Negotiation-Based Routing Between Neighboring ISPs

Ratul Mahajan David Wetherall Thomas Anderson

University of Washington

Abstract – We explore negotiation as the basis for cooperation between competing entities, for the specific case of routing between two neighboring ISPs. Interdomain routing is often driven by self-interest and based on a limited view of the internetwork, which hurts the stability and efficiency of routing. We present a negotiation framework in which adjacent ISPs share information using coarse preferences and jointly decide the paths for the traffic flows they exchange. Our framework enables pairs of ISPs to agree on routing paths based on their specific relationship, even if they have different optimization criteria. We use simulation with over sixty measured ISP topologies to evaluate our framework. We find that the quality of negotiated routing is close to that of globally optimal routing that uses complete, detailed information about both ISPs. We also find that ISPs have incentive to negotiate because *both* of them benefit compared to routing independently based on local information.

1 Introduction

A defining characteristic of the Internet (and increasingly, other planetary scale distributed systems) is that it is operated by autonomous organizations with varied interests. These organizations need to cooperate to provide a useful service, but they also compete with each other, e.g., for the same set of customers. This makes protocol design challenging, as organizations tend to hide information and make selfish policy decisions. The consequence can be both poor stability and poor efficiency.

The current interdomain routing protocol (BGP) provides an example. ISPs export little internal information and make selfish routing decisions based on their local view of the internetwork. Routing can be unstable because the actions of ISPs influence each other: in the absence of knowledge about other networks, one ISP can adversely influence another, and in the worst case cycles of influence can lead to oscillations. We are aware of one such incident that involved two large ISPs and lasted for two days [12]. Routing can be inefficient because locally sound routing decisions may be globally unsound [28, 30]. For instance, early-exit routing, in which upstream ISPs use the locally optimal exit for sending traffic to the downstream ISP, may cause the downstream ISP to carry the traffic a long way [30]. The problems with the current Internet routing architecture are also betrayed by the fact that operators are often forced to work around it by manually cooperating (as was the case in the incident above) using ad hoc mechanisms to make the routing work as desired. This manual control is neither efficient nor robust [17].

In this paper, we explore negotiation as the basis for stable and efficient routing between neighboring ISPs. This limited scenario exhibits many of the problems that occur in the more general case of interdomain routing [31], while letting us study those issues in a more tractable setting. We leave for future work the extension of our approach to cover multilateral negotiations.

With negotiation, ISPs share information in a controlled manner and jointly agree on a mutually acceptable set of paths for traffic flows they exchange. The joint agreement precludes the possibility of a cycle of influence by design. We present a practical negotiation framework, Nexit, with several properties that make it a good fit for interdomain routing. It requires ISPs to share relatively little information with each other: coarse, opaque preferences rather than transparent metrics such as latency or cost. It is flexible enough for the ISPs to reach an operating point based on their specific relationship, and it enables ISPs to optimize for their own criteria, e.g., increasing performance versus reducing cost. It also allows an ISP to ensure that it is no worse off than the default case of selfish routing with local information, so that negotiating carries no risk.

We evaluate negotiation using simulation over sixty measured ISP topologies. For both distance and bandwidth metrics, we compare negotiated routing with globally optimal routing that uses complete information to optimize the two ISPs as a single larger system. We find that the quality of negotiated routing is very close to that of the globally optimal routing. For bandwidth measures, the benefit of cooperative routing is often substantial, reducing the likelihood of overload inside either ISP. For distance measures, this benefit is small in aggregate, implying that the average “price of anarchy” [23] from a distance perspective is low in practice. The main benefit of negotiation in this setting is that it can automatically optimize a small fraction of flows with circuitous default paths. Compared to routing based on local information, both ISPs benefit with negotiation, which provides a strong incentive to negotiate. In contrast, with global optimization, one ISP may lose to benefit the other; losing ISPs will be averse to global optimization.

Our work provides a case study of when negotiation might help to coordinate the actions of competing organizations that must cooperate to provide some service. In our case, negotiation is successful because the interests of the ISPs are not completely opposed. By cooperating, both of them benefit relative to selfish routing based solely on local information.[1] Further, we find that gains are possible only if the ISPs take a holistic view of traffic. Optimizing a single flow often means a gain for one ISP and a smaller loss for the other. Both the ISPs can gain when routing is optimized across a set of flows (as is the case for negotiation): each ISP gains for some flows and loses for others, with an overall positive gain. Nexit leverages these properties.

The rest of the paper is organized as follows. In Section 2, we provide a brief background of interdomain routing and motivate the need for better cooperation. We discuss our design considerations in Section 3 and describe our negotiation framework in Section 4. In Section 5, we empirically demonstrate the benefits of negotiation. We discuss some issues concerning deployment in Section 6, discuss related work in Section 7, and conclude in Section 8.

2 Background and Motivation

In this section, we provide a brief background of interdomain routing and give examples of problems that stem from selfish routing based on local information.

2.1 Background

For our purposes, the Internet is a collection of ISP networks or autonomous systems (ASes). We refer to inter-ISP links as *interconnections*. It is common for two large ISPs to have multiple interconnections, e.g., in different cities. ISPs use the BGP protocol to exchange reachability information – the list of ISPs along the path to the destination (known as the AS-path) – with each other to provide global connectivity. Routing information flows in the opposite direction to data flow, from downstream ISPs to upstream ISPs.

When multiple paths to a destination are available, ISPs use a combination of local policy, AS-path length and local resource constraints to select the path. The commercial relationship with the adjacent ISP is an important consideration for local policy. Typical relationships include *customer-provider*, *peers*, and *siblings*. In the first, the customer pays the provider ISP. Money is not exchanged in the other two, based on the assumption of mutual benefit for traffic exchange. Peers are often competitors that benefit from direct access to each other's customers, while siblings are friendly or related networks. Usually, ISPs prefer to send traffic through their customers, peers, and providers in that order [11]. Within these groups, paths are chosen based on their length and the amount of local resources consumed.

The original design of BGP [15] allowed only AS-path reachability information to be shared. This proved to be a serious shortcoming because ISPs want to optimize their networks, for instance, to balance load in their network, to improve the performance of the traffic they carry, or to reduce overall resource consumption. While ISPs could arbitrarily control their outgoing traffic, the inability to control incoming traffic hindered optimization. Over time, many ad hoc mechanisms have been added to address this problem.

Two such mechanisms that are commonly used today are multi-exit discriminators (MEDs) and AS-path prepending. MEDs are used between ISPs that connect in multiple locations. The downstream ISP attaches an integer to route advertisements to convey its preference for a specific destination (or destination prefix) to use a specific interconnections. If the upstream ISP chooses to honor these MEDs, it picks the best interconnection from the downstream's perspective. With AS-path prepending, the downstream ISP artificially increases the path length for traffic coming in from certain links by adding its own AS identifier multiple times in the path. Whether or not an upstream uses the increased path length in selecting paths depends on its local policies. One might think that the downstream ISP could completely determine the upstream ISP's choice by selectively advertising routes on only those interconnections it wants the upstream to use; this practice is usually prohibited by contractual agreement.

2.2 Example Problems

We now present two scenarios to illustrate the shortcomings of current interdomain routing mechanisms.

Our first example concerns the tuning of traffic exchanged between two ISPs to use resources more efficiently or to improve performance. Consider the two ISPs shown in Figure 1a, each using the closest interconnection ("early-exit") to transfer traffic to the downstream ISP as it minimizes resource usage in the upstream network. This is a common policy [30]. However, the gains of this strategy vanish when one considers traffic flowing in the reverse direction, if the other ISP also uses early-exit routing. This situation is shown in Figure 1a. Compared to a judicious choice of interconnection, as in Figure 1c, early-exit routing can lead to greater resource consumption for both ISPs and poorer overall performance because it may route traffic away from the ultimate destination. Under certain topological assumptions the cost of early exit routing can be up to three times

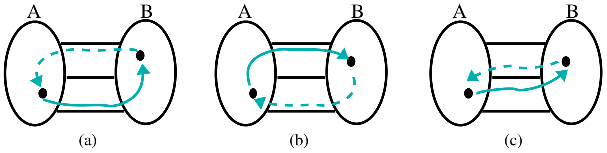


Figure 1: *Negotiation for performance tuning. (a) The default (early-exit) scenario. (b) The traffic pattern with MEDs (late-exit). (c) A mutually beneficial solution.*

that of the optimal routing [13], though we show that it is much less in practice.

There is no straightforward way to achieve the optimized routing of Figure 1c with BGP. For instance, the use of MEDs leads to late-exit routing shown in Figure 1b. When the ISPs agree to honor each other's preferences for incoming traffic, the traffic will use the link that is closest to the destination. Done consistently, this situation is simply the reverse of early-exit.

Obtaining the routing configuration of Figure 1c requires both information sharing and coordination between ISPs. The former is not sufficient by itself as an ISP has no incentive to use the middle interconnection unless the other ISP also does the same. Coordination can convince both ISPs to give up their selfish choices.

Our second example concerns managing overload after unexpected changes in the topology or traffic such as failures or flash crowds. It is adapted from the incident mentioned in Section 1. Consider the two ISPs in Figure 2a, with traffic flowing from ISP-A to ISP-B. Assume that the middle interconnection fails, and ISP-A re-routes the affected traffic based on local conditions (Figure 2b). This overloads ISP-B, which reacts by shifting some traffic to the top interconnection, using MEDs, for instance (Figure 2c). Unfortunately, ISP-B's action overloads ISP-A, and it reacts by shifting traffic to the bottom link (Figure 2d). The result is a return to the situation of Figure 2b and continue the cycle of influence.

Figure 2e shows a solution that is acceptable to both ISPs. As before, there is no straightforward way in BGP to discover this configuration. Using MEDs, ISP-B needs to specify that the preferred entrance for $f3$ is the top interconnection and for $f2$ is the bottom one. But given a purely local view, ISP-B has no basis for preferring this configuration over $f3$ on the bottom link and $f2$ on the top one. Similarly, ISP-A has little visibility into ISP-B's network to determine the acceptable routing pattern.

3 Design Considerations

In this section, we lay out the design considerations for structuring cooperation between neighboring ISPs. Our goal is to enable the negotiating ISPs to meet their individual objectives. This implies giving them control over

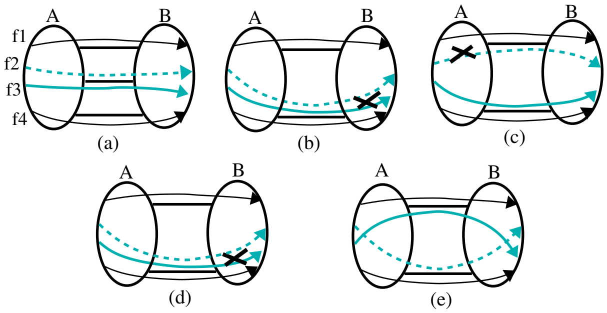


Figure 2: *Negotiation in response to failures. (a) The stable (no failure) scenario. (b) ISP-A's response to the failure of the middle interconnection congests ISP-B. (c) ISP-B's reaction of moving some traffic from the bottom interconnection to the top one congests ISP-A. (d) ISP-A reacts to its congestion by moving the traffic back to the bottom link, which again congests ISP-B. (e) A mutually acceptable solution.*

both their incoming *and* outgoing traffic. This control should be mutual, i.e., both upstream and downstream ISPs should be able to influence path selection. Absolute control for either the upstream or the downstream leads to problems mentioned earlier.

Our solution is based on the following key considerations that we extracted from the problem domain.

• **Limited information disclosure:** Competitors are often reluctant to disclose detailed internal information to each other. Thus, we must work with inputs that do not directly disclose unwanted information. For an ISP, this precludes disclosing information on the topology and performance of its network. This sensitivity also extends to cost information, since an ISP may not wish to tell its competitor the true cost of carrying traffic. We handle this concern by working with opaque preference classes, rather than transparent metrics such as latency or cost. An alternative approach is mechanism design, in which the best strategy for an ISP is to reveal its true cost regardless of what it knows about the other ISP's cost [8]. But this cost information can be abused outside of the solution framework, such as when an ISP adds capacity along the competitor's profitable routes [9].

• **Support for heterogeneous objectives:** Different ISPs have different goals and hence different optimization criteria. To work across such systems, cooperation mechanisms should be agnostic towards the internal objective functions used by individual entities. For instance, while ISPs with capacity constraints may aim to avoid overload, ISPs with overprovisioned networks may aim to improve performance by reducing latency and jitter. Yet others may want the best routes for their preferred customers. There are bound to be further considerations of which we cannot be aware. While economic cost could be used as a unifying metric, it can be very difficult, if not impossible, for ISPs to quantify their internal considerations in terms of true cost [29]. As before, we handle this concern by working with opaque preferences. ISPs map their internal objectives to these preferences. This is relatively easier because ISPs already quantify these objectives for intradomain optimization.

A consequence of the above two considerations is that achieving social goals such as social optimality or fairness is not a pre-requisite for negotiation. Social goals are usually defined only when entities have comparable objectives. For instance, both social optimality and fairness are undefined when one ISP optimizes for latency and the other for link utilization.

• **Flexible outcomes:** As we noted in Section 2.1, different pairs of ISPs have different relationships that govern their interaction, e.g., ISPs treat customers and competitors differently. Instead of designing a mechanism that produces a deterministic output given some input, we should provide a flexible framework for ISPs to compute outcomes based on their relationship [6].

Flexibility requires that all kinds of outcomes should be possible but the most interesting space is that of "win-win" outcomes where both ISPs gain. The social optimum that treats both ISPs as a single larger system may cause one to lose compared to the situation with no negotiation. Profit-maximizing entities will not negotiate if they lose. Side payments between ISPs can alter this balance but in this paper we focus on protocols that compute win-win solutions without side payments and leave exploring the use of side payments for future work.

It is desirable that the outcomes, whatever they might be, come close to *Pareto-optimal*. An outcome is Pareto-optimal if all other outcomes are worse for at least one of the entities. Pareto-optimality rules out outcomes with obvious wastage, i.e., those that are worse for both. (The current Internet is often not Pareto-optimal as illustrated in Figure 1.) There can be multiple Pareto-optimal solutions in the system.

• **Incentive compatibility vs. efficiency:** A concern when competing entities interact is that one of them may try to manipulate the outcome in its favor by lying. Incentive compatible mechanisms, in which truth telling is provably the best strategy for all entities, guarantee interactions that are robust against manipulation. However, incentive compatibility often runs counter to efficiency. It is known that in the absence of a third party acting as a subsidizer, appraiser or arbitrator, there does not exist a mechanism that is both incentive compatible and able to implement all mutually acceptable solutions for bilateral trading [21].

Faced with the trade-off between incentive compatibility and efficiency, we favor efficiency for two reasons. First, we believe that cooperation will be the common case because parties tend to act honestly while seeking joint gains over a default contract [26]. Even today, ISPs often cooperate using ad hoc mechanisms that are not robust against manipulation. We want to compute efficient solutions when ISPs cooperate. Second, even if we were to pick incentive compatibility, it is not clear that a mechanism design approach can be used to yield flexible outcomes. Usually, for such approaches the objective of the interaction, for instance, computing least cost paths [8], is fixed by design. But we want to leave the objective up to the ISPs.

However, we will see that favoring efficiency over incentive compatibility does not necessarily imply that a cheating ISP can infinitely game the system (Sections 4.2 and 5.4).

• **Information exchange model:** Careful attention needs to be paid not only to what information is disclosed but also to how it is shared. It is virtually impossible to give ISPs mutual control over traffic in the routing information exchange model used today because routing information flows only from downstream to upstream [18]. If this information is obeyed, e.g., MEDs by contract, then the upstream loses control over outgoing traffic. If obeying this information is optional, e.g., as with AS-path prepending, whether the upstream follows it depends on its local policies. Since these policies are not known to the downstream, it cannot effectively control its incoming traffic. We use a two-way information exchange in which both the upstream and downstream ISPs provide their preference.

• **Scope of optimization:** Mutual control implies that entities have to compromise over certain issues for gain in others. What is the most effective way to arrange this mutual compromise? Economic and political negotiations tell us that better solutions are obtained when the entities negotiate over a larger set of issues [26, 3]. We find this to be true in our two-ISP scenario and encourage ISPs to keep all the traffic on the negotiating table to increase the chances of finding mutual compromises. For systems where simultaneous, mutual compromises are hard to find, compromises can be decoupled in time using "credits," a topic we leave for future work.

• **Efficient computation:** Finally, computing the result of the negotiation should be efficient in time and the number of messages. This excludes trial-and-error protocols, such as where each ISP blindly proposes to re-route a subset of the traffic at a time, in the hope that it is acceptable to the other ISP. We propose that ISPs exchange sets of preferences to efficiently discovery a mutually acceptable operating point.

4 The Nexit Framework

The goal of Nexit (short for negotiated exit) is to enable a pair of ISPs to agree upon an interconnection for each traffic flow they exchange. A *flow* is a stream of packets from a source node in one ISP to a destination node in the other ISP. There may be multiple flows between the same pair of nodes but all packets in a flow take the same path through the two networks. We discuss how ISPs establish identifiable flow signatures in Section 6. We assume that ISPs are capable of source-destination routing, i.e., flows with the same destination but different sources can be routed independently, e.g., using MPLS. By using more flexible flow definitions, Nexit can be extended to destination-based routing but for ease of exposition we focus on source-destination routing in this paper.[2]

Nexit is guided by the following observation. While improving the path of an individual flow might hurt one of the ISPs, a set of improvements will lead to a win-win solution if each improvement brings a large benefit to one ISP at a smaller cost to the other. Identifying such changes only requires the ISPs to disclose a rough measure of the cost or benefit of the change.

Conceptually, Nexit consists of two steps. First, as is the case for any negotiation, parties internally evaluate their routing choices. Second, they participate in a protocol that uses these evaluations to arrive at a mutually acceptable solution. We discuss these steps below.

1. ISP-internal evaluation of routing choices Each ISP maps flow *alternatives* to opaque preference classes based on its internal optimization criterion. An alternative corresponds to an interconnection for a flow. For example, there are three alternatives per flow for ISPs with three interconnections. Instead of using transparent metrics such as latency, Nexit works with opaque preference classes in the integral range $[-P, P]$. Internal ISP metrics are mapped to this range as described below. P is chosen to be large enough to differentiate alternatives with substantially different quality but small enough to avoid unnecessary information leakage. Opaque preferences provide a basis for negotiation between ISPs with different objectives and disclose less internal information as neither the objective nor the mapping process is revealed. But if the ISPs are interested in a social goal, they must decide in advance on the common metric and the mapping process. We consider this to be a special case for negotiation between friendly ISPs.

The mapping to preferences is done based on the default alternative for the flow, which is the alternative that the ISP reckons the flow will use in the absence of negotiation. The two ISPs need not agree on the default alternative for the flow. The ISPs map the default to preference class 0 and non-default alternatives to preferences that reflect their relative goodness.

One requirement for the mapping process is that preferences compose over addition. That is, an ISP should be happy to use two alternatives each with preference -1 if that enables it to use an alternative with preference $+3$. ISP optimization objectives of which we are aware can be mapped within this constraint. For instance, mapping per-flow objectives, such as minimizing the distance a flow traverses inside the ISP network, is straightforward as the preferences for different alternatives are independent. Network-wide objectives, such as minimizing maximum link load, can be mapped using linear program formulations [10] that optimize the sum of the individual-path preferences. Preferences for metrics that are external to an ISP network, such as those based on end-to-end path quality (gathered using measurements, for instance) can be considered mutually independent.

Preference classes are similar to BGP MEDs in terms of information disclosure, but their relative magnitude reveals some extra information. Individual ISPs can control the extent of information disclosed by using either ordinal preferences or fewer than P classes.

2. Negotiation protocol Next, the ISPs exchange their preference lists and agree on an interconnection for each flow using a protocol that proceeds in rounds. In each round, one ISP proposes an alternative and the other decides if it is acceptable. This is accomplished in several steps. The exact implementation method of each step is agreed upon contractually in advance by the ISPs.

• *Decide turn:* Decide which ISP proposes an alternative in the current round. The method we use in our experiments is that the ISPs alternate. Another is that the ISP with the lower cumulative gain (as measured using the sum of preferences for the flows negotiated so far) gets the next turn. Yet another possibility is a coin toss.

• *Propose an alternative:* The ISP whose turn it is proposes an alternative based on local and remote preferences. The method we use picks from the set that maximizes the sum of preferences of the two ISPs, breaking ties using local preferences. An alternative is to propose the best local alternative with minimal negative impact on the other ISP.

• *Accept alternative?* The other ISP decides whether to accept the proposal. This gives ISPs veto power over the proposal, which they might use if the preference for this alternative has changed since last advertised or if they perceive that the proposer is not playing by the mutually agreed rules. We always accept proposed alternatives in our experiments. Accepted flows are removed from the preference lists.

• *Reassign preferences?* Reassignment occurs when one of the ISPs wants to update its preference list. This is needed when the preferences are based on constraints such as available bandwidth that may change after some flows have been negotiated. We reassign preferences af-

Initial preference lists

	$f2_{top}$	$\mathbf{f2_{bot}}$	$f3_{top}$	$f3_{bot}$
(A, B)	(-1, 0)	**(0, 0)**	(0, 0)	(0, 0)

Reassignment after $f2_{bot}$

	$f2_{top}$	$f2_{bot}$	$\mathbf{f3_{top}}$	$f3_{bot}$
(A, B)			**(0, 1)**	(0, 0)

Figure 3: *Preference lists for the example in Figure 2. The column headings correspond to flow alternatives; the subscripts correspond to the interconnections. The tuples represent the two ISPs' preferences for that alternative. The alternative selected at each step is shown in bold.*

ter negotiating each 5% of the traffic for bandwidth experiments and do not reassign preferences for distance experiments.

• *Stop?* ISPs decide whether they want to continue negotiating over more flows. In our experiments, ISPs stop when they perceive no additional gain in continuing. We call this *early termination*. Alternately, ISPs may continue as long as their cumulative gain is positive even though it may be lower than that with early termination. We call this *full termination*. It will be preferred in interest of social welfare. The socially best outcome occurs when ISPs negotiate for all the flows, even if that means a reduction in one of the ISPs' gain.

4.1 An Example

We illustrate the working of Nexit using the second scenario of Section 2, shown in Figure 2. We simulate negotiation over the two flows, $f2$ and $f3$, impacted by the failure. Each flow has two alternatives – the top and bottom interconnections. Assume that the preference class range is [-1, 1] and the ISPs propose alternatives that maximize the total gain, breaking ties at random.

The top table in Figure 3 shows the initial preferences lists for the two ISPs. These are relative to the default of both flows traversing the bottom link. The subscripts for the flows denote the interconnection. Recall that, in that example, ISP-A is averse to $f2$ traversing the top interconnection, and ISP-B is averse to both flows coming in via the bottom interconnection. Initially, all the alternatives for ISP-A are as good as the default except $f2$ going over the top link. ISP-B is initially indifferent to all the alternatives because preference classes to flows are assigned independently of each other. ISP-B can handle either of the flows entering via the bottom link; the problem arises only when they both do. Suppose that ISP-A gets the first turn and it proposes $f2_{bot}$ by randomly picking out of the three equally good options. ISP-B accepts.

Next, the ISPs reassign preferences as shown in the bottom table: ISP-B prefers $f3_{top}$ over the default. Reassignment takes into account the expected state of the network, assuming that the first accepted choice was implemented. ISP-B takes the next turn and proposes $f3_{top}$. This alternative is accepted by ISP-A, leading to the desirable final solution shown in Figure 2e that could not be found by BGP.

Of course, Nexit may not always arrive at an exactly optimal solution. In the example, this occurs if ISP-A happens to pick $f3_{bot}$ the first time. At this point, whichever way $f2$ is routed, one of the ISPs suffers: ISP-A does not want $f2$ to use the top link and ISP-B does not want it to use the bottom link. It is possible to prevent such sub-optimality if ISPs disclose resource dependency among flows. But we opt for simplicity in the design of Nexit; we will see that for realistic scenarios, this does not lead to much efficiency loss.

4.2 Discussion

In this section, we make two observations about Nexit. First, it can be used to obtain a wide variety of outcomes. Computing exact socially optimal or Pareto optimal outcomes in our problem setting is NP-hard. The hardness for load-dependent metrics stems from the inability to split a flow across multiple paths. For load-independent metrics, computing Pareto-optimal solutions in which both ISPs do better than the default is NP-hard. This follows from a simple reduction from PARTITION; we omit this reduction due to space constraints. Nexit approximates those outcomes using its hill climbing (or greedy) structure. Socially optimal solutions are approximated when the ISPs' metrics and the mapping process are compatible (e.g., both ISPs optimize for latency and map a gain of 20ms to the same preference class), ISPs select alternatives that maximize the combined gain, and continue negotiating until all flows have been negotiated. Max-min fair solutions (that maximize the minimum gain) are approximated when the metrics and the mapping process are compatible and the ISP with lesser cumulative gain proposes alternatives, giving it a chance to catch up with the other ISP. Finally, Pareto optimal solutions are approximated when the ISPs (with possibly incompatible metrics) propose alternatives that maximize the combined gain.

Second, even though Nexit is not strictly strategy-proof in that an ISP can lie about its preferences, its structure is such that it cannot be infinitely gamed. First, a cheating ISP can never cause the truthful ISP to lose, only gain less, because the truthful ISP will not accept solutions that are worse than the default. Second, the combination of truthful ISPs that terminate negotiation when they see no more self-gain and ISPs that take turns to pick flow alternatives may lead to premature termination of negotiation. When this occurs, it hurts the cheating ISP compared to it being truthful. Third, various modes of Nexit make cheating harder. For instance, if the alternative selection criterion is decided in advance, lying might hurt the cheater because its choices will be limited by the (incorrect) preference list that it discloses to the other ISP. In Section 5.4, we evaluate these arguments empirically. Analytically understanding the impact of cheating is an interesting avenue for future work.

5 Evaluation

In this section, we evaluate negotiated routing by comparing it with today's default routing and globally optimal routing. While the former is based on local information, the latter is based on complete information sharing and treats both ISPs as a single larger system; as such it ignores the legitimate differences in ISP interests.

We answer three high-level questions:

- *How much of the gain of globally optimal routing can be realized using negotiation, given the restrictions such as limited information sharing?* We show that negotiated routing is very close to globally optimal routing. We also show that negotiating over a large set of flows is necessary to achieve that gain.
- *Compared to the default routing, how do individual ISPs fare with globally optimal routing and with negotiation?* We show that, while the global optimal often benefits one ISP but hurts the other, negotiation always benefits both ISPs.
- *How much can a cheating ISP gain by lying about its preferences?* We show that a cheating ISP may lose compared to being truthful.

The answers to these questions depend on many aspects of ISP networks, some of which are hard to model. Our approach is to use measured data where it is available and experiment with a range of models drawn from the literature where it is not. In this way, we hope to focus on realistic rather than theoretical best- or worst-case bounds, while avoiding results that are sensitive to incidental choices in our setup.

As measured input, we use a dataset of PoP (city)-level topologies of 65 ISPs, along with geographic coordinates of PoPs and estimated inter-PoP link weights that model routing internal to an ISP [30]. This dataset is diverse in terms of ISP sizes and geographical presence.

We consider two kinds of ISP optimization criteria. The first, based on a distance metric, explores the steady-state reduction in overall network resource usage that can be achieved, implicitly assuming that the network capacity is well-matched to the traffic it carries. The second, based on a bandwidth metric, explores how negotiation can reduce the possibility of overload that might occur when the traffic is no longer well-matched to the network, e.g., due to a failure. The results from these two criteria are presented in Sections 5.1 and 5.2. Since we are interested in evaluating the potential of negotiation by comparing it to the globally optimal, which is well-defined only when ISPs use the same optimization criteria, both ISPs use the same criteria in these two sections. In Section 5.3, we evaluate the case where the two ISPs have different optimization criteria. Finally, in Section 5.4, we consider a scenario where one of the ISPs cheats by lying about its preferences.

We used Nexit as follows in our experiments. Preference class range is [-10,10]; we found that increasing the range does not lead to noticeable increase in performance. ISPs take turns to propose alternatives and pick the alternative that maximizes the gain across both of them, breaking ties using their own preferences. Proposals are always accepted as our goal is to evaluate the benefit of negotiation when ISPs cooperate fully. Preferences are not reassigned for the distance experiments and are reassigned for bandwidth experiments after negotiating each 5% of the traffic. Negotiation stops when one of the ISPs cannot gain more.

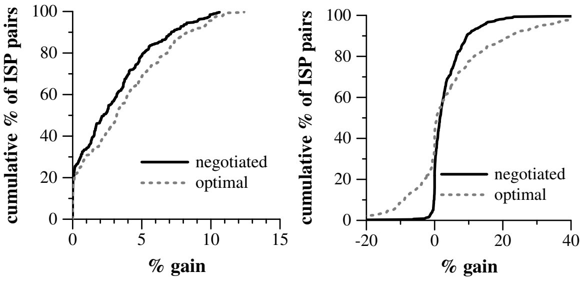
5.1 Distance and Cost

In this section, we evaluate negotiation for improving steady-state routing.

Methodology We assess the quality of steady-state routing using a metric that reflects the total resource consumption in the network. This is the sum of path lengths of all flows. There is a flow from each PoP in one ISP to each PoP in the other ISP. The length of a path is the sum of the lengths of its constituent links; we estimate link length using the geographical distance between its endpoints [22]. Our metric attempts to capture the motivation behind early-exit routing: to reduce overall network resource consumption by minimizing the distance a flow travels inside the upstream network, allowing a smaller or thinner network to support a given set of external traffic demands. Admittedly, it is a crude approximation of ISP objectives because it does not capture many factors, such as flow sizes, that ISPs might consider in practice.

For this evaluation, we use pairs of ISPs that connect at two or more locations to allow a choice of interconnections. We exclude eight ISPs whose measured topologies are a logical mesh because their geographic distance is not reflective of true distance. In all, we have 229 ISP pairs, each with traffic flows going in both directions.

We compute routing for the three methods as follows. The default routing uses the early-exit policy: the interconnection chosen by the upstream ISP for that flow is the one that is closest to the source PoP. The globally optimal routing uses the interconnection that minimizes the total distance for each flow. Negotiated routing is



(a) total gain

(b) individual gain

Figure 4: *The benefit of the optimal and negotiated routing. The x-axis is the percentage reduction in the distance relative to the default routing.*

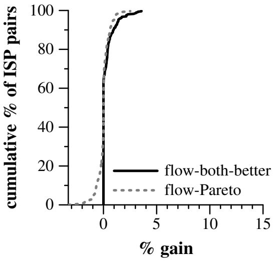


Figure 5: *The gain, relative to the default routing, of two alternate routing strategies that simply discard bad alternatives. Neither achieves nearly the potential benefit of the negotiated or optimal routing.*

computed using Nexit; ISPs use the distance of flows inside their network to map interconnection to preference classes.

Results Figure 4 shows the results of this experiment by plotting the gain of the optimal and negotiated routing relative to the default routing. The left graph plots the cumulative distribution function (CDF) of the total gain across the two ISPs. Each point corresponds to an ISP-pair. The graph shows that negotiated routing is very close to the globally optimal routing. In other words, the ISPs do not lose much by insisting that all solutions be win-win. Interestingly, however, this gain is little on average: roughly 4% for half of the ISP pairs. This suggests that the aggregate cost of early-exit routing is low, i.e., the "price of anarchy" is low for pairs of ISPs. This is well below the theoretical bound [13], probably because the topological assumptions made for computing the bound do not hold in practice. The main value of optimization in this setting is to automatically improve the performance of individual flows that suffer significantly under default routing; we consider flow-level gains shortly. We also find that, in general, ISPs with more interconnections gain more through negotiation. We omit this analysis due to space constraints.

Figure 4b plots the gain for individual ISPs in the pair. With globally optimal routing, roughly a third of the ISPs actually lose, with some losing by more than 30%. These ISPs will have little incentive to move to the globally optimal solution. In contrast, individual ISPs do not lose with negotiated routing, providing a strong incentive to negotiate.[3]

Next, we show that the gains for both ISPs depend on negotiation across a set of flows. A simpler alternative strategy would be to restrict it to pairs of flows going in the opposite direction and discard bad routing paths. We experimented with two strategies – *flow-Pareto* and *flow-both-better*. The former rejects paths that are worse than the default for both ISPs, while the latter rejects those that are worse for any one ISP. For example, in Figure 1, using the top link for $A \rightarrow B$ and the middle link for $B \rightarrow A$ is flow-Pareto, and using the middle interconnection for both directions is flow-both-better. If multiple paths satisfy the required criterion, one is picked at random. Figure 5 plots the gains from these strategies. It shows that these seemingly reasonable strategies which avoid obvious wastage at flow-level are not effective, and their cost is close to that of the default itself. We also experimented with breaking down the set of flows into several groups and negotiating within each group separately. We find that this does not provide as much benefit as negotiating over the entire set. Thus, for mutual gain to be realized, negotiation must be done across flows and ISPs must be willing to trade minor losses on some flows for significant gains on others.

We close this section with a flow-level view of negotiation. Figure 6 shows the gain for individual flows with globally optimal and negotiated routing. Some individual flows gain significantly: 7% of the flows gain over 20%, and 1% of the flows gain over 50%. We speculate that the flows that suffer heavily due to the default routing are the ones that are manually optimized by operators today. Spring *et al.* observed that a small fraction of Internet flows were routed along non-default paths between ISP-pairs [30]. Negotiation can automatically improve the performance of these flows, thus saving precious operator time. Further, the proximity of the negotiated curve to the optimal one suggests that negotiation catches almost all of the flows that need optimization.

Another interesting conclusion that can be drawn from the graph is that only a fraction of flows – roughly 20% in our experiment – need to be non-default routed to get most of the gain.

5.2 Bandwidth and Congestion

We now evaluate the benefit of negotiation in a setting where the ISPs are interested in controlling congestion and overload. Even when ISP networks are well-

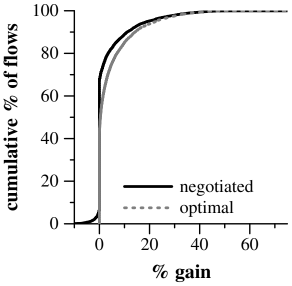


Figure 6: *A flow-level view of optimal and negotiated routing. This graph aggregates all flows across all ISP pairs.*

engineered, overload can occur during failures and sudden changes in traffic demands, as might be caused by a flash crowd [4].

Methodology We consider a scenario where an interconnection fails and simulate negotiation for the flows that are impacted by the failure; in the interest of stability, ISPs are likely to reroute only such flows. Our results may also apply to internal link failures and changes to traffic matrices.

For this experiment, we consider only ISP pairs with three or more interconnections because negotiation after a failure requires at least two working interconnections. There are 247 such ISP pairs in our dataset.

Overload is difficult to evaluate for two reasons. First, calculating bandwidth measures requires estimates of ISP link utilizations and workloads, none of which is readily available. Second, the choice of metric to represent overall ISP cost in terms of individual, congested links is less clear. For both of these, we experimented with a range of plausible models. We first describe the models used for the results presented in this paper and then list the alternate models we tried. While our results are limited to our modeling choices, we found them to be qualitatively similar for these alternate models as well, suggesting that they are not overly sensitive to our specific models.

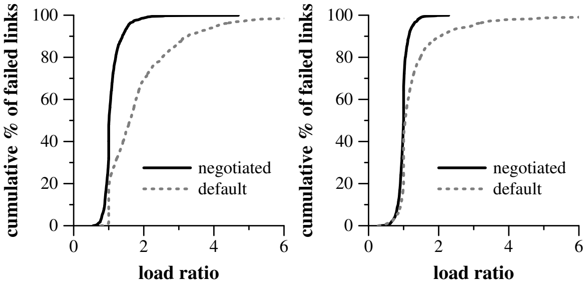
First, we need a workload model. We assume that there is one flow from each upstream-ISP PoP to each downstream-ISP PoP; we consider only one direction of traffic at a time. To determine flow sizes we use a gravity model [19, 32], which predicts that the amount of traffic between a pair of PoPs is proportional to the product of the "weight" of the PoPs. We assume that the weight of a PoP is proportional to the population of its city. Using data from CIESIN [5], we estimate this as the number of people in a 50×50 square mile grid centered on the geographical coordinates of the city. This model leads to a skewed traffic matrix with larger cities consuming more bandwidth, both hallmarks of real Internet traffic [14, 2].

Second, to model link capacities, we assume that they are proportional to the load on the link before the failure [32], i.e., in steady-state a well-designed network tends to be roughly matched to its traffic so that links that carry more traffic tend to be of higher capacity. The traffic matrix combined with the routing within an ISP lets us compute the load on each link. But this method does not assign capacity to links in the topology that do not carry any traffic before the failure. We should not remove these links since they may be used after failures. To such links we assign a capacity that is the median of the links with non-zero load. The intuition here is that the unused links are backup links, and their capacity varies between the minimum and maximum among the links in use. Finally, to preclude our results being dominated by links that carry little traffic, we "upgrade" all links below the median to the median.

Finally, as the choice of the ISP optimization metric, we use a measure based on the intuition that ISPs prefer routing that does not significantly increase the load on links after a failure. All ISPs overprovision to some extent, so the link capacity of well-engineered networks is likely to be some small multiple of its average load. A much higher offered load after a failure implies that either the link becomes congested or it must have been significantly overprovisioned, which is expensive. Thus, our metric should penalize large increases in link load after a failure. We measure the quality of routing using maximum excess load or *MEL*, which is the maximum ratio of load after and before the failure on any link in the topology. Higher MELs are undesirable as they reflect a higher offered load on the link after the failure.

We experimented with the following alternate models. For workload, we tried identical weights for all PoPs and weights drawn from a uniform random distribution. For link capacities, we used discrete capacities by rounding them up to the nearest power of two. For assigning capacities to unused links, we used other measures such as the maximum and average load. Finally, as an alternate ISP optimization metric, we used a metric based on a linear programming formulation of optimal routing [10]. This metric minimizes the sum of link costs, where the cost is a piecewise linear function of load with increasing slope.

We reroute the impacted flows after a simulated failure using the three routing methods as follows. The default routing is early-exit over the new topology. The globally optimal is computed by solving an optimization problem that minimizes the maximum increase in link load. For computational tractability, we allow flows to be fractionally divided among interconnections; thus, the quality of this routing is an upper bound on the global optimal without fractional routing. Negotiated routing is computed using Nexit, with both ISPs using the maximum



upstream ISP

downstream ISP

Figure 7: *The quality of negotiated routing when managing overload. The x-axis is the MEL relative to the MEL of optimal routing.*

increase in link load along the path to map flows to preferences. The preferences are recomputed after each 5% of the traffic is negotiated.

Results Figure 7 shows the results of this experiment by plotting the ratio of the MEL of the default and negotiated routing to that of the optimal routing. Each data point corresponds to one hypothesized interconnection failure; so there are four distinct points for ISP pairs with four interconnections. The MEL for the default routing is often significantly larger than the optimal routing, implying that the default routing tends to overload certain links in the topology even when this overloading is avoidable. For the upstream ISP, the ratio of the two MELs is more than two for half of the cases, and more than five for 10% of the cases. Though not shown in the graph, the MEL ratios are high even when the optimal MEL is high, suggesting that overload with default routing is not limited to thin links in the network. The overload tendency is more for the upstream because many previously unused paths inside the upstream are used to send traffic from the sources to the interconnections that continue working after the failure.

The graphs also show that negotiated routing is very close to the optimal routing (since most of the MELs are one) even though the amount of information used to compute it is much less, the procedure to compute it is much simpler, and the routing itself is restrictive (compared to optimal routing which can fractionally divide a flow among interconnections).

As for distance, negotiation leads to non-default paths for only a fraction of the traffic. In our experiments, negotiation for 20% of the flows brings most of the benefit. We omit this analysis due to space constraints.

A natural question is what happens if, instead of negotiating with the downstream, the upstream unilaterally load balances outgoing traffic. It is possible that this will not hurt or may even benefit the downstream, coming close to optimal in the process. We evaluate this hypothesis by simulating the upstream ISP optimizing the routing for its own network.

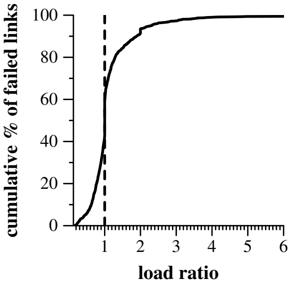


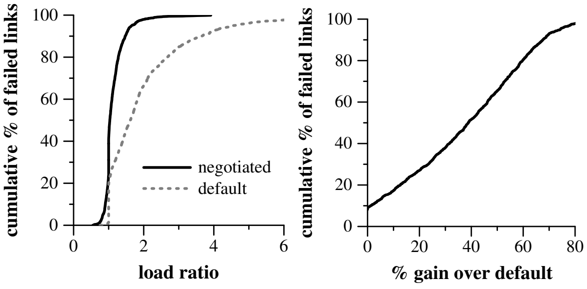
Figure 8: *The impact on downstream ISP of unilateral routing optimization by the upstream ISP. The x-axis is the ratio of the MELs for the upstream-optimized and default routing; values more than one imply that upstream-centric optimization was harmful for the downstream ISP.*

Figure 8 shows the impact of upstream-centric optimization on the downstream ISP. It shows the ratio of MELs in the downstream ISP with upstream-centric optimization versus early-exit routing. The result is unpredictable: while in some cases, upstream-centric optimization helps the downstream (left end of the graph), in others the downstream ISP suffers (right end of the graph). In 10% of the cases, the MEL for upstream-centric optimization is more than twice of that for the default routing. Thus, the unilateral adjustment of routing by the upstream is undesirable because that may end up causing more congestion in the downstream. This is similar to the second example in Section 2.

5.3 Diverse Optimization Criteria

So far we have shown the quality of negotiated routing when both ISPs use the same optimization criteria, but many negotiating ISPs will use different criteria. We evaluate this case using an experiment similar to that in Section 5.2 except that the downstream ISP uses the distance metric from Section 5.1.

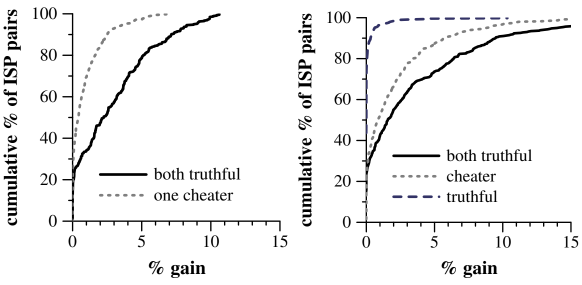
Figure 9 shows the results. The left graph shows how successfully the upstream ISP controls overload in its network. It plots the MEL for the default and negotiated routing relative to the MEL of optimal routing in which overload is optimized across both ISPs. The right graph shows the distance reduction in the downstream ISP relative to the default routing. Both ISPs are able to optimize for the metric of their interest. The upstream can effectively control overload and the downstream can significantly reduce the distance that the traffic traverses in its network.



upstream ISP

downstream ISP

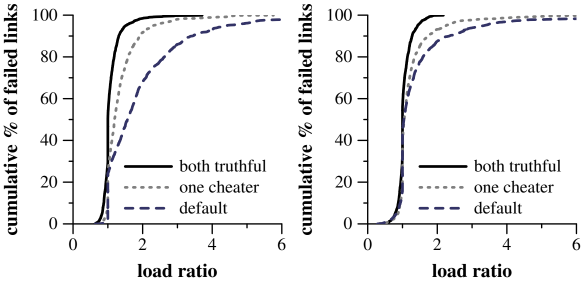
Figure 9: *Negotiation with different optimization criteria. The upstream ISP optimizes for bandwidth and the downstream ISP for distance. The x-axis of the left graph is the MEL relative to the MEL of optimal routing and that of the right graph is distance reduction relative to the default routing.*



(a) total gain

(b) individual gain

Figure 10: *The impact of cheating for the distance experiment. The x-axis is the reduction in distance compared to the default routing. (a) Reduction in distance across both ISPs. (b) Reduction in distance for individual ISPs.*



upstream ISP

downstream ISP

Figure 11: *The impact of cheating for the bandwidth experiment. The upstream ISP is the cheater. The x-axis is the MEL relative to that of the optimal routing.*

5.4 The Impact of Cheating

In this section, we empirically evaluate the impact of cheating on the results produced by Nexit. We do so using a cheating strategy that, on the surface at least, appears to help the cheater. Experiments with a few other strategies yielded similar results. We do not claim that this is the case for *all* possible cheating strategies.

The cheating ISP uses the following strategy in this experiment. Assume that the cheating ISP has perfect knowledge of the other ISP's preferences, which overestimates the cheater's ability because in practice some uncertainty can be introduced in this knowledge. The criterion for selecting alternatives is to maximize the sum of preferences across the two ISPs, breaking ties at random. The cheater uses the knowledge of the other ISP's preferences to inflate the preference of its best alternative for each flow just enough so that it corresponds to maximum sum. This is a better strategy than blindly maximizing preferences because as far as possible the relative ordering of the cheater's original preferences are preserved, which is useful for ensuring that better alternatives are picked first. When the other ISP's preferences are such that inflating the best alternative does not lead to maximum sum, the cheater decreases the preferences for the other alternatives accordingly.

We use the above cheating strategy for both the distance and bandwidth experiments of Sections 5.1 and 5.2. Figure 10 shows the impact of cheating for the distance experiment. The right graph shows that while cheating significantly reduces the gain of the truthful ISPs, it is unattractive as it also reduces the gain for the cheating ISPs. Figure 11 shows the impact of cheating for the bandwidth experiment with the upstream ISP acting as the cheater. As before, cheating reduces the benefit for not only the truthful, downstream ISPs but also for the cheating, upstream ISPs.

In both the experiments above, the cheating ISP loses because the negotiation terminates prematurely as the truthful ISP stops when it sees no benefit for itself. Assuming that the cheating ISP is interested in maximizing its gain, rather than minimizing the other ISP's gain, this provides a disincentive against cheating. Further, recall that even if there exist strategies by which a cheater gains, the structure of Nexit is such that an honest ISP can always protect itself by not negotiating loses.

6 Deployment Considerations

In this section, we outline how Nexit might be integrated into the current Internet. While we do not present a detailed design, we discuss several key issues concerning a practical deployment to argue for its plausibility.

Integration with ISP routing Instead of an in-band integration with BGP, we advocate an out-of-band integration with routing as shown in Figure 12. Negotiation agents use the current state of the network to map routing alternatives to preference classes and disclose these preferences. Once the path has been negotiated, low-level BGP mechanisms such as *local-prefs* are used to implement it. This architecture is similar to RCP [7] and has three important advantages in our context. First, negotiation requires a holistic view of traffic (with ISPs losing on some flows and gaining on others) which is more cleanly accomplished with a centralized approach. Second, it avoids overloading an already fragile BGP. Third, it does not require ISPs to modify the bulk of deployed routers to benefit from negotiation.

Identifying flows for negotiation The ISPs partition the traffic they exchange into flows. Recall that a flow is a stream of packets from a node in one ISP to a node in the other. We need identifiable flow signatures because ISPs typically do not know where packets enter or leave the other ISP's network. Routing prefixes provide a basis for such a signature. Assume that the two ISPs agree on a common set of prefixes, for instance, the union of the prefixes they announce to each other through BGP. Also assume that if the prefix is aggregated, the subprefixes attach to the aggregating network at the same place.

A flow is uniquely identified using the (most specific) source and destination prefixes of its packets and an identifier that corresponds to its ingress into the upstream. When traffic that is not covered by an existing flow is observed, the upstream signals the arrival of a new flow. It informs the downstream of the two prefixes (which the downstream can also observe independently), its choice of the identifier, and the estimated flow size. To prevent information leakage, the upstream chooses different identifiers for different flows that enter at the same place.

The upstream periodically refreshes the information on active flows and flows that are inactive for a certain period are timed out.

As a practical matter, to improve scalability ISPs can decide to negotiate over only the set of long-lived and high-bandwidth (or important) flows. For this, the upstream will trigger a new flow only if its size stays above a threshold for a certain period of time. Optimizing the small fraction of high-bandwidth flows can optimize most of the traffic [31].

Input data The input data required for negotiation depend on the ISP optimization criteria but most of it can be obtained using today's technology. For instance, the network path of a given flow can be computed using the current routing state (e.g., OSPF weights or MPLS configuration). The distance of a flow through the network can be computed using the distance of individual edges. Link utilization can be obtained using SNMP probes. Information on existing flows and their sizes can be gathered using NetFlow or similar tools.

When to negotiate? While we have conceptually described negotiation as being a one-shot event between neighboring ISPs, in practice, it will be a continuous process. ISPs inform each other of their updated preferences for each flow being exchanged. These would be used to continually find routing patterns that benefit both ISPs.

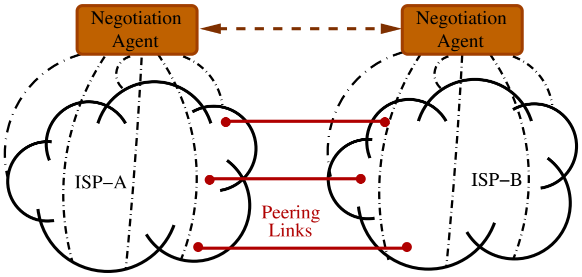


Figure 12: *Integrating negotiation with current ISP routing. Logically, the negotiation agents sit on top of the routing infrastructure. They collect data concerning the state of the network as inputs to negotiation and appropriately configure the routers to implement the negotiated solution.*

Dealing with changes Certain events, such as failures or increase in traffic quantity, require an ISP to change where or how much traffic it sends to its neighbor. There are two ways to address such events. First, the ISP informs its neighbors of the upcoming changes, negotiates with them and routes accordingly. This ensures that ISPs do not violate each other's resource constraints. However, waiting for negotiation to end before routing the flows could lead to heavy packet loss or delay if there are no alternate paths. Thus, this method is more appropriate when alternate paths exist. The second method is more suited for unplanned changes: the ISP simultaneously routes the flows and opens up a negotiation channel with the other ISP. The two ISPs then negotiate, at the end of which the flows may be rerouted. This enables faster restoration of service, with the danger that one ISP might initially overload the other. ISPs can protect other traffic from such transient overloads by lowering the priority of non-negotiated traffic entering their network.

ISPs can easily verify whether the traffic exchange complies with what was negotiated. If unilateral changes are detected (without a renegotiation request as described above), the ISP can partially or fully rollback the compromises made in return.

7 Related Work

Two works have addressed the problem of enabling neighboring domains to cooperatively manage the traffic they exchange. First, Machiraju and Katz use secure multi-party computation (SMPC) to enable ISPs to select interconnections without directly disclosing private information [16]. In contrast to our approach, they assume that both ISPs optimize for the same objective and do not consider the possibility of an ISP losing through cooperation. In the future, we plan to explore combining SMPC with negotiation to design mechanisms that are both flexible and do not require the disclosure of internal metrics.

Second, Winick *et al.* propose a method in which before moving traffic an upstream ISP informs the downstream of changes it intends to make [31]. The downstream decides if those changes are acceptable. This method is one form of negotiation. Instead, Nexit uses preference lists to compute a mutually acceptable solution. This is useful because the solution space is exponential in the number of flows and it is hard for one ISP to find good solutions without input from the second.

Mechanisms that enable autonomous entities to cooperate have received much attention in recent years. A popular approach is distributed algorithmic mechanism design (DAMD) [9]. The two applications of this approach, of which we aware, are both *direct* mechanisms in which the entities disclose their costs [8, 24]. These costs are used to compute solutions that satisfy the desired property such as social optimality. Competitive concerns are addressed by proving that a cheating entity cannot manipulate the solution in its favor even with the knowledge of others' costs. However, this may not capture all real-world competitive concerns [9]. For instance, an ISP can use the knowledge of the competitor's cost to plan its own network in a way that undercuts the competitor's profits. We address this concern by disclosing only coarse grained, opaque preference classes. Another advantage of our approach is that the ISPs do not need to map their optimization metric to true cost, which can be difficult, if not impossible [29].

Researchers have advocated the use of money as the basis for interdomain traffic control [1, 20]. These works propose that downstreams advertise the price of carrying traffic as part of routing announcements and upstreams pay this price while sending traffic. This approach assumes that ISPs are able and willing to advertise path prices. Additionally, this approach appears to be incompatible with the charging model in the Internet, in which monetary payments flow from customer to provider ISPs irrespective of the direction of the traffic.

Our work is another piece in the research theme that examines the "price of anarchy" in the Internet. While other researchers have studied selfish routing by individual users [27, 25], we study selfish routing by ISPs. Johari and Tsitsiklis use a graphical argument to show that the latency of early-exit routing can be three times that of optimal routing [13]. Our results over real ISP topologies show that this is much lower in practice.

Finally, we draw broadly on concepts in negotiation theory [3, 21, 26], though our specific techniques are geared towards our problem domain.

8 Conclusions

In this paper, we have explored negotiation as the basis for cooperation between competing entities. Our focus has been two neighboring ISPs with multiple interconnections, which forms the base case for interdomain routing in the Internet. We presented Nexit, an inter-ISP negotiation framework in which ISPs disclose only coarse, opaque preference classes, much like BGP MEDs, to each other and jointly decide paths for the flows they exchange.

Using simulation with over sixty measured ISP topologies, for both bandwidth and distance metrics, we showed that the quality of negotiated routing is close to that of globally optimal routing which considers both ISPs to be part of one larger system. The success of negotiation stems from the fact that ISPs can trade small losses for significant gains; when applied across flows this leads to a net gain for both ISPs. While globally optimal routing can lead to both winners and losers, both ISPs benefit with negotiation, providing a strong incentive to negotiate. The benefit is often substantial for bandwidth measures, lessening the likelihood of congestion in either network. The benefit for distance is small on average, suggesting that the overall "price of anarchy" is low in practice. We also showed that because of the trading nature of negotiation, an ISP that lies can lose compared to being truthful.

In the Internet routing context, negotiation has advantages beyond the more easily quantifiable performance benefits. It can increase stability as ISPs do not inadvertently violate each other's resource constraints in a way that might set off a reactionary chain of events. It relieves operators from some of the time-consuming and error-prone tasks related to route optimization. It enables ISPs to jointly optimize traffic for profitable services such as VPNs (virtual private networks). Today, such services are limited to individual providers, and thus have limited reach. As part of ongoing effort we are working towards a prototype implementation of Nexit that will work in concert with BGP.

Our work is a first step towards designing an Internet-wide negotiation mechanism and, more broadly, understanding the trade-offs involved in the design of protocols between competing yet cooperating entities. Stability and efficiency of such systems requires that the participants have a global perspective while making local decisions. Our study shows that negotiation can be highly effective towards that goal.

9 Acknowledgements

We thank David Kempe and John Zahorjan for useful discussions and Anukool Lakhina for help with computing city population. We also thank the anonymous reviewers and our shepherd, Jon Crowcroft, for their feedback.

This work was supported in part by NSF (Grant No. ANI-0133495), Cisco Systems, and Microsoft Research.

References

[1] M. Afergan and J. Wroclawski. On the benefits and feasibility of incentive based routing infrastructure. In *ACM SIGCOMM PINS*, Sept. 2004.

[2] S. Bhattacharyya, C. Diot, J. Jetcheva, and N. Taft. PoP-level and access-link-level traffic dynamics in a Tier-1 PoP. In *ACM SIGCOMM IMW*, Nov. 2001.

[3] S. J. Brams. *Negotiation Games: Applying game theory to bargaining and arbi tration*. Routeledge, 1990.

[4] C.-N. Chuah. A Tier-1 ISP perspective: Design principles & observations of routing behavior. http://sahara.cs.berkeley.edu/jun2002-retreat/chuah_talk.pdf, 2002 June.

[5] Center for International Earth and Science Information Network. http://www.ciesin.columbia.edu.

[6] D. Clark, J. Wroclawski, K. Sollins, and R. Braden. Tussle in cyberspace: Defining tomorrow's Internet. In *ACM SIGCOMM*, Aug. 2002.

[7] N. Feamster, H. Balakrishnan, J. Rexford, A. Shaikh, and J. van der Merwe. The case for separating routing from routers. In *ACM SIGCOMM FDNA*, Aug. 2004.

[8] J. Feigenbaum, C. Papadimitriou, R. Sami, and S. Shenker. A BGP-based mechanism for lowest-cost routing. In *ACM PODC*, July 2002.

[9] J. Feigenbaum and S. Shenker. Distributed algorithmic mechanism design: Recent results and future directions. In *the 6th International Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications*, Sept. 2002.

[10] B. Fortz and M. Thorup. Internet traffic engineering by optimizing OSPF weights. In *IEEE INFOCOM*, Apr. 2000.

[11] L. Gao. On inferring autonomous system relationships in the Internet. In *IEEE Global Internet Symposium*, Nov. 2000.

[12] V. Gill. Private Communication, Nov. 2003.

[13] R. Johari and J. N. Tsitsiklis. Routing and peering in a competitive Internet. Technical Report P-2570, MIT LIDS, Jan. 2003.

[14] A. Lakhina, J. Byers, M. Crovella, and I. Matta. On the geographic location of Internet resources. *IEEE JSAC*, 2003.

[15] K. Lougheed and Y. Rekhter. A border gateway protocol (BGP). RFC 1105, IETF, June 1989.

[16] S. Machiraju and R. Katz. Reconciling cooperation with confidentiality in multi-provider distributed systems. Technical Report CSD-4-1345, UC Berkeley, Aug. 2004.

[17] R. Mahajan, D. Wetherall, and T. Anderson. Understanding BGP misconfiguration. In *ACM SIGCOMM*, Aug. 2002.

[18] R. Mahajan, D. Wetherall, and T. Anderson. Towards coordinated interdomain traffic engineering. In *HotNets-III*, Nov. 2004.

[19] A. Medina, N. Taft, K. Salamatian, S. Bhattacharyya, and C. Diot. Traffic matrix estimation: Existing techniques and new directions. In *ACM SIGCOMM*, Aug. 2002.

[20] R. Mortier and I. Pratt. Incentive based inter-domain routeing. In *Internet Charging and QoS Technology Workshop (ICQT'03)*, Sept. 2003.

[21] R. B. Myerson and M. A. Satterthwaite. Efficient mechanisms for bilateral trading. *Journal of Economic Theory*, 29(2), Apr. 1983. Cited in Brams [3].

[22] V. N. Padmanabhan and L. Subramanian. An investigation of geographic mapping techniques for Internet hosts. In *ACM SIGCOMM*, Aug. 2001.

[23] C. Papadimitriou. Algorithms, games, and the Internet. In *ACM STOC*, July 2001.

[24] G. Philips, S. Shenker, and H. Tangmunarunkit. Scaling of multicast trees: Comments on the Chuang-Sirbu scaling law. In *ACM SIGCOMM*, Aug. 1999.

[25] L. Qiu, Y. R. Yang, Y. Zhang, and S. Shenker. On selfish routing in Internet-like environments. In *ACM SIGCOMM*, Aug. 2003.

[26] H. Raiffa. *The art and science of negotiation*. Harvard University Press, 1982.

[27] T. Roughgarden and E. Tardos. How bad is selfish routing? *Journal of the ACM*, 49(2), Mar. 2002.

[28] S. Savage, A. Collins, E. Hoffman, J. Snell, and T. Anderson. The end-to-end effects of Internet path selection. In *ACM SIGCOMM*, Aug. 1999.

[29] S. Shenker, D. Clark, D. Estrin, and S. Herzog. Pricing in computer networks: Reshaping the research agenda. *ACM SIGCOMM CCR*, 26(2), Apr. 1996.

[30] N. Spring, R. Mahajan, and T. Anderson. Quantifying the causes of path inflation. In *ACM SIGCOMM*, Aug. 2003.

[31] J. Winick, S. Jamin, and J. Rexford. Traffic engineering between neighboring domains. http://www.research.att.com/~jrex/papers/interAS.pdf, July 2002.

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

Notes

[1] In game theory parlance, the two-ISP situation is not zero-sum but is akin to prisoner's dilemma because both players benefit from cooperation.

[2] Empirical evaluation with destination-based routing yields results similar to those in Section 5.

[3] A subtle advantage of opaque preferences is that it makes negotiation "jealousy free" because one ISP cannot determine whether the other profits more in any meaningful terms.