

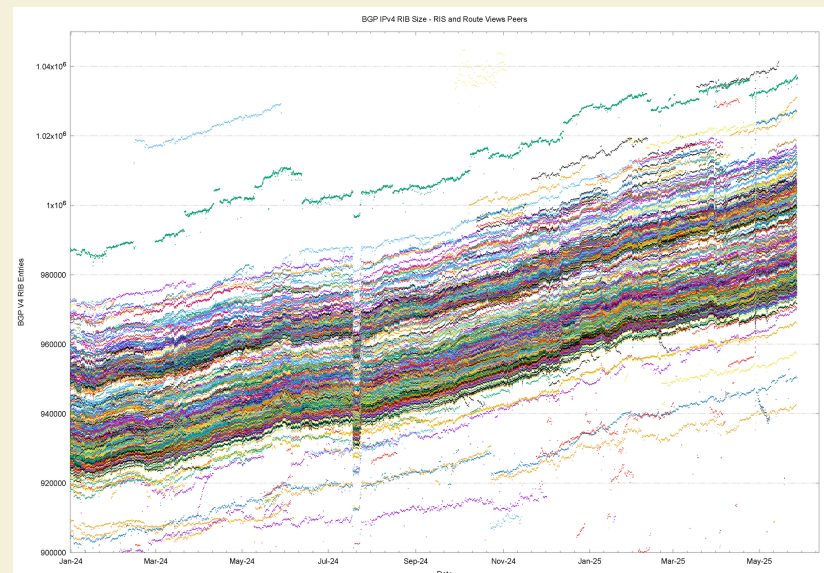
June 2025
Geoff Huston

A Day in the Life of BGP

The Internet is, as its name suggests, a network of networks. The glue that holds this together is the inter-domain routing protocol, BGP, or Border Gateway Protocol. BGP is a *flooding* protocol whose objective is to ensure that all the BGP speakers across the Internet see the same picture of reachable address prefixes. The paths of how to reach to each prefix is relative to each BGP speaker, so the paths contained in each local view of the Internet all differ to some extent, but the intention of the protocol itself is to ensure that everyone has a similar set of reachable destinations.

Of course, there are some differences between the various views of BGP in terms of the collection of reachable address prefixes, as there may be local routing policies that moderate the basic BGP behaviour of route flooding.

The following figure is the plot of the number of IPv4 route objects reported by each peer of Route Views and RIS since 1 January 2024, showing that there is a variance of some 50,000 route objects across this set of BGP speakers.



BGP is by no means a recent protocol, and it was first described in an RFC in June 1989 (RFC 1105). The current specification is BGP-4, described in RFC 4271, January 2006.

In looking at the way BGP operates in keeping the Internet self-connected, there are many approaches. Some tools gather the BGP state from a number of locations and use such a collection of data sets to look at systemic behaviour, such as speed and patterns of information propagation, or variations and potential inconsistencies in the routing information between locations. Other analytical approaches can use a smaller scale, looking at a single BGP data set gathered from a single location using a single BGP peer session. The advantage of this minimalist approach lies in its relative simplicity and clarity.

Here I want to look at just one day of the operation of the Internet’s BGP network by looking at the behaviour of a single BGP session. The day we’ll use for this study is the 8th May 2025, and the BGP vantage point used here is an unremarkable network at the edge of the network, AS 131072.

Nothing special or extraordinary happened on that day. There were no large-scale power blackouts, no major faults in the world’s submarine cable network, nor in the terrestrial trunk cable systems. No headlining-grabbing cyber attack took place on that day, as far as I’m aware. It was just an ordinary Thursday on the Internet, just like any other day, and I selected this data due to its very ordinariness!

Some BGP Measurements

At the start of the day, at midnight UTC on the 8th of May, there were 1,001,332 unique IPv4 address prefixes in the BGP routing table and 221,004 unique IPv6 prefixes. One day later the IPv4 routing table contained 1,002,006 IPv4 prefixes and the IPv6 table contained 221,144 IPv6 prefixes, a net gain of 674 IPv4 prefixes and 140 IPv4 prefixes.

There are 83,739 unique Autonomous System (AS) numbers in the BGP table on this day. Most of these AS numbers, 72,181 or them, are used exclusively on the edge of the network to originate address prefixes to pass into BGP (so-called *stub* networks). Some 11,189 AS numbers are used to both originate address prefixes and operate as a transit provider, carrying the routes of other ASes. The remaining 369 AS number are exclusively transit networks and originate no address prefixes.

While the IPv6 address prefixes are one quarter of the number of visible IPv4 prefixes, in terms of ASes, IPv6 penetration in the routing system is slightly under one half of the combined network. The relatively higher AS count in the IPv4 network is generally attributed to *fragmentation* of the IPv4 address space, where the long-term effects of address scarcity in IPv4 have resulted in reducing the span of address space announced by each network.

There are 76,917 ASes that originate or transit IPv4 prefixes and 34,766 ASes are used for IPv6. This approximate 50% ratio between the two protocols holds for stub networks, transit-only and mixed ASes (Table 1).

IPv4	Address Prefixes	1,001,332		IPv6	Addresses Prefixes	221,004
	AS Count	76,917			AS Count	34,766
	Origin ASes	65,959			Origin ASes	29,076
	Transit ASes	557			Transit ASes	342
	Mixed ASes	10,401			Mixed ASes	5,348

Table 1 – AS Counts in BGP

These numbers show that an average of 13.1 IPv4 address prefixes per originating AS. The distribution of advertisements per originating AS is highly skewed with a heavy tail, and in IPv4 the mean standard deviation of this distributions of the measure of advertisements per AS is 120.8, and the maximum number is 13,283 address prefixes, originated by AS 16509, Amazon-O2.

The comparable numbers for the IPv6 network are an average of 6.4 address prefixes per originating AS, with a mean standard deviation of 73.4, and the maximum number of 6,598 prefixes by AS 9808, originated by China Mobile.

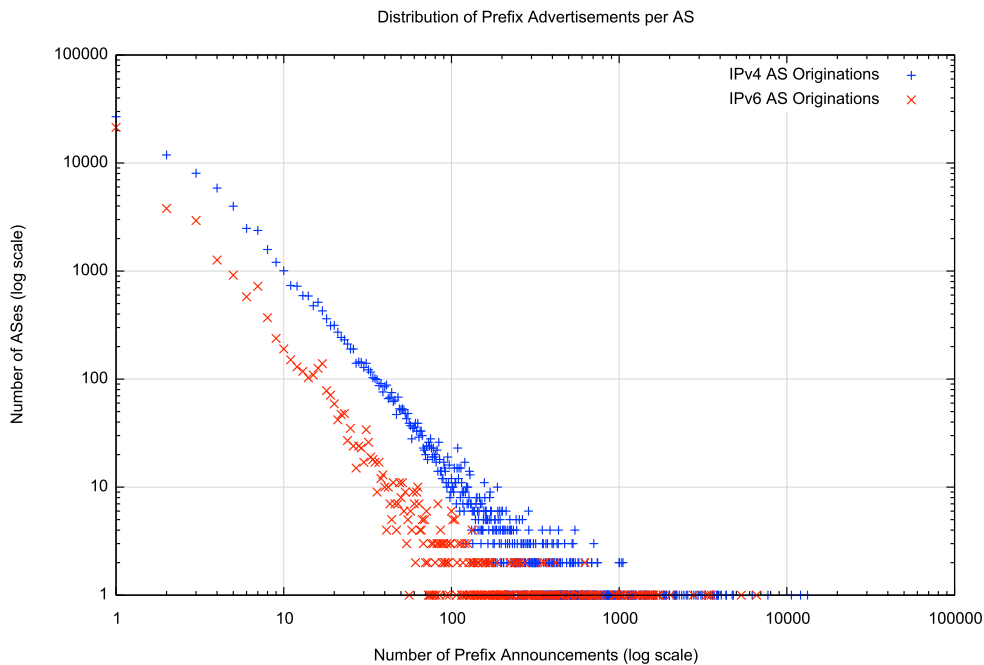


Figure 1 – Distribution of Prefix Announcements per Originating AS

Figure 1 shows the distribution of the number of prefixes advertised by each originating AS. Figure 2 shows the same data using a cumulative distribution, showing more clearly that the majority of ASes originate 10 or fewer address prefixes in both IP protocol address families.

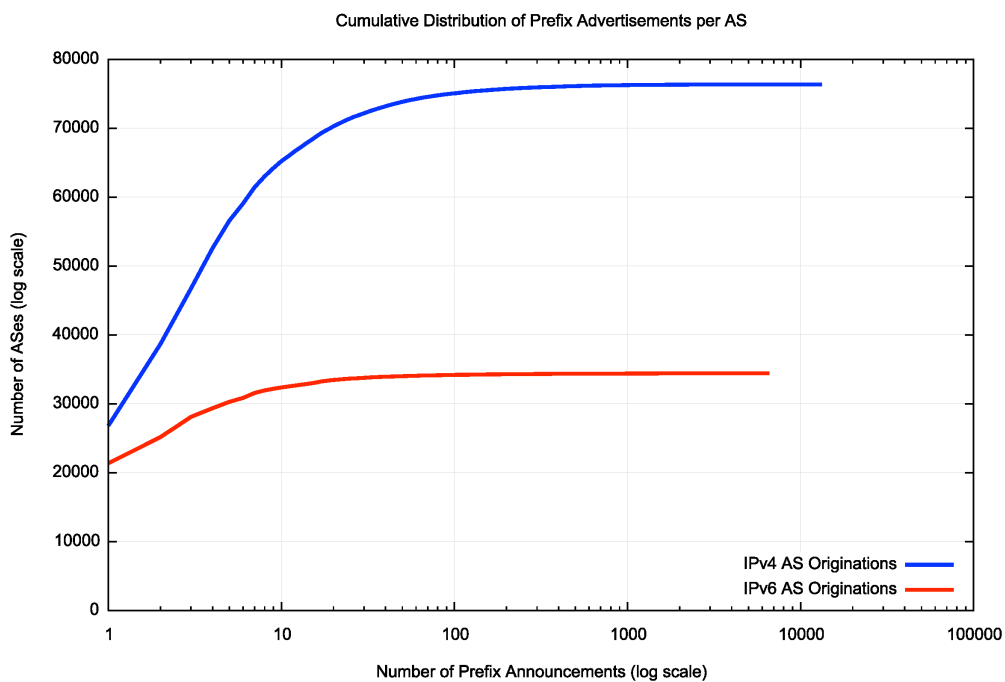


Figure 2 – Cumulative Distribution of Prefix Announcements per Originating AS

Let's also look at the *span* of addresses announced by each AS. Here we differentiate between addresses that are *more specific* (or *covered*) by existing aggregate addresses and addresses that are not covered.

The total span of advertised IPv4 addresses was 3,110,649,600 addresses, or some 72% of the total IPv4 address space, or 84% of the pool of IPv4 usable addresses (addresses that have not been held aside as a reserved addresses). The distribution of announced address spans per origin AS in IPv4 is shown in Figure C. The networks announcing the largest span of IPv4 addresses is shown in Table 2.

AS	Addresses	% of Total	AS Name
AS17676	224,720,128	6%	DNIC DoD, US
AS16509	167,196,672	5%	AMAZON-02, US
AS4134	110,719,744	3%	CHINANET, CN
AS7018	93,760,512	3%	ATT, US
AS721	72,110,336	2%	DNIC DoD, US
AS7922	68,017,664	2%	COMCAST, US
AS8075	65,683,968	2%	MICROSOFT, US
AS4837	57,570,048	2%	CHINA UNICOM, CN
AS4766	47,198,208	1%	Korea Telecom, KR
AS701	41,418,496	1%	UUNET, US

Table 2 – 10 Largest Spans of Advertised IPv4 Addresses

IPv6 address counts are expressed as extremely large integers, so here I will use conventional prefix notation, where, for example, a /32 is the same as a block of 2^{96} individual IPv6 addresses. The total span of advertised IPv6 addresses is a /14, or 0.004% of the total IPv6 address pool. The networks announcing the largest span of IPv6 addresses in shown in Table 3.

AS	Addresses (/64s)	Prefix	% of Total	AS Name
AS7922	36,318,243,520,512	(/18)	5.3%	COMCAST, US
AS3320	35,364,965,122,048	(/18)	5.2%	DTAG, DE
AS17676	18,695,992,639,488	(/19)	2.7%	SoftBank, JP
AS9808	17,888,572,604,416	(/19)	2.6%	CHINAMOBILE, CN
AS7303	17,630,840,750,080	(/19)	2.6%	Telecom Argentina, AR
AS23910	17,605,070,946,304	(/19)	2.6%	CERNET2, CN
AS5713	17,596,481,077,248	(/19)	2.6%	SAIX-NET, ZA
AS4134	17,592,187,682,816	(/19)	2.6%	CHINANET-BACKBONE, CN
AS148000	17,592,186,306,560	(/19)	2.6%	National Knowledge Network, IN
AS1221	17,592,186,109,952	(/19)	2.6%	Telstra, AU

Table 3 – 10 Largest Spans of Advertised IPv6 Addresses

The distribution of address prefix sizes for IPv4 is shown in Figure 3. The majority of announced prefixes (61%) are /24 in size, and slightly over one half of those prefixes are more specific of larger aggregate address prefixes. Less than 0.5% of advertised prefixes are a /15 or larger.

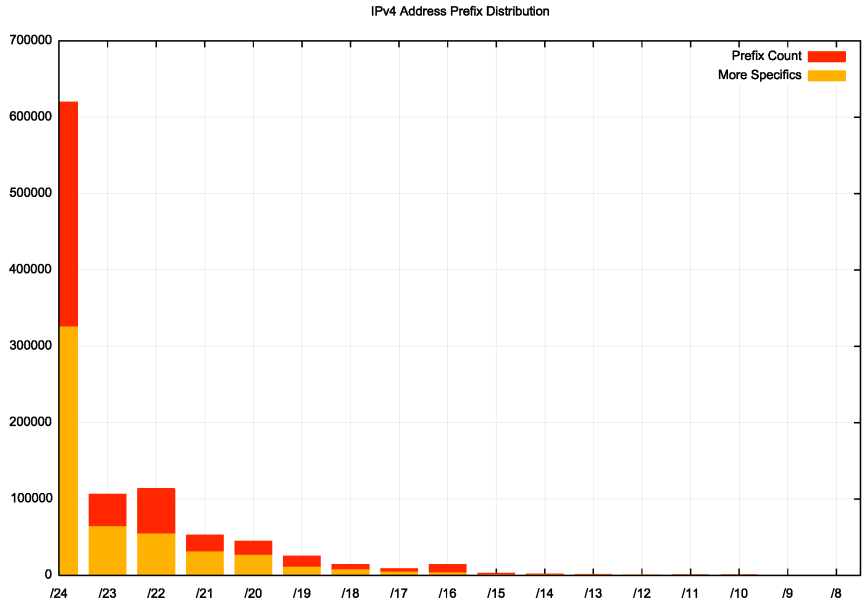


Figure 3 –Distribution of IPv4 Prefix Announcements by Prefix Size

There is a similar skew in the distribution of IPv6 prefix sizes. As shown in Figure 4, the most commonly seen address prefix is a /48 (45% of prefixes) followed by a /32 (12%), then /40 (9%) and /44 (10%). For address prefixes smaller than a /32, the majority of prefixes of each size are more specific prefixes of covering aggregates, while for /32 prefixes and larger the opposite is the case. In IPv4 522,163 prefixes are more specific, or 52% of the total route count, covering a total span of 1,126,753,488 addresses, or 36% of the advertised IPv4 address span. In IPv6 125,262 prefixes are more specific, or 56% of the total IPv6 route count, covering a total span of a /18, or 6% of the advertised IPv6 address span.

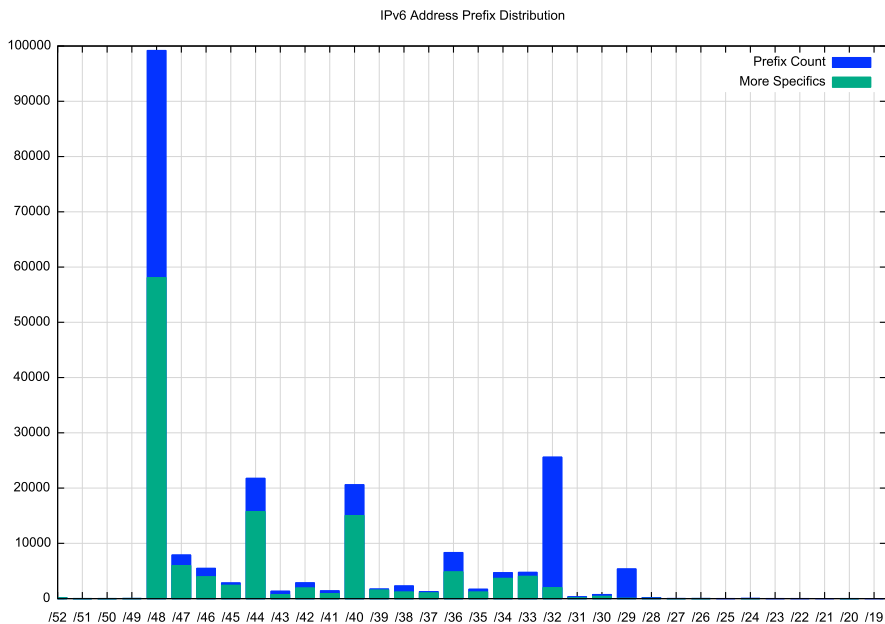


Figure 4 –Distribution of IPv6 Prefix Announcements by Prefix Size

This use of more specific prefixes highlights the number of distinct roles being performed by the routing system. Firstly, it is a *topology maintenance* protocol essentially maintaining a collection of local maps of the network, where each network is aware of the optimal path through the network

to reach every reachable destination, and the local first hop egress point that sets a packet on the path to its intended destination. Secondly, it's a *traffic engineering* protocol that allows each network some degree of capability to balance incoming traffic across a collection of viable paths to reach this network. Thirdly, it's a *policy reconciliation* protocol, where the actions of the protocol reconcile the policy preference of traffic exporters and traffic importers in terms of path selection. The use of more specific routes is not germane to topology maintenance, it is an intrinsic element of traffic engineering and policy reconciliation, in that BGP speakers prefer to use more specific route advertisements over larger aggregate route advertisements. The high incidence of more specifics in these prefix length distributions is indicative of the degree of use of BGP as a traffic engineering tool.

Finally, let's look at the Internet's inter-AS topology by looking at the characteristics of the AS Paths as seen in the routing table. The AS Path is a record of the sequence of networks that have processed the route advertisement as it is being propagated through the network. Each network adds its AS number to the AS Path of a route advertisement as it forwards the advertisement to its BGP peers.

The primary role of the AS Path is to prevent the formation of routing loops, in that a network will not accept an incoming route if the route's AS Path already contains the AS number of the local network. It is also used in the selection of optimal paths, in that each BGP speaker, all other factors being equal, will select the route with the shortest AS Path as the route to use to reach the associated destination. To bias this selection process, it's common for network operators to use *AS prepending*, adding its own AS to the AS path multiple times before forwarding it on to local BGP peers, with the intention of signalling less preferred routes to other BGP speakers.

At the start of the 8th May there were 130,922 unique AS Paths in the IPv4 network, and 46,758 unique AS paths in the IPv6 network. The average length of these IPv4 AS Paths is 3.8 ASes, and when we strip out prepending, then the average length falls to 3.4 ASes. The average length of IPv6 AS Paths is 4.0 ASes and when prepending is stripped this value falls to 3.6 ASes. This is a curious outcome, in that if one assumes that the underlying topology of the inter-AS network is a constant then these two values should be the same. The slightly higher IPv6 value tends to indicate a relative prevalence of IPv6-enabled networks closer to the "edge" of the network.

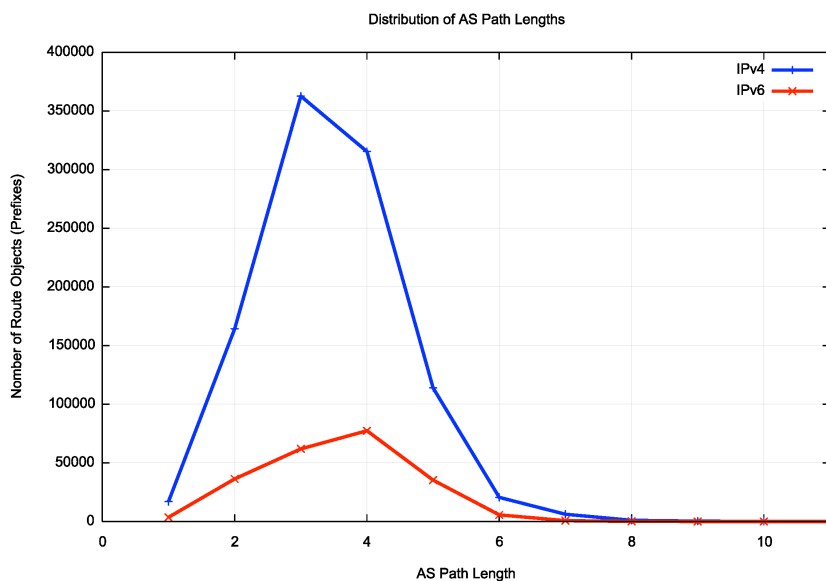


Figure 5 –Distribution of AS Path Lengths

The IPv4 BGP network is highly clustered around a common core of connectivity, and most ASes are positioned no more than two AS hops away from this core. This core clustered topology is not quite the same in IPv6, where a common core is somewhat more diffuse. It is worth bearing in mind that the underlying physical connectivity is the much the same for both protocols, so what we are seeing here is the IPv4-only edge networks are clustered more tightly around this common core than is the case in IPv6.

Finally, let's quickly look at the distribution of AS adjacencies. In a highly clustered system, we'd see a very small set of ASes with a high number of inter-AS adjacencies, and a large set of ASes with 5 or fewer adjacencies, and that's what we see in both IPv4 and IPv6 (Figure 6). A plot of the cumulative distribution makes this cluster-heavy distribution more evident (Figure 7). The most connected 10 ASes account for 82% of all AS adjacencies in the IPv4 net, and 89% in IPv6.

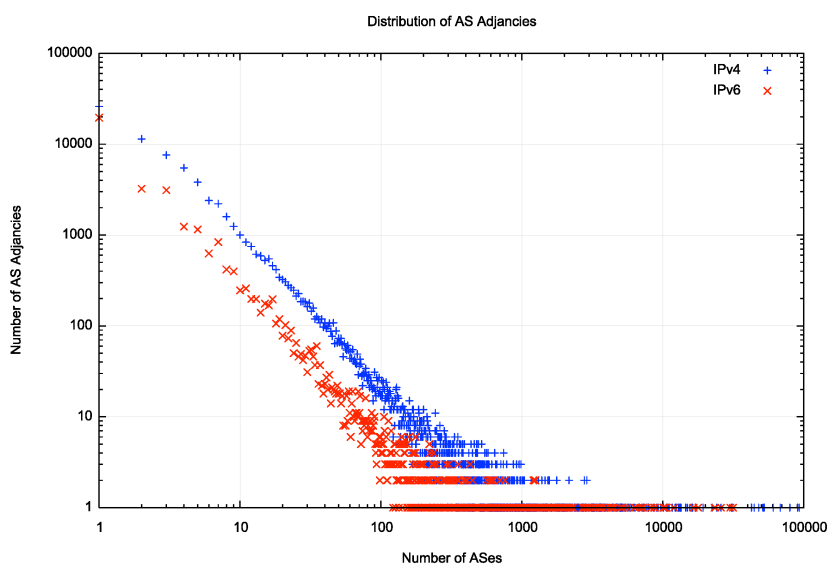


Figure 6—Distribution of AS Adjacencies

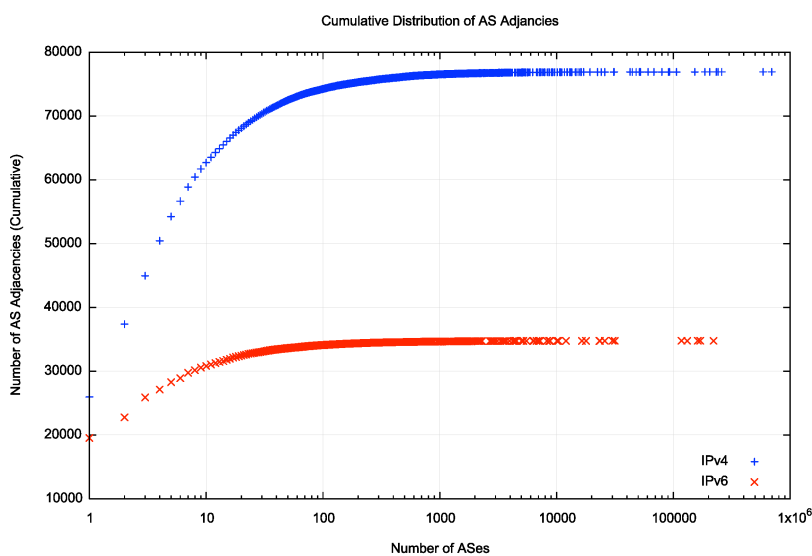


Figure 7—Cumulative Distribution of AS Adjacencies

For the operation of a distance vector protocol this densely clustered topology is close to optimal. Such protocols perform poorly over long sequences (“strings”) as updates necessarily explore all alternatives before arriving at a converged state. In a dense cluster once an update has reached the inner core the new state is sent to all the attached networks, and the network has reached a converged state. Accordingly, we would expect the level of dynamic updates in BGP to be relatively small and protocol convergence to be relatively fast.

Dynamic Behaviour of BGP

Let’s look at the dynamic behaviour of BGP on this day, the 8th May 2025.

BGP is a *distance vector* routing protocol and uses a process of progressive refinement to reach a converged state. Our BGP router received a total of 722,489 IPv4 address prefix updates over the 24-hour period, or an average of 8.3 such address prefix update operations per second. For IPv6 the total is 270,380 IPv6 address prefix updates, or an average of 3.1 updates per second. The distribution of these updates over the 24-hour period is shown in Figure 8.

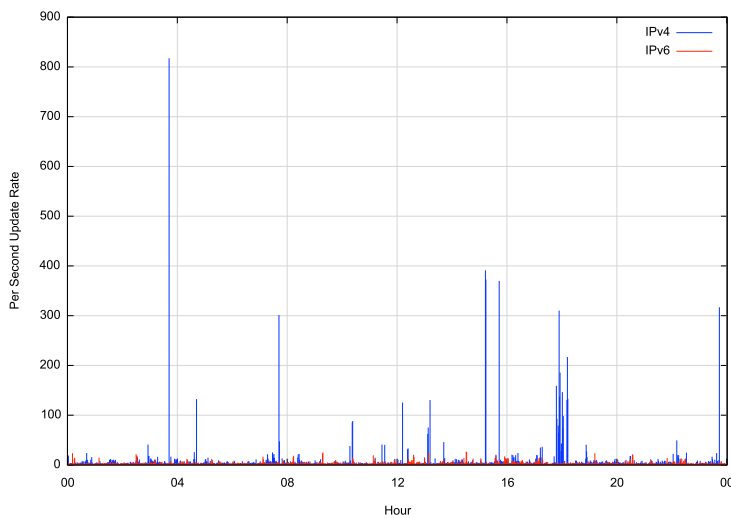


Figure 8 – BGP Update Rate

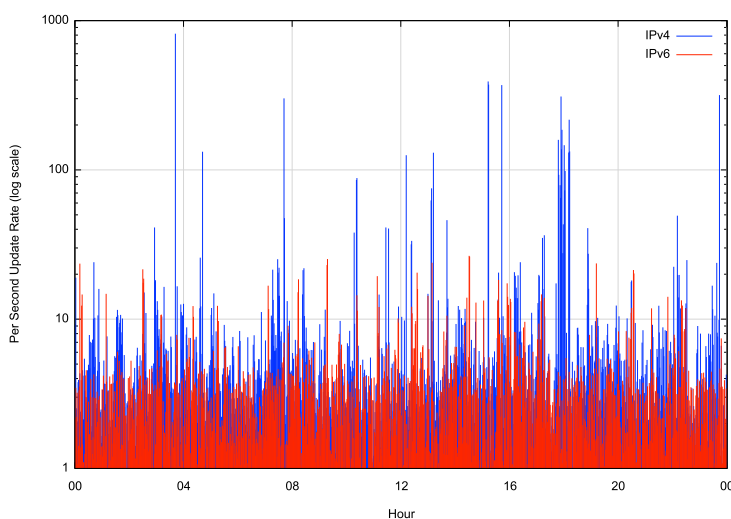


Figure 9 – BGP Update Rate – Log Scale

There were a small number of periods in this 24-hour interval when the IPv4 update rate exceeded 150 updates per second, while there were no such exceptional periods in IPv6. To gain a better

insight into the remainder of this day when the update rate was lower, we can view the same data using a log scale for the update rate (Figure 9). The steady state is a continual update rate of between 4 to 6 updates per second for IPv4, and approximately one half that, at 2 – 3 updates per second in IPv6.

We can further categorise these updates.

- There are “new” announcements of an address prefix that was not in the routing table.
- Change of origin AS, presumably where a multi-homed address prefix changes its primary access path.
- Change of the next-hop AS where a multi-homed network changes its access path.
- Change of AS path where the preferred path changes within the inter-AS network
- Change of prepending of the AS Path where the level of AS prepending in the AS path changes, presumably to change traffic path selection.
- Change of path attributes
- Withdrawal of a prefix

The relative occurrence of these update types is shown in Table 4.

Update Type	IPv4		IPv6	
Announce	47,544	6.6%	28,558	10.6%
Origin AS	45,136	6.2%	7,336	2.7%
Next Hop	135,516	18.8%	104,249	38.6%
AS Path	257,460	35.6%	99,220	36.7%
Path Prepend	10,307	1.4%	1,595	0.6%
Attributes	208,727	28.9%	3,856	1.4%
Withdrawals	17,799	2.5%	25,566	9.5%

Table 4 – BGP Update Types

It is unexpected to see a high level of attribute change updates in IPv4 updates (28.9%) that are not also seen in IPv6 updates (1.4%). Also, the change of origin AS in IPv4 (6.2%) is significantly higher than the level seen in IPv6 (2.7%).

The source of the BGP updates is not uniformly spread across all ASes. In the IPv4 network some 23,337 ASes (of a total of 76,917 ASes) contribute to the observed updates, and of those, the top 10 ASes account for 14% of the total BGP update volume. These ASes are listed in Table 4.

	AS	Updates	Proportion	Cumulative	AS Name
1	8151	30,496	4.22%	4.22%	UNINET, MX
2	16509	16,042	2.22%	6.44%	AMAZON-02, US
3	45899	11,394	1.58%	8.02%	VNPT-AS-VN VNPT Corp, VN
4	9829	10,681	1.48%	9.50%	BSNL-NIB National Internet Backbone, IN
5	21003	6,474	0.90%	10.39%	GPTC-AS, LY
6	8551	5,642	0.78%	11.17%	Bezeqint Internet Backbone, IL
7	36903	5,446	0.75%	11.93%	MT-MPLS, MA
8	367	4,959	0.69%	12.61%	DNIC-ASBLK-00306-00371, US
9	18403	4,883	0.68%	13.29%	FPT Telecom Company, VN
10	8866	4,872	0.67%	13.96%	VIVACOM-AS BULGARIA, BG

Table 5 – Top 10 ASes generating BGP Updates in IPv4

Looking at the most voluminous update source, AS8151, the major component of the update load was a change in the Next Hop AS, which accounted for 27,231 BGP updates. The routes are shifting between the transit ASes 174, 1299, 2914, and 6453. This network, the Mexican national academic and research network originates some 12,067 IPv4 address prefixes, where 1,610 address prefixes are a /23 and 4,864 are a /24. This use of disaggregated small prefixes combined with a multihoming arrangement across a minimum of four transit networks is an ideal candidate for a large volume of routing updates when the advertised prefixes are moved between the transit networks in an effort to balance incoming traffic volumes across the various traffic paths.

For the IPv6 network the 10 most active Origin ASes account for 49% of the total number of updates for the day (Table 6).

1	151194	56,311	20.83%	20.83%	STELIGHT Zhu Yucheng, CN
2	202256	19,116	7.07%	27.90%	LAWLIETNET, CN
3	16509	13,907	5.14%	33.04%	AMAZON-02, US
4	53667	10,283	3.80%	36.84%	PONYPNET, US
5	132884	8,189	3.03%	39.87%	Summit Communications Limited, BD
6	40138	5,666	2.10%	41.97%	MDNET, US
7	42298	5,276	1.95%	43.92%	GCC MPLS peering, QA
8	11014	4,529	1.68%	45.59%	CPS, AR
9	27947	4,224	1.56%	47.16%	Telconet S.A, EC
10	52257	3,853	1.43%	48.58%	Telconet S.A, EC

Table 6 – Top 10 ASes generating BGP Updates in IPv6

The most active IPv6 Origin AS, AS151194 had the majority of updates in AS Path changes.

The distribution of update volumes across origin ASes for IPv4 and IPv6 is shown in Figure 10. The IPv6 update activity is skewed by a small number of origin ASes that have an unstable AS path where the entire block of announced prefixes changes paths.

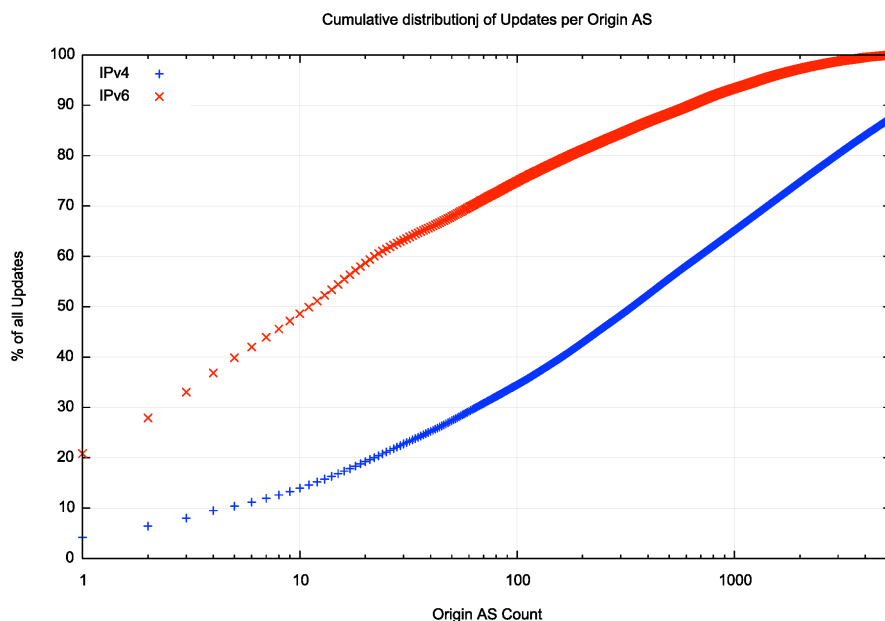


Figure 10 – Cumulative Distribution of Updates per Origin AS

During the course of the day, some 2,114 prefixes were added to the IPv4 routing table and 1,440 prefixes were withdrawn. In IPv6 539 prefixes were added and 403 were withdrawn by the end of the day. There were 76,917 ASes visible at the start of the day, and 62 were added and 83 were dropped, leaving 76,896 ASes in the IPv4 network at the end of the day. For the IPv6 network there were 34,766 ASes at the start of the day, 43 additional ASes were added and 33 were withdrawn, leaving 34,776 at the end of the day.

Observations

BGP was intended to be a highly efficient flooding protocol. By sending only incremental changes in the inter-AS topology of the network and additions and removals of address prefixes, BGP was intended to be a very lightweight protocol. If nothing changed in the network, then BGP was supposed to be essentially quiet.

The day we've examined here was a very ordinary day in BGP terms, 2,553 address prefixes were added to the BGP routing system in both IPv4 and IPv6, and 1,843 address prefixes were withdrawn, and this might lead you to think that that day could've elapsed with some 5,000 BGP updates, not the million BGP updates we observed. Where have the other 995,000 BGP updates come from?

Now it's certainly true that most of the network is mostly quiet, most of the time. Figure 8 shows that the general level of protocol activity is some 2 – 3 updates per second, which the occasional bursts of between 300 to 1,000 updates per second. Of the 1,001,332 IPv4 address prefixes, 213,674 were listed in one or more BGP updates, some 20% of the total address prefix population. The remaining 80% of prefixes were completely quiet for the entire 24-hour period. It appears that a small population of networks generates most of the BGP noise.

The issue here is that BGP does one task remarkably well, and most other tasks quite poorly. The flooding of information across the network is remarkably efficient and is performed in a timely manner. BGP is really good at this task. However, the roles that BGP does not perform well are the roles of traffic engineering and policy mediation. BGP has no language to express desired traffic outcomes, nor any way to express where or how any such measures are intended to apply. The result is that more specific address prefix announcements are the majority of announcements in both protocols and there is no direct way to limit the propagation of such more specific announcements beyond the immediate locus of effect. The result is a noisy protocol where the result of applying the stream of changes to a local BGP's speakers' forwarding table is essentially no change whatsoever. A BGP implementation spends most of its time churning through these voluminous updates, each of which make no absolutely tangible change in the local forwarding decisions being performed by the router.

However, all this extraneous noise in BGP has not been enough of an issue to motivate enough of us to seek alternative mechanisms. Somehow, we've been able to keep BGP under some manageable threshold of acceptability in terms of scalability, cost and stability for more than forty years now. That does not mean that all has been quiet across that period, and from time-to-time concerns surface that the BGP-based routing system is crossing this threshold of viability or cost efficiency (as was the case at the time of the 2006 IAB workshop on Routing and Addresses, documented in RFC 4984). However, but such periodic ringing of the alarm bell has not galvanised a sufficient reaction at any time so far to tackle what a new generation of inter-domain routing should look like. To put it another way, the dire predictions that the Internet's routing system would grow far faster than the capabilities of what we can afford in routing hardware never eventuated, and the desire to load more functions into the routing system, such as quality of service differentiation, multihoming and service resilience, and even the desire to implement tight

integration of routing security have all foundered on the lack of sufficient support on the part of network operators. Nobody has wanted such features, or at least insufficient numbers of potential users exist who would pay vendors for such network features and the operational issues we encounter in running something as simple as BGP tend to act as a strong deterrent in adding further operational complexity into the routing environment. Instead, the network, and its BGP routing infrastructure, has followed almost the opposite path of progressive commoditisation and feature stripping, and the locus of development and investment has moved further up the protocol stack into the area of end-to-end transport controls, applications and content delivery.

What that leaves us with is much the same BGP as we had four decades ago. It's a distance-vector flooding protocol that implements a simple best path selection function. It works in a manner that is best summarised as "well enough!"

Disclaimer

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

Author

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

www.potaroo.net