing). Moreover, the topology is constrained by the coarse resolution we have for link capacities: there are only a small number of available link capacities, *e.g.*, OC3, OC48, OC192. The net result is that there is considerable redundant bandwidth in the network. So capacities rarely match the traffic we are expected to carry other than a few bottleneck links, and there are only a small number of local hot spots. The above two factors make selfish routing perform close to optimal.

Effects of network topologies: Next we examine the effects of network topologies on the latency of selfish routing. Fig. 5 compares the latency of different routing schemes when the link latency function is M/M/1, the load scale factor is 1.0, and the links' bandwidth in random topologies is OC48. The links' bandwidth in $ISPTopo$ is according to the actual topology.



(a) $ISPTopo$ (b) Sprint from Rocketfuel (c) PowerD10 from BRITE

Fig. 4. Selfish source routing: comparison of user latency using M/M/1 link latency under various network loads.

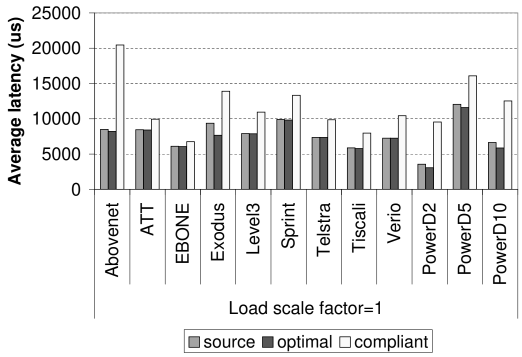


Fig. 5. User latency for all topologies with the M/M/1 latency function and load scale factor 1. Selfish stands for selfish source routing; optimal stands for optimal routing; compliant stands for optimized-compliant OSPF routing. The other figures in this section use the same notation.

As Fig. 5 shows, network topologies have a pronounced effect on the relative performance of selfish and compliant routing. For example, in the Abovenet and power-law topologies, the latency achieved by selfish routing is less than half of that incurred by compliant routing. A detailed look at these two topologies shows that these two topologies have mesh-like connectivity. Therefore selfish routing is likely to find more paths, and achieves much lower latency. However, in all topologies, we observe that selfish routing consistently yields close to optimal latency. Similar results are observed when the links' bandwidth in random topologies is changed to OC3.

Effects of latency functions: Finally, we study how different latency functions affect the latency of selfish routing. From Fig. 6, we observe similar latency across different latency functions. When comparing the latency achieved by different routing schemes, we see that the performance of selfish routing is close to that of optimal routing and noticeably better than that of compliant routing.

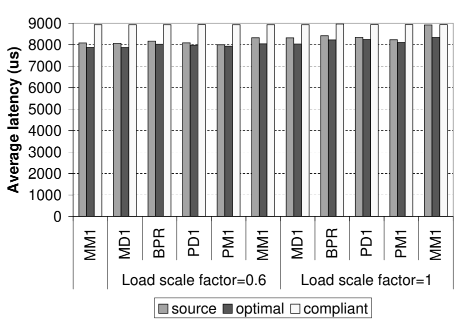


Fig. 6. User latency for $ISPTopo$ under various latency functions.

B. What is the system-wide cost for selfish source routing?

The previous subsection shows that unlike theoretical worst cases, selfish source routing in Internet-like environments incurs low latency. A natural question is whether the low latency comes at the expense of increased system-wide cost. We examine this issue by comparing different routing schemes based on two metrics: (i) maximum link utilization and (ii) network cost, both defined in Section VI-D.

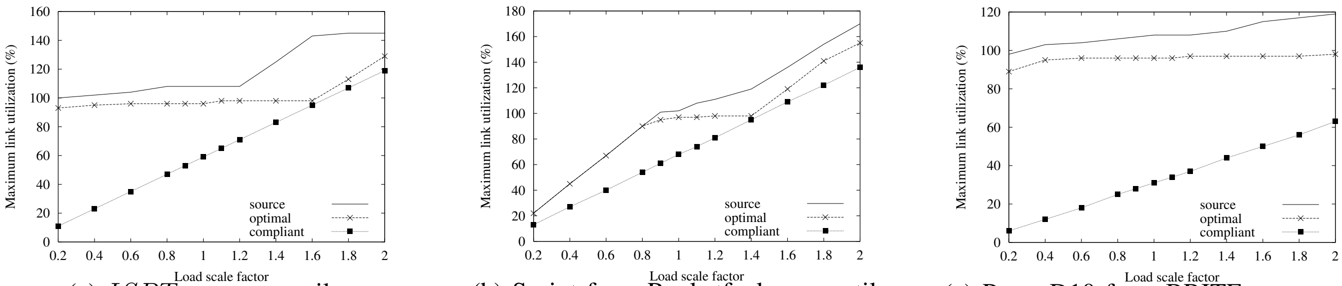
Effects of network load: We start by examining the impact of network load. Fig. 7 shows the maximum link utilization for the same network configurations as those in Fig. 4. From Fig. 7, we observe that in compliant routing, maximum link utilization increases linearly with offered load. This is expected since we use the same set of weights to scale the traffic (see Section VI-B). In comparison, both optimal routing and selfish routing can cause high link utilization even when the overall offered load is low. For example, in both $ISPTopo$ and PowerD10 topologies, at a load factor of 0.2, the maximum link utilization of optimal routing is close to 90% and that of selfish routing is close to 100%. This result occurs because both optimal routing and selfish routing aim to choose shortest paths; thus they are more likely to cause congestion there, whereas compliant routing more uniformly spreads traffic across the entire network to avoid link overloads at the cost of longer end-to-end paths. The high network utilization is undesirable, since many backbone networks are kept at a load well below 50% so that there are enough backup paths during link or router failures [19].

Effects of network topologies: Next we verify the above observations by varying the network topologies. As shown in Fig. 8, selfish routing consistently yields the highest maximum link utilization and network cost in all topologies. For example, in the Exodus network, the maximum link utilization achieved by selfish routing is 40% higher than that of optimal routing and 80% higher than that of compliant routing. For the same network, the network cost of selfish routing is over an order of magnitude higher than that of optimal routing or compliant routing. Similar results are observed when the links' bandwidth in random topologies is changed from OC48 to OC3.

Effects of latency functions: The results based on other latency functions are qualitatively the same, as shown in Fig. 9. Since both latency and network cost/utilization are not very sensitive to latency functions for the topologies that we consider, in the following sections we focus on the M/M/1 latency function. Moreover, we show only the maximum link utilization, since it is more intuitive and it gives consistent results as network cost.

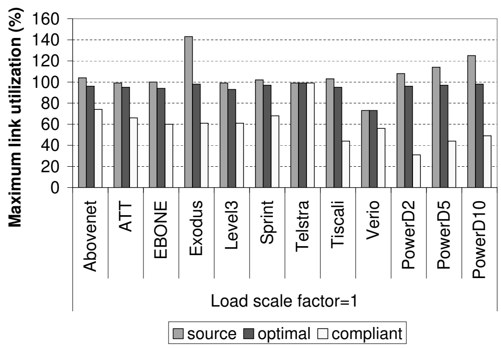
C. Summary

To summarize, in this section we compare the performance of different routing schemes using realistic network topologies and traffic demands. Our results show that unlike the

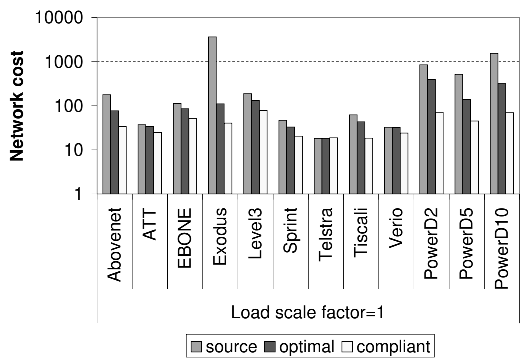


(a) $ISPTopo$, max. util. (b) Sprint from Rocketfuel, max. util. (c) PowerD10 from BRITE, max. util.

Fig. 7. Selfish source routing: comparison of maximum link utilization using M/M/1 link latency under various network loads.

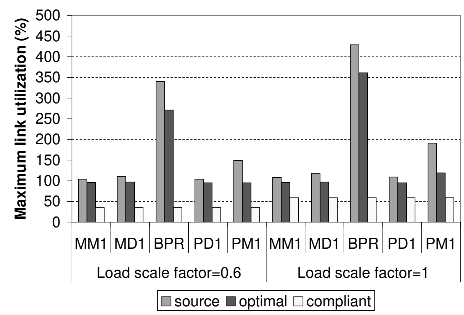


(a) Maximum link utilization (%)

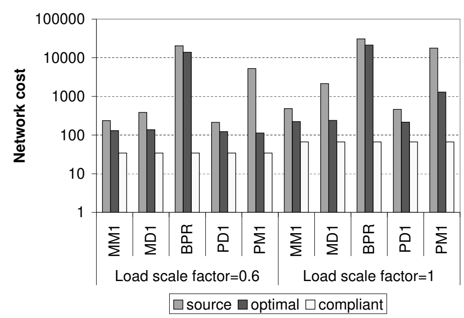


(b) Network cost

Fig. 8. Selfish source routing: comparison of maximum link utilization and network cost using M/M/1 link latency across different network topologies. The links' bandwidth in random topologies is OC48, and the bandwidth in $ISPTopo$ is based on the actual setting.



(a) Maximum link utilization



(b) Network cost

Fig. 9. Selfish source routing: comparison of maximum link utilization and network cost across different latency functions.

theoretical worst cases, selfish source routing in Internet-like environments is effective in choosing shortest paths, and yields close to optimal average latency. On the other hand, this may come at the cost of overloading links on the shortest paths, which suggests that selfish routing may potentially have a negative impact on traffic engineering. We will further investigate the issue in Section X.

VIII. Selfish Overlay Routing

The previous evaluations consider selfish source routing. However, as we discussed in Section I, in practice, end users often do not have complete routing control. We initially expected that reducing routing flexibility would increase both latency and link utilization, since users lose fine-grained control over routing. However, as we will see, this is often not the case.

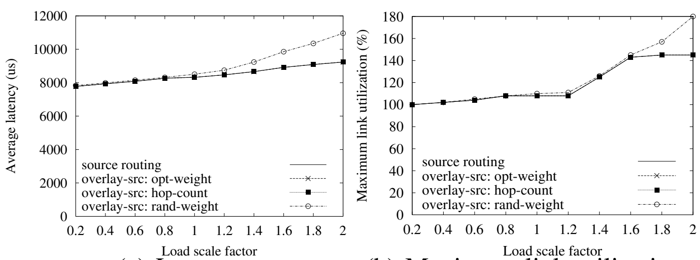
A. Does selfish overlay routing perform well when every node is in the overlay?

We first consider an overlay that consists of all network nodes. Note that even if the overlay includes all network nodes, routing on an overlay is still different from routing on the physical network in that the latter has access to all network resources, but this may not be the case for the former. For example, the network-level routing can easily prevent any overlay traffic from using a particular link by setting its corresponding column in the routing matrix to 0 (in OSPF this can be achieved by assigning a large weight to the link). As a result, certain physical routes cannot be implemented by any overlay routing schemes.

In our evaluation, we use the same network setting as before, except that the routes between any pair of overlay nodes are no longer determined by end users, but by the network-level routing. We adopt OSPF for network-level routing and use the three OSPF weight assignments as described in Section III.

Fig. 10 shows the performance of overlay source routing for the $ISPTopo$ network, as we vary network load. In both figures, three of the four curves overlap, namely source routing, overlay source routing when the network-level routing uses optimized-compliant OSPF weights, and overlay source routing when the network-level routing uses hop count. This suggests that routing constraints, whether based on hop-count or optimized-compliant weights, have little effect on user latency or system-wide cost. This result came as quite a surprise since our initial conjecture was that routing constraints would degrade performance. In contrast, when the network-level routing uses random weights, we observe much higher delay and link utilization.

As we explained in [31], when an overlay covers all network nodes and link weights satisfy triangular inequality, routing on the overlay has as much routing flexibility as directly routing on the underlying physical network. Since hop-count-based OSPF weights satisfy triangular inequality and optimized-compliant OSPF weights satisfy triangular inequality to a large extent, they both perform well. When triangular inequality is violated, this essentially prunes out the link with the highest weight in the triangle. This reduces the network resources available to the selfish overlay and can result in higher latency and link utilization. With random OSPF weights, violations of triangular inequality are common and therefore the network resources available to the overlay are significantly reduced. This explains why we see substantially higher latency and maximum link utilization with random OSPF weights. We will show later in Section X that selfish overlay routing interacts poorly with OSPF optimizer for exactly the same reason.



(a) Latency (b) Maximum link utilization

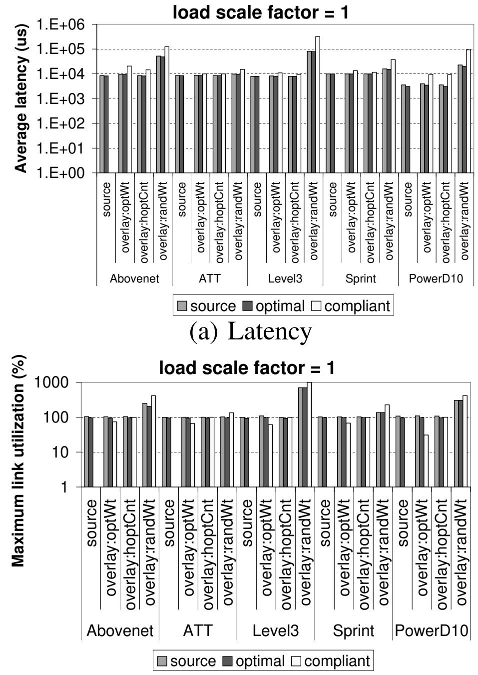
Fig. 10. Selfish overlay routing: comparison of user latency and maximum link utilization for the $ISPTopo$ topology.

We further verify the above observations by using different network topologies; the results are shown in Fig. 11. As before, random OSPF weights continue to yield substantially higher delay and maximum link utilization, while the performance of the other three is close to each other. This confirms our previous findings. When comparing the performance across different routing schemes, we observe that selfish routing continues to result in close to optimal average latency. Moreover, it yields noticeably lower latency than compliant routing in most cases. However, this lower latency often comes at the cost of higher maximum link utilization.

B. Does selfish overlay routing perform well when only some nodes are in the overlay?

The previous evaluation includes all of the network nodes in an overlay. In practice, an overlay may only have *partial coverage*, *i.e.*, only a fraction of the nodes are in the overlay. In such a case, the routing choice is further constrained, which may have an impact on the performance. Below we investigate this issue in detail.

Effects of only covering edge nodes: In our first experiment, we form an overlay from all of the edge nodes in $ISPTopo$, and route all demands among these edge nodes

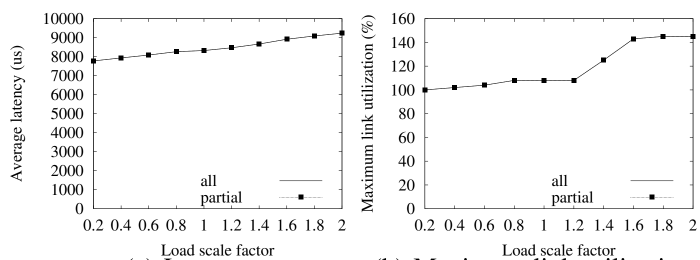


(a) Latency

(b) Maximum link utilization

Fig. 11. Selfish overlay routing: comparison of user latency and maximum link utilization for different network topologies.

through the overlay. We then compare the performance with what we achieve when the same set of demands is routed through an overlay that includes all of the network nodes. As shown in Fig. 12, the curves of full overlay coverage almost completely overlap with those of partial coverage, in terms of both latency and maximum link utilization. These results are likely due to the fact that the Internet backbone is fairly well-connected and well-provisioned; therefore, even though end users can only forward traffic through edge nodes, they do not lose much flexibility in controlling their routes.



(a) Latency (b) Maximum link utilization

Fig. 12. Effects of partial coverage ion the performance of selfish overlay routing. Here edge nodes in $ISPTopo$ belong to an overlay, and OSPF weights are set according to hop count.

Effects of random partial coverage: In our second experiment, we uniformly choose a fraction of network nodes to form an overlay. We observe that the latency is similar as the overlay coverage changes from 20% to 100%. On the other hand, as shown in Fig. 13, full overlay coverage incurs a slightly higher maximum link utilization than partial coverage, because as more nodes and links are included, it becomes more likely that the overlay has popular shortcuts that get overloaded.

C. Summary

To summarize, in this section we investigate the effects of overlay routing constraints. We show that if the physical

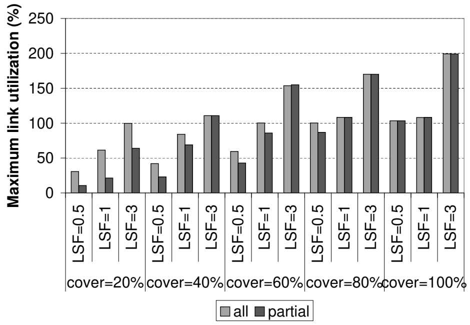


Fig. 13. Effects of partial coverage in $ISPTopo$ with random node selection on maximum link utilization.

network uses a routing scheme that satisfies triangular inequality, the overlay has full control over how its traffic is routed through the physical network. In the context of OSPF, the only way in which OSPF can affect overlay traffic is by violating triangular inequality, which effectively reduces network resources and may therefore degrade both user and system-wide performance. We also show that like source routing, overlay source routing reduces latency at the expense of higher network cost. Finally, we observe that the effects due to partial coverage are small in backbone topologies.

IX. Interactions among Competing Overlays

So far we have only considered either a large number of independent, small users using source routing (Section VII) or a single selfish overlay (Section VIII). In practice, it is possible that multiple overlays and background traffic will share the same physical network, and different traffic will compete against one another for the shared network resources. We call such interactions *horizontal interactions*.

A. What is the relative competitiveness of two routing schemes?

We start by looking at the interactions between any two types of traffic. The objective of this subsection is to evaluate the "friendliness" of different types of routing schemes. We use R_1/R_2 to denote that the routing scheme of the foreground traffic is R_1, and that of the background is R_2. Here R_i is either overlay source routing, overlay optimal routing, or compliant routing.

We study how traffic using two different routing schemes compete against each other in different topologies. In this set of experiments, we put the competing demands at the same nodes, and we set both the foreground and background traffic to be 50%. In other words, the two types of competing traffic have the same amount of traffic and the same set of overlay nodes. Fig. 14 shows the results. We make two observations. First, the performance difference between compliant routing and the competing overlay routing scheme varies across different topologies. For example, the performance difference is larger in the Abovenet and power-law topologies. This is consistent with Fig. 5 and can again be explained by the better connectivity of these topologies (see Section VII-A for details). Comparing the results in Fig. 14 with those in Fig. 5, we observe that the latency of the compliant traffic is not substantially increased, which indicates that selfish routing does not hurt the performance of compliant routing in this environment. Second, overlay source routing achieves similar performance compared to overlay optimal routing. This suggests that the performance gain of cooperative overlay optimal routing over uncooperative overlay source routing is not significant.

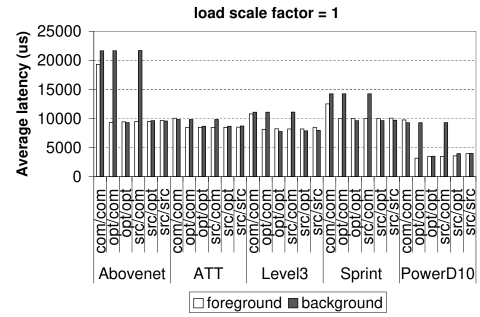


Fig. 14. Coexistence of two routing schemes: varying network topologies.

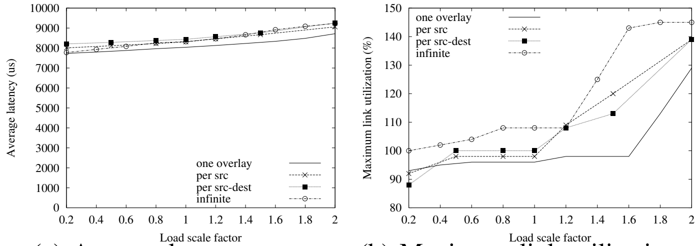
We also explored the impact of network-level routing schemes on the horizontal interactions as follows. We set both the foreground and background traffic in $ISPTopo$ to be 50%, and we vary how OSPF weights are set. We found that the foreground and background traffic experience similar latency in most cases, except when OSPF weights are set randomly. When OSPF weights are set randomly, compliant traffic incurs about twice as much delay as that of the competing overlay source routing or overlay optimal routing. This indicates that inappropriate OSPF weights can significantly degrade the performance of compliant traffic. In comparison, a selfish overlay is able to reduce the latency of its traffic, as it looks for better alternative paths. Interestingly, this also has a positive side effect: it helps to reduce the load on the links used by the competing compliant traffic, thereby cutting the latency of the latter by half. When the network-level routing scheme is configured reasonably, different overlay routing schemes can coexist well.

B. Can many overlays coexist well?

Next we study horizontal interactions by varying the number of overlays. Each overlay uses overlay optimal routing and covers all network nodes. Fig. 15 shows the result for $ISPTopo$, when the number of overlays is changed in the following ways: (i) one overlay, which includes all the demands; (ii) overlay per source, where each overlay includes all demands originated from a source; (iii) overlay per source-destination pair, where each overlay includes all demands between a source and destination pair; (iv) an infinite number of overlays, where each overlay has infinitesimal demands. We use the relaxation framework to compute the traffic equilibria for (ii) and (iii). For (iv), we note that having an infinite number of overlays with infinitesimal demands is equivalent to having all the infinitesimal demands on a single overlay, each of which tries to minimize its own latency. In other words, (iv) is equivalent to having a single overlay using overlay source routing. Thus we do not need to use the relaxation framework. From Fig. 15, we observe that there is only a slight difference in user latency due to variations in the number of overlays. Results from other topologies confirm this finding, which suggests that performance degradation due to competition among overlays is not significant.

C. Summary

To summarize, with reasonable OSPF weights (*e.g.*, hop-count), different routing schemes can share network resources



(a) Average latency (b) Maximum link utilization

Fig. 15. Coexistence of multiple overlays in $ISPTopo$.

reasonably well without hurting each other; with bad OSPF weights, selfish overlays improve performance both for themselves and for compliant traffic.

X. Selfish Routing vs. Traffic Engineering

So far all of our experiments assume that the network-level routing is fixed. We find that while selfish routing can achieve close to optimal latency, it often increases maximum link utilization and network cost. In practice, the network-level routing may be constantly changing since one principal goal of traffic engineering is to reduce network cost by adapting the network-level routing in response to varying traffic patterns. This motivates us to examine the interactions between selfish routing and traffic engineering, which we term *vertical interactions*. This vertical interaction can be considered as an iterative game played between overlay networks and traffic engineering. More specifically, we ask the following basic question: *Will the system reach a state with both low latency and low network cost, as selfish routing and traffic engineering each tries to minimize its own cost function by adapting to the other process?*

Below we evaluate vertical interactions in the context of OSPF and MPLS route optimization. As we will see, OSPF route optimization provides little control over selfish traffic. As a result, the system performance, both in terms of user latency and network cost, is no better than using hop-count-based OSPF routing. In contrast, MPLS provides fine-grained control and can potentially lead to better performance.

A. Specification of vertical interactions

We specify vertical interactions as an iterative process between the two players: traffic engineering and selfish overlays.

Traffic engineering adjusts physical routing based on network traffic patterns, which are usually in the form of a traffic matrix. More specifically, let T_t denote the estimated traffic matrix for time slot t, then $T_t(s,d)$ represents the total traffic from source s to destination d during the time slot t. Traffic engineering takes T_t as input, and computes a routing matrix R_t to optimize network performance. In our study, we assume T_t is given. In reality, T_t can either be obtained through direct measurements [10] or be estimated based on link loads [44].

Selfish routing interferes with traffic engineering by changing the traffic matrix. More specifically, after traffic engineering installs the routing matrix R_t to the network, selfish routing will respond and redistribute traffic through overlay nodes, which leads to a new traffic matrix T_{t+1}. This process repeats.

Fig. 16 specifies the process of vertical interactions. We also add a relaxation option in the hope of improving stability. However, our results show that it does not yield significant performance improvement. Thus, in the interest of brevity, below we only present the results of traffic engineering without relaxation.

$\triangleright$ T_t is the estimated traffic matrix at time t.
$\triangleright$ T_t^* is the real traffic matrix at time t.
$\triangleright$ R_t is the routing matrix at time t.
$\triangleright$ Assume $\sum_t \alpha_t \to \infty$; $\alpha_t \to 0$ as $t \to \infty$.

T_t^* = Traffic matrix when routing matrix is R_{t-1}
if (relaxation)
    $T_t = (1-\alpha_t)T_{t-1} + \alpha_t T_t^*$
else
    $T_t = T_t^*$
R_t = OptimizedRoutingMatrix(T_t)
Traffic engineering installs R_t to network
Selfish routing redistributes traffic to form T_{t+1}^*

Fig. 16. One round during vertical interaction.

B. Does selfish routing work well with OSPF optimizer?

We first evaluate vertical interactions when the route controller uses OSPF. In all of our experiments, the traffic engineering process uses an OSPF optimizer to optimize link weights as described in [13], and the starting routing matrix of the interactions is computed using hop-count-based OSPF. We choose this starting point to model a scenario in which selfish routing initially has full control over the routing of its traffic in the physical network (see Section VIII), and then the network decides to start using traffic engineering.

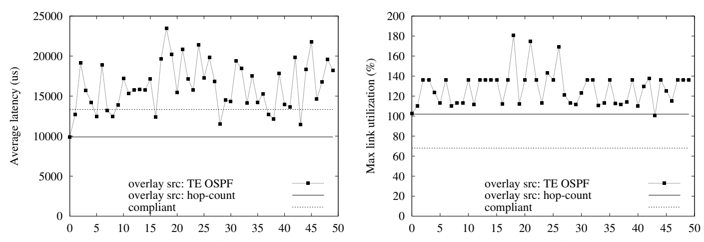


Fig. 17. Vertical interaction with OSPF optimization for the Sprint topology.

Fig. 17 shows the dynamics of vertical interactions for the Sprint topology. The results indicate that the response of OSPF traffic engineering could yield considerably worse performance than compliant routing using optimized-compliant OSPF weights (*i.e.*, traffic engineering without selfish traffic), and worse than overlay source routing on top of hop-count-based OSPF (*i.e.*, selfish routing without traffic engineering). We observe qualitatively similar results as we vary network topologies, the fractions of selfish traffic, and the sizes of selfish overlays.

These results suggest that the interactions between the two separate routing control processes is so ineffective that each individual control process, when applied alone, can yield better performance than having such interactions.

Such inefficiency is partly due to the fact that the adaptive nature of selfish traffic creates considerable variability in traffic demands and therefore makes it harder to do traffic engineering. Another important reason is the limited control of OSPF over selfish overlay traffic. Recall in Section VIII we have shown that when all network nodes belong to an overlay, the only way in which OSPF can affect the selfish overlay traffic is by violating triangular inequality, which effectively reduces available network resources. As a result, both latency and network cost could be worse than those of hop-count-based OSPF, which gives the overlay full access to all available network resources.

C. Does selfish routing work well with MPLS optimizer?

The poor interactions between selfish routing and the OSPF optimizer motivates us to look for alternative solutions. In this subsection, we examine vertical interactions between selfish routing and the MPLS optimizer, which allows one to implement general multi-commodity routing. Given a traffic matrix and a piece-wise linear, increasing, convex network cost function, the MPLS optimizer can find the optimal routing matrix R that minimizes the network cost by solving a linear programming problem.

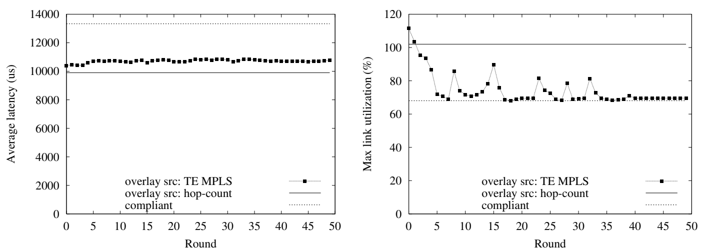


Fig. 18. Vertical interaction with MPLS optimization for the Sprint topology.

Fig. 18 shows the average latency and maximum link utilization for the Sprint topology. We observe that the routing performance is noticeably better than that of OSPF. It allows the system to reach a state in which the network cost is close to that of optimal traffic engineering without selfish routing, and the average latency is only marginally higher than what selfish routing can achieve in the absence of traffic engineering. This is important because the traffic engineering process can choose to stop at any moment and settle on a routing matrix that gives a satisfactory result. That is, the traffic engineering process can be considered as a type of Stackelberg game. We observe similar results from other topologies.

These results indicate that MPLS-based traffic engineering can interact much more effectively with selfish routing. This is likely due to the fact that MPLS has much more fine-grained control over selfish overlay traffic. Specifically, unlike OSPF, MPLS can adjust the routing matrix R without having to reduce available network resources.

Despite the encouraging results, however, we note that there are a number of practical challenges in applying MPLS-based traffic engineering, or traffic engineering in general, in the presence of selfish traffic. In our evaluation we assume that we know the perfect traffic matrices, which need to be estimated in practice. The adaptive nature of selfish traffic can make it difficult to accurately estimate traffic matrices. Another challenge is that MPLS-based traffic engineering requires solving a large linear programming problem. For large networks, the problem may contain millions of unknowns, which is infeasible to solve using software available today. A thorough exploration of these subjects is outside the scope of this paper, so we defer it to future work.

D. Summary

To summarize, in this section we examine the interactions between selfish routing and traffic engineering. We find that OSPF route optimization interacts ineffectively with selfish routing, largely due to its limited control over selfish traffic. In contrast, MPLS route optimization has more fine-grained control, and therefore interacts with selfish traffic more effectively. However, further research is required to investigate such interactions in more detail.

XI. Effects of Multi-AS

In this section, we conduct a preliminary investigation on whether the multi-AS structure of the Internet affects the routing performance. We start by describing our evaluation methodology, and then present the performance results.

A. Evaluation methodology

We evaluate the effects of multi-AS nature based on inter-domain traffic traces and an inter-domain topology. We obtain Abilene traces, which contain netflow data from a number of universities and enterprises on the Internet-2 during October 2003. We select traffic traces from the organizations, shown in Table III, for our evaluations. To speed up our evaluations, during each 5-minute interval, we use only the 2000 destination prefixes with the largest volumes. We call these prefixes *top prefixes*. Note that in different time intervals, the sets of top prefixes are different, but they always account for over 90% of the total traffic in an interval.

AS	Organization	Traffic Rate (Mbps)
3582	University of Oregon	215.576 (202.527)
3	MIT Gateways	64.598 (64.587)
52	UCLA	52.245 (52.234)
59	University of Wisconsin, Madison	33.333 (33.253)
237	NSF (MERIT-AS-14)	117.366 (108.621)
6629	NOAA Silver Springs Lab	62.340 (62.335)
70	National Library of Medicine	72.810 (72.691)
1701	NASA/GSFC (Goddard Space Flight Center)	37.451 (37.448)
22753	Red Hat Inc.	33.241 (33.238)
Anonymized	Commercial Web Server	156.231 (64.124)

TABLE III

TRAFFIC TRACES USED IN OUR EVALUATION, WHERE THE LAST COLUMN SHOWS THE ORIGINAL TRAFFIC RATES AVERAGED OVER 91 DAYS, AND THE TRAFFIC RATES AFTER FILTERING, WHICH ARE SHOWN IN PARENTHESES.

We construct an inter-domain topology using the Rocketfuel data [32]. To make our simulations scalable, we select 6 ASes in the United States from the Rocketfuel data to construct a network topology of over 363 nodes and 1639 edges. For each intra-domain link, we use the inferred OSPF weight and propagation delay from the data; and for each peering link, we use the estimated propagation delay from the data. Since most of the ASes for which we have traffic traces do not have corresponding topology data, we randomly map the ASes in Abilene traces to the ASes in the Rocketfuel topology. We compare compliant routing with overlay routing. In compliant routing, the network chooses the inter-domain route based on the shortest AS hop count, and chooses intra-domain route based on the shortest OSPF path. Overlay routing allows a user to select an overlay path; meanwhile the network determines the route between two consecutive overlay nodes based on the same hierarchical routing strategy. Since the Rocketfuel data do not contain link bandwidth, we set the peering links to be OC3 (*i.e.*, 155 Mbps) and intra-domain links to be OC12 (*i.e.*, 622 Mbps). We use the M/M/1 latency function for all links in the network to capture the effect of traffic load on link latency.

B. Effects of network load

First we examine the impact of varying network load on routing performance. In all cases, the overlay consists of all the peering points in the inter-domain topology. Figure 19 compares compliant routing with overlay routing, where the number of overlays is varied in the following way: (i) one overlay, which includes all traffic (This corresponds to cooperative routing, since traffic is routed to minimize the overall latency), (ii) overlay per AS, where each overlay contains traffic from the same AS in Abilene traces, and (iii) an infinite number of overlays, where each overlay contains infinitesimal amount of traffic. As we can see, overlay routing yields similar performance as we vary the number of overlays. In addition, overlay routing performs better than compliant routing, especially under heavy load.

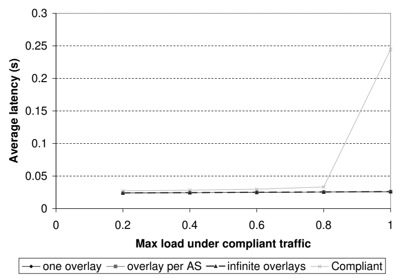


Fig. 19. Comparison of average latency between overlay routing and compliant routing using the Abilene traffic trace during the first 5-minute time window on Oct. 8, 2003, where the number of overlays varies from one overlay to an infinite number of overlays, and the load varies such that the max-utilization under compliant routing changes from 0.2 to 1.

C. Varying time windows

Next we repeat the previous experiment by varying the time window. Fig. 20 shows the results for the 288 time windows, each lasting 5 minutes, on Oct. 8, 2003. As before, different types of overlay routings yield similar performance, all significantly out-performing compliant routing.

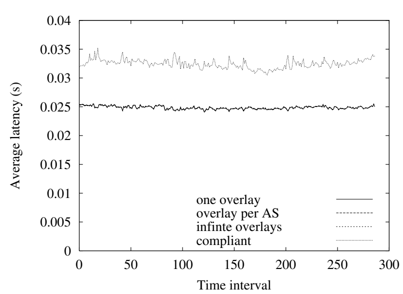


Fig. 20. Comparison of average latency between overlay routing and compliant routing, where the number of overlays varies from one overlay to an infinite number of overlays, and the max-utilization under compliant routing is kept at 0.75.

D. Effects of overlay coverage

Finally we examine the impact of overlay coverage on routing performance. Fig. 21 shows the results as we randomly pick a fraction of nodes in the topology as overlay nodes; and Fig. 22 shows the results as we randomly pick a fraction of peering points in the topology as overlay nodes. In both cases, we observe that overlay routing, regardless of the number of overlays, yields similar performance as the overlay coverage varies from 0.1 to 1. Moreover it consistently out-performs compliant routing.

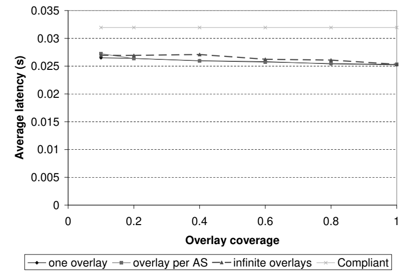


Fig. 21. Comparison of average latency between overlay routing and compliant routing using the Abilene traffic trace during the first 5-minute time window on Oct. 8, 2003, where the number of overlays varies from one overlay to an infinite number of overlays, and the fraction of nodes in an overlay varies from 0.1 to 1.

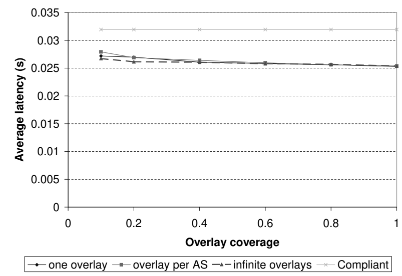


Fig. 22. Comparison of average latency between overlay routing and compliant routing using the Abilene traffic trace during the first 5-minute time window on Oct. 8, 2003, where the number of overlays varies from one overlay to an infinite number of overlays, and the fraction of peering points in an overlay varies from 0.1 to 1.

E. Summary

In this section, we observe that as in the intra-domain, overlay routing in inter-domain out-performs compliant routing. Moreover the performance degradation due to competition among overlays is not significant.

XII. Conclusions and Future Work

In this paper, we use a game-theoretic approach to study the performance of selfish routing in Internet-like environments. Our results show that unlike the theoretical worst case, selfish routing in such environments achieves close to optimal average latency, when the network-level routing is static. Moreover, compared with compliant routing, selfish routing yields lower latency. This is true for both intra-domain and inter-domain scenarios. On the other hand, such performance often comes at the cost of overloading certain links. Moreover, when selfish routing and traffic engineering each tries to minimize its own cost by adapting to the other process, the resulted performance could be considerably worse.

There are a number of avenues for future work, *e.g.*, a better understanding and improving the interactions between selfish routing and traffic engineering; investigating the dynamics of selfish routing, *i.e.*, how equilibria are reached; and evaluating the performance of selfish routing under alternative performance metrics, such as loss and throughput.

Acknowledgments

We are grateful for the helpful comments from Jennifer Rexford, Stan Eisenstat, Joan Feigenbaum, Eric Friedman, Ted Jewell, Yanbin Liu, Jitendra Padhye, Haiyong Xie, Ellen Zegura, and the anonymous reviewers.

REFERENCES

[1] A. Akella, S. Seshan, R. Karp, and S. Shenker. Selfish behavior and stability of the Internet: A game-theoretic analysis of TCP. In *Proc. of ACM SIGCOMM '02*, Pittsburgh, PA, Aug. 2002.

[2] E. Altman, T. Boulogne, R. E. Azouzi, and T. Jimenez. A survey on networking games. *Telecommunication Systems*, Nov. 2000.

[3] D. G. Andersen, H. Balakrishnan, M. F. Kaashoek, and R. Morris. Resilient overlay networks. In *Proc. of SOSP '01*, Banff, Canada, Oct. 2001.

[4] D. O. Awduche. MPLS and traffic engineering in IP networks. *IEEE Communication Magazine*, pages 42–47, Dec. 1999.

[5] I. Castineyra, N. Chiappa, and M. Steenstrup. *The Nimrod Routing Architecture, RFC 1992*, Aug. 1996.

[6] A. Chen, D.-H. Lee, and R. Javakrishnan. Computational study of state-of-the-art path-based traffic assignment algorithms. *Mathematics and Computers in Simulation*, pages 509–518, 2002.

[7] R. Cole, Y. Dodis, and T. Roughgarden. Pricing networks with selfish routing. In *Proc. of Workshop on Economics of Peer-to-Peer Systems*, June 2004.

[8] P. Dubey. Inefficiency of Nash equilibria. In *Math Oper. Res.*, Feb. 1986.

[9] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the Internet topology. In *Proc. of ACM SIGCOMM '99*, Cambridge, MA, Aug. 1999.

[10] A. Feldmann, A. Greenberg, C. Lund, N. Reingold, and J. Rexford. Deriving traffic demands for operational IP networks: Methodology and experience. *IEEE/ACM Transactions on Networking*, June 2001.

[11] L. Fleischer, K. Jain, and M. Mahdian. Taxes for heterogeneous selfish users in a multicommodity network. In *Proc. of Foudations of Computer Science (FOCS)*, Oct. 2004.

[12] M. Florian and D. Hearn. *Network Routing*, chapter 6, Network equilibrium models and algorithms. Elsevier Science, 1995.

[13] B. Fortz, J. Rexford, and M. Thorup. Traffic engineering with traditional IP routing protocols. *IEEE Comm. Magazine*, Oct. 2002.

[14] B. Fortz and M. Thorup. Internet traffic engineering by optimizing OSPF weights. In *Proc. of IEEE INFOCOM '00*, Tel Aviv, Israel, Mar. 2000.

[15] E. Friedman. A generic analysis of selfish routing. In *Proc. of the 43rd IEEE Conference on Decision and Control*, 2004.

[16] L. Gao. On inferring autonomous system relationships in the Internet. *IEEE/ACM Transactions on Networking*, 9(6), Dec. 2001.

[17] D. Gross and C. Harris. *Fundamentals of Queueing Theory*. John Wiley, 3rd edition, 1998.

[18] C. M. Harris, P. H. Brill, and M. J. Fischer. Internet-type queues with power-tailed interarrival times and computational methods for their analysis. *INFORMS Journal on Computing*, pages 261–271, 2000.

[19] S. Iyer, S. Bhattacharyya, N. Taft, and C. Diot. An approach to alleviate link overload as observed on an IP backbone. In *Proccedings of IEEE INFOCOM '03*, San Francisco, CA, Apr. 2003.

[20] G. Karakostas and S. G. Kolliopoulos. Edge pricing of multicommodity networks for heterogeneous selfish users. In *Proc. of the 42nd Annual Symposium on Foudations of Computer Science*, Rome, Italy, Oct. 2004.

[21] R. Keralapura, N. Taft, C. Chuah, and G. Iannaconne. Can ISP's take the heat from overlay networks? In *Proc. of ACM HotNets Workshop*, Nov. 2004.

[22] Y. A. Korilis, A. A. Lazar, and A. Orda. Architecting noncooperative networks. *IEEE Journal of Selected Areas in Communications*, 13(7):1241–1251, Sept. 1995.

[23] Y. A. Korilis, A. A. Lazar, and A. Orda. Achieving network optima using Stackelberg routing strategies. *IEEE/ACM Transactions on Networking*, 5(1):161–173, Feb. 1997.

[24] E. Koutsoupias and C. Papadimitriou. Worst-case equilibria. In *Proc. of the 16th Annual Symposium on Theoretical Aspects of Computer Science*, pages 404–413, 1999.

[25] J. B. Krawczyk and S. Berridge. Relaxation algorithms in finding Nash equilibria. In Computational Economics from Economics Working Paper Archive at WUSTL, July 1997.

[26] V. S. A. Kumar and M. V. Marathe. Improved results for Stackelberg scheduling strategies. In *Proc. of the 29th International Colloquium on Automata, Languages, and Programming*, Malaga, Spain, July 2002.

[27] Y. Liu, H. Zhang, W. Gong, and D. Towsley. On the interaction between overlay routing and MPLS traffic engineering. In *Proc. of IEEE INFOCOM*, Mar. 2005.

[28] Z. M. Mao, L. Qiu, J. Wang, and Y. Zhang. On AS-level path inference. In *Proc. of ACM SIGMETRICS*, June 2005.

[29] A. Medina, A. Lakhina, I. Matta, and J. Byers. BRITE: Boston University representative Internet topology generator. Available from http://www.cs.bu.edu/brite.

[30] The network simulator: ns-2. http://www.isi.edu/nsnam/ns/.

[31] L. Qiu, Y. R. Yang, Y. Zhang, and S. Shenker. On Selfish Routing in Internet-Like Environments. In *Proc. of ACM SIGCOMM*, August 2003.

[32] Rocketfuel. Pop-level ISP maps. http://www.cs.washington.edu/research/networking/rocketfuel/maps/policy-dist.tar.gz.

[33] T. Roughgarden. Stackelberg scheduling strategies. In *Proc. of the 33rd Annual Symposium on Theory of Computing*, pages 104–113, 2001.

[34] T. Roughgarden. *Selfish Routing*. PhD thesis, Cornell University, May 2002.

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

[36] S. Savage, A. Collins, E. Hoffman, J. Snell, and T. Anderson. The end-to-end effects of Internet path selection. In *Proc. of ACM SIGCOMM '99*, pages 289–299, Cambridge, MA, Aug. 1999.

[37] Y. Sheffi. *Urban Transportation Networks: Equilibrium Analysis with Mathematical Programming Methods*. Prentice-Hall, 1985.

[38] N. Spring, R. Mahajan, and D. Wetherall. Measuring ISP topologies with Rocketfuel. In *Proc. of ACM SIGCOMM '02*, Pittsburgh, PA, Aug. 2002.

[39] L. Subrmanian, S. Agarwal, J. Rexford, and R. Katz. Characterizing the Internet hierarchy from multiple vantage points. In *Proc. of IEEE INFOCOM '02*, New York, NY, June 2002.

[40] H. Tangmunarunkit, R. Govindan, S. Jamin, S. Shenker, and W. Willinger. Network topology generators: Degree-based vs. structural. In *Proc. of ACM SIGCOMM '02*, Pittsburgh, PA, Aug. 2002.

[41] H. Tangmunarunkit, R. Govindan, S. Shenker, and D. Estrin. The impact of routing policy on Internet paths. In *Proc. of IEEE INFOCOM '01*, Anchorage, AK, Apr. 2001.

[42] S. Uryas'ev and R. Y. Rubinstein. On relaxation algorithms in computation of noncooperative equilibria. *IEEE Transactions on Automatic Control*, 39(6):1263–1267, June 1995.

[43] X. Xiao, A. Hannan, B. Bailey, and L. Ni. Traffic engineering with MPLS in the Internet. *IEEE Network Magazine*, Mar. 2000.

[44] Y. Zhang, M. Roughan, N. Duffield, and A. Greenberg. Fast accurate computation of large-scale IP traffic matrices from link loads. In *Proc. of ACM SIGMETRICS '03*, June 2003.

Lili Qiu Lili Qiu received her Ph.D. degree in Computer Science from Cornell University in 2001. She spent 2001 - 2004 as a researcher at Microsoft Research. She is currently an Assistant Professor in the Department of Computer Sciences at the University of Texas at Austin. Her research interests are in Internet and wireless networking, with special focuses on Internet measurement, overlay routing, content distribution, and wireless network performance and management.

Yang Richard Yang Yang Richard Yang received his Ph.D. degree in Computer Science from the University of Texas at Austin in 2001. Since 2001, he has been with the Department of Computer Science, Yale University, New Haven, CT, where currently he is an Associate Professor. He leads the LAboratory of Networked Systems (LANS) at Yale University. His current research interests are in computer networks, mobile computing, sensor networks, and network security.

Yin Zhang Yin Zhang is an Assistant Professor in the Department of Computer Sciences at the University of Texas at Austin. His research interests span several areas of computer networks, including network measurement, anomaly detection, data stream computation, traffic engineering, overlay routing and multihoming. Dr. Zhang received his Ph.D. in Computer Science from Cornell University in 2001. Between 2001-2004, he was a member of the Networking Research Department at AT&T Labs - Research.

Scott Shenker Scott Shenker spent his academic youth studying theoretical physics but soon gave up chaos theory for computer science. Continuing to display a remarkably short attention span, his research over the years has wandered from computer performance modeling and computer networks to game theory and economics. Unable to hold a steady job, he currently splits his time between the U. C. Berkeley Computer Science Department and the International Computer Science Institute (ICSI).