Negotiation-based Routing

Ratul Mahajan

University of Washington

This paper argues for an interdomain routing architecture based on dynamic negotiation between the source, intermediate, and destination ISPs.

1 Motivation

Interdomain route selection is a complex process driven by constraints arising from topology, policy (e.g., commercial relationships), traffic engineering (e.g., load balancing), and performance. BGP was designed to find (a single) policy conformant route through the network for each source, destination pair. There are two fundamental problems with the route selection process of BGP.

First, edge ISPs have too few options for selecting routes. Their choices are limited to the paths selected by their provider(s). As a result, there are many valid paths in the topology that cannot be used. Customers suffer when their providers select poor (from the customer's perspective) routes, for instance, when the performance metrics for the provider and customer are different. This shortcoming is evident in trends such as increased multi-homing and the use of "intelligent routing" solutions such as those provided by RouteScience.

The second problem, which is at the other extreme, is that the senders unilaterally select routes from those available to them. This ignores the traffic engineering needs of the destination and intermediate ISPs. We use the term traffic engineering to loosely refer to the process of controlling the paths through the network (driven by a high-level goal such as efficient use of the network). As a result of this route selection methodology, the destination ISP has no control over which of its upstreams gets used. Similarly, intermediate ISPs have no control over whether the upstream ISP would use the exported route, and if yes, for how much traffic. Over time hooks such as MEDs, communities, and extended communities have been added to the protocol to address this shortcoming. However, these hooks are both insufficient (e.g., ISPs have limited control over incoming paths) and have an unpredictable impact (e.g., MEDs can lead to persistent oscillations [4]).

The solution to the first problem is to give more routing choices to the edge ISPs. The solution to the second problem is getting the intermediate and destinations ISPs involved in the route selection process such that the selected route suites all the ISPs. This route negotiation should be explicit and transparent, and its outcome predictable. Otherwise, the result may be a situation much like the current world in which ISPs try to manipulate the outcome to their benefit by trying to second-guess the actions of others.

It is important to address both the problems together. Solving only the first exacerbates the second problem – it would be even harder for downstream ISPs to manage and provision their network, and may also have an impact on the overall stability of the Internet. Similarly, solving only the second exacerbates the first by further limiting the choices at the source ISP (downstream ISPs can reject certain routes).

2 Architecture

We argue that a better routing architecture can be designed by proposing the following strawman.

1. Distribute the topology information using a link state protocol (for instance). There are three types of edges in the topology – external peering edges (EBGP; between two ISPs), internal peering edges (IBGP), and edges between ISPs and prefixes.

2. Optionally, ISPs advertise the policy associated with their edges. Policy representation in an advertisement is compact and does not have to be precise (can be looser than the real policy).

3. Using the topology and policy information, the source ISP computes all valid paths to the destination. Edges for which no policy is advertised are assumed to be free to use for all kinds of traffic. From this pool, the source selects the route(s) it wants to use. Route selection is influenced by performance (measured by the ISP itself or an Internet weather service) and traffic engineering goals.

4. The source ISP sends each selected route towards the destination to get the approval from all the downstream ISPs in the route. The approval request is optionally accompanied

by an estimate of the shape of the traffic that would be sent along this route, and the duration for which route approval is sought. Each downstream ISP's decision depends on its policy and traffic engineering goals. Rejected requests can be accompanied by a hint as to which alternate paths are likely to be successful.

5. The source ISP starts sending traffic along the approved route(s). Forwarding can be achieved either by inserting the full route (or its hash [3]) in the packet or in a manner similar to label switching.

The distribution of topological information (and not just the paths that the provider happens to use, as in BGP) presents a wider selection to the source ISPs. Requiring route approvals before traffic can be sent implies that all the ISPs know beforehand what they are getting into, which gives them more control over their networks. Destinations can discourage traffic over their overloaded (or expensive) links while encouraging it over the lightly loaded link. Similarly, the transit ISP can discourage more traffic over routes that contain overloaded links.

2.1 Notes

- An easy extension of the above approach can be used to implement backup routing so that failover in case of failures is prompt. Approval requests for backup routes will be marked as being backup so that the downstream ISP knows that it would get traffic along this route only in case of failures.

- Policy advertisement helps to prune the source ISP's search for policy compliant paths. The advertised policy does not have to be precise as ISPs have the option to refuse a route later (in Step 4). ISPs that advertise their policy receive no (or fewer) requests that are not compliant with their policy. Advertised policies can also be used to prune the link database [5].

- Incidentally, the above architecture avoids the destructive interference of policies. Currently, it is hard to predict the end result of the combination of the policies of various ISPs, which in some cases leads to persistent oscillations [2]. In the architecture above, routing oscillations due to policy fluctuations can be easily detected.

2.2 Open Issues

- We have ignored cost concerns in the negotiation phase. Different options used by an edge ISP have different cost implications for its provider. A complete solution would take these concerns into account.

- It is important to ensure that negotiations converge to a solution acceptable to all concerned ISPs. Robust techniques based on contractual obligations, financial incentives or mechanism design [1] are needed for this purpose.

3 Summary

Two key problems with BGP's route selection process are too few routing choices at the edge ISPs, and the source ISP's ability to unilaterally select any route. The first has performance implications for the edge ISP, and the second makes traffic engineering harder for the destination and intermediate ISPs. We proposed an architecture that solves these problems by increasing routing choices at the edge ISP and selecting routes based on negotiation between the source, intermediate and destination ISPs.

References

[1] J. Feigenbaum, C. Papadimitriou, R. Sami, and S. Shenker. A BGP-based mechanism for lowest-cost routing. In *Symposium on Principles of Distributed Computing*, July 2002.

[2] T. Griffin and G. T. Wilfong. An analysis of BGP convergence properties. In *ACM SIGCOMM*, pages 277–288, August 1999.

[3] H. T. Kaur, S. Kalyanaraman, A. Weiss, S. Kanwar, and A. Gandhi. BANANAS: An evolutionary framework for explicit and multipath routing in the Internet. In *FDNA workshop at SIGCOMM*, Aug. 2003.

[4] D. McPherson, V. Gill, D. Walton, and A. Retana. BGP persistent route oscillation condition. Internet Draft draft-mcpherson-bgp-route-oscillation-01.txt, IETF, Mar. 2001.

[5] X. Yang. NIRA: New Internet routing architecture. In *FDNA workshop at SIGCOMM*, Aug. 2003.