BGP Table Fragmentation: What & Who?

Julien Gamba[1], Romain Fontugne[2], Cristel Pelsser[1], Randy Bush[2] et Emile Aben[3]

[1] *University of Strasbourg*
[2] *IIJ Research Lab, Tokyo*
[3] *RIPE NCC, Amsterdam*

BGP routing table growth is one of the major Internet scaling issues, and prefix deaggregation is thought to be a major contributor to table growth. In this work we quantify the fragmentation of the routing table by the type of IP prefix. We observe that the proportion of deaggregated prefixes has quasi doubled in the last fifteen years. Our study also shows that the deaggregated prefixes are the least stable; they appear and disappear more frequently. While we can not see significant differences in path prepending between the categories, deaggregated prefixes do tend to be announced more selectively, indicating traffic engineering. We find cases where lonely prefixes are actually deaggregation in disguise. Indeed, some large transit ISPs advertise many lonely prefixes when they own the covering prefix. We show the extents of this practice that has a negative impact on the routing table even though it could usually be avoided.

Mots-clefs : Border Gateway Protocol, BGP, Internet scaling, Routing table size

1 Introduction

BGP routing table growth is a major Internet scaling issue. While some prefixes are announced for traffic engineering (TE), prefix hijacking protection, or traffic blackholing, they are often not necessary for reachability. These prefixes put a burden on the overall routing system by providing additional features to a few players. We aim to quantify their presence and highlight their purpose. We would like to understand the problem space to be able to design appropriate solutions for these features and provide an up-to-date perspective on the magnitude of the problem.

We use data from the Level3 (AS3356) BGP feed into the Route Views project [4], a full feed from a large transit provider. We found no notable differences among Route Views or RIS [3] BGP peers for our purpose, so consider this feed representative.

We classify prefixes using the taxonomy in [6]:

- **lonely**: a prefix that does not overlap with any other prefix;
- **top**: a prefix that covers at least one smaller prefix but is not itself covered;
- **deaggregated**: a prefix covered by another less specific prefix originated by the same AS;
- **delegated**: a prefix covered by another less specific prefix originated by a different AS.

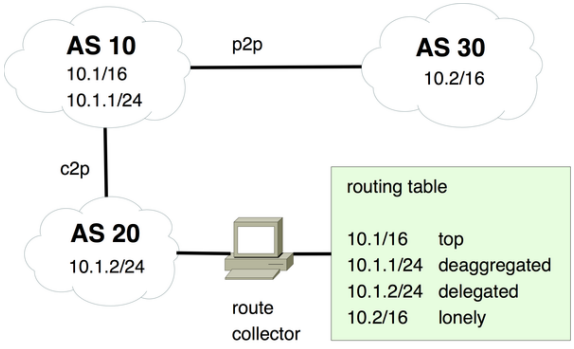

Figure 1: BGP prefixes classification, taken from [6]

Figure 1 shows an example of this classification. In this configuration, we have three ASes. A route collector gathers all prefixes announced by AS 20 and classifies them. Here AS 20 hands over a customer view to the monitor. 10.2/16 is not overlapping with any other prefix, so it is classified as lonely. However, 10.1/16 is covering two other prefixes: it is classified as top. 10.1.1/24 is announced by the same AS as its covering prefix (AS 10), so it is deaggregated. Finally, 10.1.2/24 is covered by 10.1/16, but it is announced by another AS (AS 20): the prefix is delegated.

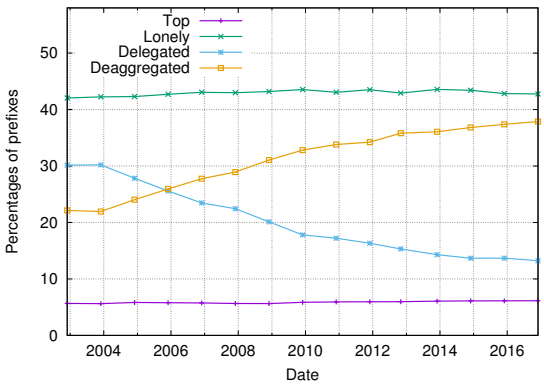

Figure 2: Prefix classification over time

We classify the data from the last fifteen years. As shown in figure 2, the proportion of deaggregated prefixes increases from 22% in 2003 to 38% in 2017, while the proportion of delegated prefixes declines from 30% to 13%. Top and lonely prefixes are stable, at 42% and 6% respectively. Recently, the rate of increase of deaggregated prefixes is not as steep as observed previously [6]. As the routing table increases, the absolute number of deaggregated prefixes increases but the proportion of these prefixes does not increase as much. This slow down is also visible in the rate of decrease of delegated prefixes.

2 Is it Traffic Engineering?

We look at AS path prepending and at selective announcement, as these are signs of traffic engineering. Here, we use data from all the Route Views route collector peers. We see a 3% increase in popularity of AS path prepending: going from 9% of all prefixes in 2001 to 12% in 2016. We note that this increase of AS-path prepending is not specific to deaggregated or lonely prefixes; it occurs for all classes of prefixes.

To detect selective announcements we count, given a prefix, how many peers see it. We consider a prefix to be selectively announced if it is seen by less than half of the route collector peers. Figure 3 shows our results. It appears that lonely and deaggregated prefixes follow the same trend, and are the most selectively announced. In 2016, 13% were selectively announced, where the proportion was 5% for delegated prefixes and 1% for top prefixes. We conclude that not only deaggregated but also lonely prefixes are used for traffic engineering.

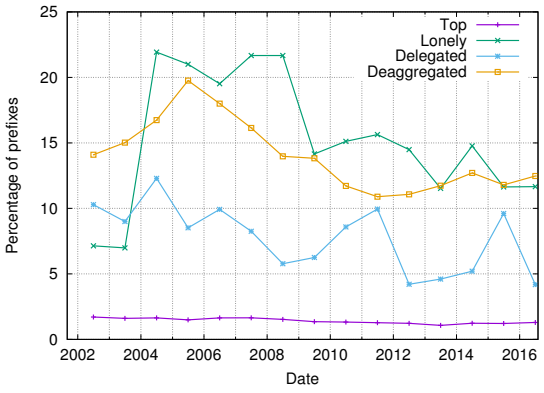

Figure 3: Proportion of selectively announced prefixes by category over time

3 Who is engineering traffic?

To see what kinds of ASes use TE, we first select the tier ones ASes using CAIDA's AS ranking [2] and select the known content providers based on the presence of keywords in their name. We extract 30 large transit providers and 137 content providers out of 53731 ASes. The complete list of keywords is available online for review [1].

The large transit providers announce 2.33% of the prefixes: these prefixes correspond to 9.97% of the address space. Similarly, the content providers announce 1.25% of the prefixes, which correspond to 0.67% of the address space.

As shown in figure 4, we find that large transit providers (followed by content providers) tend to announce more deaggregated prefixes. 45.22% of all prefixes announced by large transit providers are deaggregated. To a smaller extent, they also advertise more lonely prefixes than other AS types: 28.73%. Large transit providers may be splitting their address space into smaller prefixes to do traffic engineering or segregate PoPs.

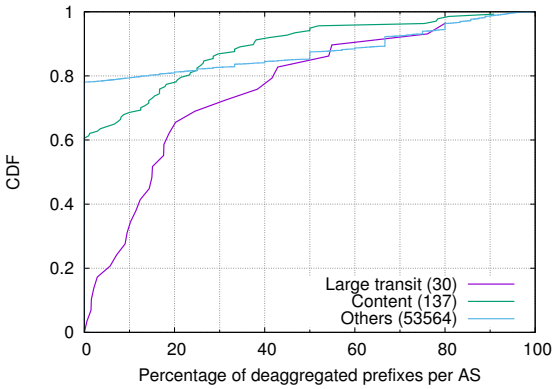

Figure 4: Distribution of the percentage of deaggregated prefixes per AS

4 Hidden Deaggregation

We also try to detect prefixes moving between categories between two months. We observe a significant number of lonely prefixes becoming deaggregated and vice versa. We believe that these movements happen when an AS which announces multiple lonely prefixes starts announcing a large covering prefix, and the reverse. This hints that our original postulate for deaggregated prefixes may also apply to the class of lonely prefixes. Some ASes may hold a large prefix but announce it only in smaller pieces. Why would ASes do that? Likely for the same reasons as the advertisement of deaggregated prefixes; we saw that both types of prefixes are selectively announced to the same extent. In the case of lonely prefixes, operators may assume sufficient redundancy not to need the advertisement of the covering prefix. The aggregation of the prefixes from these two classes could help reduce the size of the routing table. This is a new finding, [6] does not consider lonely prefixes to be aggregatable.

5 Reducing the routing table size

We saw in Section 2 that both lonely and deaggregated prefixes are used to do traffic engineering. In this section we want to estimate to what extent these prefixes could be aggregated. We consider two prefixes to be aggregatable if (1) they have the same AS origin, (2) they are consecutive and (3) the aggregate falls on a power of two boundary. This lets us estimate the level of pollution of the global routing table. This estimate does not take into account that some content providers may split their address space because they lack of a proper backbone. In this case, replacing the prefixes by an aggregate would not be an option as traffic may not reach the proper geographic location.

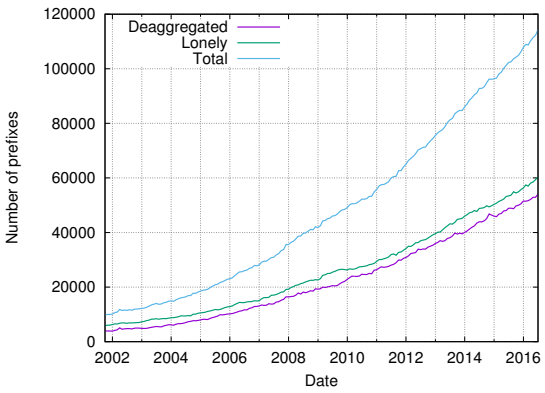

Figure 5: Amount of aggregatable prefixes

Figure 5 shows the lonely and deaggregated prefixes that could be removed from the Level3 feed after aggregation. This amount is increasing: going from 9.5% aggregatable prefixes in 2001 to 19.5% in 2016. Clearly, there would be a real benefit to aggregate lonely and deaggregated prefixes as much as possible. The gain is less than in [9] but our criteria allow one to aggregate prefixes and still be able to do route origin validation [5].

6 Related work and conclusion

Others have studied the extent of pollution of the routing table. In this work we extend [6] and study the state of deaggregation from 2001 to 2016. In [8] Michaelson et al. compute the required allocation if

prefixes could be redistributed optimally. Dragon [9] is a solution to dynamically filter the deaggregated as well as to aggregate prefixes conditionally while ensure BGP convergence and correctness. Krenc and Feldmann [7] looking at only delegated prefixes, find both provider-to-customer delegation and the reverse.

In this work, we observe that the proportion of deaggregated prefixes has increased over the last fifteen years. While we see only 3% growth in path prepending, deaggregated and lonely prefixes tend to be announced more selectively, indicating traffic engineering. Our work extends the results from [6] by a decade. We also find cases where lonely prefixes are actually deaggregation in disguise.

Traffic engineering is not solely used by stub ASes. We found that some large transit ASes heavily fragment their address space. Aggregating these prefixes would reduce the routing table size by roughly 20%. The key question is which traffic do they try to control? If it concerns the traffic from a few ASes, one could imagine TE being negotiated separately, outside the routing protocol.

References

[1] ASs business type classification.
`https://github.com/romain-fontugne/ASclassification/`.

[2] CAIDA's ranking of autonomous systems.
`http://as-rank.caida.org/`.

[3] RIPE NCC routing information service.
`https://www.ripe.net/analyse/internet-measurements/routing-information-service-ris/`.

[4] University of oregon route views project.
`http://www.routeviews.org/`.

[5] Why origin validation can't help DRAGON routing.
`https://archive.psg.com/150705.why-no-dragon.html`.

[6] L. Cittadini, W. Mahlbauer, S. Uhlig, R. Bush, P. François, and O. Maennel. Evolution of internet address space deaggregation: Myths and reality. *IEEE journal on selected areas in communications*, 2010.

[7] T. Krenc and A. Feldmann. BGP prefix delegations - a deep dive. In *IMC 2016*, 2016.

[8] G. Michaelson, E. Aben, and R. Bush. Reducing the BGP table size - a fairy tale.
`https://labs.ripe.net/Members/ggm/reducing-the-bgp-table-size-a-fairy-tale`.

[9] J. L. Sobrinho, L. Vanbever, F. Le, and J. Rexford. Distributed route aggregation on the global network. *ACM CoNEXT 2014*, 2014.