Characterisation of AS-level Path Deviations and Multipath in Internet Routing

Eric Elena, Jean-Louis Rougier, Stefano Secci
Institut Telecom, Telecom ParisTech, LTCI CNRS, France.
E-mail: {elena, rougier, secci}@telecom-paristech.fr

***Abstract*—Although significant efforts have been devoted in the literature to measurements of the Internet topology, little attention has been given to the qualification of routing deviations and multipath dynamics at the Autonomous System (AS) routing level. We observed Internet AS-level routes toward thousands of destinations for several weeks between 2009 and 2010 to have a better understanding of these phenomena, nowadays. Using a modified form of traceroute at the state of the art called paris-traceroute, able to detect load-balancing, we sampled thousands of AS-level routes for several weeks. By an extensive analysis, we found that between 15% and 17% of the monitored destinations experiences AS-path deviations, and that load balancing at the AS-level (i.e., forms of BGP multipath) is not widespread at all, and, when frequently used, it is limited to a sort of AS confederation.**

***Index Terms*—routing stability, Internet measurement, multipath routing**

I. Introduction

The Internet is nowadays an interconnection of more than thirty thousand Autonomous System (AS) networks. Each AS uses the Border Gateway Protocol (BGP) to exchange routing information with its neighbours and to implement interconnection policies. In the corresponding AS-level graph, a node represents an AS and a link one or several Border Gateway Protocol (BGP) sessions between two ASs.

The Internet topology is not known in practice as the information available in Internet registries about providers' relationships is incomplete and not up to date. Many works have thus been conducted in the past decade to infer the Internet graph structure. There are basically two methodologies to draw the Internet AS-level graph: either you passively monitor one or more backbone BGP routing tables, or you actively discover the topology using traceroute-like tools. In both cases, the visibility is truncated and only reflects the point of view of the router or the monitor. Traceroute-like measurements are usually less reliable but allow to cover a wider scope (as the access to BGP tables is not always available). Characterizing the Internet AS-level graph dynamics is an interesting open challenge that can help to better understand the flaws of the current Internet architecture.

The assessment of how much route deviations[1] affect Internet routing can qualify the stability of the current inter-domain routing protocol and may also qualify the level of path diversity in the Internet core. On the other hand, the assessment of how much AS-level multipath routing is used today can give an index on the willingness of AS carriers to migrate toward multipath routed future Internet architectures [4].

The Internet topology dynamics has been widely studied over the last few years. A survey on relevant achievements in this area can be found in [1] and [20]. Some work addressed the characterisation of AS-level and router-level Internet routing graphs using traceroute-like analysis; however, it is now commonly known that traceroute analysis can be strongly biased by load balancing [2]. To overcome these and also other anomalies, a tool called "paris-traceroute" has been developed to perform more accurate measurements [3]. Sending many probes with modified TCP, UDP and IP header fields, the tool is able to detect load-balanced paths, giving paths significantly more reliable than with traditional traceroute. However, even if the Internet graph one can build using such a tool has an acceptable pertinence, we are still incapable to obtain the real full connectivity [17] [19].

With the aim to detect route deviations and multipath routing at the AS-level, we performed extensive route sampling from the Telecom ParisTech network (AS 1712) toward thousands of destination hosts for several weeks using the paris-traceroute tool. We jointly focused our attention to route deviations and multipath routing detection because the two aspects are strictly related. Indeed, behind an AS-level route deviation there may be the announcement of a new available BGP path, then chosen either as the new best-path (deviation from a best-path to another) or as part of a BGP multipath routing solution (deviation from a best multipath solution to another multipath or mono-path one, or viceversa). We found that an important part of the Internet routes frequently deviate, and that BGP multipath is practically not used today.

The paper is structured as follows. Section II presents our measurement methodology and reminds some details on the BGP routing decision process. Section III presents the characterisation of AS-level multipath routing. Section IV presents the characterisation of AS-level route deviations. Finally, Section V concludes the paper.

II. Methodology

For our measurement campaign, we used two different destination datasets, one collecting thousands of IP addresses

[1] From now on, with "route deviation" we mean a change of the best BGP route(s) toward a given destination.

coming from a LIP6 measurement project and publicly available [6] (we call it "LIP6" from now on), and one created starting from the CAIDA ranking [9] (we call it "CAIDA" from now on). We also used the Route Views BGP routing table [7] for IP to AS Number (ASN) conversion and path cross-checking.

In order to perform multi-thread parallel measurements, the destination datasets were split into four groups. Indeed, paris-traceroute can require many seconds per destination. We sampled the Internet routes toward these destinations every 18 hours to cover four different daytimes. In total, 81 sampling rounds were executed (i.e., about 61 days). In the following, we give more details on the datasets, recall principles of BGP routing and present how we managed some routing anomalies.

A. *LIP6 dataset*

In the frame of a project on the characterisation of router-level Internet graph dynamics, the LIP6 laboratory launched a route measurement campaign towards thousands of IP hosts for some months using PlanetLab nodes [8]. They release all the collected data on the project website; those measurements have been performed using a special traceroute, called *tracetree*, that in fact traces in a compact way the route tree toward hundreds of destinations at the same time. Their destination sets are composed of randomly picked IP addresses.

It is worth mentioning in the following the rationale that conducted to the present work. In a previous work, we used those routing samples to assess the importance of AS-level route deviations [5]. In that preliminary work, we concentrated on the deviations for the routes crossing top-tier frontiers only (the CAIDA top 50 in fact), finding that those routes are particularly instable. However, since the tracetree is as inaccurate as the classical traceroute for routing analysis with respect to the load-balancing issue, we did not submit those results for publication because of the risk of excessive bias due to load balancing. Therefore, we then run our own measurements to collect pertinent data for routing analysis with paris-traceroute [3], object of this paper.

For our measurements, we select 6741 unique destination IP hosts among the destination datasets used for the measurements in [8]. These selected destinations are those among the original datasets that frequently passed through top-tier AS frontiers in the previous work [5]. It is worth underlining that some destinations of the LIP6 dataset may belong to the same AS and even to the same prefix.

B. *CAIDA dataset*

For this dataset, we concentrate on destination IP addresses that are assigned to stub ASs, i.e., ASs that are at the periphery of the Internet, that do not belong to carrier providers, and that do not have customers but only providers. We focus on stub ASs because they represent the large majority of the global Internet ASs [10]. In order to pick potential stub ASs, we use the well-known CAIDA ranking [9]; it is based on the pruning customer cone and thus the connectivity, and ranks nowadays about 34000 different ASs. To concentrate on ASs that are likely stub, we use the number of prefixes announced by an AS, as well as the equivalent announced IP address space in terms of /24 prefixes; both such parameters are given by the CAIDA ranking.

We arbitrary promote an AS as stub AS if it announces at most five different prefixes (whatever their length is), including at least a prefix with a maximum length of /24. This last is the maximum length a stub AS should advertise. After applying this filter, 23587 different ASs were remaining, and to decrease this number to a reasonable value, we keep one prefix out of three, which gives a total of 7863 different ASs. The next step is to select a destination IP address for each AS. The mapping is done using the Route Views routing table; if an AS announces more than a prefix, the last prefix seen in the routing table is kept, and the first available IP address is selected as destination host for the AS.

After this procedure, we obtain a set of 7863 IP destinations all belonging to different ASs. Unlike the LIP6 data-set, destinations belonging to the CAIDA data-set were not filtered based on the crossing of a top-tier frontier. As shown in Table I, if the number of monitored destinations is similar for the CAIDA and the LIP6 datasets, there is a huge difference in the corresponding number of destination ASs. For CAIDA dataset, the number of different destinations corresponds to the number of different ASs. For the LIP6 dataset, there are 6741 different destinations for 1,100 different ASs. As a consequence, an AS has on average more than 6 destinations; however, they may belong to different prefixes. If all the destinations are different (14604), there is overlapping for 92 ASs, with "only" 8871 different Ass.

	nb. of different IP hosts	nb. of different ASs
Total	14,604	8,871 (61%)
CAIDA	7,863	7,863 (100%)
LIP6	6,741	1,100 (16%)

TABLE I
CHARACTERISATION OF THE MONITORED DESTINATIONS

C. *BGP, route selection and multipath*

We briefly remind the BGP decision process. Each BGP router announces its IP network addresses to its neighbours. A BGP router may receive multiple announces towards the same destination network. Before forwarding the received announces to its neighbours, the router applies a list a criteria to select a single best path. The first one is the "local preference", which indicates the preferred egress path. This policy is mainly guided by economic issues. The other criteria are guided by operational network issues, such as the smallest AS hop count, the late exit criterion for routes toward the same downstream provider (based on the Multi-Exit Discriminator attribute), the early exit criterion (or hot potato), and the tie-breaking criterion choosing the routes by the router with the smallest IP address [15].

When some of the more priority BGP decision criteria are equivalent, the load might be balanced on the equivalent routes. However, by default only one route is retained, and some inefficient rules such as tie-breaking are often the ones used to take a decision. Otherwise, one would have forms of *BGP multipath*, which have been discussed in standardization fora, but finally not standardised; however, some recommendations have been published [11], and some vendors now implement it (see, e.g., [12] and [13]).

For our measurement analysis, we have been using the same snapshot of a BGP routing table from Route Views 3, issued at November 12, 2009 at 12:00 (UTC), in particular to map IP network addresses to ASNs. Sometimes, a prefix inserted in the routing table is just an aggregation of smaller prefixes [14]; when this happens, since it is not possible to know exactly which AS is announcing it, to complete the routing table we used some looking glasses (a looking glass is a web interface allowing a user to run some commands on a router) and some whois servers. In this way, we were able to translate a router-level paris-traceroute in an AS-level path.

D. Anomaly management

Some cases have not been taken into account for the analysis because an anomaly was detected. Since our goal is to use data with the highest degree of confidence, we preferred removing those paris-traceroute data that looked like incorrect to allow for an accurate statistical characterisation. Namely, we discarded the data for which the IP2ASN mapping and the observed destination ASN were different, some routes for which AS-level routing loops appeared, some traces for which some IP2ASN mapping could not be solved. Finally, when an answer from a router could not be obtained, and the previous and the next routers belong to different ASs, we discarded the trace. It is also possible, however, that some valid data were discarded.

III. AS-LEVEL MULTIPATH CHARACTERISATION

We present in this section our characterisation results about AS-level multipath. All in all, among all the 14,604 different destinations, we discovered multipath routing towards only 70 of them. Grouping these by their AS number, 19 different ASs remain. To characterise the AS-level multipath routing for this little number of destinations, we observe two performance factors:

- width: the number of different paths used to reach the destination;
- delta: the AS-hop difference between the longest path and the shortest path.

All the multipath routes had a width equal to 2. Moreover, for 17 destination ASs, the routes had a delta equal to 1, while the remaining two ASs had routes with a delta equal to 0. This means that, for all the cases, there were two different paths to reach the destination and that, for most of the cases, a path was longer than the other path. For the routes of the remaining cases, BGP multipath was used.

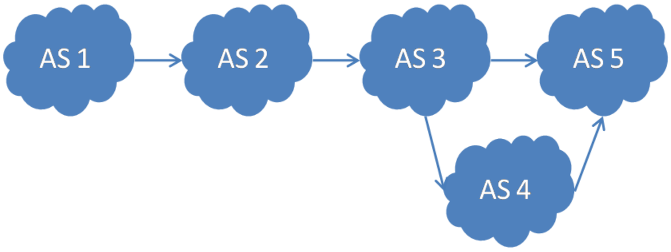

Fig. 1. Typical example of detected multipath topology.

Figure 1 shows a typical example of topology where multipath routing is detected. Out of the 17 ASs with multipath routes and with a delta equal to 1, 13 match exactly this multipath topology, and the 4 remaining have a very similar topology where AS4 is located between AS2 and AS3, instead of AS3 and AS4 (hence AS-level load balancing appears close to the destination). However, this routing configuration should not happen because not allowed in BGP multipath (we have AS paths with different lengths, and BGP multipath is at least executed on equal-length AS path; it is worth noting that in these cases AS path prepending was not used as observed cross-checking with the BGP routing table).

Figure 2 reports the frequency of multipath occurrence for the destinations, during the whole observation period, in terms of number of rounds with multipath. For better pointing out the behaviour, we split the destination set into two groups, a first one for which multipath occurs sporadically, and a second one for which it is practically permanent.

For the first group, the sporadic occurrence of BGP multipath routing can be due to a low routing visibility of some BGP routers (i.e., the routers have not received many route alternatives for the same destination prefix). Indeed, even with enabled multipath mode, a router may seldom receives multiple routes for the same destination network prefix.

For the second group, multipath routing is instead more steady and appears for the large majority of the observation period. However, as above mentioned, the multipath topology is as in Figure 1 and should not allow a native execution of BGP multipath because the available paths are not equivalent. Looking more precisely into these situations, for each of the ASs involved in these cases, we checked the administrative owner via whois requests. We found that always the ASs involved in the multipath branch either had the same registered name or belonged to the same AS provider or AS company (stub AS). Those ASs were thus functionally acting as a single AS, which can also be enforced technically using AS confederations [16]. Technically, it is worth citing two configurations that could have been the cause of these phenomena.

- Multi-Protocol Label Switching (MPLS) is used in the AS core network, and an MPLS load-balancer is used within the branching AS, toward the destination AS and an intermediate transit AS.
- A shared prefix is used between two adjacent ASs, on the

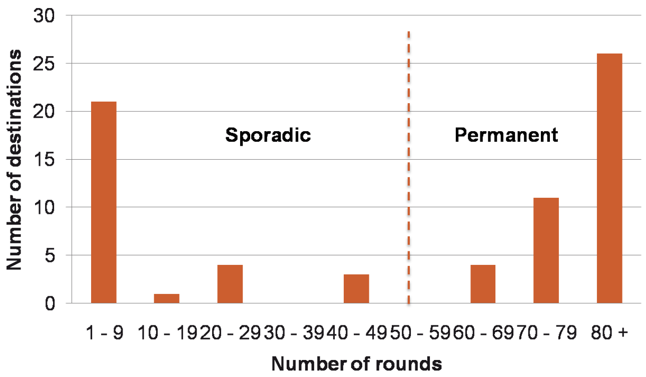

Fig. 2. Per-destination frequency of multipath occurrence

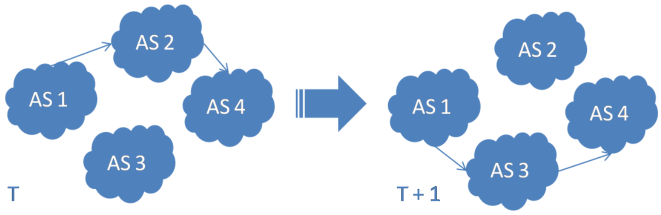

Fig. 3. Example of an AS-level route deviation

way toward a third AS. This typically happens between a customer AS with a large address space partially unused and a provider AS, or between an AS and an Internet eXchange Point (IXP) that has an ASN assigned. Our cases would probably fall into the last example class, where an IXP AS answered at a single router-hop to the paris-traceroute using an interface addressed with an IP address belonging to its AS customer. These pitfalls about traceroute artefacts are detailed in [17] and [18].

To summarize, multipath in inter-AS routing is a phenomenon that we could observe as significant only for ASs belonging to the same company or AS confederation, in a form that does not seem related to an implementation of BGP multipath mode. Inter-AS BGP multipath routing appears very seldom for a few ASs. Moreover, these AS have likely a low routing visibility.

IV. AS-LEVEL ROUTE DEVIATION CHARACTERISATION

AS-level route deviations (see, e.g., Figure 3) typically happen when a better AS-path than the one currently used for a destination is received by a BGP router, or when the currently used best path is withdrawn or is no longer reachable and an alternative one in the local routing information base is used.

In our measurement campaign, it is worth mentioning that Telecom ParisTech (AS 1712) is multi-homed with two AS providers. For this reason, each provider has then been considered as an independent source of the paris-traceroute. Thus, we monitored 14604 destinations, but virtually they produced 29208 different paths during the observation period (surely, with irregular sampling steps). Among all these destinations, for none of them only one provider was used during the observation period because of periodical default egress rerouting of AS 1712.

		Monitored destinations	with path deviations
Whole set	Mixed	29,208	4,150 (14%)
	Unique	0	2,932 (71%)
CAIDA	Mixed	15,726 (54%)	2,605 (17%)
	Unique	0	1,663 (64%)
LIP6	Mixed	13,482 (46%)	1,545 (11%)
	Unique	0	1,269 (82%)

TABLE II
REPARTITION OF THE AS-LEVEL ROUTE DEVIATIONS

Out of the 29208 monitored paths, 14% of them faced at least an AS-level route deviation during the observation period. There are 71% of the paths which faced at least a deviation by using only one of the providers. For the remaining 29%, deviations occur for both provider sources. Because of our distinction, there are 609 destinations which experienced an AS-path deviation by using the two providers during the observation. According to these numbers, we can assume that most of the deviations occur far from the destination. Table II reports more precise numbers, differentiating between the LIP6 and the CAIDA datasets. A "mixed destination" in the table is a destination which deviates by using both providers. A "unique destination" is instead a destination which deviates by using only one provider. As expected, we observe more deviations for CAIDA dataset (17%) compared to LIP6 dataset (11%), which is mainly due to the fact that the CAIDA dataset covers a higher number of unique ASs. On the other hand, the most part of the deviating paths of the LIP6 dataset is due to only one provider, while this number drops to 64% for the paths of the CAIDA dataset.

It is possible to analyse the AS-level route deviations looking at how much AS-level path diversity has been observed toward a destination during the whole observation period. At the end of the measurements, we can say that the observed paths toward each destination during all the observation have traced a diamond topology that is meaningful to analyze. Such AS-level deviation diamonds normally assume a regular shape, with a divergence point followed by a convergence point a few AS-hops later. To characterise the AS-level deviation diamonds, we used the same two metrics presented in the previous section for the multipath analysis, i.e., the *width* (number of different paths used to reach the destination) and the *delta* (the difference in terms of AS-hop between the longest path and the shortest path). The delta deviation can be positive, negative or null, depending on the path becoming longer (increase), shorter (decrease) or keeping a constant AS-level length. To enrich a bit the path diversity of the diamonds, we also included in this analysis the detected multipath routes (however, we just considered if there was multipath or not,

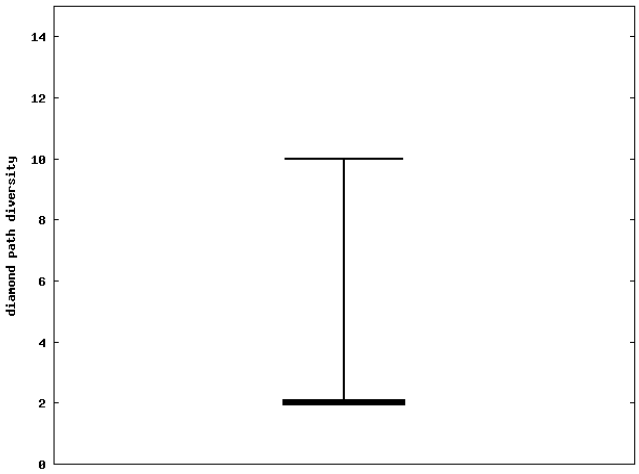

Fig. 4. Path diversity boxplot statistics of AS-level route deviation diamonds

without trying to characterise the way multipath could have interacted with the deviations).

The statistical behaviour of the width and delta parameters of the AS-level route deviation diamonds is reported in Figure 4 and Figure 5, with boxplot statistics (with minimum, lower quartile, median, higher quartile and maximum). Note that for Figure 5 for all the boxes the median (the line in bold) coincides with the first quartile and the minimum). For Figure 4, also the third quartile coincides with the first quartile, the median and the minimum. We can assess that:

- most of the paths which deviate during our observations have a width equal to 2 (Figure 4) and faced a single deviation (Figure 5). This means that during more than 70 days, we used only two different paths to reach more than half of the destinations for which we observed an AS-level deviations (obviously, with only a deviation, there can be only two different paths).
- looking more carefully at the boxplot statistics of the width (Figure 4), at least 75% of all the deviating paths generate a width equal to 2. For a path with more than a deviation, it results in a temporary use of a spare path while the main path is unreachable.

Figure 6 gives an insight on the number of deviations for each observed path. The distribution shows the number of paths having the same number of observed deviations. We can observe that a significant number of the deviating paths suffer from a number of deviations between 2 and 10.

Globally, about 15% of the destinations we monitored faced at least a deviation during our observations, and a significant number frequently deviates. This value is higher (17%) for the CAIDA dataset than for the LIP6 dataset, mainly because the LIP6 destinations belong to a fewer number of ASs, this value is lower (11%).

An important factor behind this significant percentage of AS-level route deviations likely descend from the choice of the destination sets; the LIP6 destinations belong to routes that pass through top-tier AS borders, while the CAIDA destinations belong to stub ASs at the very border of the Internet. Both types of datasets have been chosen because are good candidates to suffer from AS-level deviations. The first one is a good candidate because it is a current belief that top-tier borders, which likely are (have been, or will become) peering interconnections, are now representing the real bottleneck of the Internet due to lack of coordination among peering carriers (see [21] where the authors propose a peering management framework to overcome this issue). The second one, the CAIDA dataset, is also a good candidate because its destinations are reachable with AS paths longer than the average AS path lenght (the source of the traces is a stub AS too, the AS 1712) and because operationally speaking Internet carriers may not be interested in stabilizing stub routes; if they could implement such traffic management procedures, they would probably prefer stabilizing first the routes toward AS content providers and AS eyeball providers (with a lot of Internet users).

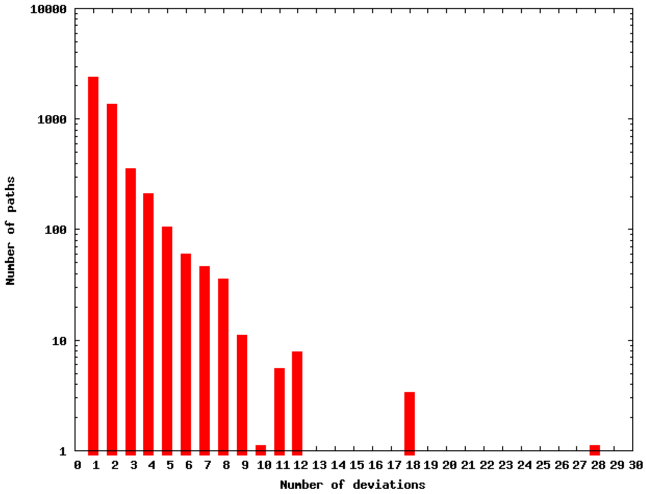

Fig. 6. Number of paths as function of the number of AS-path deviations

V. Summary and conclusions

We performed for several weeks between November 2009 and February 2010 a routing measurement campaign. We sampled the routes toward thousands of Internet destinations using the paris-traceroute tool, which is able to detect load balanced paths. The objective of our measurement campaign was to characterise the occurrence of multipath routing and of route deviations in Internet routing at the Autonomous System (AS)-level.

We performed the measurements toward two different datasets, one containing thousands of destination IP hosts for which the observed route passed by a top-tier inter-carrier frontier, and one containing IP addresses belonging to stub ASs only.

All in all, our analysis clearly tells that BGP multipath is practically not used today in Internet routing and appears only

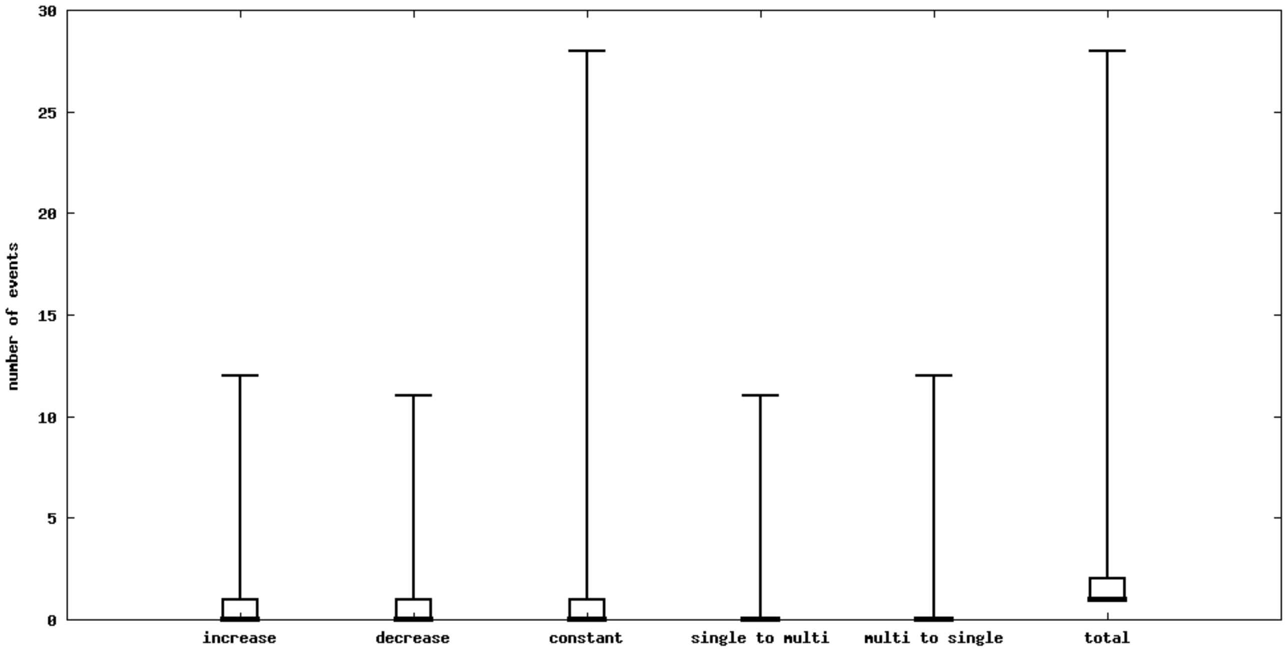

Fig. 5. Boxplot statistics of the delta variation of AS-level route deviation diamonds

seldom in very specific cases. Moreover, between 15% and 17% of the monitored destinations suffer from AS-level route deviations. This performance may be seen as a symptom of a lack of path diversity in the Internet core; otherwise, it may be seen as the result of lack of routing coordination in the Internet core. Many research activities on Internet routing insist solely on one of these aspects as a possible field of improvement for future Internet protocols, while probably a mix of them is the cause of the observable Internet routing instability.

Would be good to have multipath routing for future Internet? Surely, a larger path diversity in Internet routing is a desirable feature (which could be implemented quite easily at different extents; see, e.g., [22][23]). With more available paths, the Internet reliability can be increased; if AS-level multipath routing were broadly used, it could take benefit from a larger path diversity. However, BGP route deviations are already a critical phenomenon, and shall be better controlled, instead than increasing, in the future, as it could lead to bad end-to-end performance. From our measurement analysis, one warning that emerges is that there is, and there will be, the need for a carefully coordinated multipath routing decision. Further work shall try to define intelligent and rational ways to fine-select multipath routes in a multi-lateral (multi-router or multi-AS) coordinated fashion.

Acknowledgments

The authors would like to thank the student Matteo Marinoni, Guido Maier and Achille Pattavina for their attentive feedbacks on the contents of this study and for their previous contributions on this research activity. The authors also thank Clemence Magnien and Matthieu Latapy from the LIP6 for releasing their radar datasets and captures.

This work was partially funded by the European FP7 Network of Excellence EuroNF ("Anticipating the Network of the Future - From Theory to Design") and by the FP7 IP European Project ETICS (Economies and Technologies for Inter-Carrier Services).

References

[1] B. Donnet, T. Friedman, "Internet topology discovery: a survey", *IEEE Communications Surveys and Tutorials*, Vol. 9, No. 4, Pp:56-69 (2007).

[2] B. Augustin, T. Friedman, R. Teixeira, "Measuring Load-balanced Paths in the Internet", in *Proc. of IMC 2007.*

[3] B. Augustin, X. Cuvellier, B. Orgogozo, F. Viger, T. Friedman, M. Latapy, C. Magnien, R. Teixeira, "Avoiding Traceroute Anomalies with Paris Traceroute", in *Proc. of IMC 2006.*

[4] J. He, J. Rexford, "Towards Internet-wide multipath routing" in *IEEE Network magazine*, March 2008.

[5] S. Secci, J.-L. Rougier, A. Pattavina, G. Maier, M. Marinoni, E. Elena, "Detection of BGP route deflections across top-tier interconnections", ENST res. report nb 9287-2009, http://www.tsi.enst.fr/publications/enst/techreport-2009-9287.pdf

[6] LIP6 data, http://data.complexnetworks.fr/Radar/.

[7] Route Views, http://www.routeviews.org/.

[8] M. Latapy, C. Magnien, F. Ouédraogo, "A Radar for the Internet", in *Proc. of ADN 2008.*

[9] CAIDA, http://www.caida.org.

[10] The CIDR report, www.cidr-report.org

[11] A. Lange, "Issues in Revising BGP-4", draft-ietf-idr-bgp-issues (2003)

[12] "Configuring BGP to Select Multiple BGP Paths", JUNOS document.

[13] "BGP Best Path Selection Algorithm", Cisco documentation.

[14] "Understanding Route Aggregation in BGP", Cisco documentation.

[15] "BGP Best Path Selection Algorithm", Cisco documentation.

[16] P. Traina, "Autonomous System Confederations for BGP", RFC 1965.

[17] Y. Zhan et al., "Quantifying the Pitfalls of Traceroute in AS Connectivity Inference", in *Proc. of PAM 2010.*

[18] M. Crovella, B. Krishnamurthy, *Internet measurement: infrastructure, traffic and applications*, 2006, John Wiley & Sons, Inc. New York, NY, USA.

[19] R. Oliveira et al., "The (in)Completeness of the Observed Internet AS-level Structure", in *IEEE/ACM Transactions on Networking*, 2010 (to appear).

[20] R. Oliveira, B. Zhang and L. Zhang "Observing the Evolution of Internet AS Topology" in *Proc. of SIGCOMM 2007*.

[21] S. Secci, J.-L. Rougier, A. Pattavina, F. Patrone, G. Maier, "PEMP: Peering Equilibrium MultiPath routing", in *Proc. of 2009 IEEE Global Communications Conference (GLOBECOM 2009)*, 30 Nov. - 4 Dec. 2009, Honolulu, USA.

[22] V. Van den Schrieck, P. Francois, O. Bonaventure, "BGP Add-Paths : The Scaling/Performance Tradeoffs", *IEEE Journal on Selected Areas in Communications*, 2010 (to appear).

[23] V. Van den Schrieck, P. Francois, "Analysis of paths selection modes for Add-Paths", Internet draft draft-vvds-add-paths-analysis-00, July 2009.