BGP in 2013

The Border Gateway Protocol, or BGP, has been toiling away, literally holding the Internet together, for more than two decades and nothing seems to be falling off the edge of the Internet so far. As far as we can tell everyone can still see everyone else, assuming that they want to be seen, and the distributed routing system appears to be working smoothly. All appears to be working within reasonable parameters, and there is no imminent danger of some routing catastrophe, as far as we can tell.

If that's the case, then why should we be interested in BGP? Isn't this just a classic case of "Nothing to see here, move along?"

One cause for concern is the inexorable growth of the Internet's routing system. Does this constant growth in routing imply that our routing system is growing faster than our capacity to afford ever larger and faster routers, assuming of course that we can keep on building ever larger and faster routers in the first place?

Here's some possible reasons why an analysis of BGP can be useful for folk in Internet business.

For the ISP network operator, this information may be help in figuring out how big a router should you buy today if you want it to cope with the full BGP routing load in 3 - 5 years time. Perhaps you might want to understand what FIB size is necessary in that time, and what level of TCAM size might be appropriate, in which case you may want to have a conservative estimate of the anticipated number of entries in the routing table over that period.

The same consideration applies to a vendor of routing equipment: How big a router should a vendor build to cope with the BGP load over the next 3 - 5 years? What growth factors for the routing system should be added into the product design phase? What are the Internet's scaling factors at play here?

Underlying these questions are a more basic set of questions about BGP itself.

Is BGP scaling or failing?

Do we need to develop a new Inter-Domain Routing protocol to take over from BGP?

If BGP is failing, when will this failure become critical? How much time do we have before a new approach is needed?

And if we are going to head down this path of attempting to slide in an entirely new routing protocol, is the routing problem simply one of routing over an ever larger and more diverse population, or is this an expression of a more fundamental scaling limitation of the Internet's current concepts of names and addresses?

In other words, if we are facing a major problem with routing scalability do we need now to examine alternate architectural models of identity and location separation in order to build truly massive and highly diverse networks? Or, is routing scaling an intractable problem within the confines of the current architecture and we need to shift around the basic building blocks of the Internet architecture in order to allow a different routing architecture that has radically different scaling properties?

Or, on the other hand, is BGP coping just fine and there is no current expectation of its imminent demise?

The BGP Measurement Environment

In trying to analyse long baseline data series the ideal approach is to keep as much of the local data gathering environment as stable as possible. In this way the changes that occur in the collected data reflect changes in the larger environment, as distinct from changes as a result of changes in the local configuration of the data collection equipment.

In this case the measurement point being used is a BGP router configured as AS131072. This AS generates no traffic and originates no routes in BGP. It's a passive measurement point that has been logging all received BGP updates since 1 July 2007, and is the successor to an earlier setup located in AS1221. The router is fed with a default-free eBGP feed from AS 4608, which is the APNIC network located in Australia, and AS 4777, which is the APNIC network located in Japan, for both IPv4 and IPv6 routes.

There is also no iBGP in this particular measurement setup. While it has been asserted at various times that iBGP is a major contributor to BGP scalability concerns in BGP, the consideration here in trying to objectively measure this assertion is that there is no "standard" iBGP configuration, and each network has its own rather unique configuration of Route Reflectors and iBGP peers. This makes it hard to generate a "typical" iBGP load profile, let alone analyse the general trends in iBGP update loads over time.

In this study the scope of attention is limited to a simple eBGP configuration that is likely to be found at a "stub" AS at the edge of the Internet. This AS is not an upstream for any third party, it has no transit role, and does not have a large set of BGP peers. It's a simple view of the routing world that I see when I sit at an edge of the Internet.

The measurement system that I'm using takes a snapshot of the BGP RIB every hour, as well as logging all received BGP updates.

The Data

IPv4 Routing Table

The following figures show some of the vital statistics for IPv4 in BGP since the start of 2009 .

Figure 1 shows the total number of routes in the routing table over this period. This is a classic "up and to the right" Internet curve, but it should be noted that the days of the strong "J curve" that indicated growth at an exponential rate that doubled every year, or even growth that doubled every few months, are over. The growth elements in the Internet today are more strongly aligned to a far more modest linear growth model.

Over this period there is one significant event, namely the exhaustion of the IPv4 address space pools in APNIC in April 2011 (serving the Asia Pacific region) and in the RIPE NCC in September 2012 (serving Europe and the Middle East). The three year period since the start of 2011 has seen the span of addresses advertised in the routing system slowing down (Figure 5). However, at the same time there has been a higher level of growth in the number of entries in the routing table over the 2011 – 2013 period than was visible in the preceding two year 2009 – 2010 period. The result of these two factors is that the average announcement in the IPv4 routing table is spanning fewer addresses, or, to put it another way, the granularity of the IPv4 routing space is getting finer. As Figure 4 shows, while the period 2009 – 2010 the average announcement size was a steady 7,000 addresses, in the period since the start of 2011 the average announcement size has fallen to 5,500 addresses at the start of 2014.
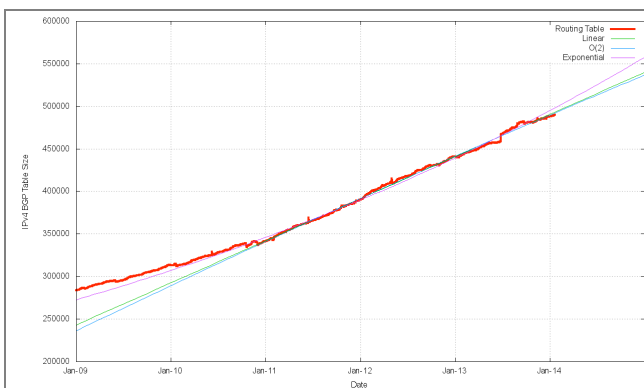

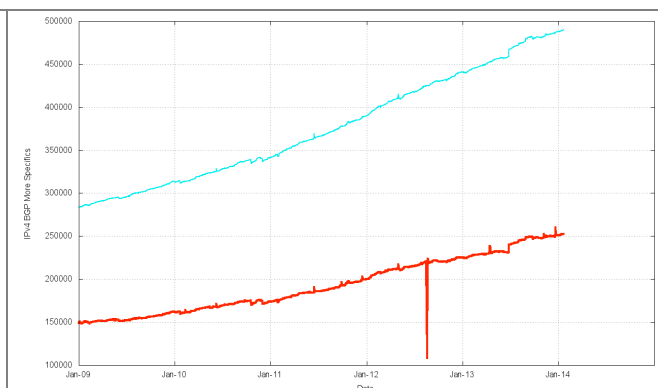
Figure 1 - IPv4 BGP Routing Table Size (RIB)
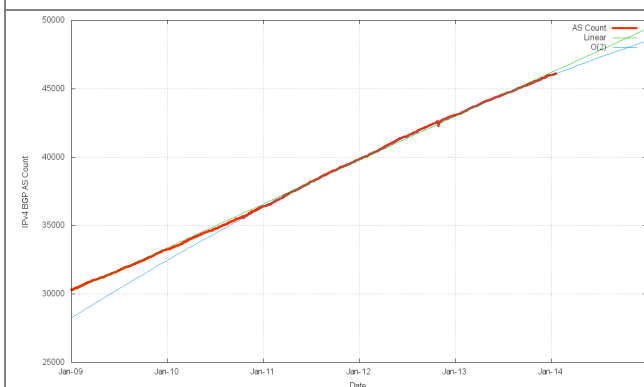

Figure 2 - IPv4 More Specific Entries
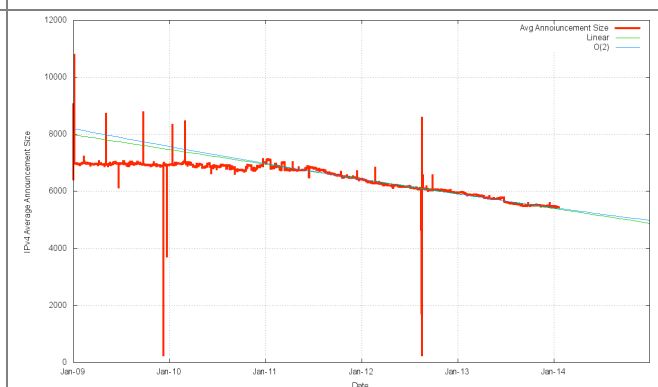

Figure 3 - IPv4 AS Count
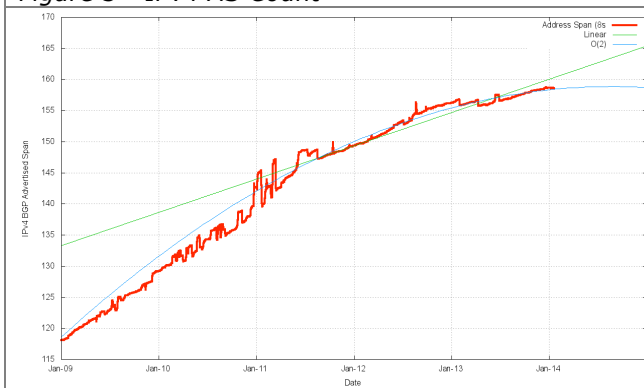

Figure 4 – Average Announcement Size

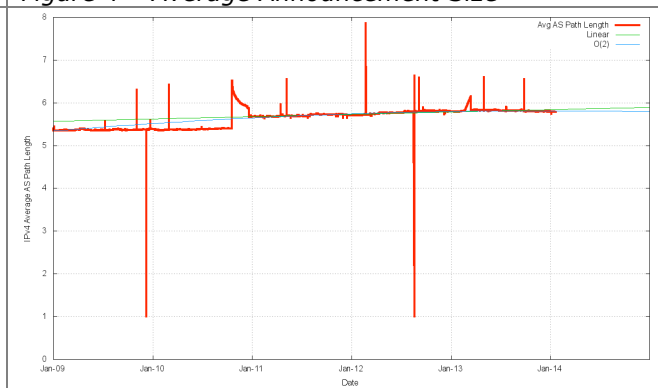*Figure 5 - IPv4 Advertised Address Space*

Figure 6 - IPv4 Average AS Path Length

Figure 7 shows the day-by-day progress of the size of the routing table through 2013. There was one notable event, the advertisement of some 10,000 more specific routings on the 25th June. There were four origin ASes that were responsible for the bulk of these additional advertisements: AS7029 (Windstream Communications, US), responsible for an additional 3,606 routes on this day, AS33363 (Bright House Networks, US), responsible for

1,651 additional routes, AS4323 (TW Telecom Holdings, US), for some 1,351 additional routes, and AS9116 (Goldenlines, Israel), for some 348 additional routes. This single day was responsible for some 20% of the entire year's growth in the size of the routing table for the year. It appears somewhat unlikely that this deaggregation was a coordinated action across a number of origin AS's, so a more likely explanation is that a change in the routing environment for these particular AS's, or their upstreams or peers, admitted a set of more specifics routes that had been previously filtered. When we look at the daily view of the routing table from each of the peers of Route Views (www.routeviews.org) (Figure 8), it appears that this is indeed the case, and other peers of Route Views did not see this jump in the number of routes. The data from Route Views suggests that in this case it's AS6939 (Hurricane Electric) that is allowing this set of more specifics through into some parts of the global routing table. A closer examination, and some followup with some of the AS's involved here reveals that in the case of AS7029 in particular (Windstream Communications, US), they announce 4,429 specific routes to their peers, while they announce 460 aggregates to their upstreams. While this tends to bias incoming traffic to peer connections, the incremental load that is imposed on the network is a further 4,000 routes on the set of AS's that have paths that reach AS7029's peers.

This illustrates an important principle in BGP, that there is no single authoritative view of the routing table – all views are in fact relative to the perspective of the BGP speaker. It also illustrates that at times the cause of changes in routing is not necessarily a change at the point of origination of the route, but it may well be a change in transit arrangements within the interior of the network that may expose, or hide, collections of routes. And thirdly, this illustrates the prevalent use of more specifics to affect traffic engineering. It is often the case that these more specifics are advertised with a limited scope, and if the changes to the transit arrangements move a BGP speaker in or out of this scope, then one can expect changes in the set of visible routes as a consequence.

This could be seen as an instance of a "tragedy of the commons," (http://en.wikipedia.org/wiki/Tragedy_of_the_commons) where the self interest of one actor in attempting to optimise its incoming traffic loads in order to minimise its transit service costs becomes an incremental common cost that is borne by all other actors. To quote the Wikipedia article on this topic "In absence of enlightened self-interest, some form of authority or federation is needed to solve the collective action problem." This appears to be the case in the BGP realm, where there is an extensive reliance on enlightened self interest to be conservative in one's own announcements, and the actions by a smaller set of actors, such as Windstream in AS7029, are prominent because they fall well outside of the conventional "norm" of inter-domain routing practices.
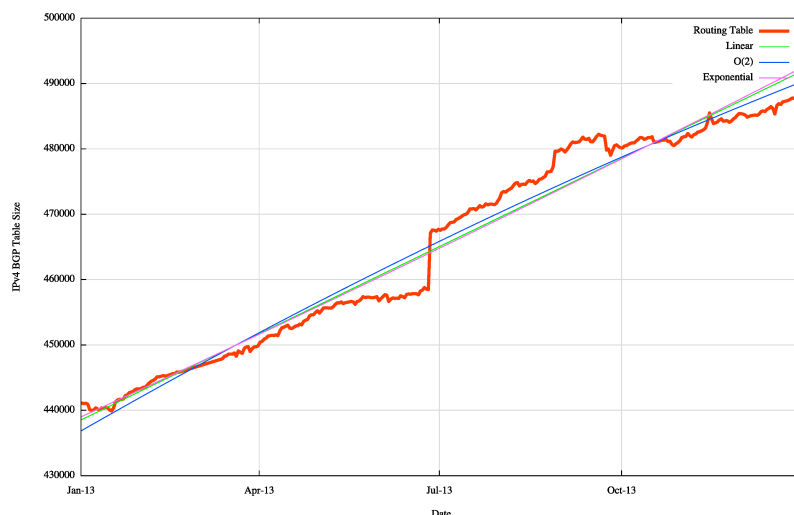

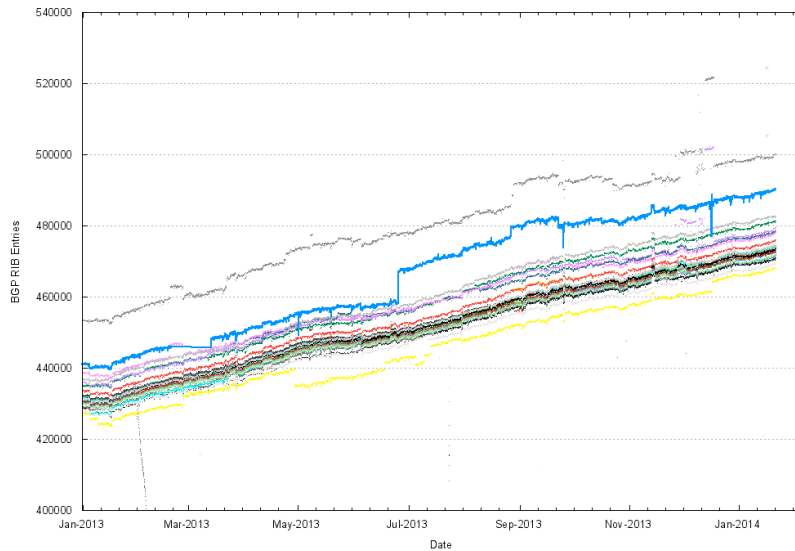*Figure 7 – IPv4 BGP Table in 2013, as seen by AS131072*

*Figure 8 – IPv4 BGP Table in 2013, as seen by peers of Route Views
(the solid blue line traces AS131072, the AS used as the source of
the data in this analysis, while other lines track the daily size of
the routing table as presented by other peers of Route Views.)*

The summary of the IPv4 BGP network over the 2011-2013 period is shown in Table 1.

	Jan-11	Jan-12	Jan-13	Jan-14	2012 growth	2013 growth
Prefix Count	341,000	390,000	441,000	488,000	13%	11%
Root Prefixes	168,000	190,000	215,000	237,000	13%	10%
More Specifics	173,000	200,000	226,000	251,000	13%	11%
Address Span (/8s)	142.7	149.3	156.2	158.6	5%	2%
AS Count	36,400	39,800	43,100	46,000	8%	7%
Transit AS Count	5,000	5,500	6,000	6,400	8%	7%
Stub AS Count	31,400	43,300	37,100	39,000	8%	7%

Table 1 – IPv4 BGP Table Growth Profile

What this table indicates is that the IPv4 network continues to grow steadily at around 10% for the past two years.

Other observations about the trends of the IPv4 BGP routing system include noting that the use (or abuse!) of more specifics in the routing system has not improved, nor has it become significantly worse. The average size of advertisements getting smaller in terms of address span per routing table entry, as is the span of originating addresses per origin AS, which appears to be an outcome of IPv4 address exhaustion, which is forcing a number of providers to increase their efficiency of use of public IPv4 addresses.

The average AS path length is constant at around 6 AS hops (which would translate to 4 AS hops if the measurement setup overhead was removed). The number of AS's is increasing at a very regular pace, and the interconnection degree of AS's is getting higher.

The overall conclusions from this collection of observations is that the V4 network continued to grow, but as the supply of new addresses is slowing down, what is now becoming evident is more efficient use of addresses, which results in the granularity of the IPv4 inter-domain routing system increasing. The density of inter-AS interconnection continues to increase. The growth of the Internet is not "growth at the edge" as the network is not getting any larger in terms of average AS path change. Instead, the growth

is happening by increasing the density of the network by attaching new networks into the existing transit structure and peering at established exchange points. This makes for a network whose diameter, measured in AS hops, is essentially static, yet whose density, measured in terms of prefix count, AS interconnectivity and AS Path diversity, continues to increase. This denser mesh of interconnectivity could be potentially problematical in terms of convergence times if the BGP routing system used a dense mesh of peer connectivity, but the topology of the network continues along a clustered hub and spoke model, where a small number of transit ASs directly service a large number of stub edge networks. This implies that the performance of BGP in terms of time and updates required to reach convergence continues to be relatively static.

IPv6 BGP Table Data

A similar exercise has been undertaken for IPv6 routing data, and the comparable figures are shown below.
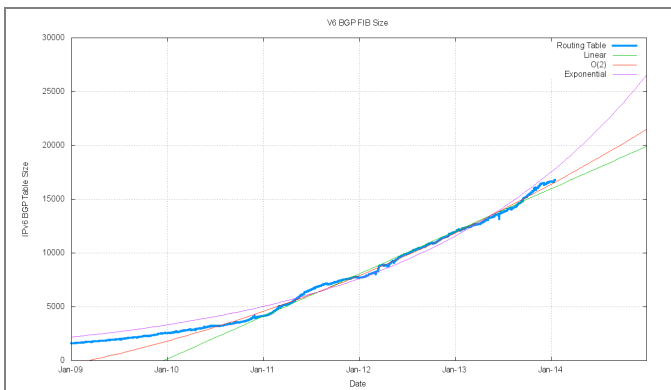

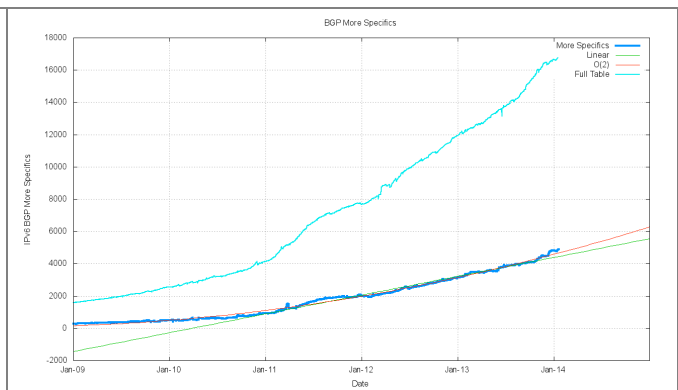Figure 9 - IPv6 BGP Routing Table Size (RIB)
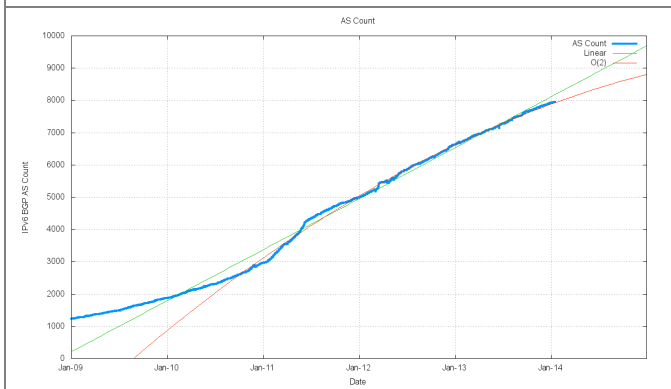

Figure 10 - IPv6 More Specific Entries

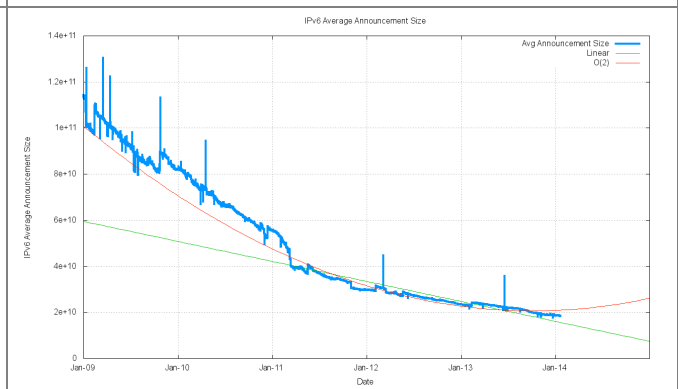Figure 11 - IPv6 AS Count

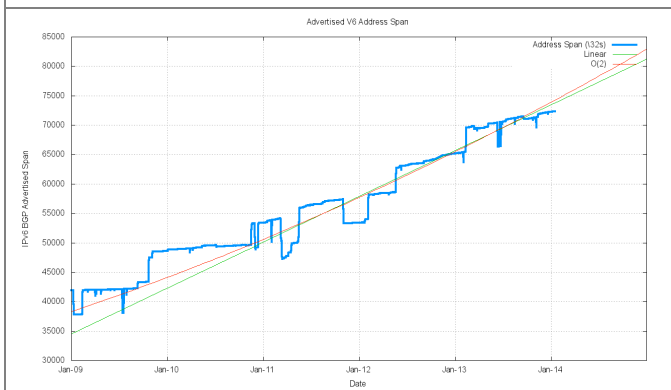

Figure 12 – Average Announcement Size

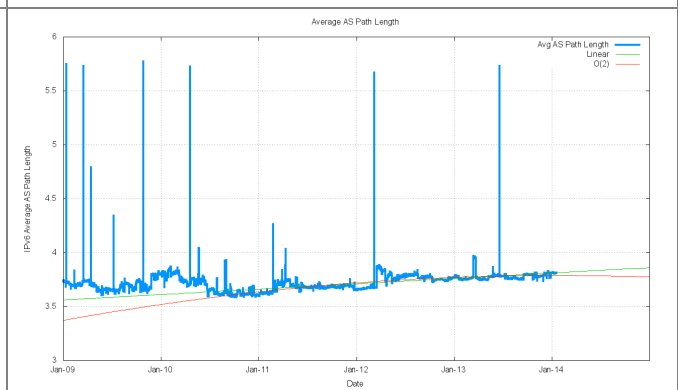Figure 13 - IPv6 Advertised Address Space

Figure 14 - IPv6 Average AS Path Length

The summary of the IPv6 Internet for the period 2011-2013 is as follows:

	Jan-11	Jan-12	Jan-13	Jan-14	2012 growth	2013 growth
Prefix Count	4,100	7,700	11,900	16,700	55%	40%
Root Prefixes	3,200	5,800	8,600	11,400	48%	33%
More Specifics	900	1,900	3,300	5,300	74%	61%
Address Span (/32s)	53,400	53,300	65,100	72,200	22%	11%
AS Count	3,000	5,000	6,600	7,900	32%	20%
Transit AS Count	600	1,000	1,300	1,600	30%	23%
Stub AS Count	2,400	4,000	5,300	6,300	33%	19%

Table 2 – IPv6 BGP Table Growth Profile

There are a number of interesting aspects to this growth where the characteristics of IPv4 look to be appearing in IPv6. The number of more specific advertisements of existing aggregate announcements is rising faster than the number of aggregate announcements. More specifics are now 31% of the total IPv6 routing table. Also the number of transit AS networks is falling, and its now 20% of all AS's in IPv6, as compared to 14% in IPv4.
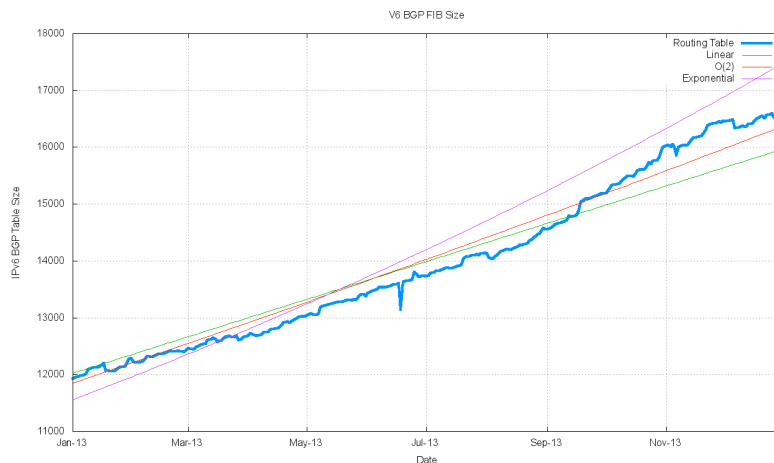

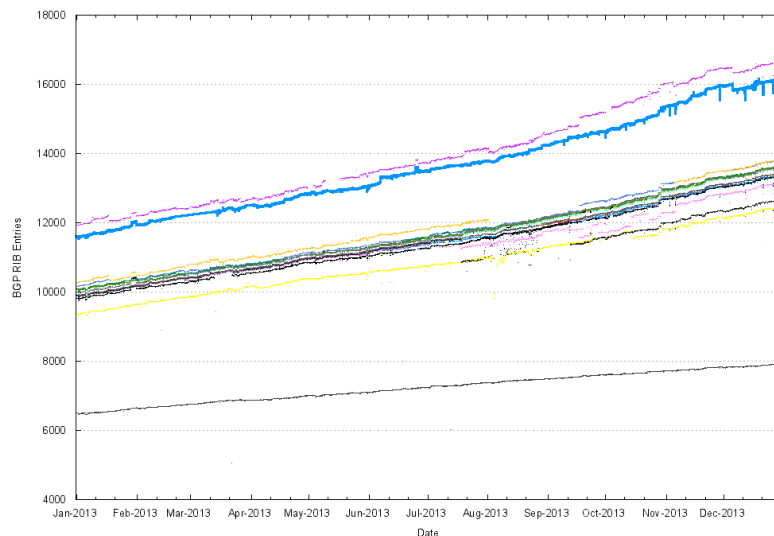*Figure 14 – IPv6 Table Growth in 2013*

Figure 15 – IPv6 BGP Table in 2013, as seen by peers of Route Views (the solid blue line traces AS131072, the AS used as the source of the data in this analysis, while other lines track the daily size of the routing table as presented by other peers of Route Views.)

Again, as with IPv4, this observation provokes the question as to whether this is something that is seen only in the locale of AS131072, where the data is collected, or whether this is a general factor that is visible across the IPv6 network. When we compare the data against a comparable set of data gathered by Route Views, its evident that AS131072 sees a routing table which is of comparable size to one peer of Route Views, but the remainder of the Route Views feed AS's see a IPv6 network which has some 2,000 fewer entries. The difference lies mainly with the more specifics, but the underlying concern here is that not all IPv6 addresses are announced by all peers of Route Views. In other words this data points to the observation that its not always the case that in IPv6 anyone can send a packet to anyone else.

The Predictions

What can this data tell us in terms of projections of the future of BGP in terms of BGP table size?

At the outset it's appropriate to observe that this is a time of extreme uncertainty in the BGP prediction business! The run out of the IPv4 address pool in the Asia Pacific region and in Europe and the Middle East will be followed by a similar run down event for North and South America, currently anticipated to occur in late 2014. What was observed in the Asia Pacific region was an initial dampening on the growth in the address table, but then an acceleration in growth coming from the advertisement of smaller prefixes. It remains to be seen what will occur in the Americas, post exhaustion.

So with the caveat that we are now heading deep into highly speculative areas, and the associated warning that the predictions being made here come with a very high level of uncertainty, lets look at the predictions for the Internet's routing system for the coming few years.

Forecasting the IPv4 BGP Table

The technique used here is to use the least squares curve fitting approach to the historical data (http://en.wikipedia.org/wiki/Least_squares). Figure 16 shows the data set for BGP from the 1st January 2009 until January 2014, and also shows the fit of the most recent 3 years of data to various models.

The first order differential, or the rate of growth, of the BGP routing table is shown in Figure 17. The rate of growth of the routing table appears to be increasing in the period 2009 to 2012. With the exception of the mid-2013 deaggregation event, the table growth in 2013 is at much the same level as 2012. If the first order differential matches a flat line, then the data set matches a linear slope. This is strongly indicative of a recent growth trend that is closer to a constant growth model than that of compound growth.

The consequent predictions of IPv4 BGP table size using this constant growth model are shown in Table 3.
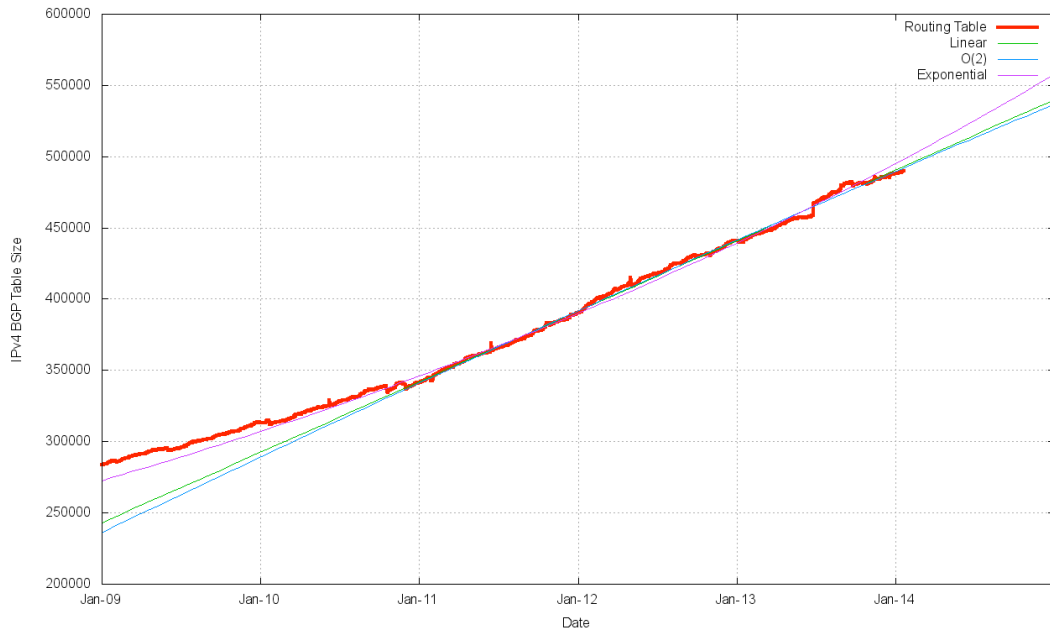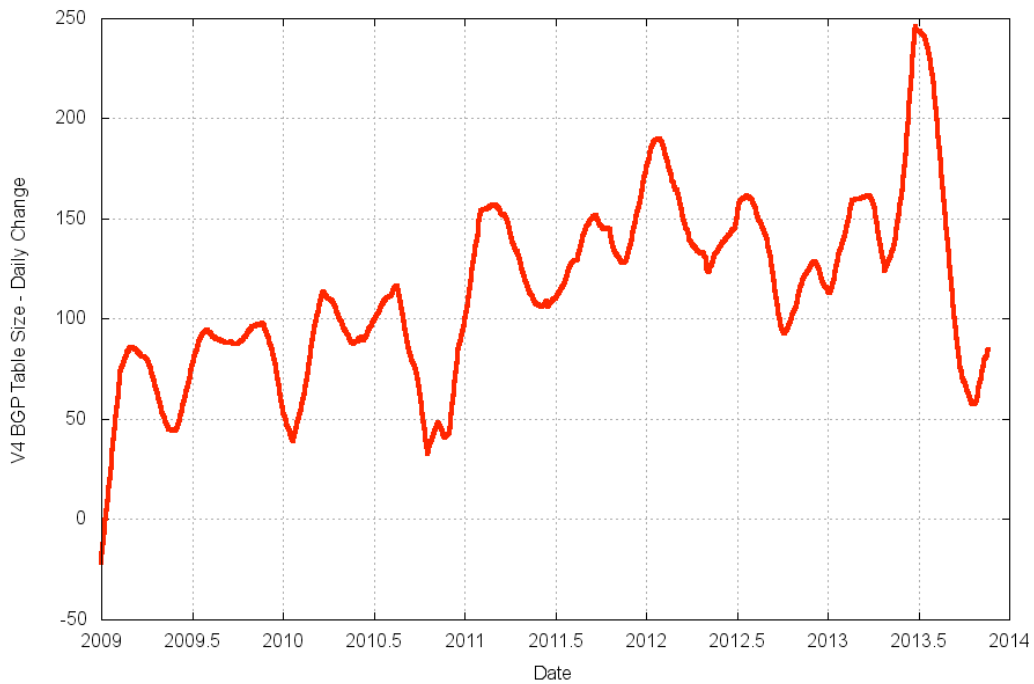
Figure 16 – IPv6 BGP Table 2009 - 2014

Figure 17 -First Order Differential of Smoothed IPv4 BGP Table Size

	IPv4 Table	**IPv4 Prediction**
Jan 2011	341,475	
Jan 2012	390,014	
Jan 2013	441,115	
Jan 2014	487,939	
Jan 2015		540,000
Jan 2016		590,000
Jan 2017		640,000
Jan 2018		690,000
Jan 2019		740,000

Table 3 – IPv4 BGP Table Size Prediction

With the caveat that this prediction is based on the assumption that tomorrow is a lot like today and that the influences that shape tomorrow have already shaped today, then its reasonable to predict that the IPv4 routing table in a little over five years time, at the start of 2019, will contain an additional 250,000 entries, making a total for IPv4 of some 740,000 entries in the BGP routing table at that time.

However I'm not all that confident in the predictions generated by this model. It is simply not possible to use the current models of BGP growth to peer into the longer term aspects of this post-exhaustion IPv4 routing environment. Further growth in the IPv4 routing environment is going to be fuelled in ever-finer granularity in the size of prefix announcements, rather than the release of "new" address space. To what extent this tend of increasingly finer levels of granularity in the routing system is feasible remains an open question.

Forecasting the IPv6 BGP Table

The same technique can be used for the IPv6 routing table. Figure 18 shows the data set for BGP from the 1st January 2009 until January 2014.

The first order differential, or the rate of growth of the IPv6 BGP routing table is shown in Figure 19. The picture for IPv6 was relatively modest in early 2009, with the table growing in size by an average of 2 new entries per day. The rate of growth has increased in the intervening period to the current to the current level of some 15 to 20 new entries per day. Obviously this is far lower than the equivalent figure in the IPv4 domain, which is growing by some 100 - 150 new entries per day, but it does show a consistent level of increasing growth.

This implies that a linear growth model is inappropriate for modelling growth in IPv6. A better fit to the data is a compound growth model, with a doubling factor of some 22 months. It is possible to fit a linear model to the first order differential of the data, which can be used to derive an O(2) polynomial fit to the original data. The fit of a linear, O(2) polynomial and an exponential model against the data is also shown in Figure 20.
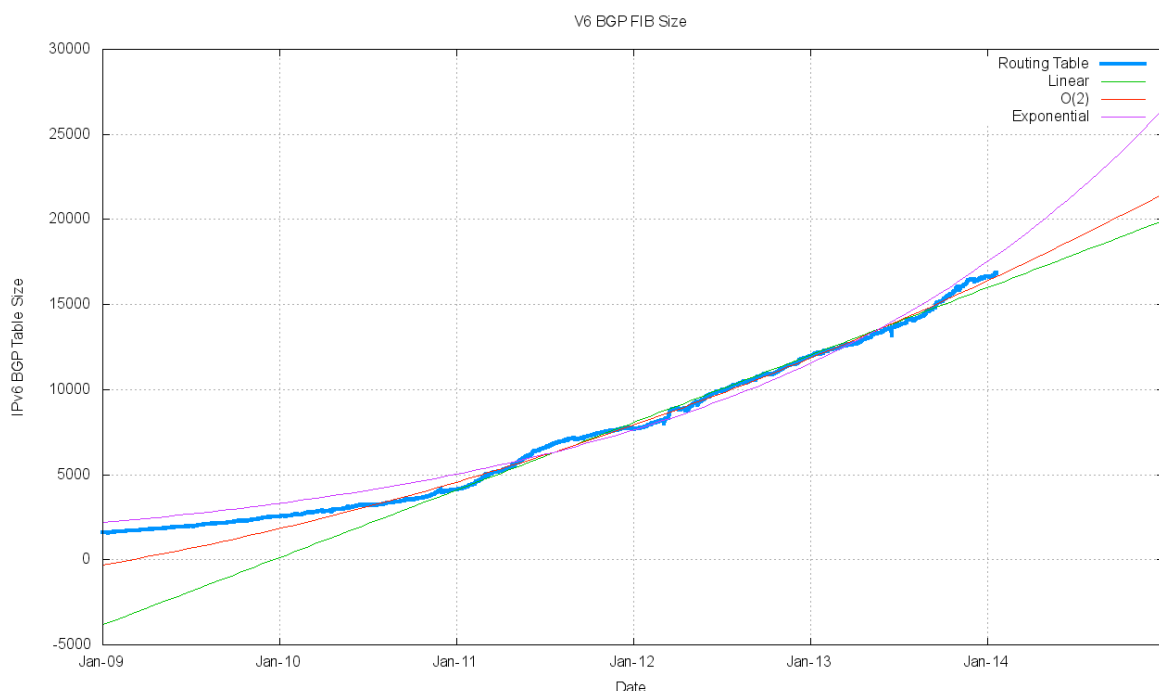

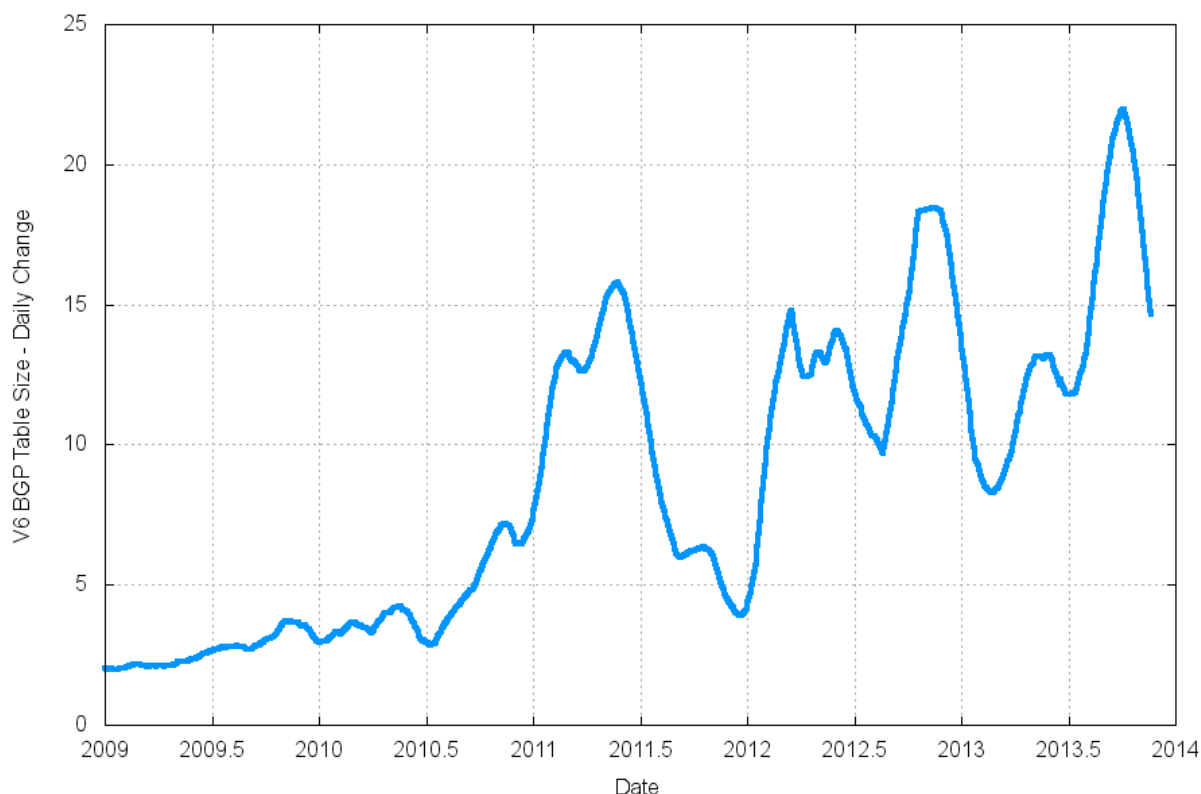*Figure 20 – IPv6 BGP Table Size from 1 January 2004*

Figure 21 -First Order Differential of Smoothed IPv6 BGP Table Size

The projections for the IPv6 table size are shown in Table 4.

	IPv6 Table	IPv6 Prediction Exponential	IPv6 Prediction Polynomial
Jan 2011	4,122		
Jan 2012	7,667		
Jan 2013	11,932		
Jan 2014	16,658		
Jan 2015		26,254	21,534
Jan 2016		40,180	27,276
Jan 2017		60,937	33,637
Jan 2018		92,311	40,582
Jan 2019		139,839	48,128

Table 4 – IPv6 BGP Table Size Prediction

The exponential and polynomial projections in Table 4 provide a reasonable estimate of the high and low bounds of the growth of the IPv6 BGP routing table in the coming years. It should be noted that while the high bound of this growth curve tends to a growth profile that is close to the parameters of Moore's Law (http://en.wikipedia.org/wiki/Moore's_law), with a doubling interval of some 22 months, the mitigating factor is that these numbers are not exactly massive numbers, and in five years time we would be expecting a IPv4 BGP table size of 740,000 entries, or some of times the size of the higher bound of expectations for the IPv6 routing table.

It appears from these projections that for the next five years, the significantly larger size of the IPv4 network will continue to drive the overall costs of BGP routing, and the IPv6 BGP network will live, in effect, in the margins of oversupply in meeting the demands of IPv4. It

will be some time before there is significant change in the relativities of the two protocols from this particular perspective.

Conclusions

These predictions are highly uncertain. Despite this uncertainty, nothing in these figures indicates any serious cause for alarm. There is no clear signal of the need to invent a new inter-domain routing protocol in the near future, assuming that the motivation to do so was based on a reasonable prediction of the imminent collapse of BGP.

None of the metrics indicate that we are seeing such an explosive level of growth in the routing system that it will fundamentally alter the viability of carrying a complete eBGP routing table in the near future. In terms of the projections of table size in the IPv4 and IPv6 networks, the BGP sky is not about to fall on our heads any time soon.

However, if you are in the market for BGP routing systems, you should probably factor in a likely eBGP FIB table size of one million entries across a five year operational timespan of the equipment.

But this is only half the story of BGP growth. The other half of the story is concerned with the characteristics of the dynamics of the BGP protocol, and the rate of growth of updates in BGP and the performance of BGP in terms of the routing convergence properties of the routing system. We'll look at that in the next article.

Disclaimer

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

Author

Geoff Huston B.Sc., M.Sc., is the Chief Scientist at APNIC, the Regional Internet Registry serving the Asia Pacific region. He has been closely involved with the development of the Internet for many years, particularly within Australia, where he was responsible for the initial build of the Internet within the Australian academic and research sector. He is author of a number of Internet-related books, and was a member of the Internet Architecture Board from 1999 until 2005, and served on the Board of Trustees of the Internet Society from 1992 until 2001.

www.potaroo.net