Geoff Huston
June 2017

More Specifics in BGP

The Border Gateway Protocol (BGP) is the routing protocol that literally keeps the Internet glued together. The public Internet is composed of some 58,000 component networks (BGP calls them "Autonomous Systems" (AS's)), many of which are very small, while some are very large both in terms of geographical coverage and numbers of users. BGP is used to bind these networks together in a form of a shared network map. This virtual map essentially tells each AS how to pass on a packet that is addressed to a reachable destination so that in passing it on it will get closer to its intended destination. BGP tries to do a little better than mere functional adequacy of getting "closer" to each destination. BGP tries to construct the map such that this AS level forwarding decision is the best of all possible forwarding decisions for each destination, so that the path to the intended destination is, in inter-AS terms, as short as possible, but at the same time consistent with the traffic handling policies of each component network.

> BGP is an instance of the now venerable Bellman-Ford group of distance vector distributed routing protocols. The basic principle of BGP is simple: each router tells its adjacent AS neighbours everything it knows, and each router prunes from the potential multiple ways to reach a destination by selecting what it believes is the best AS neighbour to use to reach each distinct destination. What defines "best" in BGP is also simple: BGP will select the path through an AS neighbour that has the fewest number of AS transit hops to reach the destination. A detailed description of BGP can be found at http://www.potaroo.net/ispcol/2006-05/bgp.html.

While the BGP protocol's definition of a "best" network path is one that transits a minimum number of networks, a truly "best" path from a network operations perspective or even a user's perspective is not necessarily based on the same metric. A network operator may wish to bias the path choices made by other BGP speakers to direct traffic to links with greater capacity, lower latency or lower cost, for example. One way a network operator can bias this simple protocol metric is to artificially lengthen the number of AS transit hops on undesired network paths. This practice is termed "AS prepending". Another technique uses the observation that a BGP router will always prefer to use the most specific routing entry to a packet to its intended destination. For example, if a router contains an entry for the address prefix 10.0.0.0/8 and another entry for the more specific prefix 10.0.1.0/24, then the router will always select the 10.0.1.0/24 route when handling a packet addressed to, say, 10.0.1.15 or any other address drawn from the more specific address prefix. This use of more specific routing entries can be used to direct traffic to use particular network paths.

The number of more specific advertisements in the IPv4 Internet is more than 50% of all advertisements, and the comparable picture in IPv6 has more specific advertisements approaching 40% of all network advertisements. It is tempting to label this use of more specifics as part of the trashing of the Internet commons. Individual networks optimise their position by large scale advertising of more specifics, which in turn, creates an incremental cost on all other networks in terms of increased BGP

table size and increased overhead of processing BGP updates. The question I'd like to look at here is whether these more specific advertisements represent a significant imposition on everyone else, or whether they are simply unavoidable. In other words, are more specific routing advertisements in BGP an instance of senseless routing vandalism, or do they represent the considered use of routing to ensure that users enjoy the best possible service at the lowest possible cost?

Here I'll examine the collected statistics of the use of more specific routing advertisements in BGP in further detail, looking at the impact of more specifics on the growth of the routing system and the dynamics of BGP updates in the larger context of scaling BGP to meet the demands of an ever-larger Internet.

A Taxonomy of More Specific Advertisements

Not all more specific routing advertisements are the same. Some advertisements serve the purpose of advertising reachability, announcing to other networks that a particular network, as identified by a common address prefix, is connected to the network. Other advertisements attempt to qualify this basic reachability advertisement to advise other networks of the preferred path to reach the network. In looking at more specific networks, we can devise a basic taxonomy of more specific by looking at the relationship between the more specific advertisement and its immediately enclosing aggregate advertisement.

Type I - Hole Punching

A "Hole Punching" More Specific is where the origin AS of the more specific route is different from the origin AS of the covering aggregate. This is a modification to the network topology where the intended destination network of the more specific prefix is different to the encompassing aggregate address prefix.

Here is an example, taken at random from a recent BGP routing table snapshot.

```
Network            Path
72.249.184.0/21    4777 2497 3356 36024
72.249.184.0/24    4777 2497 2914 40824 394094
```

Here the aggregate route, 72.249.184.0/21, is directing traffic to the network identified by AS 36024. However, for traffic addressed to the more specific prefix 72.249.184.0/24, packets will be passed towards the network AS 394094.

It is possible to describe this situation without using more specifics by changing the aggregate announcement to be a set of prefixes that do not cover the more specific. However, it is often the case that the result requires more prefix announcements than the rather elegant hole punching approach. The example, the situation above of these two BGP advertisements could be expressed as the following set of four non-overlapping advertisements.

```
Network            Path
72.249.184.0/24    4777 2497 2914 40824 394094
72.249.185.0/24    4777 2497 3356 36024
72.249.186.0/23    4777 2497 3356 36024
72.249.188.0/22    4777 2497 3356 36024
```

Obviously, the use of a hole punching more specific is, in routing terms, often a more efficient solution.

Type II - Traffic Engineering

A "Traffic Engineering" more specific is one where the origin AS of the more specific route and its covering aggregate are the same, while the AS paths differ. They may differ only by virtue of AS prepending or may differ in terms is different transit ASs. The intent of a traffic engineering prefix its not to change the underlying inter-AS reachability, but to bias the way in which traffic is directed to the origin AS for certain destinations.

An example, again drawn at random from the routing table, is:

```
Network          Path
1.37.0.0/16      4608 1221 4637 4775
1.37.27.0/24     4608 1221 4637 4837 4775
1.37.237.0/24    4608 1221 4637 4837 4775
```

While all traffic to any address within the /16 will be passed to the same destination network, AS 4775, traffic to addresses in the two more specific prefixes will transit AS 4837 to get there.

Again, this could be described using a set of non-overlapping advertisements, but the same observation applies, namely that the collection of non-overlapping advertisements could well be a larger set of advertisements than the use of more specifics.

Type III - Overlay

Here the more specific address prefix and the enclosing aggregate share a common AS Path and a common origin AS. For example:

```
Network          Path
1.0.4.0/22       4608 4826 38803 56203 i
1.0.4.0/24       4608 4826 38803 56203 i
1.0.5.0/24       4608 4826 38803 56203 i
1.0.6.0/24       4608 4826 38803 56203 i
1.0.7.0/24       4608 4826 38803 56203 i
```

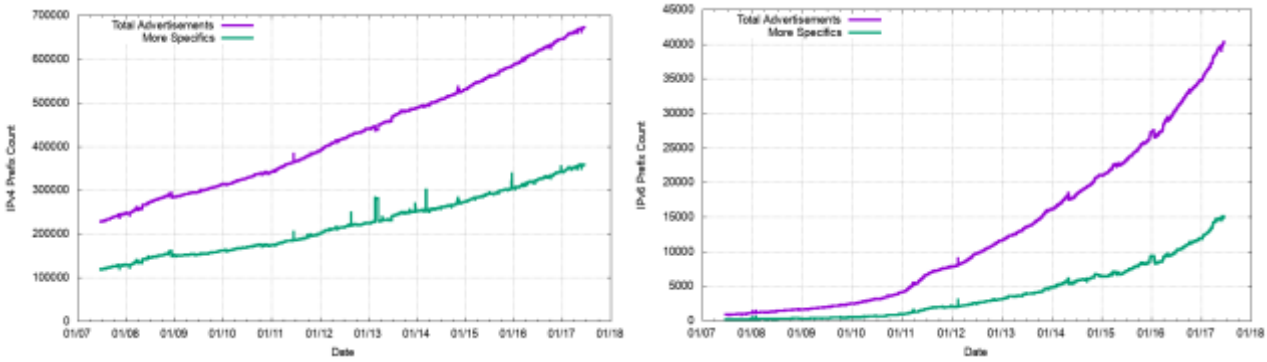
In this case, the more specific advertisement is fulfilling no function at all, as the handling of packets addressed to addresses in the aggregate prefix and the more specific are identical It is also the case that the collection of more specifics exactly span the aggregate, so in routing terms the covering aggregate could be removed with no change in functionality.

A common theory as to why network operators do this is that the more specific advertisements are intended to mitigate, to some extent, the risks of a more specific routing attack. By advertising the more specific itself, a hostile attempt to advertise a more specific would not redirect the entirety of the traffic to the attacker's site. On the other hand, the hostile advertisement would still be partially successful, so it is unclear to me what would be the actual benefit of this measure, other than some level of rather senseless routing vandalism and some rather messy occlusion of the routing attack!

We also need to consider that fact that BGP only propagates its version of the "best" route. While a remote BGP observer may only see a covering aggregate and the more specific with a common path and assume that the more specific serves no useful purpose, it conceivable that the originating network has generated a number of different advertisements for the more specific address prefix and passed them to different local peers to support a local traffic engineering outcome. What may seem somewhat pointless from a distant vantage point may not necessarily be the same for those networks close to the originating network.

More Specific Advertisements over Time

This analysis uses a 10 year data history of the eBGP routing state from June 2007 until June 2017. The BGP session used was captured from AS 131072, a stub AS located at an edge of the network. While this may differ in some level of detail from other stub networks, the picture gained from this analysis is not expected to deviate in any significant terms from that you would expect to find from any other eBGP vantage point. It is also worth noting that this is a eBGP analysis, and does not include interior BGP routes that are found in many networks.

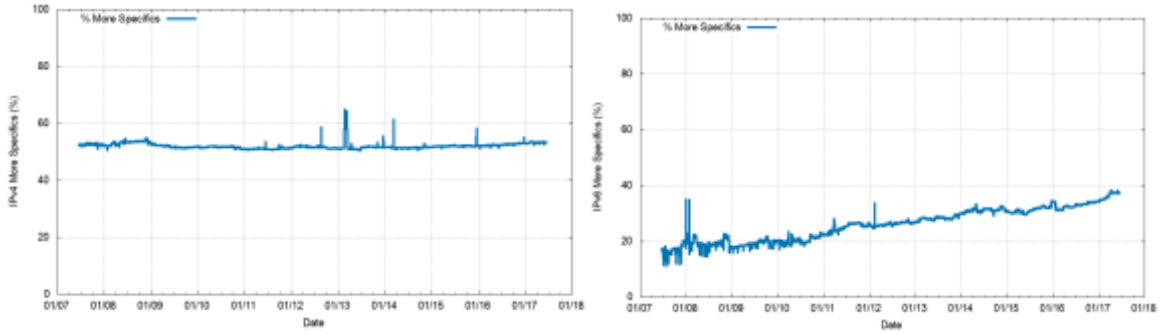


IPv4 IPv6

Figure 1 – BGP Prefix Count and More Specific Prefixes

Figure 1 shows the count of the total number of unique prefixes for both IPv4 and IPv6 in relation to the number of more specific prefix advertisements in each protocol. While both pf these plots are classic "up and to the right", the IPv6 plot on the right shows signs of accelerating growth, while the IPv4 plot shows a growth model that is only slightly higher than linear growth. The difference of course lies in the absolute scale of growth. Over the past decade, the IPv4 routing table has grown from some 220,000 prefixes to some 680,000 prefixes, while the IPv6 table has grown from 1,000 prefixes to 40,000 in the same period. In both cases the number of more specific prefixes has also grown. The question is whether this count of more specifics is growing at the same rate as the total prefix count, or whether the more specific growth rate is accelerating or slowing down.

A useful way to look at the relativity of more specifics to aggregates is to plot them as a ratio (Figure 2).



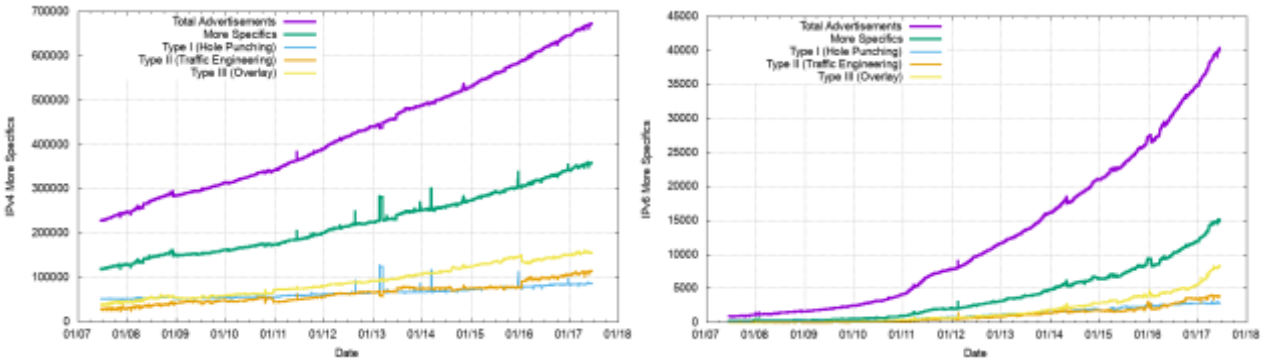
IPv4 IPv6

Figure 2 – BGP More Specifics as a % of the total Prefix Count

Figure 2 shows that for the IPv4 network, the number of more specific prefixes has been extremely stable at some 51% - 52% of the total prefix set over the entire decade. The story is somewhat different for IPv6, where the relative number of more specific prefixes has grown from some just under 20% a decade ago to the current level of close to 40% today.

Why is the IPv4 ratio so constant while the IPv6 ratio is growing? Will the IPv6 ratio stop growing when it reaches the same level as IPv4, or will it continue to grow?

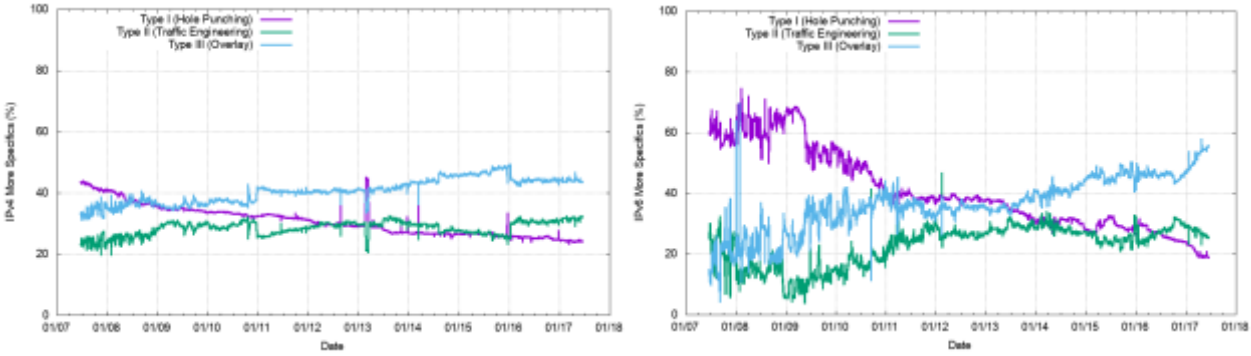
We can use the more specific taxonomy above to further categorise these more specifics, and Figure 3 shows the breakdown of the more specific pool into Types I, II and III for both IPv4 and IPv4.



IPv4 IPv6

Figure 3 – BGP More Specifics by Type category

Again, a relative view is helpful here, and this picture (Figure 4) shows some change over time, which for IPv4 is somewhat different to the picture of relative static stability of more specifics as shown in Figure 2.



IPv4 IPv6

Figure 4 – Relative Proportion BGP More Specifics by Type

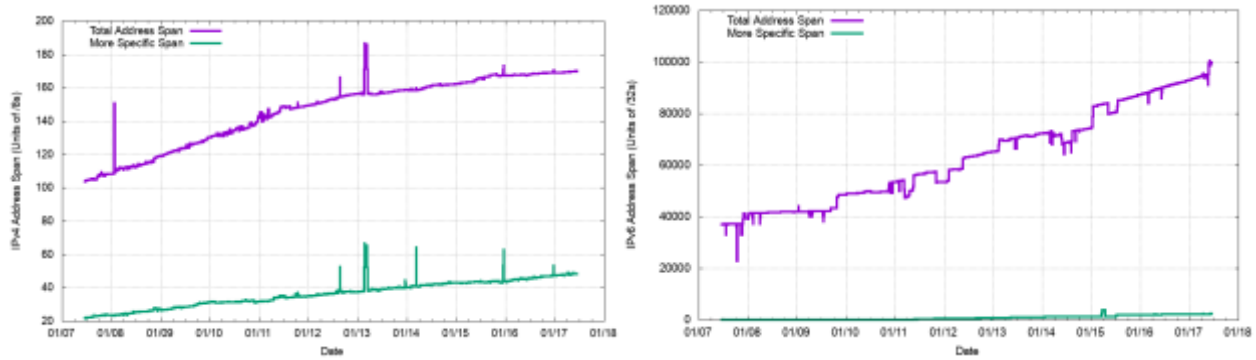
A decade ago Hole Punching more specifics were the most prevalent form of more specific advertisements, accounting for 45% of all more specifics in IPv4 and up to 70% in IPv6. This has declined over time in relative terms, and today the relative population of this form of more specifics has declined in relative terms to 25% in IPv4 and 20% in IPv6. In absolute terms the number has risen in both protocols, but the rise has been far slower than the rise of other forms of more specifics and the rise in the total table size, hence the relative decline. The relative decline of this type of more specific advertisements may be attributed to a shift in attitude on the part of service providers, who are increasingly reluctant to permit their customers to take some of the provider's address space and advertise it independently into the routing system. The common response these days is to refer the customer back to the address registry to obtain their own address space and have them advertise that independent address block, leaving the provider's address block unfragmented.

The relative number of Type II Traffic Engineering more specifics has increased in both protocols, but the overall level of change over the decade is around 6% in IPv4 and 10% in IPv6. The greater relative rise is the Type III Overlay more specifics.

The rise of the Overlay form of more specific could be attributed to the increasing awareness of the more specific routing hijack attack. If an attacker advertises more specific routes of the target prefix, all networks that see these more specifics will prefer to use them and divert all their traffic to the destination nominated by the attacker. If the site defends itself by advertising the more specific routes, then an attacker cannot usurp the entire traffic load with a more specific. But, as already noted, this is an incredibly flimsy defence, as an announcement of a competing route will still cause some level of disruption to the address holder!

Address Span

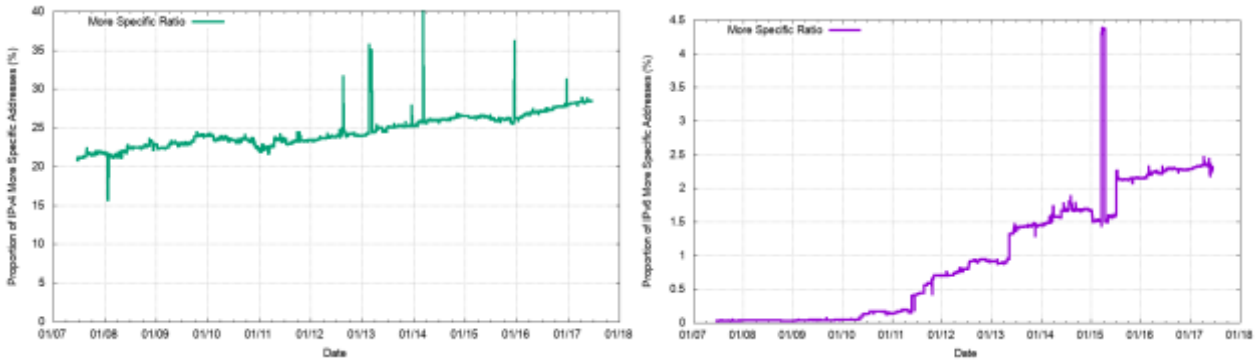
While more specifics are one half of all IPv4 advertised prefixes and two fifths of all IPv6 advertised prefixes, how much address space is covered by these more specific routing advertisements?



IPv4 IPv6

Figure 5 – Address Span of BGP More Specific

In IPv4, the total span of advertised address space is the equivalent of 170 /8s, or 2.8B /32s. More specifics encompass just 50 /8s, or less than one third of the total address pool. A decade ago this measurement was 20%, so the relative extension of addresses covered by more specifics has risen 8% over the decade.

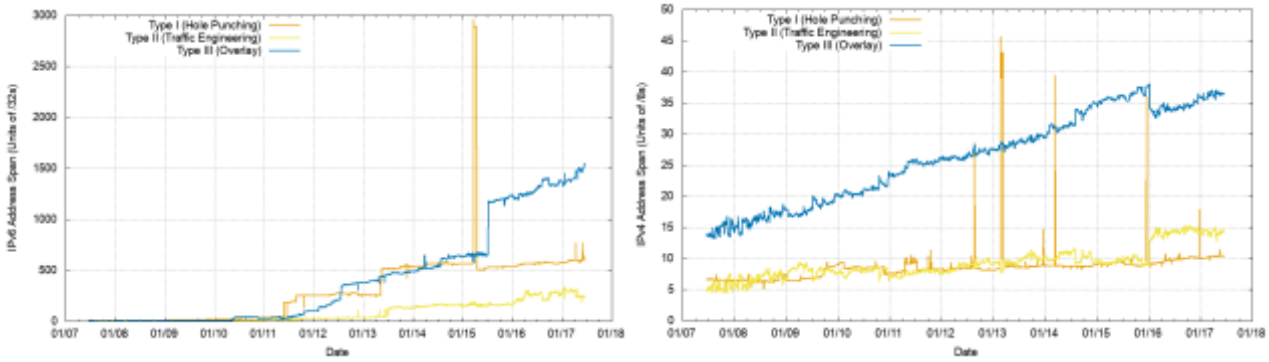


IPv4 IPv6

Figure 6 – Ratio of the Address Span of BGP More Specifics

In IPv6, more specifics encompass just some 2.5% of the total span of advertised addresses. Back in 2010 the number was essentially negligible, so the introduction of more specifics to IPv6 has occurred over the past seven years. In IPv4 the span of more specifics has increased slightly over the decade, rising from 22% to 29% of the total address span.

Figure 7 shows the breakdown of the span of more specific addresses into the types of more specifics. It is very clear in this figure that Type III Overlay more specifics span the majority of the more specific address space, covering there span of the equivalent of 35 /8s, while Type I Hole Punching is the smallest category, spanning a total of the equivalent of 10 /8s.

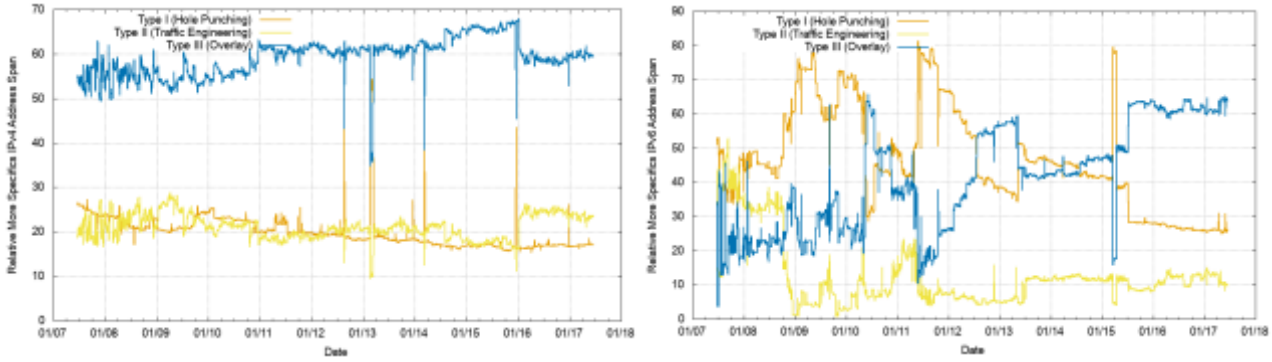


IPv4 IPv6

Figure 7 –Address Span of BGP More Specifics by Type

In IPv6 Type III more specifics, Overlays, are the largest span of more specific addresses, but in this case Type II Traffic Engineering is the smallest pool of addresses. This may reflect on the ongoing low levels of deployment of IPv6 networks and the reduced need to use the routing system to support traffic engineering at this point in time.

The relative size of these address pools as a percentage pf the total span of addresses covered by more specific routing advertisements is shown in Figure 8.

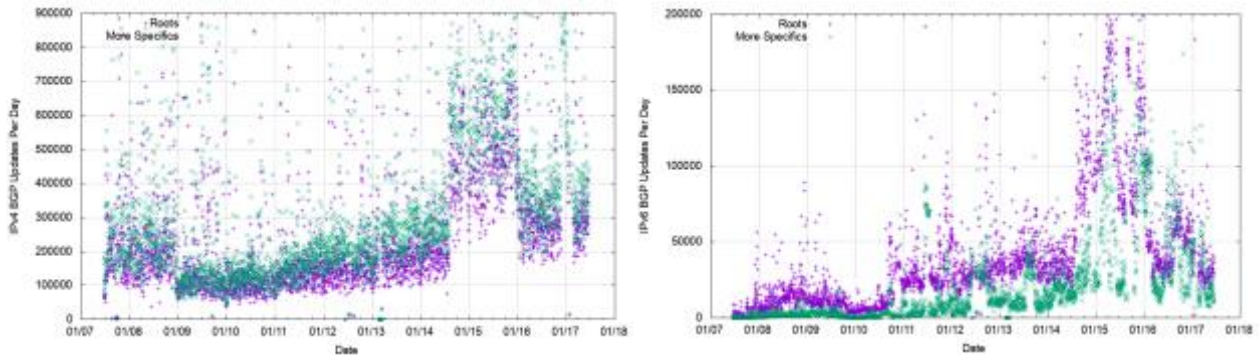


IPv4 IPv6

Figure 8 –Relative Address Span of BGP More Specifics by Type

Routing Updates

While more specifics take up space in a router's forwarding tables, do they also take up a disproportionate level of capacity a router's processing capacity? Are these more specific prefixes disproportionately "noisy" in terms of BGP updates?

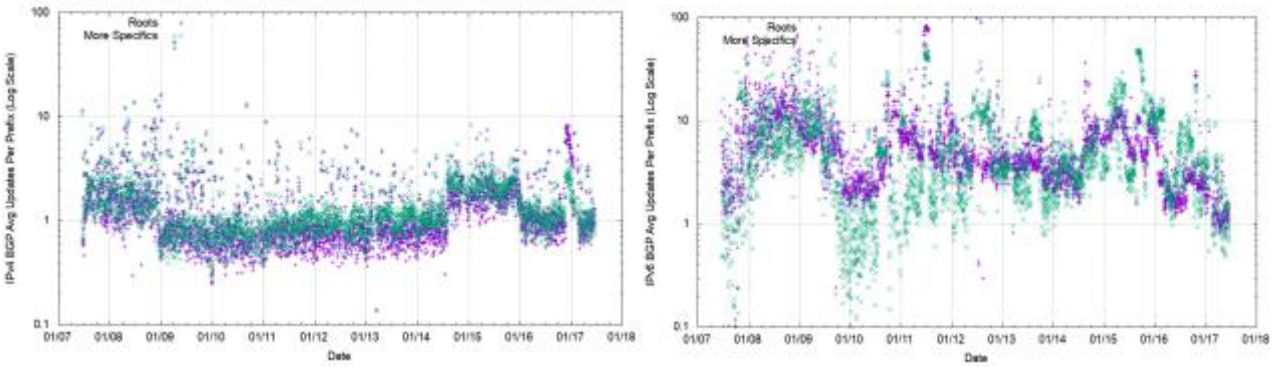


IPv4 IPv6

Figure 9 – BGP Updates per day

Figure 9 shows the total daily count of all BGP updates received by AS131072 over the past decade, broken into the classification of updates relating to root prefixes and updates of more specifics. In the

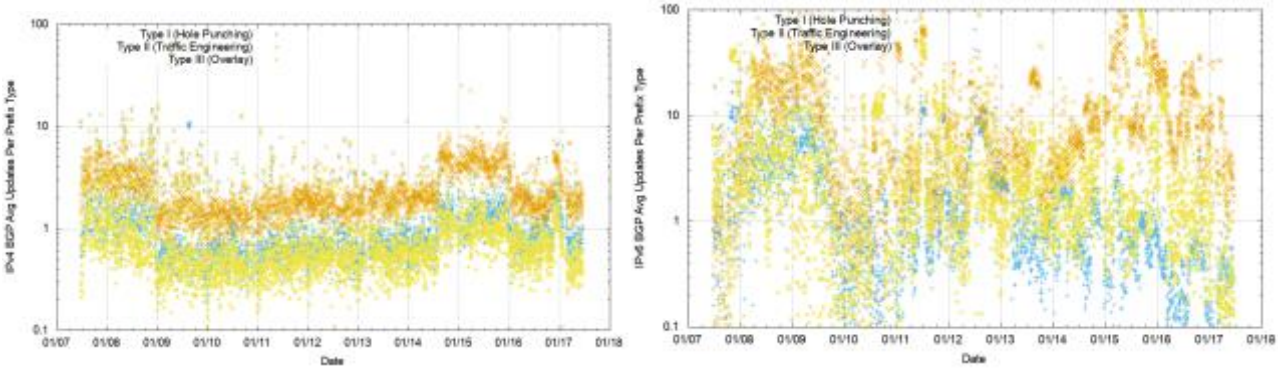
case of IPv4, more specifics are responsible for a slightly higher level of updates than root prefixes. The opposite is the case for IPv6. However, there are fewer more specific prefixes as compared to the count of root prefixes in IPv6, so the lower update count might reflect the relatively lower population of more specifics in IPv6. One way to compensate for this is to divide the daily update count in each of these classifications by the number of prefixes, yielding the average number of updates per announced prefix for both root and more specific prefixes.



IPv4 IPv6

Figure 10 –Average of Updates per advertised prefix

Using this metric, more specifics remain slightly noisier than root prefixes in IPv4, but the picture is not so clear in IPv6. There, the two rates appear to be very similar over time. There is however one very notable difference between IPv4 and IPv6. In IPv4 there is an average of around 1 update per announced prefix over this extended period. In IPv6 the relative update rate is far higher, with most of the daily readings sitting between 3 and 5 updates per announced prefix for both more specifics and root prefixes. Recent years has seen a trend of higher stability, with the average update rate approaching that of IPv4 in recent months.



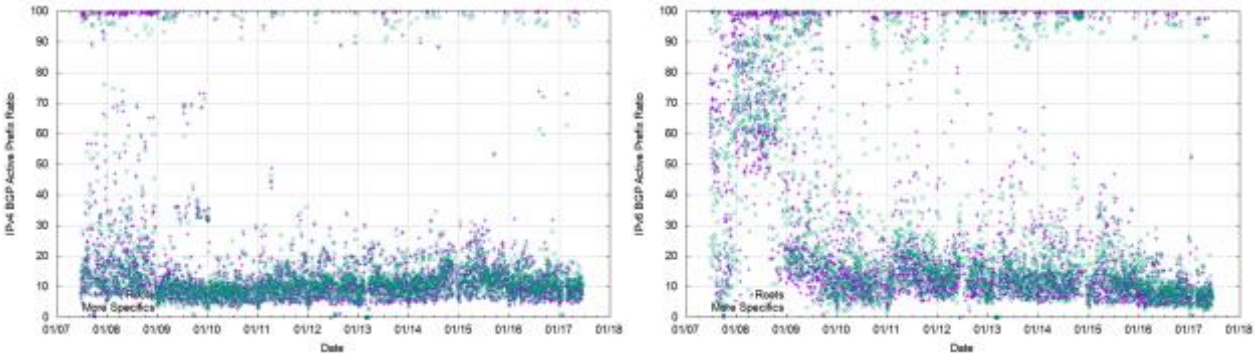
IPv4 IPv6

Figure 11 –Average of Updates per More Specific Type

We can break down the more specifics into the three categories, and perform a similar measurement to see which of these types of more specific is the noisiest in terms of BGP updates. Type II Traffic Engineering more specifics are clearly responsible for a higher number of BGP updates per announced prefix in both IPv4 and IPv6. In IPv4 these updates occur at an average rate of 2 updates per prefix of the 10-year period, while Type I and III more specifics are relatively far quieter, with average update rates of less than 1 per announced prefix in both categories. The IPv6 picture is noisier, but there is a discernible signal that Type II traffic engineering prefixes are relatively less stable, and Type I Hole punching prefixes are relatively more stable, by up to a factor of 10 on some days.

Looking at the count of average number of updates per prefix is often accompanied by an implicit assumption that updates are well distributed across the set of announced prefixes. In BGP this is definitely not the case. A small proportion of prefixes are extremely noisy, while the majority of

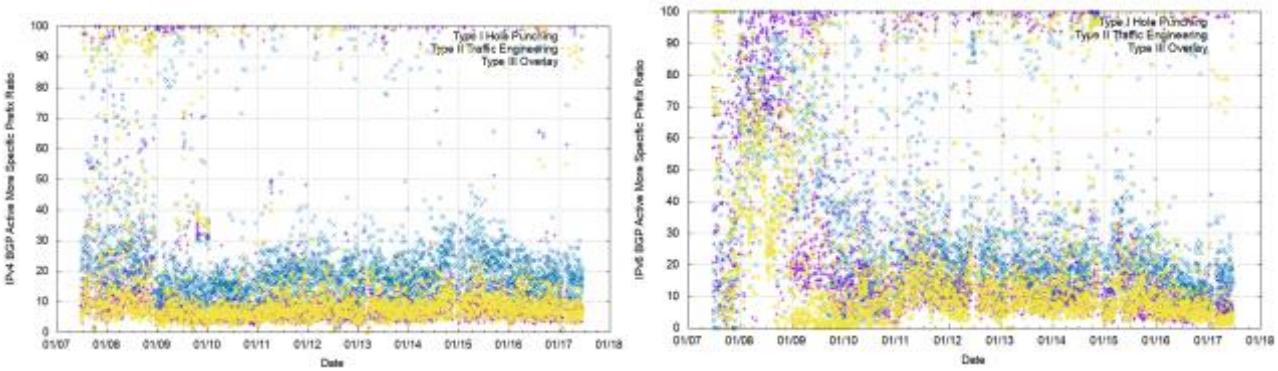
prefixes are quiet on a day-to-day basis. The number of active prefixes per day is shown as a proportion of the total prefix count in Figure 12, distinguished by root and more specific category.



IPv4 IPv6

Figure 12 – Proportion of Active Prefixes

In IPv4 the activity is relatively steady at 8-9%, although this is falling in the most recent two years. There is a slight signal that a greater proportion of more specifics are updated than roots in IPv4. There is a comparable picture in IPv6, with the most recent 6 months showing a daily activity level of between 5-10% for both roots and more specifics, shown in Figure 12.

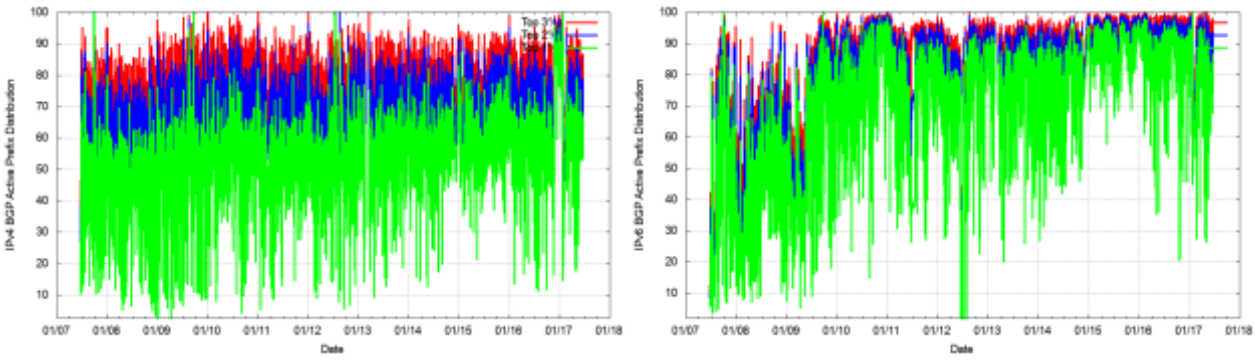


IPv4 IPv6

Figure 13 – Proportion of Active More Specific Prefixes

Type I Hole Punching and Type III Overlay more specifics tend to be the most stable, and in both protocols less than 10% of these more specifics are the subject of BGP updates on any day. Type II Traffic Engineering prefixes tend to be more active and 20% of these prefixes are updated each day in both IPv4 and IPv6.

These figures look at averages across prefixes, but it is useful to bear in mind that the distribution of updates is highly skewed.

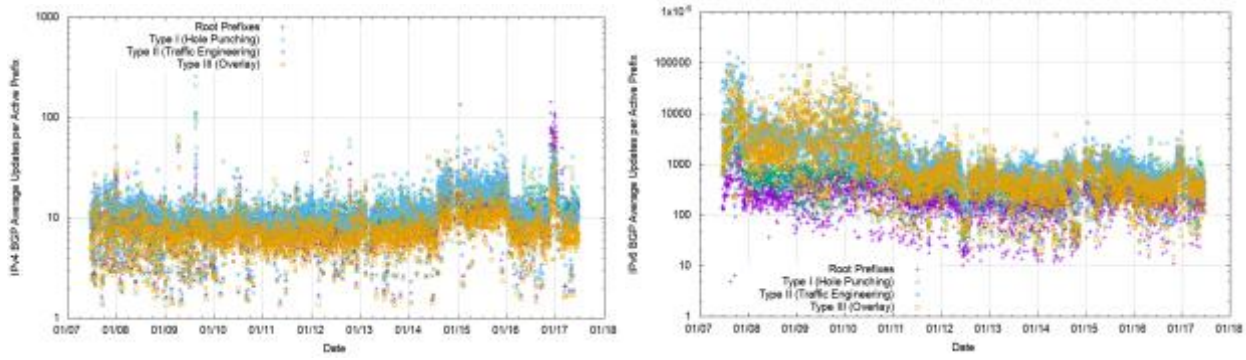


IPv4 IPv6

Figure 14 – Distribution of BGP Updates

Figure 14 looks at the proportion of BGP updates from the noisiest 1%, 2% and 3% of prefixes. In IPv4 some 90% of all BGP updates are associated with just 3% of all announced prefixes, and the busiest 1% of prefixes are responsible for 60-70% of all updates. This uneven skew is even more evident in IPv6, where the busiest 1% of prefixes are responsible for 90-95% of all IPv6 BGP updates, and the busiest 3% of all prefixes are responsible for 98%-99% of all IPv6 BGP updates. It appears that in IPv6 when a prefix is unstable, it becomes highly unstable!

We can partially accommodate for this by looking at only the unstable prefixes on any day, and disregarding all others. We can now look once more at the number of updates per prefix, but in this case look only at the active prefixes rather than all prefixes.



IPv4 IPv6

Figure 15 – Updates per active prefix per day

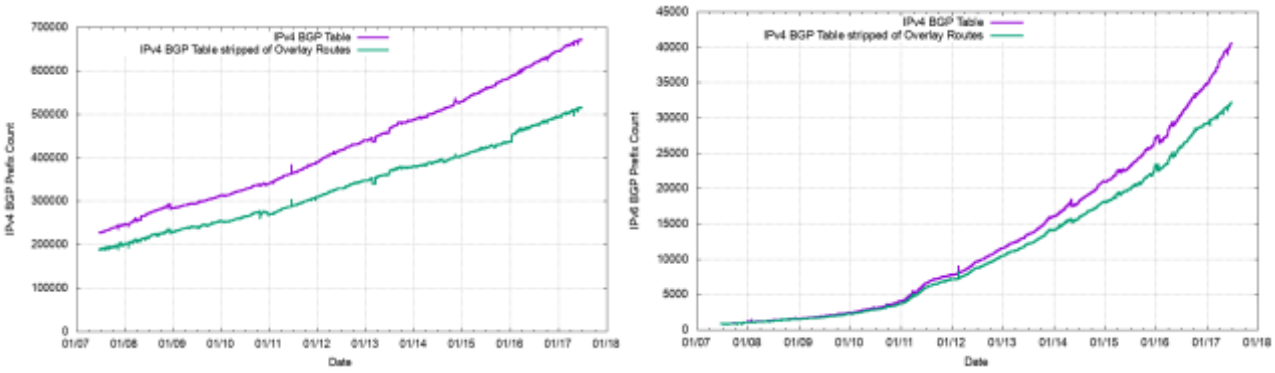
In IPv4 Type II Traffic Engineering more specifics appear to consistently be responsible for 10 updates per day per active prefix, while the other two categories of more specifics, and the root prefixes are generally quiet, with a count of between 5 and 8 updates per active prefix.

In IPv6 the relatively instability is 100 times greater than IPv4, with traffic engineering prefixes being responsible for some 1000 updates per day per active prefix. Other forms of more specifics generate some 500 - 800 updates per active prefix, while the root prefixes generate some 200 updates per day per active prefix.

Cleaning Up More Specifics

From time to time there is an effort to motivate network operators to clean up their BGP advertisements and clear up the more specific advertisements.

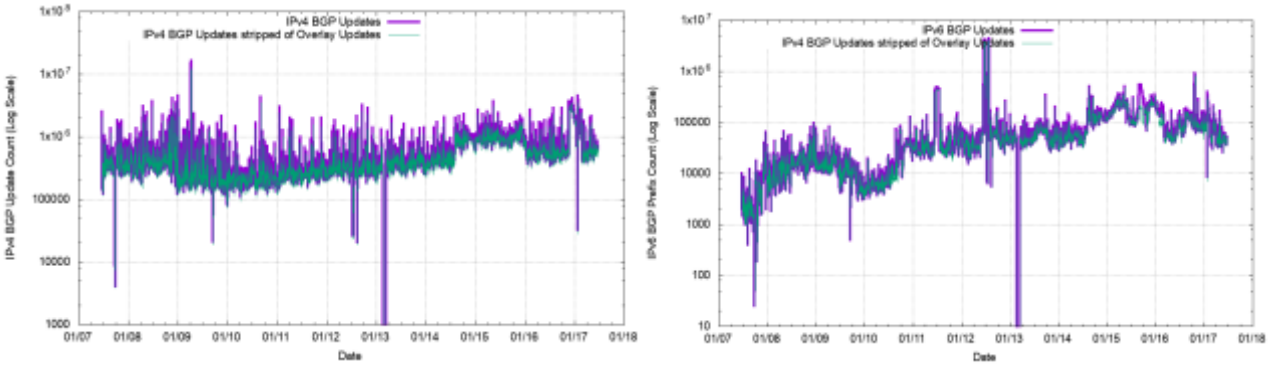
The CIDR Report (http://www.cidr-report.org and http://www.cidr-report.org/v6) is a continuously generated report that lists BGP aggregation potential over a rolling 7 day basis, looking at IPv4 and IPv6. The report lists those AS's that advertise the most more specifics and the potential 'gain' if the Type III Overlay more specifics were removed. In the most recent IPv6 CIDR report it is interesting to note that the operational community is tending to regard the /48 prefix size as the equivalent of an IPv4 /24. Of the 822 changes to the IPv6 routing table in the third week of June 2017, more than one half (56%) of the prefix changes (additions and withdrawals) were /48s, and the next most active prefix size was /32s (17%).



IPv4 IPv6

Figure 16 – BGP Table Size showing potential reduction through the removal of Type III more specifics

Figure 16 shows the change in size of the BGP routing table on the assumption that Type III Overlay more specifics were removed from the table. As already noted, these prefixes, where the AS Path of the more specific exactly matches the AS path of its covering aggregate, including AS prepending, add no additional routing information to the overall routing table, yet occupy space in the routers' forwarding tables.

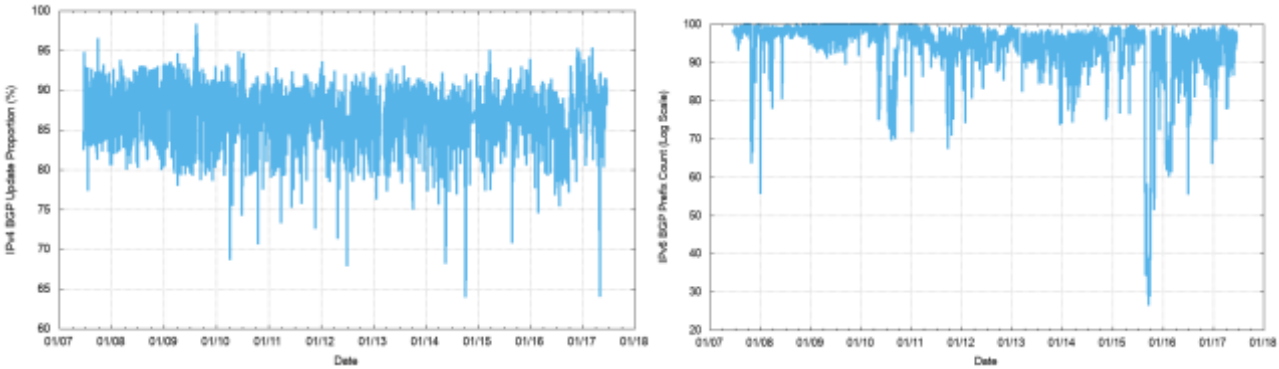


IPv4 IPv6

Figure 17 – BGP Update Counts showing potential reduction through the removal of Type III more specifics

A similar exercise has been performed for BGP updates, and Figure 17 compares the total BGP update count against the update count were Type III more specifics removed from the table.

Another way to look at the updates is to take the ratio of updates with the Type III prefix updates removed and compare that to the total number of updates per day. This is shown in Figure 18



IPv4 IPv6

Figure 18 – BGP Update Counts showing relative potential reduction through the removal of Type III more specifics

What we see from this view is that the removal of these forms of exact match overlay more specifics has the potential to reduce the update rate in IPv4 by between 10% and 15%. In IPv6 the situation is

quite different, and the removal of the updates relating to these particular more specifics would reduce the BGP update rate in IPv6 by between 1% and 3%.

Conclusion

What can we say about more specifics in BGP?

More Specific address advertisements appear to be a source of potential inefficiency by adding to the total table size and the dynamic update load without providing further routing information.

But this is not the case. More specifics are used to the redirect some of the traffic for an address prefix to a different AS, and other more specifics attempt to steer a part of incoming traffic down a different network path. For the networks that advertise these more specifics, these are very useful routing techniques and cannot be readily dismissed as a form of abuse of routing.

It could be argued that only the overlay form of more specifics is more questionable. Network operators have been heard to defend this practice as a rudimentary form of routing security. It is argued that by advertising the more specific themselves they somehow stop the advertisement of more specifics by some third party. This is a somewhat specious justification. In an insecure routing system, such as the one we use to support the Internet today, there is no intrinsic protocol-based control that would prevent anyone from also injecting an advertisement for the same more specific address prefix, and there would still be damage as a consequence of such an attack.

But were we to motivate all network operators to remove these more specifics from their advertised routes it is unclear how effective this would be in overall terms, and unclear what level of benefit would be passed to the BGP routing system. In IPv6 it's pretty clear that overlays are a minor factor in both table size and BGP updates, so any action in this space may well have no visible result. There is a slightly larger margin for improvement in IPv4, but it's still the case that the potential benefits would be small, and probably not worth the effort in the first place.

We can conclude that Hole Punching and Traffic Engineering More Specifics play a useful role for network operators, and could not be arbitrarily removed from the Internet without some other practice taking their place, and like the enumeration equivalent of more specifics, it's likely that the alternative would result in more advertised prefixes rather than less.

It could be argued that only Overlay More Specifics have no visible useful role, but within the larger picture of BGP the size of this sub-class of more specifics, and their update rate, while irritating to some, remain mostly harmless to BGP.

What can we say about More Specifics? Right now, the data suggests that the use of more specifics in the Internet is valuable for some, and at worse mostly harmless for the rest of us!

Author

Geoff Huston B.Sc., M.Sc., is the Chief Scientist at APNIC, the Regional Internet Registry serving the Asia Pacific region.

www.potaroo.net

Disclaimer

The above views do not necessarily represent the views or positions of the Asia Pacific Network Information Centre.