Metrics to Evaluate the Cost of Maintaining Diverse BGP Routes

Cristel Pelsser Tomonori Takeda*
NTT Network Service Systems Laboratories, NTT Corporation

1 Introduction

In a Service Provider (SP) network, routes for external destinations are distributed on iBGP sessions. A common practice is to make use of Route Reflectors (RR). Such a practice is more scalable in the number of iBGP sessions to be configured in a SP network than a full-mesh of iBGP sessions. However, it has been shown that RRs have a negative impact on the diversity of routes available in the network. This is an important issue as routers may not be able to quickly use an alternate route in case of a route failure. In a previous work, we proposed an algorithm to design iBGP session topologies with improved route diversity. In addition, we have shown that this is achieved with a low number of iBGP sessions, compared to a full-mesh.

In this paper, we first propose additional metrics to evaluate the cost of a sparse iBGP topology ensuring NH diversity. The first new metric is a measure of the size of the BGP routing tables maintained in the SP network. The second and third metrics indicate the IGP and the peering costs of the solution, respectively. Then, we estimate the cost of our NH-diverse iBGP design algorithm by means of these metrics.

2 Proposed metrics

While BGP route diversity enables to achieve fast recovery in the case of a failure, it has a cost. First, additional routes have to be supported in the routers. Second, upon a failure, alternate routes which do not have a minimal administrative cost are used instead of the failed and preferred routes. Lastly, Sending traffic along the alternate routes may increase the peering expenses of the SP. These routes may cross SPs that are more costly than the SPs crossed by the failed routes.

Different diversity solutions may lead to different costs. While diversity is ensured when routers know two routes for each prefix, some solutions may require routers to maintain a large number of additional routes. Moreover, a solution may provide alternate routes with lower IGP cost or lower peering cost than other solutions. The metrics that we propose in this section take these costs into account in the comparison of BGP route diverse solutions.

The first metric that we propose measures the total number of routes maintained in the routers. The other two metrics reflect the cost of using the alternate routes upon the occurence of a failure.

1. **Router memory cost** This metric consists of the minimum, average and maximum number of BGP routes found in the routing tables of the routers. We note this triplet as follows $M = (M_{min}, M_{avg}, M_{max})$.
2. **Administrative cost** This metric consists of the sum of the IGP costs of the best alternate routes available in the routers. Let $r \in R$, where R is the set of BGP routers in the SP network. Let P be the set of external prefixes advertised by BGP to the routers of the SP network. The administrative cost A is:

$$A = \sum_{p \in P} \sum_{r \in R} c_{rp} \quad , \tag{1}$$

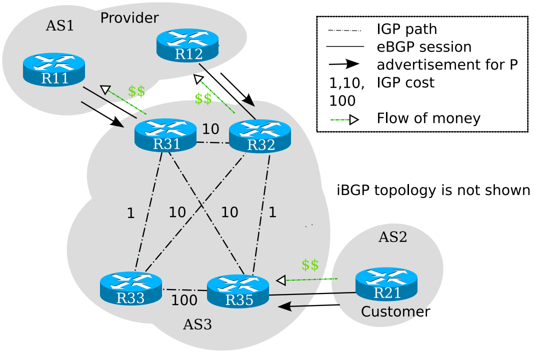

Figure 1: Example network.

where c_{rp} is the IGP cost of the path from r to the Next-Hop (NH) of the best alternate BGP route toward prefix p. If the traffic matrix is available, A can be weighted by the amount of traffic to be supported on the alternate paths. We define this weighted IGP cost metric as

$$A_W = \sum_{p \in P} \sum_{r \in R} c_{rp} w_{rp} \quad , \tag{2}$$

where w_{rp} is the amount of traffic arriving at r and destined to p.

3. **Peering cost** This metric consists of the sum of the costs of the peering links that are along the shortest alternate paths available in the routers. The peering cost E is

$$E = \sum_{p \in P} \sum_{r \in R} e_{rp} \quad , \tag{3}$$

where e_{rp} is the peering cost when r sends traffic to its best alternate BGP route toward prefix p. Similarly to the administrative cost, E can be weighted by the amount of traffic to be supported on the alternate paths, if the traffic matrix is available. We define the weighted peering cost, E_W, as

$$E_W = \sum_{p \in P} \sum_{r \in R} e_{rp} w_{rp} \quad , \tag{4}$$

We illustrate the the computation of these metrics by means of the example network presented in figure 1. In this example, AS3 is the considered AS. AS1 is a provider of AS3. On the contrary, AS2 is a customer of AS3. That is, the traffic exchanged between AS1 and AS3 costs money to AS3 while the traffic exchanged between AS2 and AS3 brings revenue to AS3.

Let us assume that R11, R12 and R21 advertise the same prefix, p, to AS3. Moreover, a single unit of traffic is sent from each router of AS3 to p. Figure 2 and figure 3 show the routes that are available in the routers of AS3 with two different NH diverse solutions. For each router, we show the NH of the BGP routes, the type of the route (best route or alternate route), the IGP cost of sending traffic to this NH and the peering cost of the inter-domain link used to reach the NH. The values of c_{rp} and e_{rp} are in bold, in the 4th and 5th column respectively.

---

*The authors thank Kohei Shiomoto and Ichiro Inoue for their comments.

Solution 1

Router	NH	Type	IGP cost	peering cost
R33	R11	alt	**1**	**1**
	R21	best	11	-1
	R12	alt	10	1
R35	R12	alt	**1**	**1**
	R21	best	0	-1
	R11	alt	10	1
R31	R11	alt	**0**	**1**
	R21	best	10	-1
	R12	alt	10	1
R32	R12	alt	**0**	**1**
	R21	best	1	-1
	R11	alt	10	1

A=1+1+0+0=2
E=1+1+1+1=4
M=(3,3,3)

Figure 2: Solution1.

Solution 2

Router	NH	Type	IGP cost	peering cost
R33	R11	alt	**1**	**1**
	R21	best	11	-1
R35	R21	best	0	-1
	R11	alt	**10**	**1**
R31	R21	best	10	-1
	R11	alt	**0**	**1**
R32	R21	best	1	-1
	R12	alt	**0**	**1**
	R11	alt	10	1

A=1+10+0+0=11
E=1+1+1+1=4
M=(2,2.25,3)

Figure 3: Solution2.

We see in figure 2 that R33 knows three routes for p. The best route is via R21 in AS2. There are also two alternate routes, one via R11 and the other via R12. The route via R11 is the best atlternate route because it has the smallest IGP cost. If there is a failure of node R35, R21 or link R35-R21, the route via R11 will be used. Thus, for R33, $c_{R_{33}p} = 1$ and $e_{R_{33}p} = 1$. In this example, A is equal to 2 and E is equal to 4. Moreover, the minimum, average and maximum number of routes at the routers of AS3 are equal to 3, in figure 2.

Figure 3 illustrates the values of c_{rp} and e_{rp} for a different NH diverse solution and a different set of routes in the routers. Here, we see that A is higher than in figure 2. In addition, we note that the number of routes in the routers is slightly lower than in the previous example.

3 Illustration

In this section, we use our metrics to illustrate the cost of different iBGP topologies. Each of these iBGP topologies provides NH diversity in the routers. We show that while the number of iBGP sessions and the number of routes in the routers is low, in such iBGP topologies, the IGP and peering costs are slightly higher than in a full-mesh. We rely on a model of a real network for our evaluation.

3.1 Ensuring BGP route diversity

In [1], we proposed a solution for reaching NH diversity at the routers of an AS. This proposal ensures NH diversity for all routes that are learned at, at least, two peering points. We proposed an algorithm to be used in the design phase of iBGP topologies. As input, the algorithm takes the eBGP routes received at the AS Border Routers (ASBR), the IGP topology and an initial iBGP route reflection topology. We improve NH diversity for a router through the addition of iBGP sessions with ASBRs.

We have shown in [1] that, as a result, our algorithm determines a small number of iBGP sessions to add to an existing iBGP route reflection topology.

3.2 The cost of a BGP route diverse solution

For our evaluation, we relied on the model of a real network. We generated NH diverse iBGP topologies as follows. First, we used classical iBGP design principals to obtain an initial iBGP topology. Then, we applied the algorithm described in [1]. As a result, all the topologies used in this section are NH diverse.

The iBGP topologies named “Bates1+” and “Bates2+” are constructed based on the recommendations expressed in[BCC06]. The “No init” topology is solely generated by the algorithm proposed in [1]; the initial iBGP topology does not contain any iBGP sessions. All these topologies consist of 17 nodes. “Zhang+” is obtained from the recommendation in [ZB03]. This topology is composed of 51 nodes. All the nodes in the original network have been split into three nodes because the recommendations in [ZB03] apply to large networks.

In table 1, we show the administrative cost (A), peering cost (E), minimum (M_{min}), average (M_{avg}) and maximum number (M_{max}) of routes, for each iBGP topology with diverse BGP routes in the routers. In the last column, we show the number of iBGP sessions, S, in each iBGP topology.

Table 1: Cost of different NH diverse solutions.

	Cost					
Name	A	E	M_{min}	M_{avg}	M_{max}	S
Full-mesh17	364	-4	79	79	89	136
Bates1+	382	-4	33	38	55	88
Bates2+	405	-1	33	39	55	57
No init+	387	-1	28	33	55	69
Full-mesh51	11794	-57	79	79	89	1275
Zhang+	13094	-54	31	40	77	189

Negative values for E indicate that the Service Provider earns money by sending traffic to its peers. In table 1, we see that the SP earns more money along backup routes with a full-mesh of iBGP sessions than with a sparse iBGP topology. This results from the fact that more BGP routes are visible with a full-mesh than with a sparse topology. The best alternate routes may not be known by the routers in a sparse iBGP topology.

We also observe in table 1 that a full-mesh of iBGP sessions leads to smaller administrative costs than sparse iBGP topologies. There is between 4 and 11% of additional administrative costs with sparse iBGP topologies.

A full-mesh requires to store more routes and maintain more iBGP sessions in the routers, compared to sparse iBGP topologies. The less iBGP sessions there are, the more costly the solution is in terms of IGP and peering costs. There is less chances that the best alternate routes, with regard to IGP and peering costs, are available at the routers.

4 Discussion

In this paper, we proposed a set of metrics to evaluate the cost of maintaining alternate BGP routes in the routers. These metrics take into account the cost of sending traffic along the alternate routes in addition to the cost of storing these routes in the routers.

We have seen that, with the NH diverse iBGP topologies generated by the algorithm proposed in [1], there is a trade-off between the number of sessions and routes, on one side, and the administrative and peering costs, on the other side. The algorithm proposed in [1] solely aims at minimizing the number of iBGP sessions. It does not try to optimise the peering cost and the administrative costs.

If the administrative and peering costs are considered of paramount importance, one could think to continue adding iBGP sessions until an administative and peering cost threshold is reached, when designing route diverse iBGP topologies with the algorithm in [1]. Another approach would be to consider these costs in the selection of the iBGP sessions that will be part of the iBGP solution ensuring NH diversity in the routers.

References

[1] C. Pelsser, T. Takeda, E. Oki, and K. Shiomoto. Improving route diversity through the design of iBGP topologies. In *IEEE Conference on Communications 2008, ICC 2008*, pages 5732 – 5738, 19-23 May 2008.

[BCC06] T. Bates, E. Chen, and R. Chandra. BGP route reflection - an alternative to full mesh internal BGP (IBGP). RFC 4456, April 2006.

[ZB03] R. Zhang and M. Bartell. *BGP Design and Implementation.* Cisco Press, first edition, December 2003.