Enke Chen
Tony Bates
MCI

June 1995

Current Practice of Implementing Symmetric Routing and Load Sharing in the Multi-Provider Internet

<draft-ietf-idr-symm-multi-prov-01.txt>

Status of this Memo

Abstract

In the current multi-provider Internet, it is common for an entity to have multiple service providers. Symmetric routing becomes increasingly important for various reasons. This memo documents and analyzes the current practice in implementing symmetric inter-domain routing using BGP for several representative topologies of Internet connections.

1. Introduction

In the multi-provider Internet, it is common for an entity to have multiple connections to the Internet. For example,

- o A Regional Service Provider (RSP) may be connected to multiple transit Internet Service Providers (ISPs).

- o A service subscriber may be connected to multiple RSPs or ISPs.

- o Subscribers of different providers may wish to backup each other.

These connections would provide for the capability of load sharing, path diversification and backup. The Internet is a mesh of ISPs, RSPs and service subscribers and is generally sparsely connected.

Symmetric routing is generally preferred as it facilitates problem resolution, and provides for better resource (especially network capacity) planning and utilization. Routing symmetry is also desirable in achieving optimal traffic flow in terms of reliability, delay character, cost and other QoS metrics. Several applications such as NTP, RSVP and MBONE rely upon routing symmetry. In the multi-provider Internet, routing asymmetry, especially at the inter-domain level, may have serious economic and legal ramifications.

This paper presents several representative topologies of Internet connection and their inter-domain routing requirements. It then documents and analyzes the current practice in implementing symmetric inter-domain routing in these cases using BGP.

This paper assumes that in general an ISP treats other ISPs equally (in terms of the "local_pref" parameter) in the route selection process. It also assumes that the following order of preference is followed for the purpose of route selection: first the "local_pref" parameter, followed by the shortest AS-path, the MED, and the IGP metric.

It is noted that the length of the AS-path has not been specified in the BGP document [1] as a route selection criteria. However, it has been included in more than one implementations, and has been widely used as such.

2. Internet Connection and Routing

The Internet is a mesh of transit Internet Service Providers (ISPs), Regional Service Providers (RSPs), and service subscribers. In general this mesh is rather sparsely connected with loose hierarchy. In the multi-provider Internet, a good routing plan for an entity (i.e., autonomous system) requires good understanding of its internal network topology, its connection to direct providers (and neighboring ASs), and its path to the major interconnection points (or network

access points, NAPs).

In this section, we present several typical topologies of Internet connections, and their inter-domain routing requirements.   Although these cases are not meant to be exhaustive, they are expected to cover the vast majority of Internet connection topologies.

2.1 An Entity with a Single Direct Provider

```
                +-----+         +-----+         +-----+
                | ISP |         | ISP |         | ISP |
                +-----+         +-----+         +-----+
                   |               |               |
                +-----+         +-----+         +-----+
                | AS1 |         | RSP |         | RSP |
                +-----+         +-----+         +-----+
                                                   |
                                                +-----+
                                                | AS1 |
                                                +-----+

                (a)               (b)             (c)

                              Figure 1
```

The routing is always symmetric at the inter-domain level. Routing policies can be achieved using the current version of BGP. AS1 can either take full routing or use default.

2.2 Backup of Entities with Different Direct Providers

Several topologies are shown in Figure 2. Both AS1 and AS2 have their direct provider(s), and they would like to backup each other. That is, if the link between AS1 and its direct provider is down, the link between AS1 and AS2 would be used to reach the Internet.

```
     +-----+                +------+                  +------+
     | ISP |                | ISP3 |                  | ISP3 |
     +-----+                +------+                  +------+
      /   \                 /     \                  /       \
     /     \               / (NAP) \                /  (NAP)  \
+-----+    +-----+   +------+      +------+    +------+       +------+
| AS1 |----| AS2 |   | ISP1 |------| ISP2 |    | ISP1 |-------| ISP2 |
+-----+    +-----+   +------+      +------+    +------+       +------+
                        |             |            |              |
                     +------+      +------+     +------+      +------+
                     | AS1  |------|  AS2 |     | RSP1 |      | RSP2 |
                     +------+      +------+     +------+      +------+
                                                   |              |
                                                +------+      +------+
                                                | AS1  |------| AS2  |
                                                +------+      +------+

      (a)                      (b)                          (c)
```

Figure 2

Note that in Figures 2(a)-2(b), AS1 and AS2 could be RSPs.

In all cases of Figure 2, in order to provide for backup, AS1 shall permit the acceptance of AS2's routes from both AS2 and AS1's direct provider, and permit their announcements to its direct providers. Similar configuration is required for AS2.

There are two common routing policies depending upon how the link between AS1 and AS2 is used.

Policy 1: Used solely as a backup link

The routing policy can be implemented by coordinating AS-based "local_pref" values between neighboring ASs.

In all cases of Figure 2, the "local_pref" value for the peer of AS1 or AS2 with its direct provider shall be higher than that for the peer between AS1 and AS2. Either full routing or partial

routing can be configured.

For example, in Figure 2(a), AS1 can take full routing from ISP and AS2. An alternative is for AS1 to take only AS2's routes from ISP and AS2, and configure default routes (with different weights) at its border routers and then propagate them into its own AS (via, e.g., iBGP). The ISP needs to make sure that the "local_pref" values are equal for the peers with AS1 and AS2 so that the shorter AS-path would be selected.

Policy 2: Used for traffic between AS1 and AS2, and as backup in general

In general this routing policy can be implemented by coordinating AS-based "local_pref" values among the neighboring ASs and direct providers.

In Figure 2(a), equal "local_pref" values could be configured for all the peers. Then the length of AS path would be used as tie-breaker in the route selection. AS1 can either take full routing from AS2 and its direct provider. It can also choose to take only AS2's routes from its direct provider and AS2, and configure default routes (with a different weights) at its border routers and then propagate them into its own AS (via, e.g., iBGP). Similar configuration for AS2.

In Figures 2(b)-(c), AS1 can either take full routing from AS2 and its direct provider, and configure the "local_pref" parameter so that traffic to AS2 prefers the AS1 - AS2 link over the link to its direct provider. AS1 can choose to take only AS2's routes from its direct provider and AS2, and configure default routes (with different weight) at its border routers and then propagate them into its own AS (via, e.g., iBGP). Similar configuration for AS2.

AS1's direct provider (and possible its ISP) needs to configure the "local_pref" parameter so that traffic to AS2 does not prefer the link to AS1. Similar configuration is required for AS2's direct provider.

2.3 An Entity with Multiple Direct Providers

As shown in Figure 3, AS1 has two direct providers. X and Y are routes of AS1. Note that AS1 could be an RSP.

```
          +------+                 +-----+                +------+
          | ISP3 |                 | ISP |                | ISP3 |
          +------+                 +-----+                +------+
          /      \                 /     \                 /    \
         /  (NAP) \               /       \               / (NAP)\
   +------+      +------+  +------+       +------+  +------+     +-----+
   | ISP1 |------| ISP2 |  | RSP1 |       | RSP2 |  | ISP1 |-----| ISP2|
   +------+      +------+  +------+       +------+  +------+     +-----+
         \        /               \       /            |            |
          \L1    /L2               \L1   /L2           |            |
          +------+                +------+          +------+     +-----+
          |  AS1 |                | AS1  |          | RSP1 |     | RSP2|
          |X    Y|                |X    Y|          +------+     +-----+
          +------+                +------+                \       /
                                                           \L1   /L2
                                                           +------+
                                                           | AS1  |
                                                           |X    Y|
                                                           +------+

           (a)                      (b)                        (c)
```

Figure 3

Depending upon the quality of these links and the internal network topology of the AS, there are several common routing policies.

Policy 1: One link is used as primary, the other as pure backup

This policy is common when the quality of these links differ dramatically.

This policy can be implemented by coordinating AS-based "local_pref" values between the entity and its direct providers. The AS can either take full routing or use default routes. In the case of default, each border router can configure a default route and then propagate it into the AS (via, e.g., iBGP).

Policy 2: Each link is used for traffic with the respective direct provider. In general one link is used as primary, and the other as

backup.

If the traffic between AS1 and its direct providers (and their customers) shall take the direct link, AS1 needs to be configured:

- either with partial routing (only routes of the direct providers and their customers) and defaults with different weights.

- or with full routing and configure AS-based "local_pref" values.

The difficulty is that the indirect providers (e.g., ISP3 in Figure 3(a)) may have to be involved to achieve symmetric routing. More specifically,

- In Figure 3(a), ISP3 would receive routes X and Y from both ISP1 and ISP2 with identical length of AS paths. In order for L1 to be favored by ISP3 to AS1, ISP1 would need to manipulate the AS-path length, which is discussed in Section 3.1. Another approach is for ISP3 to configure AS-based "local_pref" parameter, which certainly does not scale well as there are many ISPs at an NAP. In addition, it is almost impossible to do as AS1 is not a customer of ISP3.

- In Figure 3(b), ISP would receive routes X and Y from both RSP1 and RSP2 with identical length of AS paths. Either RSP1 needs to manipulate the AS-path length, or ISP needs to configure AS-based "local_pref" parameter.

- In Figure 3(c), ISP1 would receive routes X and Y from both RSP1 and ISP2. In order for traffic from ISP1 to AS1 to favor the ISP1 - ISP2 link, either RSP1 needs to manipulate the AS-path length, or ISP1 needs to configure AS-based "local_pref" parameter.

Another problem is that it is difficult for AS1 to implement "full routing", as AS1 needs to update the AS list for the "local_pref" parameter each time its direct providers acquire a new AS as a customer. Nevertheless, some entities still prefer full routing.

Policy 3: Partial load-sharing among these links

That is, the direct link is used for traffic between AS1 and its direct providers including its customers. However, the closest

exit point would be taken for traffic beyond these direct providers and their customers. For example, in Figure 3(a) traffic between AS1 and ISP1 (and its customers) would use the direct link between AS1 and ISP1; traffic between AS1 and ISP2 (and its customers) would use the direct link between AS1 and ISP2. For traffic destined to ISP3, either ISP1 or ISP2 would be used depending on where the traffic is originated in AS1.

AS1 can take full routing. It can also take partial routing (routes of direct providers and their customers), and configure equal-weight default routes at its border routers and propagate them into its AS.

The problem is how to make sure the return traffic from a 4th party (e.g., ISP3 in Figure 3(a)) is symmetric.

Policy 4: Complete load-sharing among these links

That is, each network in AS1 sends packets to the closer (in terms of internal route preference) border router that peers with a direct providers. The return traffic is expected to take a symmetric path. For example, in Figure 3(a) a packet, which is originated from network X and is destined outside AS1, would be forwarded to ISP1, even when the destination is in ISP2.

The simpler approach is for AS1 to use default. That is, AS1 would first configure default route at each connection to a provider and propagate (e.g., via iBGP) them into the AS. Then, each network in AS1 choose the closest exit point (determined by IGP metric). The problem is how to make sure the return traffic to X and Y takes symmetric paths. Currently this is achieved by manipulating the AS-path length or other approaches detailed in the following section.

If AS1 still prefers to take full routing, more coordination would be required for using the AS path manipulation or other techniques as described in Section 3.

3. Current Practices

Currently there are mainly three approaches to implement Polices 2-4 for Figure 3. This section presents analysis and critique of these approaches. Without loss of generality, Figure 3(a) is used as an example.

3.1 Manipulation of AS Path Length

Although the length of the AS path was not specified in [1] as a parameter in the route selection process, it has been widely used as such.

Some router software offers the ability of prepending AS numbers to the AS path for the purpose of influencing the route selection. Here is how the feature can be used. First, AS1 categorizes all of its routes and prepends an AS number (either its own AS or a different AS number):

	AS1 Prepend		AS Path	
Route	To ISP1	To ISP2	ISP1	ISP2
X		AS1	AS1	AS1 AS1
Y	AS1		AS1 AS1	AS1

In general the different AS paths can be used by ISP1 and ISP2 to configure AS-based "local_pref" values to implement the desired routing policy. The "local_pref" configuration would not be necessary if there are sufficient number of ASs inserted.

With this approach the AS that originates the preference has full control, and only that AS needs to manipulate the AS path on a per-route basis.

The drawbacks of this approach includes:

- It extends the AS path with superfluous information. In particular, the superfluous information in the AS path would be propagated upstream and to the whole Internet.

- The number of ASs that need to be preprended is in general proportional to the number of direct providers.

- Compatibility with other BGP implementation may be a problem.

3.2 Splitting AS

This approach requires an AS to be split into multiple ASs and run external BGP between these ASs (possibly with MEDs configured for load balancing among these ASs). Then the cases in Figure 3 can be reduced to the cases in Figure 2, which have been discussed in Section 2.

This is probably the cleanest approach the current BGP version can offer. However, there is a great deal of reluctance in using this approach. Practical problems with this approach include:

- It does not work with the partial load-sharing case where the connections to multiple providers originate from one router.
- Splitting ASs and having them maintained could be quite involved depending upon the internal network topology.
- Extra AS numbers are required [7]. It would be necessary to apply for AS numbers at the InterNIC.
- The number of split ASs is proportional to the number of direct providers.
- Wasting of AS numbers. The exhaustion of the AS number space could become real with the ever-increasing number of such needs.
- The increased number of ASs would add complexity to the Internet topology, and therefore complicate problem resolution.
- An AS number has been traditionally tied to an organization. Splitting AS means loss of coherence for some customers.

3.3 NLRI-based Preference Specification

This is the approach that has been used for the NSFNET. Here is how it is done with Figure 3(a). ISP1 and ISP2 configure net-based preference on their routers, according to preference provided by AS1. For example, for route X, ISP1 would configure higher preference for its direct link with AS1, and lower preference for its direct link with ISP2.

This approach requires NRLI-based customization with the direct providers and sometimes indirect providers as well as the originating AS. The NSFNET configuration experience has shown that this approach requires non-trivial administrative coordination and full topology information. In addition, it places a burden on routers with limited memory capacity. For these reasons, the NLRI-based preference configuration should be avoided at the provider level if possible. Instead, such a configuration should be pushed as close to the originating AS as possible.

3.4 Perfect Aggregation and Addressing

In the cases that all routes in the AS are covered by an aggregate and address assignment is completely consistent with the network topology, the rule of the longest prefix match can be used to help achieve routing symmetry and load sharing. That is, a portion of the aggregate, along with the aggregate itself, can be announced to one direct provider. The remaining portion of the aggregate, along with the aggregate itself, can be announced to the other direct provider.

It does not work with the partial load-sharing case where the connections to multiple providers originate from one router. More importantly, the requirement of this approach is not likely to be met in practice. So, this approach is listed just for the sake of completeness.

4. Discussion

As has been illustrated in Section 3, it is not easy to implement routing symmetry and load sharing for an entity with multiple direct providers using the current functionality of BGP. There are many drawbacks with the current practice of implementation. Even the implementation of the AS-based "local_pref" parameter can sometimes be quite involved. The difficulty is caused by the lack of a globally transitive preference an AS (with multiple direct providers) can specify, and be used in the route selection process.

A new BGP attribute termed "Destination Preference Attribute" (DPA) has been proposed in [3] to address such need. As illustrated in [4], the routing policies presented in Section 2 can be implemented with ease by using the DPA attribute. In particular, only the AS that originates this preference needs to specify this preference on a per-route basis.

5. Security Considerations

Security considerations are not discussed in this memo.

6. Acknowledgments

The authors would like to thank Roy Alcala, Dennis Ferguson, John Stewart, and Jack Waters of MCI for the many interesting hallway discussions related to this work. We also acknowledge helpful comments and suggestions by Yakov Rekhter of Cisco.

7. References

[1] Rekhter, Y., and Li, T., "A Border Gateway Protocol 4 (BGP-4)", RFC1771, March 1995.

[2] Y. Rekhter, and P. Gross, "Application of the Border Gateway Protocol in the Internet", RFC1772, March 1995.

[3] Chen, E., and Bates, T., "Destination Preference Attribute for BGP", INTERNET-DRAFT, <draft-ietf-idr-bgp-dpa-01.txt>, June 1995.

[4] Chen, E., and Bates, T., "Application of the BGP Destination Preference Attribute in Implementing Symmetric Routing", INTERNET-DRAFT, <draft-ietf-idr-dpa-application-01.txt>, June 1995.

[5] Antonov, V., "BGP AS Path Metrics", INTERNET DRAFT, <draft-ietf-idr-bgp-metrics-00.txt>, March 1995.

[6] Rekhter, Y., "Routing in a Multi-provider Internet", RFC1787, April 1995.

[7] Hawkinsin, J., and Bates, T., "Guidelines for creation, selection, and registration of an Autonomous System (AS)", INTERNET-DRAFT, <draft-ietf-idr-autosys-guide-03.txt>, May 1995.

8. Author's Addresses

Enke Chen
MCI
2100 Reston Parkway
Reston, VA 22091

phone: +1 703 715 7087
email: enke@mci.net

Tony Bates
MCI
2100 Reston Parkway
Reston, VA 22091

phone: +1 703 715 7521
email: Tony.Bates@mci.net