Leveraging Network Performances with IPv6 Multihoming and Multiple Provider-Dependent Aggregatable Prefixes *

Cédric de Launois, Bruno Quoitin, and Olivier Bonaventure

Université catholique de Louvain
Department of Computing Science and Engineering
http://www.info.ucl.ac.be
{delaunois,quoitin,bonaventure}@info.ucl.ac.be

Abstract. Multihoming, the practice of connecting to multiple providers, is becoming highly popular. Due to the growth of the BGP routing tables in the Internet, IPv6 multihoming is required to preserve the scalability of the interdomain routing system. A proposed method is to assign multiple provider-dependent aggregatable (PA) IPv6 prefixes to each site, instead of a single provider-independent (PI) prefix. We show that the use of multiple PA prefixes not only allows route aggregation but also can be used to reduce end-to-end delays by leveraging the Internet path diversity. We quantify the gain in path diversity, and show that a dual-homed stub AS that uses multiple PA prefixes has already a better Internet path diversity than any multihomed stub AS that uses a single PI prefix, whatever its number of providers. The benefits provided by the use of IPv6 multihoming with multiple PA prefixes are an opportunity to develop the support for quality of service and traffic engineering.

Key words: BGP, IPv6 Multihoming, Path Diversity.

1 Introduction

Today, the Internet connects more than 20000 *Autonomous Systems* (AS) [2], operated by many different technical administrations. The large majority of ASes are *stub* ASes, i.e. autonomous systems that do not allow external domains to use their infrastructure. Only about 20% of autonomous systems provide transit services to other ASes [3]. They are called *transit* ASes. The Border Gateway Protocol (BGP) [4] is used to distribute routing announcements among routers that interconnect ASes.

The size of the BGP routing tables in the Internet has been growing dramatically during the last years. The current size of those tables creates operational issues for some Internet Service Providers and several experts are concerned about the increasing risk of instability of BGP [5]. Part of the growth of the

* This paper is an extended version of the paper published in the proceedings of the QoSIP 2005 conference [1].

BGP routing tables [6] is due to the fact that, for economical and technical reasons, many ISPs and corporate networks wish to be connected via at least two providers to the Internet. For more and more companies, Internet connectivity takes a strategic importance. Nowadays, at least 60% of those domains are multihomed to two or more providers [3], and this number is growing. Many sites are expected to also require to be multihomed in IPv6, primarily to enhance their reliability in the event of a failure in a provider network, but also to increase their network performances such as network latency. In order to preserve the scalability of the interdomain routing system, every IPv6 multihoming solution is required to allow route aggregation at the level of their providers [5]. The IPv6 multihoming solution promoted by the IETF is to assign multiple provider-dependent aggregatable (PA) IPv6 prefixes to each site, instead of a single provider-independent (PI) prefix [7]. Both IPv4 and IPv6 multihoming methods are described in section 3.

We show in this paper that the use of multiple PA prefixes introduces other benefits than simply allowing route aggregation. We first explain in section 4 how stub ASes that use multiple PA prefixes can exploit paths that are otherwise unavailable. In other words, we explain how the use of PA prefixes increases the number of concurrent paths available. Next, we show that lower delays can often be found among the new paths. Our simulations suggest that a delay improvement is observed for approximately 60% of the stub-stub pairs, and that the delay improvement could be higher in the actual Internet.

In section 5, we quantify the gain in terms of Internet path diversity. We propose a new, fine-grain metric to measure the AS level path diversity. This metric is used to show that a dual-homed stub AS that uses multiple PA prefixes has already a better Internet path diversity than any multihomed stub AS that uses a single PI prefix, whatever its number of providers.

2 Related Work

A work about IPv4 multihoming path diversity appears in [8], where the authors define two path diversity metrics to quantify the reliability benefits of multihoming for high-volume Internet servers and receivers. They notice however that their metrics have an undesirable bias in favour of long paths. Their study draws empirical observations from measurement data sets collected at servers and monitoring nodes, whereas our work focuses on IPv6 multihoming and is based on inferred and generated global-scale AS-level topologies.

A comparison of Overlay Routing and Multihoming Route Control appears in [9]. In that study, the authors demonstrate that an intelligent control of BGP routes, coupled with ISP multihoming, can provide competitive end-to-end performance and reliability compared to overlay routing. Our results agree with this finding. In addition, our work will explicitly express the impact of the path diversity on performances. It will show that IPv6 multihoming with multiple PA prefixes is able to provide these benefits.

It is well known that the use of provider-dependent aggregatable prefixes preserves the scalability of the interdomain routing system [10]. To our knowledge, this is the first study that shows that the use of multiple PA prefixes also increases network performances by leveraging the Internet path diversity, compared to the use of traditional multihoming with a single prefix.

3 IPv4 and IPv6 Multihoming

This section provides some background on traditional IPv4 multihoming and on IPv6 multihoming.

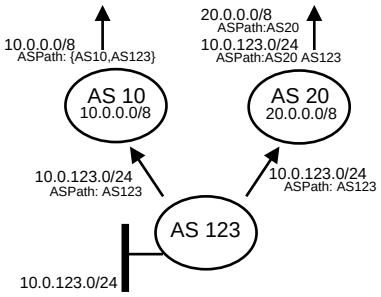

Fig. 1. IPv4 Multihoming using a provider-aggregatable prefix

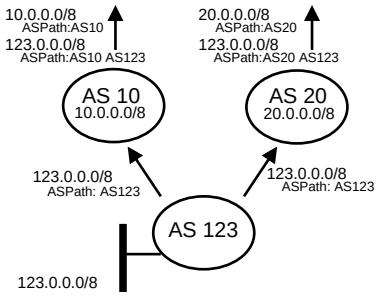

Fig. 2. IPv4 Multihoming using a provider-independent prefix

In the current IPv4 Internet, the traditional way to multihome is to announce, using BGP, the single site prefix to each provider, as depicted in Figure 1 and Figure 2. In Figure 1, AS 123 uses provider-aggregatable addresses. It announces prefix 10.0.123.0/24 to its providers AS 10 and AS 20. AS 10 aggregates this prefix with its 10.0.0.0/8 prefix and announces the aggregate to the Internet. In Figure 2, AS 123 announces a provider-independent prefix to its providers. This prefix is then propagated by BGP routers over the Internet. Throughout this paper, we will refer to this technique as *traditional IPv4 multihoming*, or simply *IPv4 multihoming*.

The way stub ASes multihome in IPv6 is expected to be quite different from the way it is done currently in IPv4. Most IPv6 multihoming mechanisms proposed at the IETF rely on the utilisation of several IPv6 provider-aggregatable prefixes per site, instead of a single provider-independent prefix, see [7, 11] and the references therein. Figure 3 illustrates a standard IPv6 multihomed site.

In Figure 3, AS 10 and AS 20 provide connectivity to the multihomed site AS 65001. Each provider assigns to AS 65001 a site prefix, respectively 2001:10:1::/48 and 2001:20:1::/48. The two prefixes are advertised by the site exit routers RA and RB to every host inside AS 65001. Finally, these prefixes are used to derive

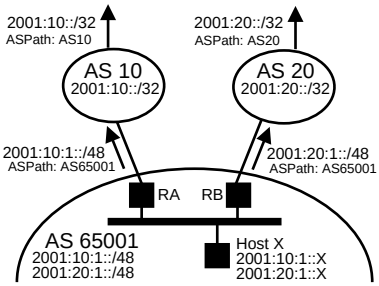

Fig. 3. IPv6 Multihoming

one IPv6 address per provider for each host interface. In this architecture, AS 65001 advertises prefix 2001:10:1::/48 only to AS 10, and AS 10 only announces its own IPv6 aggregate 2001:10::/32 to the global Internet. This new solution is expected to be used only by stub ASes. Transit ASes are not concerned by these solutions since they will receive provider-independent IPv6 prefixes. Consequently, in this study, we focus only on stub ASes.

The use of multiple PA prefixes is natural in an IPv6 multihoming environment. However, it is not impossible to use the same multihoming technique in IPv4, i.e. to delegate two IPv4 prefixes to a site. Unfortunately, due to the current lack of IPv4 addresses, the need to delegate several IPv4 prefixes to a multihomed site makes this solution less attractive. Therefore, throughout this document, the new multihoming technique presented here for IPv6 is simply called *IPv6 multihoming*; although the same concept could also be applied to IPv4 multihomed sites, and although other IPv6 multihoming techniques exist.

4 Improving Delays with Multiple Prefixes per Site

We show in this section how the use of multiple PA prefixes can reduce the end-to-end delay by leveraging the Internet path diversity.

Section 4.1 explains how stub ASes that use PA prefixes can exploit paths that are otherwise unavailable when a single PI prefix is used. Among the newly available paths, some offer lower delays. In section 4.3, we rawly estimate how often this improvement in network latency occurs. The topology used for the simulation is presented in section 4.2.

4.1 Impact of PI and PA Prefixes on Available AS Paths

In this paper, we focus on the paths announced by BGP between each pair of stub ASes in a given topology. These paths depend on the topology but also on the commercial relationships between ASes, together with their BGP routing policies. The commercial agreements between two ASes are usually classified as

customer-provider relationships or shared-cost peerings [12, 13]. The BGP routing policies basically define that an AS announces all the routes to its customers, but announces to its peers and providers only the internal routes and the routes of its customers. In addition, the policies are usually defined so that an AS prefers routes received from a customer, then routes received from a peer, and finally routes received from a provider [12, 13]. These filters ensure that an AS path will never contain a customer-to-provider or peer-to-peer edge after traversing a provider-to-customer or peer-to-peer edge. This property is known as the the *valley-free* property [12].

Figure 4 shows an AS-level interdomain topology with shared-cost peerings and customer-provider relationships. An arrow labelled with "$" from AS x to AS y means that x is a customer of y. A link labelled with "=" means that the ASes have a shared-cost peering relationship [12]. For instance, both S and D are dual-homed ASes in Figure 4.

In IPv4, D typically announces a single provider-independent prefix to each of its providers. This PI prefix is propagated by BGP routers all over the Internet. In particular, if S is single-homed, it will receive a single route from its provider to reach the dual-homed AS D. This route is the best route known by the provider to join D. If S is also dual-homed, as illustrated in Figure 4, S will receive two BGP routes $ECAD$ and $FCAD$ towards D one from each of its providers, as shown in Figure 5.

When stub ASes use IPv6 multihoming with multiple PA prefixes, additional routes exist.

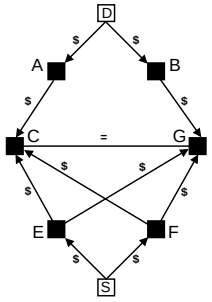

Fig. 4. Topology

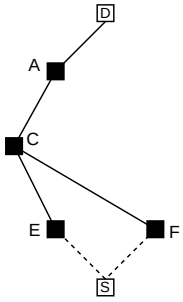

Fig. 5. IPv4 path tree

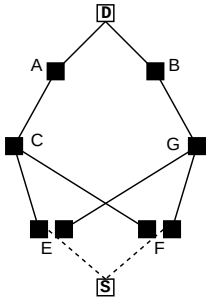

Fig. 6. IPv6 path tree

Suppose that both S and D use IPv6 multihoming with multiple PA prefixes. Every host in S has two IPv6 addresses. One is derived from the prefix allocated by E to S , while the other one is derived from the prefix allocated by F to S. Similarly, every host in D has two IPv6 addresses. When selecting the source address of a packet to be sent, a host in S could in theory pick any of its two addresses. However, for security reasons, IPv6 providers should refuse to convey packets with source addresses outside their address range [7, 11]. For example, E should refuse to forward a packet with a source address belonging to F. As

a consequence, the source address selected by a host determines the upstream provider used.

Using traditional IPv4 multihoming, two BGP routes towards D (e.g. $SECAD$ and $SFCAD$) are advertised by E and F to S, as illustrated on Figure 5. In an IPv6 multihoming scenario, since both S and D have two prefixes, S can reach D via A or B depending on which destination prefix is used, and via E or F depending on which source prefix is used. So, S has a total of four paths to reach D : $SECAD$, $SEGBD$, $SFCAD$ and $SFGBD$. These four routes are illustrated on Figure 6.

4.2 A Two-Level Topology with Delays

We detail in this section the topology that we use to rawly estimate how often lower delays can be found among newly available paths. In order to simulate delays along paths, we cannot rely on topologies provided by Brite [14], Inet [15], or GT-ITM [16] since they either do not model business relationships or do not provide delays along links.

A topology that contains both delays and commercial relationships is available at [17]. In this topology, the interdomain links and the business relationships are given by a topology inferred from multiple collected BGP routing tables [12,13]. For each peering relationship found between two domains in this topology, interdomain links are added. The different points of presence of each domain are geographically determined by relying on a database that maps blocks of IP addresses and locations worldwide. The intradomain topology is generated by first grouping routers that are close to each other in clusters, and next by interconnecting these clusters with backbone links. The delays along the links is the propagation delay computed from the distance between the routers. The IGP weights used are the delays for links shorter than 1000 km, twice the delay for links longer than 1000 km but shorter than 5000 km and 5 times the delay for links longer than 5000 km. This is used to penalise the long intradomain links and favour hot-potato routing. In this topology, 55% of the delays along the BGP route are comprised between 10 and 50ms. About 20% of the delays are below 10ms and 25% sit between 50 and 100ms. These delays can be considered as minimal bounds for delays really observed in the Internet, since only the propagation delay is taken into account. Factors that increase delays like limited bandwidths or congestion delays are not considered here. Although the simulated delays are inferior bounds to delays observed in the global Internet, their order of magnitude is preserved.

The resulting topology is described in more details in [17]. It contains about 40,000 routers, 100,000 links and requires about 400,000 BGP sessions. Since the business relationships are known for this topology, we are able to compute, for each AS, the corresponding BGP routing policies for every AS pairs. The paths for this topology are obtained by simulating the BGP route distribution over the whole topology. For this purpose, we use a dedicated BGP simulator, named C-BGP [18]. C-BGP supports import and export filters, and uses the full BGP decision process. In the absence of intradomain structures, the tie-breaking rule

used by C-BGP for choosing between two equivalent routes is to prefer the route learned from the router with the lowest router IP address, i.e. the standard rule used by BGP-4. As soon as all the routes have been distributed and BGP has converged, we perform traceroute measurements on the simulated topology, and deduce the router-router paths and the delays between multihomed stub ASes. To reduce the simulation time, we conduct the simulation for 2086 multihomed stub ASes randomly chosen among the 8026 multihomed stub ASes.

4.3 Simulation Results

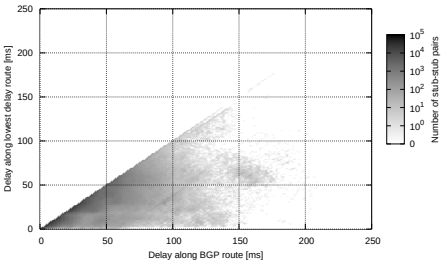

Fig. 7. Delay along the BGP route versus delay along the lowest delay route

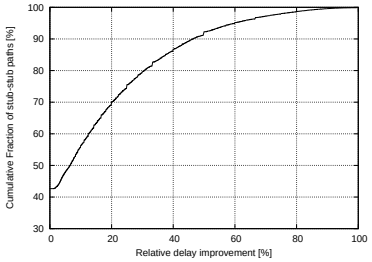

Fig. 8. Distribution of the relative delay improvement

Figure 7 plots the lowest delay obtained when stub ASes use traditional IPv4 multihoming (x-axis), against the lowest delay obtained when stub ASes use IPv6 multihoming with multiple PA prefixes (y-axis). The gray-scale indicates the number of stub-stub AS pairs, on a logarithmic scale. The diagonal line that appears represents stub-stub AS pairs for which both multihoming techniques yield to the same lowest delay.

As explained in section 4.1, the use of multiple PA prefixes provides additional paths, beside traditional paths that are still available. As a consequence, delays can only improve, and no dot can appear above the diagonal line. A dot under this diagonal line indicates that the use of multiple PA prefixes introduces a new path with a delay lower than the delay along the best BGP path obtained when a single PI prefix is used. We can see that many dots are located under this line. Sometimes, the improvement can even reach 150ms in this topology.

Figure 8 shows the cumulative distribution of the relative delay improvement. It shows that no improvement is observed for approximately 40% of the stub-stub AS pairs. However, the relative improvement is more than 20% for 30% of stub-stub AS pairs. Delays are cut by half for about 8% of stub-stub AS pairs.

As said in section 4.2, the delays observed in this topology are expected to be minimal bounds to those seen in the real Internet. Thus, we can reasonably assume that the absolute delay improvements presented in Figure 7 will not be lower in the actual Internet.

These simulation results show that improving delays is a benefit of IPv6 multihoming with multiple PA prefixes, without increasing the BGP routing tables.

5 Leveraging Internet Path Diversity with Multiple Prefixes

Section 4.1 has shown that stub ASes that use multiple PA prefixes can exploit paths that are otherwise unavailable. In other words, the use of multiple PA prefixes increases the number of paths available, i.e. the Internet path diversity. We have shown that better delays can often be found among the new paths. The path diversity also directly impacts the resilience to failure of a site, together with its ability to share its traffic load and to support quality of services. For example, a site for which all paths merge in a single AS in the Internet is dependent on the performances of this particular AS. Having a wide variety of paths to join and to be joined by other ASes ensures larger possibilities to cope with routing problems occuring in the Internet. In this section, we propose to quantify the Internet path diversity that exists when a multihomed stub AS uses either multiple PA prefixes or a single PI prefix.

First, section 5.1 introduces a new metric to measure the AS-level path diversity. Next, the topologies used for our simulations are described in section 5.2. The simulation results are presented and discussed in section 5.3. Finally, the impact of BGP and the impact of the topology on the path diversity are evaluated in sections 5.5 and 5.4.

5.1 A New Path Diversity Metric

In order to measure the path diversity for a given destination AS, we first build the tree of paths from all source ASes towards the destination AS. As explained in section 4.1, this path tree depends on the multihoming technique used. Next, we use a new, fine-grain, path diversity metric to evaluate the diversity of this tree. This metric takes into account the lengths of the paths and how much they overlap. We define this new path diversity metric, from a source AS S to a destination AS D, as follows.

Let $P_1, P_2, ..., P_n$ be the n providers of S. We first build the tree of all paths starting from providers P_i of S to destination D, for $i = 1, ..., n$. This tree represents all the BGP paths for D that are advertised by the providers P_i to S. Our path diversity metric is computed recursively link by link, from the leaves to the root. It returns a number between 0 and 1. We first assign an initial diversity of 0.5 to each link in the tree. This number is chosen in order to best distribute the values of the path diversity metric in the range $[0, 1]$. At each computation step, we consider two cases, to which all other cases can be reduced. Either two links are in sequence, or the links join in parallel at the same node.

In the first case, two links with diversity d_1 and d_2 in sequence can be merged into a single link with a combined diversity $d_{1,2} = d_1 \cdot d_2$. The combined diversity

Alg. 1. Computing Diversity Metric

```
Diversity(root)
{
    d = 0 ;
    if ( Children(root) == ∅ )
        return 1 ;

    foreach child ∈ Children(root) {
        d_child = 0.5 · Diversity(child) ;
        d = d + d_child − d · d_child ;
    }

    return d ;
}
```

$d_{1,2}$ is a number in $[0, 1]$ lower than both d_1 and d_2, so that the metric favours short paths over longer ones. This computation step also implicitly gives a higher importance to the path diversity that exists near the root of the tree, i.e. near the destination AS. This property ensures that the metric prefers trees where paths join lately near the destination node over trees where paths merge near the source node.

In the second case, when a link with a diversity d_1 and another link with a diversity d_2 join in parallel, we merge the two links into a single link with a combined diversity $d_{1,2}$, computed as $d_{1,2} = d_1 + d_2 - d_1 \cdot d_2$. The resulting diversity is greater than both d_1 and d_2, which corresponds adequately to an improvement in terms of path diversity. A recursive algorithm to compute this metric is presented in Alg. 1.

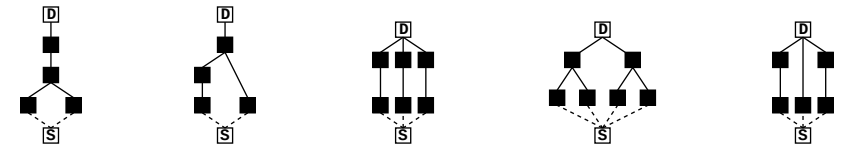

Fig. 9. Path diversity metric examples

Examples of values for d are shown in Figure 9. In figures 9(a) and (b), the source S is dual-homed and the destination D is single-homed. d in Figure 9(b) is better than d in Figure 9(a) because the tree (b) contains a path with 3 hops and a path with 2 hops, while the tree (a) contains 2 paths of 3 hops each. The diversity d is better for trees (c) and (e) than for trees (a) and (b) because the latter ones contain 3 disjoint paths instead of 2. However, d in tree (d) has a slightly better diversity than d in tree (c), even if (c) has 3 disjoint paths while (d) has only 2. The reason is that the 2 disjoint paths of (d) have 2 sub-branches each, while the diversity of the 3 disjoint paths of (c) is mitigated due to their lengths.

Other metrics exist that compute the path diversity [8, 19, 20]. In Table 1, the first metric is the one presented in this work. The second is a metric used in [8] to quantify the diversity in network paths that multihoming provides. The expected fraction of edges that are shared by two or more paths in the tree is given by $\frac{P-E}{E}$ where P denotes the sum of the hop-counts of the individual paths from the source to the destination, and E is the total number of edges in the tree. Thus $1 - \frac{P-E}{E}$ could be used to estimate the fraction of paths that are non overlapping, i.e. to estimate path diversity. The last four metrics are used in [19, 20] to characterise the path diversity of complete ISP topologies. The third and fourth metrics calculate respectively the number of node-disjoint and link-disjoint paths. A partially node- or link-disjoint path is defined as one for which there are respectively some nodes or links that appear in more than one path. These last four metrics were adapted to compute the inter-AS path diversity.

Values of these metrics for the examples illustrated in Figure 9 are indicated in Table 1. For all these metrics, a higher value suggests a better diversity.

Table 1. Path diversity values computed by different metrics.

Metric	*(a)*	*(b)*	*(c)*	*(d)*	*(e)*
1. Our metric d	0.19	0.31	0.58	0.61	0.72
2. $(1 - \frac{P-E}{E})$ [8]	0.5	0.75	1	0.67	1
3. Node-disjoint paths	1	1	3	2	3
4. Link-disjoint paths	1	1	3	2	3
5. Partially node-disjoint paths	2	2	3	4	3
6. Partially link-disjoint paths	2	2	3	4	3

For our study, the second metric has an undesirable bias in favour of long paths. Moreover, it cannot differentiate some cases, such as those illustrated in Figure 9(c) and 9(e). Finally, this second metric is unable to correctly compare other cases. For example, when comparing trees in Figure 9(b) and Figure 9(d), the metric evaluates that tree 9(b) has a better diversity than tree 9(d). This is obviously wrong. The 3rd, 4th, 5th and 6th metrics are not fine-grain enough for our analysis. For example, none of them is able to distinguish cases 9(a) and 9(b), or cases 9(c) and 9(e). Only our first metric d is able to provide a precise and fine-grained measure of the path diversity between two nodes.

5.2 Internet Topologies

IPv6 multihoming with multiple PA prefixes is currently not deployed. As a consequence, our evaluations are performed on synthetic Internet topologies, instead of conducting measurement experiments on the actual IPv4 Internet. No accurate model of the global Internet currently exists. Modelling the Internet, even only at the AS level, remains an active research topic [21]. Hence, in order to draw some conclusions about the real Internet, we perform our simulations

on several Internet-like topologies, with different properties. The simulations on these various topologies allow us to determine the impact of the topology on the results, but also to explore possible evolution scenarios for the Internet.

In section 4, we used a large router-level Internet topology that models delays. Here, we use AS-level topologies instead, for two reasons. A first reason is the computation time. The topology used in section 4 is unnecessarily complex for an AS-level simulation since it models routers and delays. A second reason is that we want to consider different types of topologies to estimate the variability of our results with respect to the topology.

We first use an AS-level Internet topology inferred from several BGP routing tables using the method developed by Subramanian et al. [13].

Next, we generate three AS-level Internet-like topologies, using a Barabási-Albert model [22]. The topologies are created level by level, from the dense core to the customer level. Nodes are added one at a time, using the Barabási-Albert preferential connectivity model, i.e. new nodes tend to connect to existing nodes that are highly connected. The generated topologies provide details about customer-provider and peer-to-peer relationships. Their numbers of Internet hierarchy levels and nodes in each level can be specified, so that we can produce small- or large- diameter Internet topologies while preserving the same number of stub ASes and transit ASes. This feature is used in section 5.4 to explore different scenarios of the Internet evolution.

5.3 Simulation Results

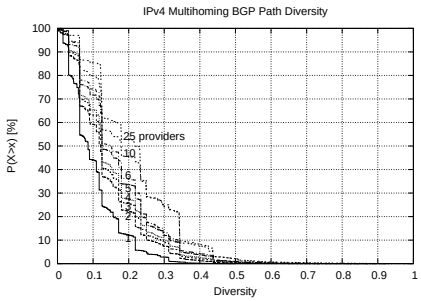

Fig. 10. AS-level path diversity for the inferred Internet topology, using traditional IPv4 multihoming

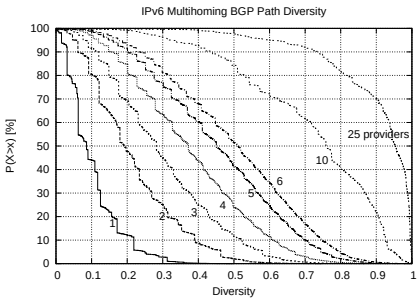

Fig. 11. AS-level path diversity for the inferred Internet topology, using IPv6 multihoming

Figure 10 presents the path diversity available to stub ASes that use traditional IPv4 multihoming in the inferred AS-level Internet topology. Figure 11 shows the path diversity when all stub ASes use IPv6 multihoming with multiple PA prefixes, in the same inferred topology.

The figures show $p(x)$: the percentage of couples *(source AS, destination AS)* having a path diversity greater than x. The results are classified according to the number of providers of the destination stub AS. The number of providers is indicated beside each curve. Figure 10 shows for example that only 12% of single-homed stub ASes using traditional IPv4 multihoming have a diversity better than 0.2. This percentage raises to 22% for dual-homed stub ASes. Figure 11 shows that about 50% dual-homed IPv6 stub ASes have a path diversity better than 0.2.

We can observe that the diversity remains the same when considering only single-homed destinations. Indeed, only one prefix is announced by a single-homed stub AS, using either IPv4 or IPv6 multihoming technique. The use of IPv6 multihoming does not introduce any benefit in this case.

When comparing figures 10 and 11, it appears that the AS-level path diversity is much better when stub ASes use multiple PA prefixes than when they use a single PI prefix. For example, when considering dual-homed IPv6 stub ASes, Figure 11 shows that the path diversity observed is already as good as the path diversity of a 25-homed stub AS that uses traditional IPv4 multihoming. The path diversity obtained by a 3-homed stub AS that uses IPv6 multihoming completely surpasses the diversity of even a 25-homed stub AS that uses traditional IPv4 multihoming.

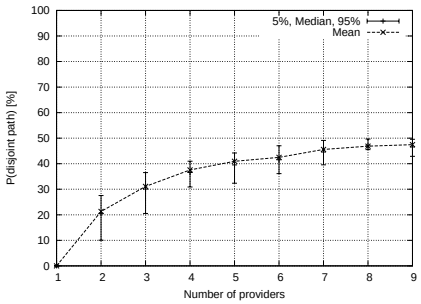

Fig. 12. Probability that a stub AS has at least two disjoint paths towards any other stub AS, when it uses a single PI prefix

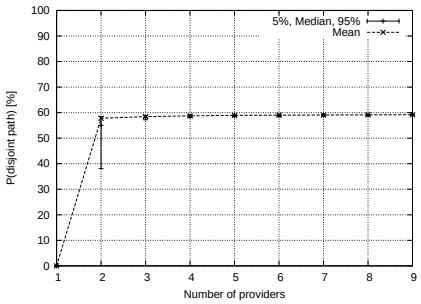

Fig. 13. Probability that a stub AS has at least two disjoint paths towards any other stub AS, when it uses multiple PA prefixes

These results are corroborated by Figure 12 and Figure 13. These figures present the probability that a stub AS has at least two disjoint paths towards another stub AS, in the inferred Internet topology. They show the mean, 5^e percentile, median and 95^e percentile of this probability. The results are classified according to the number of providers of the stub AS. The percentage of single-homed stub ASes in this topology is about 40%, and thus the probability of having disjoint paths is at most 60%, whatever the number of providers. Figure 12 shows for instance that a dual-homed stub AS has at least two disjoint paths

towards 20% of the destination ASes in average. Figure 13 considers the use of multiple PA prefixes. It shows in this case that being dual-homed is sufficient for most stub ASes to reach the maximum probability of having disjoint paths up to a destination AS. This confirms our previous finding.

5.4 Influence of Topology on Path Diversity

The way Internet will evolve in the future remains essentially unknown. In order to determine the range of variation for our simulation results, we perform simulations with three distinct generated topologies.

The first is a topology that tries to resemble the current Internet [13]. Four hierarchy levels of ASes are generated for this topology : a fully-meshed dense core, a level of large transit ASes, a level of local transit ASes, and a level of stub ASes. The proportion of nodes in each level is similar to the proportion observed for the current Internet. Figure 14 and Figure 15 show the AS-level path diversity for this generated topology. As expected, the path diversity results for this generated topology are almost identical to the results obtained for the inferred topology.

The second is a small-diameter Internet topology, consisting of stub ASes directly connected to a fully meshed dense core. This topology simulates a scenario where ASes in the core and large transit ASes concentrates for commercial reasons. At the extreme, the Internet could consist in a small core of large transit providers, together with a large number of stub ASes directly connected to the transit core. This could lead to an Internet topology with a small diameter. The AS-level path diversity for such a topology is illustrated on Figure 16 and Figure 17. As expected, the diversity in a small-diameter topology is better, since the paths are shorter than in the current Internet. When comparing the results illustrated by Figure 16 and 17, it appears that the gain in path diversity is also large for a low-diameter topology.

The third is a large-diameter topology, generated using eight levels of ASes. This topology simulates a scenario where the Internet continues to grow, with more and more core, continental, national and metropolitan transit providers. In this case, the Internet might evolve towards a network with a large diameter. The same simulations are performed. The path diversity results are presented by Figure 18 and Figure 19. These figures show a poor path diversity in comparison with the path diversity of the previous topologies. This is due to the paths being longer. Again, these two figures show that the path diversity remains low when stub ASes use a single PI prefix, whatever their number of providers. When multiple PA prefixes are used, the path diversity rises much faster with the number of providers, as shown by Figure 19. These two figures confirm that the gain in path diversity is substantial also for a large-diameter topology.

Figures 20 and 21 show the average path diversity in function of the number of providers for all topologies considered. For a given destination stub AS D, we compute the mean of path diversities from every source stub towards D. We then group the destination stub ASes according to their number of providers, and

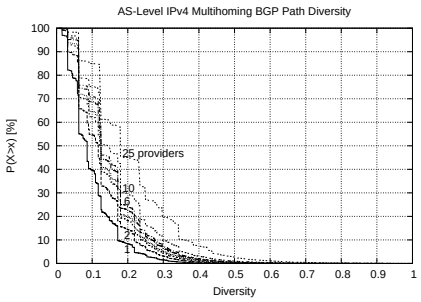

Fig. 14. AS-level path diversity d for a generated Internet-like topology, using a single PI prefix

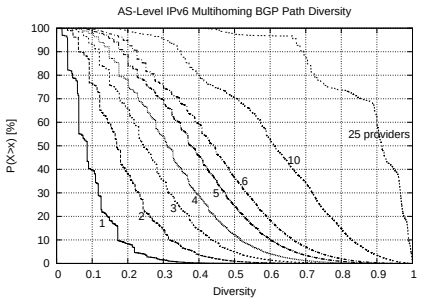

Fig. 15. AS-level path diversity d for a generated Internet-like topology, using multiple PA prefixes

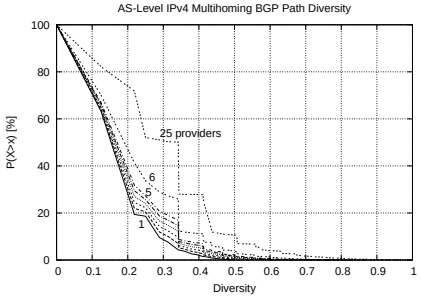

Fig. 16. AS-level path diversity d for a small-diameter generated topology, using a single PI prefix

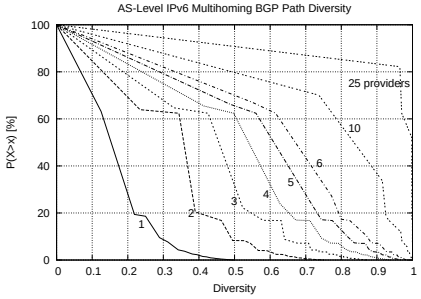

Fig. 17. AS-level path diversity d for a small-diameter generated topology, using multiple PA prefixes

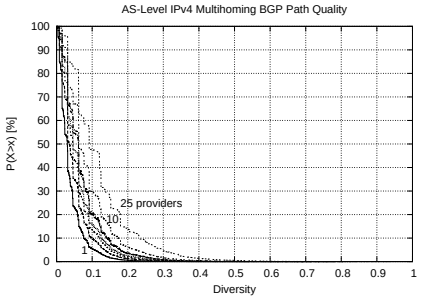

Fig. 18. AS-level path diversity d for a large-diameter generated topology, using a single PI prefix

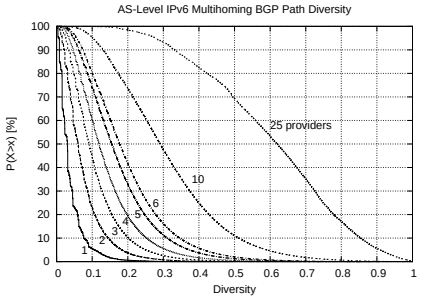

Fig. 19. AS-level path diversity d for a large-diameter generated topology, using multiple PA prefixes

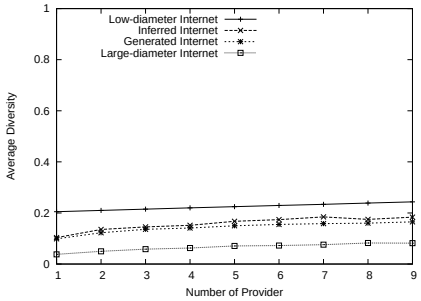

Fig. 20. Average path diversity using traditional IPv4 Multihoming

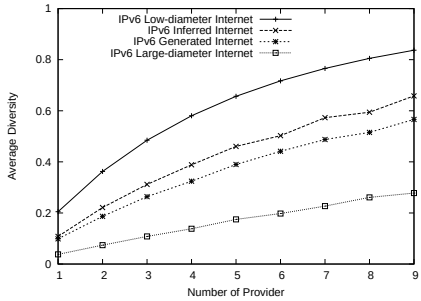

Fig. 21. Average path diversity using IPv6 Multihoming

compute the mean of their path diversities. In Figure 20 and Figure 21, we can first observe that the results obtained for the generated and inferred Internet topologies are fortunately quite close. We can also observe that the average diversity of the inferred Internet is included between the average diversities of the small- and large-diameter generated Internet topologies. Figure 20 shows that the average path diversity using a single PI prefix does not rise much in function of the number of providers, for all topologies considered. Figures 10 and 20 suggest that it is nearly impossible that a stub AS achieves a good path diversity using traditional IPv4 multihoming, whatever its number of providers. In contrast, as shown by Figure 21, the path diversity that is obtained using multiple PA prefixes is much better. Figures 20 and 21 show that a dual-homed stub AS using IPv6 multihoming already gets a higher diversity than any multihomed stub AS that uses traditional IPv4 multihoming, whatever its number of provider and for all topologies considered. In a small-diameter Internet, this diversity rises fast with the number of providers, but also shows a marginal gain that diminishes quickly. In a large-diameter Internet, the diversity rises more slowly.

Figure 22 summarises the results for the analysed topologies. It shows the path diversity benefit in percent that a stub AS obtains when it uses multiple PA prefixes instead of a single PI prefix. We can notice that the gain is obviously null for single-homed stubs, as the use of one PA prefix instead of one PI prefix has no impact on the path diversity. The figure shows that the gain is high when multiple PA prefixes are used, as soon as the stub AS has more than a single provider. Additionally, we can see that the gain does not vary much with the topology considered. Figure 22 also shows that the gain for the current inferred Internet is almost everywhere included between the gains of the two extreme cases. Hence, this figure strongly suggests that the results observed for our synthetic topologies should also hold for the real Internet. In particular, the gain curve for the real Internet should most likely lie somewhere between the two extreme cases.

So far, we have analysed the AS-level path diversity considering one router per AS. However, a factor that can impact the path from a source to a destination is the intradomain routing policy used inside transit ASes. In [23], we also

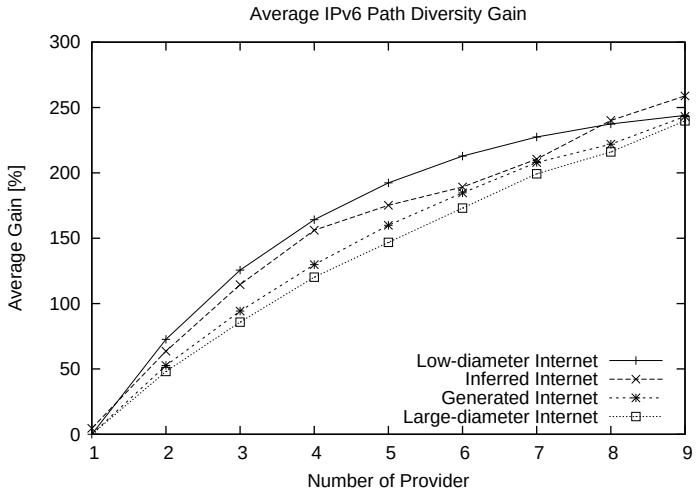

Fig. 22. Summary of path diversity gains when using multiple PA prefixes instead of a single PI prefix

evaluate the path diversity that exists when ISP routing policies in the Internet conform to hot-potato routing. In hot-potato routing, an ISP hands off traffic to a downstream ISP as quickly as possible. Results presented in [23] show that hot-potato routing has no significant impact on the AS-level path diversity.

5.5 Impact of BGP on Path Diversity

We discuss in this section how the path diversity is affected by the BGP protocol.

Multihoming is assumed to increase the number of alternative paths. However, the AS-level path diversity offered by multihoming depends on how much the interdomain routes, as distributed by BGP, overlap.

The results presented in the previous section suggest that BGP heavily reduces the path diversity, at the level of autonomous systems. Two factors can explain why the diversity is so much reduced.

The first and primary factor is that, for each destination prefix, each BGP router in the Internet receives one route from a subset of its neighbours. Based on this set of received routes, BGP selects a single best route towards the destination prefix, and next advertises this single best route to its neighbours. Therefore, each BGP router reduces the diversity of available paths. As a consequence, a single homed stub AS will receive from its provider only a single route towards each destination prefix, even if the destination site is connected to the Internet through multiple providers. Unfortunately, BGP is designed as a single path routing protocol. It is thus difficult to do better with BGP.

A second factor exists that further reduces the path diversity. The tie-breaking rule used by BGP to decide between two equivalent routes often prefers the

same next-hops. Let us consider a BGP router that receives two routes from its provider towards a destination D. According to the BGP decision process, the shortest AS path is selected. However the diameter of the current Internet is small, more or less 4 hops [2]. As a consequence, paths are often of the same length, and do not suffice to select the best path. It has been shown that between 40% and 50% of routes in core and large transit ASes are selected using tie-breaking rules of the BGP decision process [24]. In our model with one router per AS, the only tie-breaking rule used in this case is to prefer routes learned from the router with the lowest router address. This is the standard rule used by BGP-4. Unfortunately it rule yields to always prefer the same next-hop, a practice that degrades the path diversity.

The first factor suppresses paths, while the second factor increases the probability that paths overlap. An IPv6 multiaddress multihoming solution circumvents the first factor by using multiple prefixes. However, the use of multiple PA prefixes has no impact on the second factor, since it does not modify BGP and its decision process in particular.

6 Conclusion

In this paper, we have revealed that a new way to improve network performances at the interdomain level is to use multiple provider-dependent aggregatable (PA) prefixes, in an IPv6 Internet.

We have shown that stub ASes that use multiple PA prefixes can exploit paths that are otherwise unavailable. In other words, the use of multiple prefixes increases the number of paths available, i.e. the Internet path diversity. Among the newly available paths, some offer lower delays. Our simulations suggest that about 60% of the pairs of stub ASes can benefit from lower delays.

We have also proposed a new, fine-grain metric to measure the AS level path diversity. We performed simulations on various topologies to quantify the gain in path diversity when multiple prefixes are used. We have shown that a dual-homed stub AS that uses multiple PA prefixes has already a better Internet path diversity than any multihomed stub AS that uses a single provider-independent (PI) prefix, whatever its number of providers. We have observed that this gain in path diversity does not vary much with the topology considered, which suggests that the results obtained will most likely also hold for the real Internet.

Our observations show that, from a performance point of view, IPv6 multihomed stub ASes get benefits from the use of multiple PA prefixes and should use them instead of a single PI prefix as in IPv4 today. This study thus strongly encourages the IETF to pursue the development of IPv6 multihoming solutions relying on the use of multiple PA prefixes. The use of such prefixes reduces the size of the BGP routing tables, but also enables hosts to use lower delays and more diverse Internet paths, which in turn yields to larger possibilities to balance the traffic load and to support quality of service.

Acknowledgements

Cédric de Launois is supported by a grant from FRIA (Fonds pour la Formation à la Recherche dans l'Industrie et dans l'Agriculture, Belgium). Bruno Quoitin is supported by the Walloon Government within the WIST TOTEM project `http://totem.info.ucl.ac.be`. This work is also partially supported by the European Union within an E-Next project.

We thank Steve Uhlig and Marc Lobelle for their useful comments and support. We also thank the authors of [13] for providing the inferred Internet topology.

References

1. de Launois, C., Quoitin, B., Bonaventure, O.: Leveraging Network Performances with IPv6 Multihoming and Multiple Provider-Dependent Aggregatable Prefixes. In: 3rd International Workshop on QoS in Multiservice IP Networks (QoSIP 2005), Catania, Italy (February 2005)
2. Huston, G.: BGP Routing Table Analysis Reports. `http://bgp.potaroo.net` (May 2004)
3. Agarwal, S., Chuah, C.N., Katz, R.H.: OPCA: Robust interdomain policy routing and traffic control. In: Proceedings OPENARCH. (2003)
4. Stewart, J.W.: BGP4: Inter-Domain Routing in the Internet. Addison-Wesley (1999)
5. Atkinson, R., Floyd, S.: IAB Concerns and Recommendations Regarding Internet Research and Evolution. RFC 3869, IETF (August 2004)
6. Bu, T., Gao, L., Towsley, D.: On Routing Table Growth. In: Proceedings IEEE Global Internet Symposium. (2002)
7. Huitema, C., Draves, R., Bagnulo, M.: Host-Centric IPv6 Multihoming. Internet Draft (February 2004) <draft-huitema-multi6-hosts-03.txt>, work in progress.
8. Akella, A., et al.: A Measurement-Based Analysis of Multihoming. In: Proceedings ACM SIGCOMM'03. (August 2003)
9. Akella, A., et al.: A comparison of Overlay Routing and Multihoming Route Control. In: Proceedings ACM SIGCOMM'04. (August 2004)
10. Fuller, V., Li, T., Yu, J., Varadhan, K.: Classless Inter-Domain Routing (CIDR): an Address Assignment and Aggregation Strategy. RFC 1519, IETF (September 1993)
11. Huston, G.: Architectural approaches to multi-homing for IPv6. Internet Draft, IETF (October 2004) <draft-ietf-multi6-architecture-02.txt>, work in progress.
12. Gao, L.: On Inferring Autonomous System Relationships in the Internet. IEEE/ACM Transactions on Networking **vol. 9, no 6** (December 2001)
13. Subramanian, L., Agarwal, S., Rexford, J., Katz, R.H.: Characterizing the Internet Hierarchy from Multiple Vantage Points. In: Proceedings IEEE Infocom. (June 2002)
14. Medina, A., Lakhina, A., Matta, I., Byers, J.: BRITE: An Approach to Universal Topology Generation. In: Proceedings MASCOTS '01. (August 2001)
15. Jin, C., Chen, Q., Jamin, S.: Inet: Internet Topology Generator. Technical Report CSE-TR-433-00 (2000)
16. Calvert, K., Doar, M., Zegura, E.: Modeling Internet Topology. IEEE Communications Magazine (June 1997)

17. Quoitin, B.: Towards a POP-level Internet topology. http://cbgp.info.ucl.ac.be/itopo/ (August 2004)
18. Quoitin, B.: C-BGP - An efficient BGP simulator. http://cbgp.info.ucl.ac.be/ (March 2004)
19. Teixeira, R., Marzullo, K., Savage, S., Voelker, G.M.: Characterizing and Measuring Path Diversity of Internet Topologies. In: Proceedings SIGMETRICS'03. (June 2003)
20. Teixeira, R., Marzullo, K., Savage, S., Voelker, G.M.: In Search of Path Diversity in ISP Network. In: Proceedings IMC'03. (October 2003)
21. Zhang, B., Liu, R., Massey, D., Zhang, L.: Collecting the internet AS-level topology. SIGCOMM Comput. Commun. Rev. **vol. 35, no 1** (2005) 53–61
22. Barábasi, A., Albert, R.: Emergence of Scaling in Random Networks. Science **vol. 286** (October 1999) pp. 509–512
23. de Launois, C.: Leveraging Internet Path Diversity and Network Performances with IPv6 Multihoming. Research Report RR 2004-06, Université catholique de Louvain - Department of Computer Science and Engineering (August 2004) http://www.info.ucl.ac.be/people/delaunoi/diversity/.
24. Quoitin, B., Pelsser, C., Bonaventure, O., Uhlig, S.: A performance evaluation of BGP-based traffic engineering. International Journal of Network Management (Wiley) **vol. 15, no. 3** (2004)