BGP Next-hop Diversity: A Comparative Study

Jong Han Park, Pei-chun Cheng, Shane Amante, Dorian Kim, Danny McPherson, Lixia Zhang

University of California, Los Angeles
Computer Science Department
Technical Report # 100031
Oct. 11, 2010

BGP Next-hop Diversity: A Comparative Study

Jong Han Park[1], Pei-chun Cheng[1], Shane Amante[2], Dorian Kim[3],
Danny McPherson[4], and Lixia Zhang[1]

[1] University of California, Los Angeles
[2] Level-3 Communications Inc.
[3] NTT Communications Inc.
[4] Verisign Inc.

Abstract. Although a BGP router selects and propagates only one best path from multiple available paths for each destination network, an AS with multiple BGP border routers may still use multiple paths to reach the same destination. However there is a lack of understanding regarding path diversity as viewed at the AS level, or what factors impact such diversity. In this paper we measure BGP path diversity at AS level by using the iBGP data collected from two Tier-1 ISPs, each with a different iBGP architecture. Our results show that destination prefixes with highest degrees of path diversity are due to the topological location and connectivity of their origin ASes, and that the first two criteria in BGP best path selection, *i.e.* LOCAL_PREF and AS_PATH length comparison, are dominant contributors in eliminating alternative paths while specifics of iBGP topology have little impact on an AS's path diversity.

1 Introduction

BGP is the de-facto routing protocol used to exchange reachability information in the Internet. By design a BGP router selects and propagates only one best path for each destination network from multiple available paths. However an AS with multiple BGP border routers may still use multiple paths to reach the same destination. As the topological connectivity of the Internet grows denser over time [6], it becomes increasingly desirable to fully utilize multiple available paths to the same destination in order to improve both data delivery performance and network robustness. Several recent activities in IETF explored solutions that enable each BGP router to propagate multiple paths for the same destination [7, 4, 11, 3]. However despite the fact that a few measurement and analysis studies on the *existing* BGP path diversity have appeared recently [10, 2, 5], there is a lack of general understanding regarding path diversity from the view of an AS, or what factors impact such diversity.

To provide a comprehensive understanding on BGP path diversity in today's operational networks, in this paper we perform a BGP path diversity measurement study using iBGP routing data collected from two global ISPs, referred to as ISP_{FM} and ISP_{RR} based on their internal *full-mesh* iBGP and *route reflection* iBGP topology, respectively. For each of the destination prefixes in the global routing table, we measure its path diversity by the number of distinct next-hop POPs and next-hop ASes. Our main results and contributions can be summarized as follows.

- The number of prefixes with a high degree of path diversity in both ISPs is small. A considerable number of prefixes are reached through only a single next-hop POP (10.17% and 34.02% of all prefixes in ISP_{FM} and ISP_{RR}, respectively) and a single next-hop neighbor AS (about 63.08% and 84.42% respectively).
- The prefixes with highest degree of BGP path diversity are due to the topological location of their origin ASes. More specifically, we observed that the prefixes with the highest degree of diversity are originated from ASes that are more than one AS hop away from ISP_{FM} or ISP_{RR} , and that these origin ASes have multiple equally preferred paths from ISP_{FM} and ISP_{RR}.
- The dominant contributors in reducing path diversity in both ISP_{FM} and ISP_{RR} are the first two criteria in BGP best path selection, LOCAL_PREF and AS_PATH length comparison. They together hide up to 37% of all alternative paths. On the other hand, different iBGP topology affected the overall path diversity by less than 3.3%.

2 Background

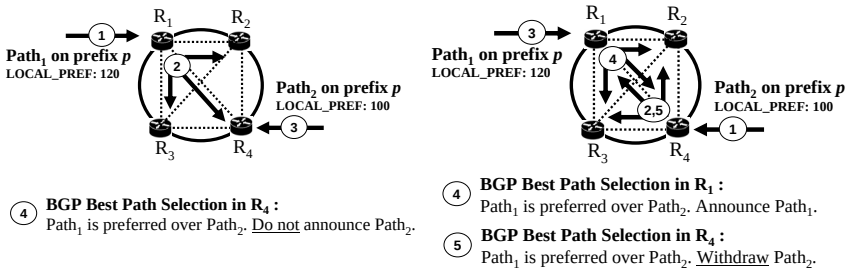
In this section, we provide a brief overview on BGP operations that are particularly relevant to our study, including a description of BGP next-hop diversity.

2.1 Routing in the Internet

The Internet is made of tens of thousands of different networks called Autonomous Systems (ASes) and BGP is used as the de-facto routing protocol. Within each AS, routers use an internal mode of BGP called iBGP (internal BGP) to distribute external routing information within the network. To avoid routing loops, iBGP requires that all iBGP routers within the same AS be connected in full-mesh, and that reachability information learned from one iBGP router must not be forwarded to any other iBGP router. This full-mesh connection requirement results in the number of iBGP sessions growing with the square of the number of iBGP routers. To mitigate this scalability problem, two alternative architectures have been proposed: AS confederations [9] and route reflection [1].

Regardless of the iBGP architecture, all BGP routers select only one best path for each destination prefix and further propagate the selected path to the neighbor routers. The best path selection considers the following criteria in the order listed: 1) highest LOCAL_PREF, 2) shortest AS_PATH length, 3) lowest ORIGIN, 4) lowest MED, 5) prefer path learned from eBGP session over path learned from iBGP session, 6) lowest IGP cost, and 7) lowest Router ID [8]. The first 4 criteria examine BGP attributes whose values are independent from an AS's internal topology, *i.e.* the preference of a path based on these 4 criteria would be the same regardless of the topological location of the router in the AS. The last 3 criteria examine values that are topology-dependent and can result in different preference by different routers depending on their topology and connectivity inside the AS.

Depending on the iBGP architecture and the internal router topology, the number of different paths learned by a router to reach a destination can differ. In either AS confederation or route reflection, only the best paths are further propagated from one



(a) Preferred path is announced first (b) Preferred path is announced later

Fig. 1. Hidden Path Phenomenon in iBGP

side of sub-AS (or route reflector) boundary to the other side. This reduction in the number of propagated paths with an AS leads to a reduction in the number of available paths in the AS.

2.2 iBGP Hidden Path Phenomenon

Although iBGP routers connected in a full-mesh can learn all the paths available to the given AS, they only forward the best path for each destination. This leads to *iBGP hidden path phenomenon*, in which a border iBGP router does not announce the learned, but less preferred paths for a given destination. Consequently these less preferred paths are known only to the border router itself; other iBGP routers are prevented from obtaining the complete view of all available paths to reach a given destination.

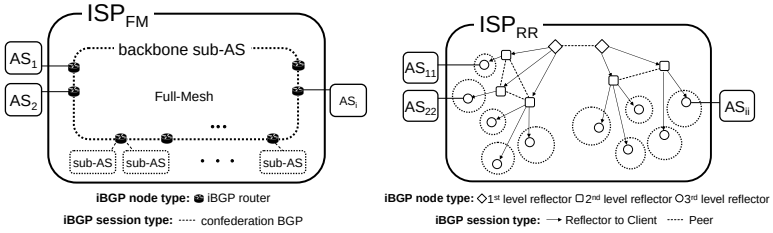
Figure 1 shows an example of a less preferred path (due to lower LOCAL_PREF attribute value in this case) being hidden in a full-mesh iBGP configuration. Note that $Path_2$ is less preferred than $Path_1$ in all iBGP routers inside the AS because the value of LOCAL_PREF associated with the path does not change as the BGP message propagates inside the AS. Thus the less preferred path ($Path_2$) will be hidden in the border router (R_4) only.

2.3 Path Diversities in iBGP

An AS connected to multiple neighbor ASes may learn the reachability to a given prefix from multiple neighbors (*i.e.* next-hop ASes). Large ASes typically interconnect with each other through multiple routers that are located at different cities, which are referred to as Point of Presence, or POPs. We measure BGP path diversity of each prefix by the number of next AS hops and the number of next POPs, which we refer as next-hop AS diversity and next-hop POP diversity in this paper.

3 Methodology

We used iBGP data collected from 2 different Tier-1 ISPs. In this section, we describe the high level network topology of the 2 ISPs, followed by data collection settings and how we measure next-hop diversity.



(a) Simplified Topology of ISP_{FM} (b) Simplified Topology of ISP_{RR}

Fig. 2. High Level iBGP Topology of Two ISPs

3.1 A Brief Description on ISP_{FM}

ISP_{FM} is a Tier-1 ISP which uses one AS number globally in the Internet. It has several hundreds of iBGP routers distributed across 14 countries in 3 different continents, and uses AS confederations [9] to scale with its network size. Figure 2(a) depicts a simplified topology of ISP_{FM} at a high level, where *backbone sub-AS* represents the backbone network of this ISP, consisting of more than one hundred iBGP routers connected in a full-mesh.

In most cases, ISP_{FM} uses one of the routers in backbone sub-AS to set up eBGP sessions directly with neighbor ASes (hence referred to as ISP_{FM}).[5] A collector (an iBGP router) is deployed in backbone sub-AS, and the collector maintains iBGP peering sessions with all other routers in backbone sub-AS to passively record all iBGP updates received.

3.2 A Brief Description on ISP_{RR}

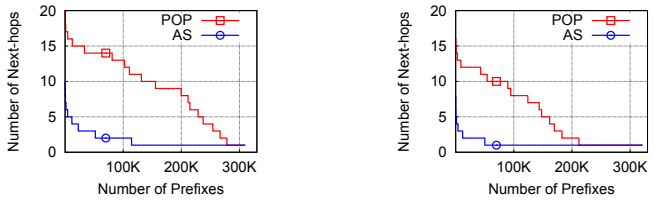
ISP_{RR} is another Tier-1 ISP which also uses one AS number globally in the Internet. It has several hundreds of iBGP routers distributed across 22 countries in 2 different continents and built a hierarchical route reflection architecture by recursively applying route reflection. Figure 2(b) depicts a simplified hierarchical route reflection system built by ISP_{RR}. The diamond-shape RRs at the top level represent continent level RRs; the square-shape RRs are at the 2^{nd} level of hierarchy, each represents a regional RR, and the 3^{rd} level circle-shape RRs represent POPs. A collector (an iBGP router) is configured as RR client to all route reflectors in the 2nd level route reflectors and passively record all iBGP updates received.

ISP_{RR} uses the top 2 levels of route reflectors for the sole purpose of distributing routing information to the rest of the network, and we refer this route reflector infrastructure in the upper 2 (1^{st} and 2^{nd}) levels of their route reflection hierarchy as backbone routers in ISP_{RR}.

3.3 Quantifying Next-hop Diversity

From ISP_{FM} and ISP_{RR}, we gathered routing table snapshots (RIBs) from all backbone iBGP routers. We first exclude 2 types of prefixes from this measurement study:

[5] In some large POPs, however, the routers are organized into several different sub-ASes. In this case, prefixes are announced from the neighbor ASes to the backbone via at least one sub-AS.



(a) ISP_{FM} (b) ISP_{RR}

Fig. 3. Next-hop POP & AS Diversity of Two Tier-1 ISPs

internal prefixes and potential bogon prefixes with their length less than 8 or greater than 24. Then, from each RIB entry, we extracted NEXT_HOP and AS_PATH attributes to measure how many distinct next-hop POPs and ASes are visible collectively in the view of the backbone routers for a given destination.[6]

In this paper, we present our measurement results based on the routing table snapshots taken on June 3rd, 2010 for clarity. To ensure that the snapshots are representative, we performed the same measurements on next-hop diversity using routing table snapshots taken on each day during the 1 week of June 3rd to 9th and on the 1st day of each month from January to May in 2010. The distributions of next-hop POP and AS diversity are very similar. In addition, we checked that the total number of prefix entries and the set of unique POPs and neighbor ASes are roughly the same.

4 BGP Next-hop Diversity

4.1 Next-hop Diversity in ISP_{FM}

We start by measuring next-hop diversity in ISP_{FM}. Figure 3(a) shows the distributions of next-hop POP (red line marked with a square) and AS diversity (blue line marked with a circle) of 309,903 prefixes.

We observe in Figure 3(a) that a considerable number of prefixes (63.08%) can be reached via only 1 neighbor AS. On the other hand, the number of prefixes with multiple next-hop ASes is quite small; only 3.78% of all prefixes can be reached via more than 4 next-hop ASes. The number of POPs to reach a given prefix is generally higher than the number of neighbor ASes, indicating that ISP_{FM} usually peers with its neighbor ASes in multiple sites. However, there are still about 10.17% prefixes that can only be reached via 1 POP.

Moreover, there exist two large groups of prefixes sharing the same degree of POP diversity. About 15% and 14% of prefixes have their POP diversity equal to 14 and 9 respectively. This is due to fact that these prefixes are received from a handful of

[6] ISP_{FM} does not use next-hop-self option. In contrast, ISP_{RR} uses next-hop-self option at the boundaries of its network. Due to such configuration difference, a direct comparison of next-hop diversity at the router level is not meaningful. Thus, we omit next-hop router from our study. When we say next-hop diversity in this paper, we mean next-hop POP and AS.

large neighbors, and thus share particular next-hop AS and POPs. For example, ISP_{FM} reaches 14% prefix via only one next-hop AS, and ISP_{FM} peers with such next-hop AS at 9 different POPs. As a result, all these prefixes would have the same next-hop POP and AS diversity, which is equal to 9 and 1 respectively.

4.2 Next-hop Diversity in ISP_{RR}

We now measure next-hop diversity in ISP_{RR} and compare the results with the next-hop diversity in ISP_{FM}. Figure 3(b) shows the distributions of next-hop POP and AS diversity of 321,432 prefixes.

Similar to what we observed in ISP_{FM}, a considerable number of prefixes can be reached via only one neighbor POP and AS; 34.02% and 84.42% of all prefixes have both their next-hop POP and AS diversity equal to 1. As in the case of ISP_{FM}, overall next-hop POP diversity is relatively higher than next-hop AS diversity, indicating that ISP_{RR} peers with its neighbor ASes in multiple POPs. Furthermore, we observe a few groups of prefixes sharing the same degree of POP diversity (*e.g.* POP diversity equal to 12 and 8). We verified that the cause for these prefix groups is the same as in ISP_{FM}; they represent the set of prefixes with the same next-hop AS(es).

Although both ISPs are classified as Tier-1, there is a noticeable difference in next-hop diversity. Overall, the number of ISP_{RR}'s next-hop POPs and ASes to reach a given prefix is lower, compared to ISP_{FM}. For example in ISP_{FM}, there are 10.17% and 62.76% of all prefixes with 1 next-hop POP and AS respectively. However in ISP_{RR}, we observe that relatively more prefixes (34.02% and 84.42%) have only 1 next-hop POP and AS respectively. There may be many reasons why the diversities differ. Later in Section 5, we further investigate on different factors and their impact on next-hop diversity and explain why there exists such discrepancy.

4.3 The Cause for Highest Next-hop Diversity

Given the high level understanding, our next goal is to examine the properties, especially the factors that increase the degree of next-hop diversity for a given prefix. In this section, we study prefixes with the highest degree of next-hop diversity to fully understand the factors which cause a given prefix to have a high diversity in the perspective of each of the 2 ISPs.

From ISP_{FM}, we identified the top 7,386 prefixes with the highest degree of next-hop diversity, announced by 1,151 unique origin ASes. The next-hop POP and AS diversity of the identified prefixes are greater than 15 and 4 respectively. Our further investigation on the origin ASes of the prefixes reveals that 1,147 origin ASes (99.65%) do not directly connect to ISP_{FM} and are a few AS hops away from ISP_{FM} [7]. Furthermore, the number of unique AS paths used to reach these prefixes is high (between

[7] The remaining 0.35% is a particular case that the origin ASes directly connect to ISP_{FM}, as well as via a few intermediate transit providers. However, due to manual configurations, both direct and indirect paths are selected as the best paths and manage to have the highest degree of diversity. This may be considered as an exception.

5 and 12), and all paths for a given prefix are tied in the BGP best path selection criteria (*i.e.* have same degree of preference [8]).

From ISP_{RR}, we also identified 6,094 prefixes with the highest degree of diversity, announced by 967 unique origin ASes. The next-hop POP and AS diversity of the identified prefixes are greater than 12 and 2 respectively. Our observation on these prefixes is essentially the same as in ISP_{FM}. None of the 967 origin ASes is directly connected to ISP_{RR}, and there is a high number of unique AS paths for each destination prefix, and the paths are tied in the BGP best path selection.

To summarize, we observed commonly from the 2 ISPs that the origin ASes which announced the prefixes with the highest degree of next-hop diversity are not direct neighbors of the 2 ISPs. Because of the dense connectivity in the Internet, the farther distance between the origin ASes and the 2 ISPs is translated into highly diverse AS-level paths which happen to include many paths with the same degree of preference. As a result, these prefixes have the highest next-hop diversity.

5 Impacting Factors on Next-hop Diversity

Based on the observation that the next-hop diversity is quite different between the 2 ISPs, we further investigate different factors to explain the observed discrepancy. In this section, we identify *external connectivity*, *iBGP hidden path phenomenon*, and *router topology and connectivity* as the 3 factors that affect the overall next-hop diversity, and focus on understanding their impact on the overall next-hop diversity.

5.1 External Connectivity

Intuitively, next-hop POP and AS diversity of a prefix are upper-bounded by the physical connectivity of the ISP with its neighbor ASes. To examine how different (or similar) these 2 ISPs are in terms of the amount of external connectivity, we infer the number of physical next-hop POPs and ASes for a set of prefixes by processing the routing update messages.

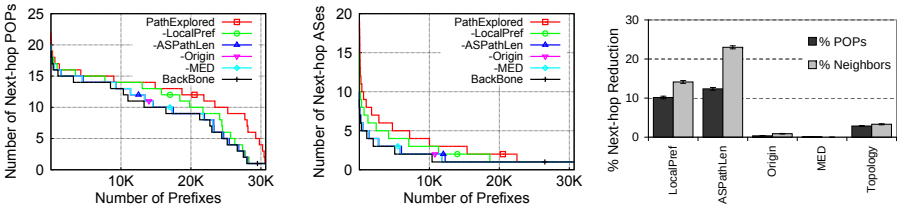
Based on the routing update messages collected during 1 week from June 3rd to June 9th in 2010, we identified prefixes that had their routes explored [9] at least once [10]. The number of such prefixes is 30,543 (9.94% of all prefixes), announced by 5,623 unique origin ASes (17.14% of all ASes: 7 Tier-1s, 713 Transits, and 4903 Stubs).

The red lines (labeled *PathExplored*) in Figure 4 and 5 show the number of next-hop POPs and ASes based on the inferred external connectivity for the identified prefixes in ISP_{FM} and ISP_{RR}. The distributions of the inferred external connectivity between the 2 ISPs reveals that there is not a significant discrepancy, indicating that the 2 ISPs are

[8] More specifically, these paths all have the same LOCAL_PREF value, AS_PATH length, ORIGIN, and MED value [8]

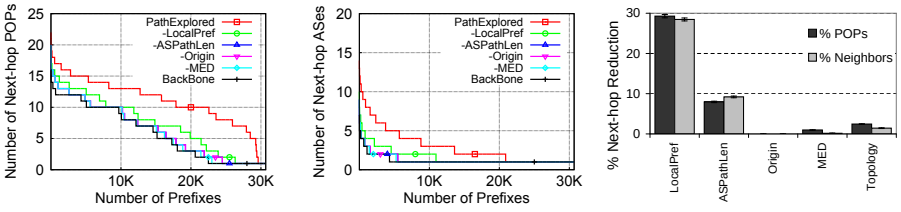
[9] In BGP, all available paths are explored before declaring that the given prefix is not reachable. Although all paths are explored in the network, not all paths are visible from a router. Thus, our estimation represents a lower bound.

[10] This value is determined empirically after measuring the number of additional next-hop POPs and ASes for a given withdrawal event.



(a) Next-hop POP Diversity (b) Next-hop AS Diversity (c) Average % Reduction

Fig. 4. Next-hop Diversity Reduction in ISP_{FM}



(a) Next-hop POP Diversity (b) Next-hop AS Diversity (c) Average % Reduction

Fig. 5. Next-hop Diversity Reduction in ISP_{RR}

rather similar in terms of the amount of external connectivity that they have with their neighbor ASes, and that this is not the dominating cause for the discrepancy observed in Figure 3.

5.2 iBGP Hidden Path Phenomenon

Given that the distribution of external connectivity of the 2 ISPs is similar, the next factor that may reduce the overall path diversity is iBGP hidden path phenomenon, which happens regardless of the iBGP architecture or router topology, as described earlier in Section 2.2. To quantify the amount of next-hop diversity reduced by iBGP hidden path phenomenon, we simulate the first 4 topology-independent criteria of BGP best path selection process and count how many external paths remain equally preferred by all routers inside the ISP after each of the criteria. The number of such remaining paths represents the optimal (*i.e.* as in the full-mesh topology) path diversity after hidden path phenomenon has happened.

Hidden Paths in ISP_{FM} Figure 4 summarizes the results for ISP_{FM}. In Figure 4(a) and Figure 4(b), each green, blue, pink, cyan colored lines show the remaining next-hop POP and AS diversity respectively after each step of the first 4 best path selection criteria in ISP_{FM}. For example in Figure 4(a), our inferred external connectivity (*i.e.* red line) indicates that there are only 0.4% of prefixes initially with their next-hop POP diversity equal to 1. After examining the 1st criterion (LOCAL_PREF comparison), the green line (labeled *-LocalPref*) shows that 7.36% of prefixes have the next-hop POP diversity equal to 1. This means, among multiple external paths to reach a given prefix, only *one* path stands out due to its higher LOCAL_PREF value, making the other (less preferred) paths hidden.

Figure 4(c) shows the average next-hop reduction (with 95% confidence intervals) after examining each of the first 4 topology-independent criteria. The first 2 criteria are identified as the dominating contributors in next-hop diversity reduction. After the 1st criterion (LOCAL_PREF comparison), about 10% of overall next-hop POP diversity is reduced. Then additional 12% next-hop POP diversity reduction happened after the 2nd criterion (AS_PATH length comparison).

Hidden Paths in ISP_{RR} Figure 5 summarizes the results for ISP_{RR}. As in the case of ISP_{FM}, the first 2 criteria of the best path selection process are identified as the dominating factors that reduce next-hop diversity. However, the amount of reduction happened by each of the 2 criteria is quite different. In case of ISP_{RR}, the 1st criterion had the most impact on next-hop diversity reduction (of about 29%), and is the main reason why the 2 ISPs have such discrepancy in the measured next-hop diversity in Figure 3. Our results reveal that although ISP_{RR} has a similar amount of external connectivity compared to ISP_{FM}, relatively less number of paths are equally preferred after examining LOCAL_PREF attribute value and the subsequent topology-independent criteria.

Figure 5(c) shows that the first 2 criteria are the dominating contributors in next-hop diversity reduction, as in the case of ISP_{FM}. The first 2 criteria together hide up to 37% of next-hop POP diversity in average.

5.3 Router Topology and Connectivity

The iBGP hidden path phenomenon due to the first 4 topology-independent criteria of the best path selection happens regardless of the iBGP topology. This implies that even in the full-mesh topology, the remaining next-hop diversity after the 4th criterion is the upper bound, and that further reduction caused by the topology-dependent criteria indicates the cost of moving away from the full-mesh topology.

Thus, we define the difference between measured diversity as seen by the backbone routers (*i.e.* black line labeled *BackBone*) and the diversity after the 4th criterion of best path selection (cyan line labeled *-MED*) as the amount of diversity reduced due to topology and connectivity between the border routers and the backbone routers.

Figure 4(c) and 5(c) show that in both ISPs, the reduction due to the topology-dependent factors is relatively small; even with ISP_{RR}'s multi-level hierarchical route reflection architecture and its topology, there is only up to 3.3% reduction.

5.4 Representativeness of the Studied Prefixes

The prefixes discussed in this section represent the most dynamic prefixes observed during the given week and may not be a good representative set. To examine the representativeness of our observation, we checked that the prefixes and their origin ASes roughly represent the different AS types, topological locations, and the overall next-hop diversity. Additionally, we performed the same measurement using 1 week of routing update messages on different months in 2010. Although the percentage in diversity reduction varies slightly, the generality of the conclusion does not change in the additional experiments. However, more studies are necessary to understand the relationship between path diversity and dynamics of a prefix.

6 Discussions and Future Work

Our measurement study based on the iBGP data from two Tier-1 ISPs quantifies the degree of path diversity in these two ISPs and reveals most influential factors on the BGP path diversity. Despite their different iBGP architectures, both ISPs exhibit a low next-hop AS diversity and a relatively higher next-hop POP diversity for majority of prefixes. Because high degrees of path diversity are due to origin ASes being distant from ISP_{FM} or ISP_{RR}, and because both ISP_{FM} and ISP_{RR} have presence in multiple continents, relatively few ASes are distant from them and with the same preference level, this explains why the number of prefixes with high degrees of path diversity is low.

Furthermore, although it has been speculated that multi-level hierarchical route reflection architecture might have an impact on reducing the overall path diversity, our results only show a minor reduction is due to such multi-level topology in ISP_{RR}. Because LOCAL_PREF and AS_PATH are two BGP attributes that are high in the BGP decision making order and that are independent from an AS's internal topological connectivity, they lead to multiple routers choose the same path for a given destination, hence contributing significantly to the reduction of path diversity in the AS regardless of the iBGP topology.

In this work, we focused on understanding the static path diversity in different iBGP architecture and in the absence of failures. It remains as an open question how different iBGP architectures may impact BGP convergence in the presence of topological changes, which is the subject of our ongoing effort.

References

1. T. Bates, E. Chen, and R. Chandra. RFC 4456: BGP Route Reflection: An Alternative to Full Mesh Internal BGP (IBGP), April 2006.
2. J. Choi, J. H. Park, P. chun Cheng, D. Kim, and L. Zhang. Understanding BGP Next-hop Diversity, UCLA CS Technical Report 100026, Aug 2010.
3. V. V. den Schrieck and P. Francois. Analysis of Paths Selection Modes for Add-paths, July 2009.
4. P. Marques, R. Fernando, E. Chen, and P. Mohapatra. Advertisement of the Best External Route in BGP, August 2010.
5. W. Mhlbauer, S. Uhlig, A. Feldmann, O. Maennel, B. Quoitin, and B. Fu. Impact of Routing Parameters on Route Diversity and Path Inflation. *Computer Networks*, 2010.
6. R. Oliveira, D. Pei, W. Willinger, B. Zhang, and L. Zhang. The (in)Completeness of the Observed Internet AS-level Structure. In *IEEE/ACM Transactions on Networking*, 2010.
7. R. Raszuk, R. Fernando, K. Patel, D. McPherson, and K. Kumaki. Distribution of Diverse BGP Paths, July 2010.
8. Y. Rekhter, T. Li, and S. Hares. RFC 4271: A Border Gateway Protocol 4 (BGP4), January 2006.
9. P. Traina, D. McPherson, and J. Scudder. RFC 5065: Autonomous System Confederations for BGP, August 2007.
10. S. Uhlig and S. Tandel. Quantifying the BGP Routes Diversity Inside a Tier-1 Network. *Networking*, 3976, April 2006.
11. D. Walton, A. Retana, E. Chen, and J. Scudder. Advertisement of Multiple Paths in BGP, August 2010.