Stable Route Selection for Interdomain Traffic Engineering[*]

Y. Richard Yang Haiyong Xie Hao Wang Li Erran Li
Yanbin Liu Avi Silberschatz Arvind Krishnamurthy [†]

March 15, 2005

Abstract

We investigate a general model of route selection for interdomain traffic engineering where the routing of multiple destinations can be coordinated. We identify potential instability and inefficiency problems, derive practical guidelines to guarantee stability without global coordination, and evaluate routing instability of interdomain traffic engineering using realistic Internet topology. We further extend our model so that the local preference of an AS can depend on not only its routes to the destinations but also its ingress traffic patterns.

1 Introduction

The global Internet consists of a large number of interconnected autonomous systems (AS), where each AS (e.g., AT&T) is administrated autonomously. Neighboring ASes exchange their routes using the Border Gateway Protocol (BGP) [13], where each route consists of a path vector of ASes from a source to a destination. After learning routes from all of its neighbors, an AS selects the best available routes, according to its local route selection policies. Recently, ASes are increasingly adopting local route selection policies to achieve their interdomain traffic engineering objectives (*e.g.*, [12]). We have recently conducted an email survey of ISPs, and the results indicate that many ISPs choose routes to achieve their interdomain traffic engineering objectives, such as satisfying the capacity constraints of links between neighboring ASes, load-balancing interdomain traffic, and/or minimizing cost due to traffic to and from ASes which are the service providers of an AS.

Despite this emerging trend, so far there are few systematic studies on the stability and efficiency of the global Internet with interdomain route selection for interdomain traffic engineering. A major breakthrough was made recently when Griffin *et al.* [8, 9, 14] proposed systematic models to study the stability of path-vector interdomain routing. In particular, they identified the existence of policy disputes as a potential reason for routing instability. By routing instability, they mean persistent route oscillations even though the network topology is stable. Although these models capture a wide range of interdomain traffic engineering objectives, they assume that the routing decisions of different destinations can be separated. Thus the models apply only to networks where there is no AS whose routing policies require it to coordinate its route selection to multiple destinations. On the other hand, a fundamental feature of route selection for interdomain traffic engineering in particular and traffic engineering in general is that route selection constraints (*e.g.*, traffic assigned to a link is within link capacity) and/or objective functions (*e.g.*, load balance) involve the route selection of multiple destinations. Thus, in route selection for interdomain traffic engineering, whether a route will be chosen for a given destination will depend on what routes are available or chosen for other destinations. For example, if an

[*] Submitted to IEEE Network — Special Issue on Interdomain Routing, November/December 2005. For questions and comments, please contact with Y. Richard Yang (yry@cs.yale.edu).

[†] Y. Richard Yang, Haiyong Xie, Hao Wang, Avi Silberschatz and Arvind Krishnamurthy are with Yale University. Li Erran Li is with Bell-labs. Yanbin Liu is with the University of Texas at Austin.

AS selects routes for each destination independently without considering the chosen/available routes of other destinations, in the worst case it may choose the same access link for all destinations, violating link capacity constraints and/or causing load imbalance.

In this article, motivated by previous models and the increasing usage of route selection for interdomain traffic engineering, we summarize our recent results on an analysis and evaluation of a general route selection model for interdomain traffic engineering; for formal models and detailed analysis, we refer the interest readers to [15]. In Section 2, we analyze a general model to capture route selection for interdomain traffic engineering where route selection for multiple destinations is coordinated. We identify that there exist networks where the interaction of the route selection of multiple destinations can cause routing instability, even though the route selection of the networks is guaranteed to converge when each destination is considered alone. In Section 3, we evaluate a set of practical guidelines, and show that if these guidelines are followed by the ASes, route selection for interdomain traffic engineering will be stable without explicit global coordination. In Section 4, using realistic Internet AS topology, we conduct simulations to show that without coordination, even with a small number of ASes coordinating route selection for just a small number of destinations, we can observe instability. In Section 5, we study a more general route selection model where the preference of an AS depends on not only its routes to the destinations but also its ingress traffic patterns. We show that there are networks which will be unstable when the ASes strictly follow AS business guidelines, and adopt any *adaptive route selection algorithms*. Our conclusion and future work are in Section 6.

2 Route Selection for Egress Interdomain Traffic Engineering

2.1 Motivation

We start with a very simple illustrative example as shown in Figure 1. The majority of the traffic of AS S goes to two destinations D_1 and D_2. Assume S wants to balance its outgoing traffic. Thus, it wants to choose a combination of routes for destinations D_1 and D_2 such that they use different neighbors, if possible, to have low utilization on the two links SA and SB. We refer to a combination of routes for D_1 and D_2 as a *route profile*. Since S may not know in advance the routes it will learn from its neighbors A and B, or the routes that A and B will export to S can be dynamic given network dynamics, S needs an automatic method to pick the best route profile, according to currently available routes. One way S specifies its preference is to define an interdomain traffic engineering objective function (*e.g.*, minimize the maximum of the utilization of the two links for this case). An advantage of using an objective function is its compact representation. Given the objective function, link capacities, and traffic demands, a traffic engineering program searches for the best route profile automatically and dynamically, according to currently available routes. The preference can also be specified by a policy language. An example policy can be: if D_1 and D_2 use different links, assign a base local preference of 100; otherwise, a base local preference of 0. If D_1 uses link SA, add 10 to local preference. If D_2 uses link SB, add 5 to local preference. The program picks the available route profile with the highest local preference. For generality, we assume a ranking table at each AS, which lists, in decreasing order, all of the potential route profiles. An example route ranking table for S is shown in Figure 1, where each row is a route profile, *i.e.*, a combination of routes for D_1 and D_2. For example, the best route profile for S is (SAD_1, SAD_2); *i.e.*, S uses SAD_1 for destination D_1, and SAD_2 for destination D_2. The worst route profile is SBD_1 and SBD_2. Thus, if the route profile (SAD_1, SAD_2) is available, S will choose it. On the other hand, if the only available route profile is (SBD_1, SBD_2), S has no choice but to use it.

2.2 Problem Definition

Now we define in more detail the problem of route selection for interdomain traffic engineering. To simplify our exposition, we make the following assumptions. We assume a connected network with a set $\mathcal{S}$ of source ASes and

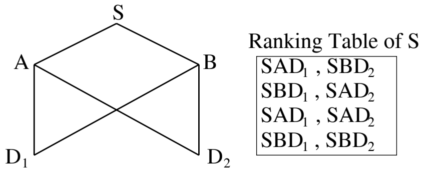


Figure 1: Egress load balancing: an example motivating the need for destination path interaction.

a set $\mathcal{D}$ of destination ASes. We assume that the underlying network infrastructure is stable so that we can focus on the effects of interdomain traffic engineering policies. We assume that there is only one link between two neighboring ASes; we leave an investigation on the interaction of intradomain routing and interdomain routing as future work (*e.g.*, [4, 10]). We focus on route selection and assume static export policies. We first consider the case that the preference of an AS depends only on the routes from the AS itself to the destinations. In other words, the ASes are conducting *egress interdomain traffic engineering*, which is one of the major tasks of ISP interdomain traffic engineering [2]. In Section 5 we will further extend this model and study route selection for general interdomain traffic engineering, in which case the route from each source to the AS itself also matters.

Now we define the stable route selection for egress interdomain traffic engineering problem. We define for AS i the set of all potential routes to all destinations as $R_i = \prod_{d \in \mathcal{D}} R_{i \to d}$, where $R_{i \to d}$ is the set of all possible routes from i to destination d. We allow empty route. We refer to an element $r_i \in R_i$ as a *route profile* of i, since r_i completely specifies the routes from i to each destination. The preference of i on routes in R_i is represented by a ranking table of route profiles. Specifically, there is a ranking function $\mathcal{R}_i$ which maps the set of routes R_i to a total order set Λ; *i.e.*, $\mathcal{R}_i : R_i \to \Lambda$. The ranking function $\mathcal{R}_i$ can be i's traffic engineering objective function. Hereafter, we assume that i's preference of routes is given in the form of a ranking table. We emphasize again that this ranking table is just a general representation of some more compact representations such as objective functions or policy languages.

An AS uses its route ranking table to select the best available routes. Figure 2 shows the standard BGP protocol/process model of interdomain route selection [7–9, 14], naturally extended to multiple destinations. Specifically, each AS maintains a routing cache for each destination of currently available routes, exported by its neighbors. AS i selects routes from its routing cache, one route $r_{i \to d}$ for each destination d, so that the chosen route profile r_i has the highest rank; *i.e.*, $\mathcal{R}_i(r_i) > \mathcal{R}_i(r'_i)$, for any other route profile r'_i available from the routing cache. This chosen route profile r_i will then be used by i to route packets. If $r_{i \to d}$ is different from the previously selected route to d, i then withdraws the previous route, and exports the new route to the neighbors that are allowed to receive this route according to i's export policy. We assume that BGP route update messages between neighboring ASes are reliably delivered in FIFO order. This is reasonable as the messages are sent via TCP. We also assume that each message will be processed with bounded delay.

A *network route selection* is a combination of route profiles, one for each AS. A network route selection is *stable* if no AS can choose a higher ranked route profile from the exported routes of its neighbors. We also call a stable network route selection a stable route solution or solution for short.

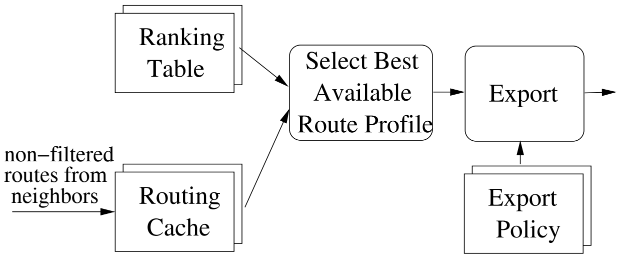
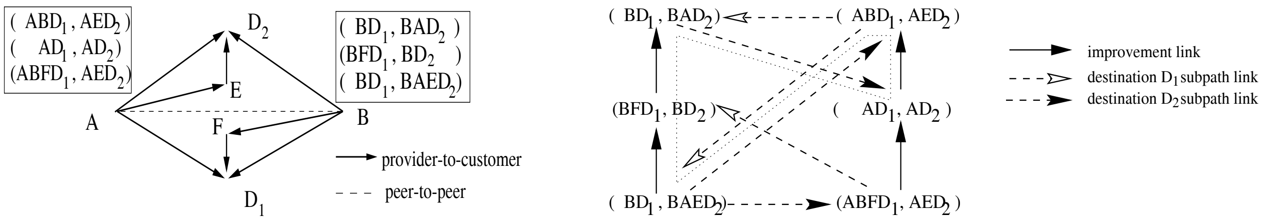

Ranking Table
Select Best Available Route Profile
Export
non−filtered routes from neighbors
Routing Cache
Export Policy

Figure 2: The protocol/process model of route selection for interdomain traffic engineering.

2.3 Multi-Destination Interactions Can Cause Instability



(a) route ranking tables of A and B.

(b) P-graph and P-cycle of network (a).

Figure 3: A network which has no stable route selection.

A somehow unexpected result is that the interaction of the routing of multiple destinations due to interdomain traffic engineering can cause routing instability. The network shown in Figure 3(a) is one such interesting example. For clarity, we show only the highest-ranked three route profiles of A and B. To make this example more realistic, the ASes export their routes according to their business relationship. There are two major types of business relationship in the current Internet. The first type is provider-and-customer relationship, where a provider provides transit service to its customers. We refer to the connection from a provider to a customer as a provider-to-customer link; such a link is represented by a directed edge from the provider to the customer. The second major type of business relationship is peer-to-peer relationship, where a pair of ASes provide transit services to the customers of each other. We refer to the connection between a pair of peers as a peer link; such a link is represented by a dashed edge between the two peers. These business relationships imply that the export policies of ASes in the Internet follow the *typical export policies* [7]: 1) each AS exports to its providers its own routes and those it learned from its customers, but does not export to its providers the routes it learned from its peers or other providers; 2) each AS exports to its customers its own routes and any routes it learned from others; 3) each AS exports to its peers its own routes and those it learned from its customers, but does not export those it learned from its providers or other peers.

We first consider each destination separately. For destination D_1, the two routes for D_1 contained in the two highest route profiles of A are ABD_1 and AD_1; the two routes for D_1 contained in the two highest route profiles of B are BD_1 and BFD_1. Consider this combination of route preference for D_1. The network has the stable route solution of ABD_1 and BD_1 for A and B, respectively. One can also verify that if we consider D_2 alone, the network has the stable route solution of AED_2 and BD_2 for A and B, respectively. Thus, if there were no interaction among destinations, A and B would settle to the stable solutions of (ABD_1, AED_2) and (BD_1, BD_2), respectively.

Next we consider destination interaction. The above solutions obtained by considering each destination alone are no longer stable. For example, B will not choose (BD_1, BD_2) since this route profile has a low rank. One can verify that the network has no stable solution at all. Specifically, we observe that the export policies of the ASes make the route profile (AD_1, AD_2) alway available to A. Thus to see that the network has no stable solutions, we just need to verify that there is no stable route solution when A chooses (AD_1, AD_2) or (ABD_1, AED_2). Clearly, there is no stable solution for (AD_1, AD_2) since if A chooses (AD_1, AD_2), B will choose (BD_1, BAD_2); this causes A to change to (ABD_1, AED_2). However, there will be no stable route selection for (ABD_1, AED_2) neither. To make (ABD_1, AED_2) available to A, B must choose BD_1 for D_1. Since (BFD_1, BD_2) is always available to B, it must be the case that B chooses (BD_1, BAD_2). However, this requires A to choose AD_2, which is inconsistent with (ABD_1, AED_2). Thus, the network has no stable route selections due to destination interaction!

2.4 Stable, Robust Route Selection and Protocol Convergence

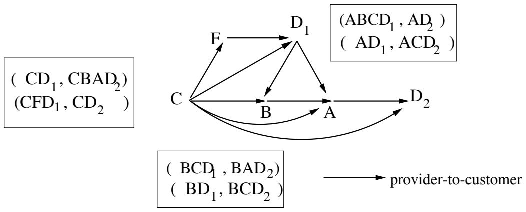
Given that multi-destination interaction due to interdomain traffic engineering can result in no stable route selection, in this section, we derive a sufficient condition that can guarantee stable, robust route selection and protocol convergence.

We first introduce the notion of a *P-graph* to capture the interaction of the interdomain traffic engineering policies of multiple ASes. The notion of a P-graph is motivated by the partial order graph of Griffin *et al.* [8], but generalized to interdomain traffic engineering. The nodes of a P-graph are from the union $\bigcup_i R_i$, namely, all route profiles of all ASes. We consider only route profiles that are allowed by export policies. There are two types of directed edges in a P-graph. The first type of edges are improvement edges. There is an improvement edge from node r_i to r_i' if i prefers route profile r_i to r_i'. The second type of edges are sub-path edges. There is a destination D_j sub-path edge from a node r_i to another node r_j if the path in r_j for destination D_j is a sub path of that in r_i. A *P-cycle* is a loop in the P-graph of the following special format: one or more improvement edges, followed by one or more sub-path edges of the same destination, followed by one or more improvement edges, and so on. For example, Figure 3(b) shows the P-graph and the P-cycle for the network of Figure 3(a). Note that there may be trivial loops in a P-graph which are not of the format of a P-cycle. For example, the loop consisting of (BD_1, BAD_2), (AD_1, AD_2) and (ABD_1, AED_2) is not a P-cycle, since there are two consecutive sub-path edges of different destinations.

As our following theorem shows, if there is no P-cycle, the BGP protocol will converge.

Theorem 1 *If the P-graph has no P-cycle, then the BGP protocol converges.*

2.5 Network with non-Pareto Optimal Solution



(a) network.

	A	B	C
Solution 1	$(ABCD_1, AD_2)$	(BCD_1, BAD_2)	$(CD_1, CBAD_2)$
Solution 2	(AD_1, ACD_2)	(BD_1, BCD_2)	(CFD_1, CD_2)

(b) two stable route selections of (a).

Figure 4: An example with two solutions but one of them is not Pareto optimal.

A network with stable solutions can have multiple solutions. The example in Figure 4(a) is one example. This example is particularly interesting in that it has two stable route solutions, as shown in Figure 4(b), and the solution at the second row is not even Pareto optimal. Specifically, a stable route solution is Pareto optimal if there does not exist another stable route solution where each AS has a higher ranked route profile. This example clearly demonstrates that to be effective, explicit coordination of route selection (*e.g.*, negotiation) may involve more than two parties.

3 Stable Egress Route Selection for Interdomain Traffic Engineering without Global Coordination

The preceding section presents a sufficient condition to guarantee the existence and convergence of route selection. The condition depends on checking P-cycle. In practice, it is difficult to obtain P-graph and check whether it contains a P-cycle. This is due to the fact that BGP is a distributed protocol, and generally ASes do not share their traffic engineering policies. Also, the preceding section considers general networks, while in the current Internet, route selection of ASes is constrained by their business-relationships. In this section, we seek rigorous, practical interdomain traffic engineering guidelines that are reasonable according to current AS business relationships, can be checked locally, and can guarantee convergence. In other words, if each AS follows these practical interdomain traffic engineering guidelines, route selection is stable and the converged route selection is unique. In particular, we are motivated by the study of Gao *et al.* [6, 7], which proposed guidelines to guarantee route convergence without global coordination in the practical setting of the Internet if each destination is considered separately. When ASes coordinate route selection for multiple destinations, we find that, to guarantee the existence and uniqueness of stable route selection for interdomain traffic engineering, we need stronger conditions.

We assume that ASes follow the typical export policies (please see Section 2.3). Such export policies imply that all valid routes have the following patterns [7]: a provider-to-customer link can be followed by only provider-to-customer links, and a peer link can be followed by only provider-to-customer links. Accordingly, we divide the routes from an AS i to a destination d into three categories:

- *Customer route*: each link along a customer route is a provider-to-customer link.
- *Peer route*: the first link along a peer route is a peer link, and the remaining links are all provider-to-customer links.
- *Provider route*: the first link is a customer-to-provider link, and the remaining route consists of zero or multiple customer-to-provider links, followed by zero or one peer link, and then zero or multiple provider-to-customer links.

We also divide the set of destinations of an AS i into three categories, according to the given network topology and export policies:

- *Customer-reachable destinations*: these destinations are direct or transitive customers of AS i.
- *Peer-reachable destinations*: these destinations are direct or transitive customers of one of AS i's peers, but they are not direct or transitive customers of AS i. We call the union of customer-reachable and peer-reachable destinations the set of *customer-peer-reachable destinations*.
- *Provider-reachable destinations*: these destinations are direct or transitive customers of one of AS i's direct or transitive providers, but they are not direct or transitive customers of either i or one of AS i's peers. We call the union of peer-reachable and provider-reachable destinations the set of *peer-provider-reachable destinations*.

Our first guideline is that customer routes are strictly preferred over peer routes, which are preferred over provider routes. We write this as customer routes $\succ$ peer routes $\succ$ provider routes. We call this guideline *the standard route preference guideline*, or *the preference guideline* for short.

Our second guideline is that routing decisions for different categories of destinations are *decoupled*. Specifically, we say that the routing decisions of an AS are *decoupled* if the following two conditions are satisfied. First, the routing decisions for its customer-reachable destinations depend only on the routing decisions for its other customer-reachable destinations, and are independent of the routing decisions for its peer-provider-reachable destinations. Second, the routing decisions for its peer-reachable destinations depend only on the routing decisions of its customer-peer-reachable destinations, and are independent of the routing decisions of its provider-reachable destinations. The decoupling of route selections can be reasonable when the destinations of different categories use different sets of access links and are economically different. When the routing decisions of AS

i are decoupled, we say that it follows *the standard route decoupling guideline*, or *the decoupling guideline* for short.

We now have the following theorem if the above guidelines are followed by all ASes.

Theorem 2 *The network has a unique stable route selection which BGP is guaranteed to converge to, if the following conditions hold:*

1. *there is no loop formed by provider-to-customer links in the network;*
2. *all ASes have fixed typical export policies;*
3. *all ASes strictly prefer customer routes over peer routes over provider routes;*
4. *the routing decisions for customer-reachable, peer-reachable, and provider-reachable destinations are decoupled.*

4 Simulation Studies of Routing Instability

The preceding sections analyze the stability of route selection for interdomain traffic engineering. In this section, we use simulations to study the likelihood of routing instability when the conditions of Theorem 2 are not satisfied.

4.1 Methodology

We construct the AS topology of the Internet using the BGP table of Univ. of Oregon Routeviews and the BGP tables of 18 Looking Glass servers. In order to make the simulations more efficient, we iteratively remove 6157 leaf ASes (degree 1 nodes) and their links from the topology. The remaining network has 13,048 ASes and 37,999 links. We infer business relationships among the ASes to produce the *AS business-relationship graph*.

An important component of our simulation studies is route ranking tables. For AS i who does not coordinate the route selection of multiple destinations, we use the subjective routing framework to construct its route ranking table [3]. The subjective routing framework is motivated by the observation that different ASes often use different performance metrics in comparing routes. Thus, in this framework, there is a set M of performance metrics assigned to each link. Each AS computes the cost of a route using its own set of weights. Specifically, AS i has a set of weights, $W_i = \{w_{i,m} | m \in M\}$, where $w_{i,m}$ is the weight associated with the performance metric m. Note that $w_{i,m} = 0$ if i is not concerned with the metric m. Let $C_l^{(m)}$ be the value of metric m at link l. Given a route $r_{i \to d}$ from AS i to destination d, AS i computes the cost of this route as $c(r_{i \to d}) = \sum_{m \in M} w_{i,m} \sum_{l \in r_{i \to d}} C_l^{(m)}$. For each destination, AS i chooses the route with the lowest subjective cost as its best route for that destination.

For an AS i who coordinates its route selection of multiple destinations, we construct its ranking table as follows. First, for each destination d, we compute the set $R_{i \to d}$ of all feasible routes from i to d, assuming all ASes have typical export policies. Then we construct the set of all possible route profiles $R_i = \prod_{d \in \mathcal{D}} R_{i \to d}$. For efficiency, we do not explicitly store R_i; instead, we store just the set of all feasible routes to all destinations (*i.e.*, $\cup_{d \in \mathcal{D}} R_{i \to d}$), and assign a unique ID to each route in this set; therefore, we represent a route profile using a set of IDs corresponding to the routes in the route profile. Finally, we construct the ranking table of AS i by randomly permuting the entries of R_i.

We implement our own event-driven simulator to study the stable route selection problem for interdomain traffic engineering. The simulator simulates BGP protocol process such as route import/export, route announcement/withdrawal, and so on. Each AS selects its routes as described above. We also add random delays to route import/export events in order to simulate network asynchronousness. In each experiment, we randomly choose a set of ASes as destinations, and all other ASes exchange routes to these destinations.

To detect instability, for each AS, our simulator keeps a history of its selected route profiles. Specifically, according to its route selection history, each AS constructs a directed stability graph with each node representing a unique route profile and each directed edge representing a temporal transition between two route profiles. An AS has no stable route selection if all nodes of the stability graph are in one single strongly connected component. Hereafter, we refer to such ASes as *unstable* ASes. Since this condition is a sufficient condition, we may underestimate the extent of instability. In order to avoid taking initial route exchanges as unstable route selection, we wait for a long enough time before checking instability. Specifically, we start to keep a history of previous best route profiles for each AS after 500 simulation steps when all ASes have routes to all destinations. We start to check the instability condition for each AS every 20 simulation steps after the routing history starts. We run the simulation for 7,000 simulation steps so that the number of ASes identified as unstable does not change any more, and take this number as the number of unstable ASes.

4.2 Routing Instability Caused by Route Coordination

We investigate routing instability caused by coordinated route selection of multiple destinations. To investigate the potential seriousness of the problem, we setup the experiments so that only a small number of ASes coordinate route selection and violate the guidelines. Specifically, we randomly choose just 53 ASes who coordinate their route selection for multiple destinations but violate the standard route preference and decoupling guidelines. Each of these 53 ASes coordinates the route selection of just 2 destinations. The way the route ranking tables of these 53 ASes is constructed is described in Section 4.1.

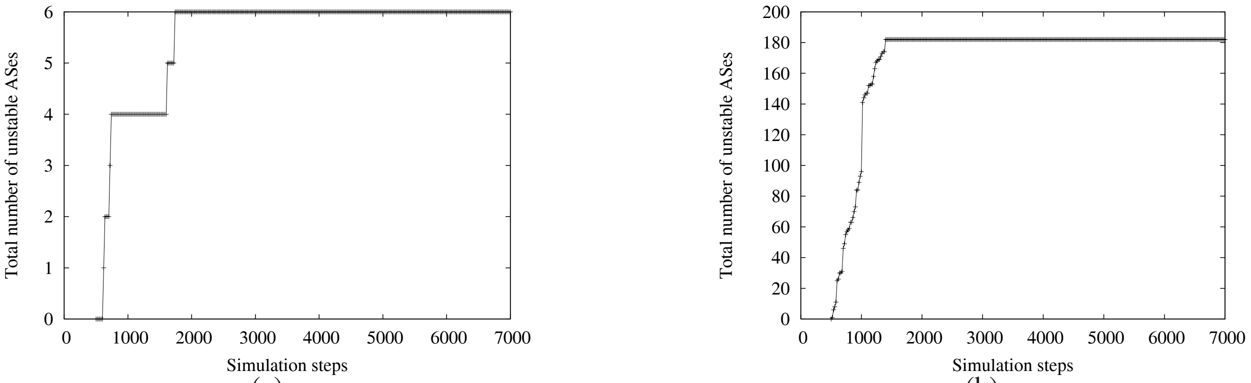
We first study the setup when the remaining ASes follow the standard route preference and select routes for each destination separately. Figure 5(a) shows the result. We observe that there are already 6 ASes who are unstable in the network. This result is surprising in that 53 ASes consist of a very small percentage (53 out of 13048) of the total number of ASes. Furthermore, 2 destinations are not many destinations. We also vary the number of ASes who coordinate route selection and the number of destinations. We observe that the number of unstable ASes further increases as the number of ASes who coordinate route selection but do not follow the guidelines increases.

In the preceding experiment, all of the ASes who select routes for each destination separately follow the standard route preference. However, from our measurements of route selection in the current Internet, we observe that there are ASes who select routes separately but violate the standard route preference condition. We vary the preceding experiment so that in addition to the 53 ASes who coordinate route selection, the remaining ASes who select routes separately now violate the standard route preference condition with a small probability of 0.03. Our simulation result is shown in Figure 5(b). We observe that the number of unstable ASes now increases from 6 to about 180.

5 Route Selection for General Interdomain Traffic Engineering

5.1 Motivation

In the preceding sections, the preference of an AS depends only on egress route profiles and is independent of ingress traffic demand patterns. As a result of this independency, we derive a set of practical guidelines which can guarantee stability for egress interdomain traffic engineering. This independency is justified when the traffic demands of an AS to its destinations are known, and thus can be considered as constants. Specifically, these demands will be used as constant parameters in the ranking function of an AS to determine the relative ranking of route profiles. Conceptually, therefore, these demands do not need to appear explicitly in the route ranking table as conditions. For example, in Figure 1, the traffic demands to D_1 and D_2 will be used in determining the relative ranking of route profiles of S. However, these traffic demands do not need to appear explicitly



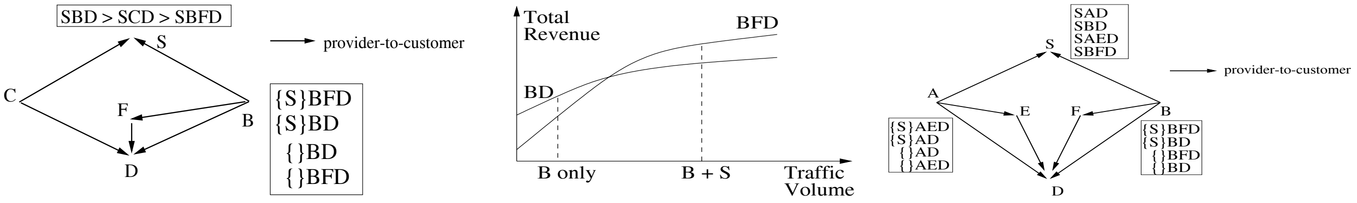
(a)

(b)

Figure 5: Total numbers of unstable ASes due to violation of the practical guidelines.

in the route ranking table. Some examples where this is true include multihomed stub ASes, and ISPs whose aggregated traffic to its major destinations is relatively stable.

However, in a more general case, the preference of an AS could include both egress route profiles and ingress traffic patterns. We call the stability problem under this model the *stable route selection for general interdomain traffic engineering problem*. Note that this problem is different from the classical single-source adaptive routing problem [1]. Route selection for general interdomain traffic engineering is likely to be important when we consider an intermediate transit ISP whose ingress traffic varies substantially with its own route selection. The objective of this section, therefore, is to investigate the stability of route selection for general interdomain traffic engineering.



(a) a network.

(b) revenue function of B in (a).

(c) an unstable network.

Figure 6: Ingress-dependent traffic engineering.

We demonstrate the challenge involved in route selection for general traffic engineering with a simple example shown in Figure 6(a). The example is constructed in such a way that the export policies and the route ranking tables of the ASes follow standard AS business assumptions: ASes follow the typical export policies, and an AS prefers customer routes over provider routes. Note that the example avoids peer links to have a clean setup. The special feature of this example is that the ranking of AS B, who is one of the two competing providers of S, now depends on *outcomes*, instead of route profiles. An outcome consists of both route selection and ingress traffic pattern for an AS. Specifically, $\{S\}BFD$ denotes the outcome that B uses the route BFD and S sends traffic for destination D through B; $\{\}BD$ denotes the outcome that B uses the route BD and S does not send any traffic through B. This example can well happen in practice. The ranking table of S is constructed according to the standard BGP decision process: S prefers routes with small hop counts; and for routes with the same hop count, it uses the next-hop ID to break the tie. As for B, note that B prefers traffic from a customer than no traffic. Thus it is a typical ISP behavior. Also note that when S sends traffic through B, the route BFD is preferred than the route BD; otherwise, the route BD is preferred. A potential revenue function that may cause this scenario to happen is shown in Figure 6(b); that is, BFD is more profitable for B when the traffic volume is high, while BD is more profitable for B when the traffic volume is low.

Given this example, it is clear that the traditional BGP route selection model, shown in Figure 2, is not enough to capture the potential behaviors of an AS when it conducts route selection for general interdomain traffic engineering. Specifically, in the traditional BGP route selection model, B will choose one of the two available routes by determining the ranking of these two routes. However, in this example, B needs to know the *outcome* of choosing BD or BFD in order to make a decision. Although B could adopt a simple, state-free strategy by keeping just the available route cache, as in the current BGP model, this could lead to trivial instability. For example, B could follow a simple greedy protocol: assume the current ingress traffic pattern, pick the available route such that the current ingress pattern and the chosen route have the highest rank. Using this protocol, assume initially S does not use B. Then B first picks BD. Since B chooses BD, S chooses the route SBD, and the traffic from S arrives at B. Since B likes to use BFD when it has high traffic volume, it switches to BFD. Then S chooses SCD, and S no longer uses B. Thus B switches back to BD and we have a loop. This loop is an instance of what we call trivial instability.

The above trivial instability is due to the fact that B does not keep state to learn the outcome of choosing BD or BFD. It appears that ISPs are already realizing the potential instability problem of not keeping state, and thus are applying learning techniques to optimize route selection for interdomain traffic engineering. When an AS keeps states about the outcomes of choosing different routes and actively seeks to optimize its traffic engineering objectives, we say that the AS is adopting an *active route selection algorithm*.

A major challenge of investigating route selection for general interdomain traffic engineering is that, unlike route selection for egress traffic engineering which has a common route selection model shown in Figure 2, there is no previous model on ISP route selection behaviors for general interdomain traffic engineering. In our email survey, ISPs admit that although they are actively conducting route selection considering ingress traffic patterns, their methods so far are based on trials-and-errors. As a result, our first major challenge in this section is to identify a general model to capture reasonable route selection behaviors. It is important to emphasize that it is not the objective of this article to claim that any specific route selection algorithms are justified by being truly rational or provably optimal. *The objective of this article is to analyze the stability and efficiency of the class of route selection algorithms that the ISPs and researchers either are currently or could potentially be using on the Internet for route selection for general interdomain traffic engineering.*

5.2 Adaptive Route Selection Algorithms

Overall, with increasing usage of BGP route selection for interdomain traffic engineering, more active route selection algorithms are bound to be designed and deployed in the Internet. Thus it is important to analyze the stability of a network running heterogeneous and reasonable active route selection algorithms. Among all possible active route selection algorithms, we identify a general class of algorithms which we call *adaptive route selection algorithms*. Our model is inspired by previous work on adaptive learning [11] and learning on the Internet [5].

Intuitively, a reasonable route selection algorithm should not choose route profiles that are shown to be inferior to other available route profiles. Consider the example in Figure 6(a). At the beginning, B does not know which route (profile) is better, BD or BFD. So it can experiment with each one several times. Later, it may learn that BD always yields a better outcome than BFD does. Thereafter, it is reasonable that B will always choose BD over BFD. The route profile BFD is an example of *overwhelmed* route profiles. Thus, an adaptive route selection algorithm is one where recursively, overwhelmed route profiles are no longer chosen.

Theorem 3 *The BGP protocol defined in Section 2 is an instance of adaptive route selection, if the following conditions are satisfied:*

1. *BGP update messages between neighboring ASes are delivered reliably in FIFO order, and have bounded delay;*

2. *Each AS sends out BGP update messages in bounded time after it updates its route profile;*

3. *Each BGP update message is processed immediately.*

5.3 Instability of Adaptive Route Selection

After introducing the notion of adaptive route selection algorithms, we study whether a network consisting of ASes running adaptive route selection algorithms for general interdomain traffic engineering (*i.e.*, ranking depends on both egress route and ingress traffic patterns) has stable route selections.

Definition 1 *A network consisting of ASes each of which is running an adaptive route selection algorithm has a stable route selection if the route selection of each AS is a single route, as time goes to infinite.*

In the above definition, we require that, in a stable route selection, the route selection of each AS be a "pure" routing decision. We do not allow mixed strategies (*i.e.*, a random combination of routes), since mixed strategies involve frequent route fluctuations, and are thus not desirable as "stable" solutions for interdomain routing.

Under the general adaptive route selection scheme, there are well-behaved network setups with no stable route selection. Figure 6(c) shows an example where no adaptive route selection algorithm can converge to a stable route selection. It is motivated by the wide spread usage of multihoming. The setup is constructed to satisfy all standard ISP business relationship constraints so that under previous route selection models [7] there is a unique stable route selection. We observe the following instability when ASes use adaptive algorithms. When AS A and B choose AD and BFD. The outcome is SAD since S ranks SAD higher than $SBFD$. A has an incentive to change from AD to AED since A ranks $\{S\}AED$ higher than $\{S\}AD$. However, AS B realizes that, it can achieve a better outcome by changing BFD to BD since S will choose SBD over $SAED$. This in turn triggers A to switch from AED back to AD. Thus we end up with A chooses AD and B chooses BFD again, and the process continues.

6 Conclusions and Future Work

In this article, we report the results of our study on the stability and efficiency of using route selection to achieve interdomain traffic engineering objectives. We identify that interdomain traffic engineering requires that route selection be coordinated among multiple destinations. We show the surprising result that the interaction of the routing of multiple destinations can cause routing instability even when the routing of each destination individually does have a unique solution. Taking into account business relationships among ISPs in the current Internet, we analyze a set of practical interdomain traffic engineering guidelines and show that if every AS follows them, the existence and uniqueness of stable route solutions in interdomain egress traffic engineering are guaranteed. Using realistic Internet AS topology, we show that if the guidelines are violated, even when a small number of ASes coordinate their routes for just two destinations, instability could happen.

Despite the success of the analysis and the guidelines, we also show that route selection for interdomain traffic engineering is an extremely important but challenging subject. In a more general model where the preference of an AS depends on both egress routes and ingress traffic patterns, we derive an important negative result: there are networks which will be unstable under any adaptive route selection algorithms where inferior routes are iteratively eliminated. There are many avenues for future work. In particular, although we propose a set of practical guidelines to guarantee convergence, ISPs may still have no incentives to follow these guidelines. How to design incentive-compatible interdomain routing protocols which can guarantee convergence in the most generic setting is a major remaining challenge. The negative result is particularly troubling in that it suggests a fundamental trade-off between local optimality of each AS and global stability. Thus to have a stable, incentive-compatible route selection protocol, the ASes must be willing to look into the future and sacrifice short-term benefits under the current BGP model.

References

[1] D. Bertsekas and R. Gallager. *Data Networks*. Prentice-Hall, Second Edition, 1992.

[2] N. Feamster, H. Balakrishnan, and J. Rexford. Some foundational problems in interdomain routing. In *Proceedings of Third Workshop on Hot Topics in Networks (HotNets-III)*, San Diego, CA, Nov. 2004.

[3] J. Feigenbaum, D. Karger, V. Mirrokni, and R. Sami. Subjective-cost policy routing. Technical Report YALEU/DCS/TR-1302, Yale University, Sept. 2004.

[4] A. Feldmann, O. Maennel, Z. M. Mao, A. Berger, and B. Maggs. Locating Internet routing instabilities. In *Proceedings of ACM SIGCOMM '04*, Portland, OR, Aug. 2004.

[5] E. Friedman and S. Shenker. Learning and implementation on the Internet. Working paper. Available at `http://www.orie.cornell.edu/~friedman/pfiles/decent.ps`, 1997.

[6] L. Gao, T. G. Griffin, and J. Rexford. Inherently safe backup routing with BGP. In *Proceedings of IEEE INFOCOM '01*, Anchorage, AK, Apr. 2001.

[7] L. Gao and J. Rexford. Stable Internet routing without global coordination. *IEEE/ACM Transactions on Networking*, 9(6):681–692, Dec. 2001.

[8] T. G. Griffin, A. D. Jaggard, and V. Ramachandran. Design principles of policy languages for path vector protocols. In *Proceedings of ACM SIGCOMM '03*, Karlsruhe, Germany, Aug. 2003.

[9] T. G. Griffin, F. B. Shepherd, and G. Wilfong. The stable paths problem and interdomain routing. *IEEE/ACM Transactions on Networking*, 10(22):232–243, Apr. 2002.

[10] T. G. Griffin and G. Wilfong. On the correctness of IBGP configuration. In *Proceedings of ACM SIGCOMM '02*, Pittsburgh, PA, Aug. 2002.

[11] P. Milgrom and J. Roberts. Adaptive and sophisticated learning in normal form games. *Games and Economic Behaviors*, 3:82–100, 1991.

[12] B. Quoitin, S. Uhlig, C. Pelsser, L. Swinnen, and O. Bonaventure. Interdomain traffic engineering with BGP. *IEEE Communications Magazine*, 41(5):122–128, May 2002.

[13] Y. Rekhter and T. Li. *A Border Gateway Protocol 4 (BGP-4), RFC 1771*, Mar. 1995.

[14] J. L. Sobrinho. Network routing with path vector protocols: Theory and applications. In *Proceedings of ACM SIGCOMM '03*, Karlsruhe, Germany, Aug. 2003.

[15] H. Wang, H. Xie, Y. R. Yang, L. E. Li, Y. Liu, and A. Silberschatz. On stable route selection for interdomain traffic engineering: Models, analysis, and guidelines. Technical Report YALEU/DCS/TR-1316, Yale University, Feb. 2005.