Geoff Huston
February 2016

On the Internet Everyone is Connected to Everyone Else - Right?

> This is a summary of a longer report on connectivity in today's Internet. The full report can be found at http://bit.ly/1TzGLL4.

We tend to make a number of assumptions about the Internet, and sometimes these assumptions don't always stand up to critical analysis. We were perhaps 'trained' by the claims of the telephone service to believe that these communicaitons networks supported a model of universal connectivity. Any telephone handset could establish a call with any other telephone handset was the underlying model of a ubiquitous telephone service, and we've carried that assumption into our perception of the Internet. On the Internet anyone can communicate with anyone else – right?

But if the Internet supports a single cohesive domain of connectivity, then why do we see divergence? One place where this is evident is in the realm of the Internet's routing system. There are so-called "route collectors" on the Internet that collect a number of views of the Internet's routing system some a set of different vantage points. If the Internet was an undistinguished 'flat' domain of connectivity then all of these views would be the same. We would all see precisely the same Internet no matter where the vantage point is located. So why is it that when we take a diverse set of perspectives of the routing system gatherred from a number of vantage points, and place them together, we get a picture that is Figure 1.
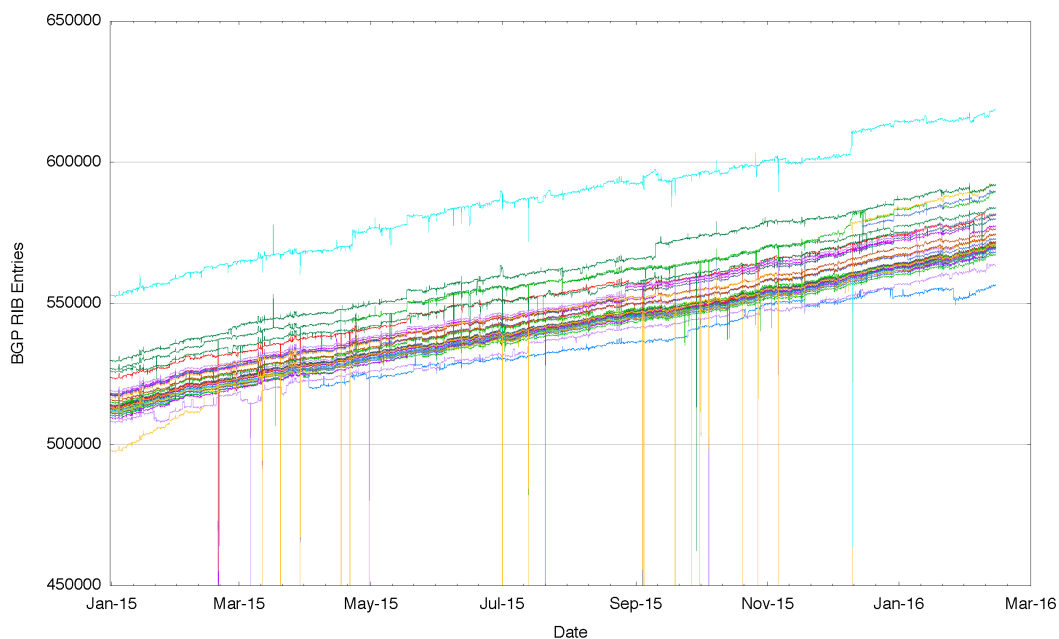

Figure 1 – Number of Routes announced by each BGP peer of Route Views for 2015

What this figure shows us is that each of these routing vantage points sees a slightly *different* Internet from their perspective. And these differences are not just temporary. Across all of 2015 these various routing vantage points see a consistently different number of route objects in their local routing system. So its not something that occured at a particular point in time and corrected via the normal operation of the routing protocol. These differences are consistent across the entire year.

It might be possible to say that these differences have occured as some property of size and age, and that the venerable IPv4 Internet has these divergent routing views solely due to its age and size. So a "younger" and smaller routing system may well show a single view by that reasoning. The second figure to consider that of the IPv6 network over 2015. Rather than some 600,000 entries there are just 20,000. And still we see that some BGP speakers see more routes and others see less, and this is persistent and stable across the entire year.
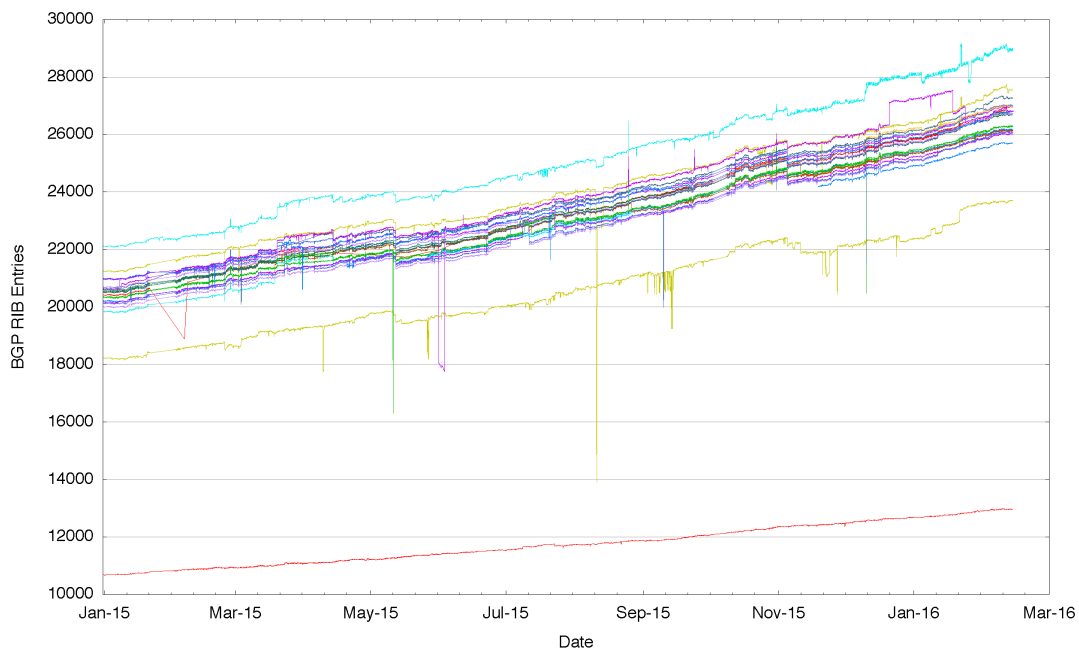

Figure 2 – Number of IPv6 Routes announced by each BGP peer of Route Views for 2015

What do these differences mean? Are we seeing evidence of a fragmented Internet where some places on the Internet cannot reach other places? Are these differences in the perspectives of various routing vantage points signs of underlying fractures of the fabric of connectivity in the Internet?

Before leaping to conclusions here, it's useful to pull together some further data. One possible explanation of the difference in the number of advertised routes is that the routing system contains two components of information: basic reachability and information relating to the policy of how to reach a destination. So it may be that the differences seen from each of these vantage points are not in fact differences in basic connectivity - but the represent differences in these more specific announcements, and they represent that in different parts of the network there may be different preferred ways to reach certain destinations.

One way to test this is to look once more at the route collector data, but this time instead of looking at the number of route objects that are visible at each routing vantage point, lets look at the total span of addresses that are advertised as reachable. If this theory is correct then the total span of addresses should be the same from each of these vantage points and we could well be justified in believing that the Internet is indeed a uniform domain of connectivity.

Again the data does not agree with this theory - when we look at the data (Figure 3) we see that there is a range of some 5 million addresses in IPv4 where some vantage points see a larger set of

addresses advertised as reachable than other vantage points. And again there is a consistency across the year: those peers that see a larger span of addresses than others appear to do so consistently across the entire year. And again, with three exceptions, those that advertise a lower span of addresses do so consistently across the entire period.
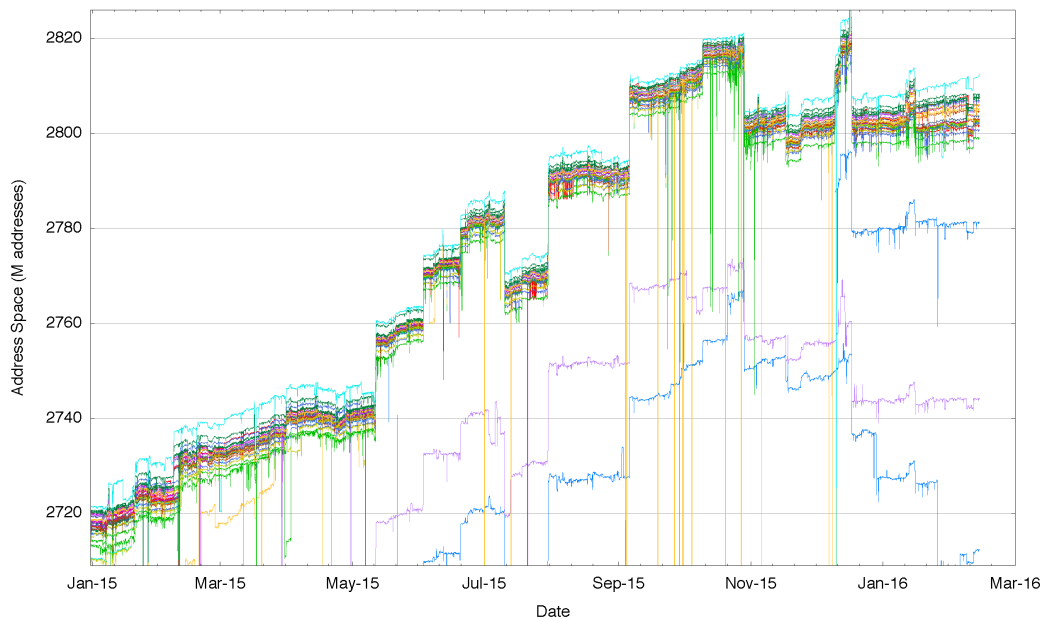

Figure 3 – Aggregate span of reachable IPv4 addresses announced by each BGP peer of Route Views for 2015

In IPv6 we see the same situation, where different vantage points see a different span of addresses. In IPv6 there are large scale differences, so that at the extremes one routing peer sees one half of the total address span of the peer announcing the largest range, while at a finer level of granularity we still see each vantage point seeing a slightly different span of addresses in the routing system.
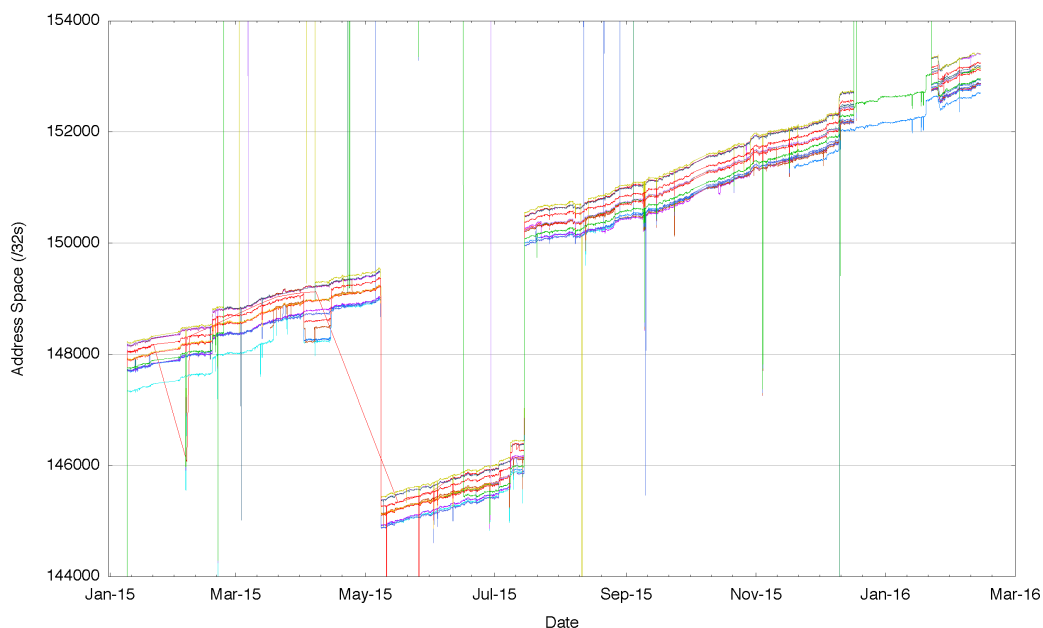

Figure 4 – Detail of Aggregate span of reachable IPv6 addresses announced by each BGP peer of Route Views for 2015

Is there a set of addresses that defines "the Internet?"

No such set of addresses exists.

There is no defined "common core" of Internet connectivity. There is no defined set of "accepted routes" that defines Internet connectivity. Perhaps the best we can do is to borrow a concept from a voting system, and look at the common core of the internet as those addresses that are announced from most peers most of the time. With such a definition we can then look at each individual vantage point and see where the peer differs from this common core. Figure 5 shows the deviation from the connom core of advertised address space, and each peer has a set of addresses that are in addition to the common core, and a second set that are in the common core but not advertised by this peer. This figure shows these numbers day-by-day for each peer of Route Views for the entire year. There is some evidence of day-to-day variability at this level. There also appears to be division in this data at the start of 2016, where a cluster of peers are advertising some 4M addresses that have not been absorbed into the common core address set.
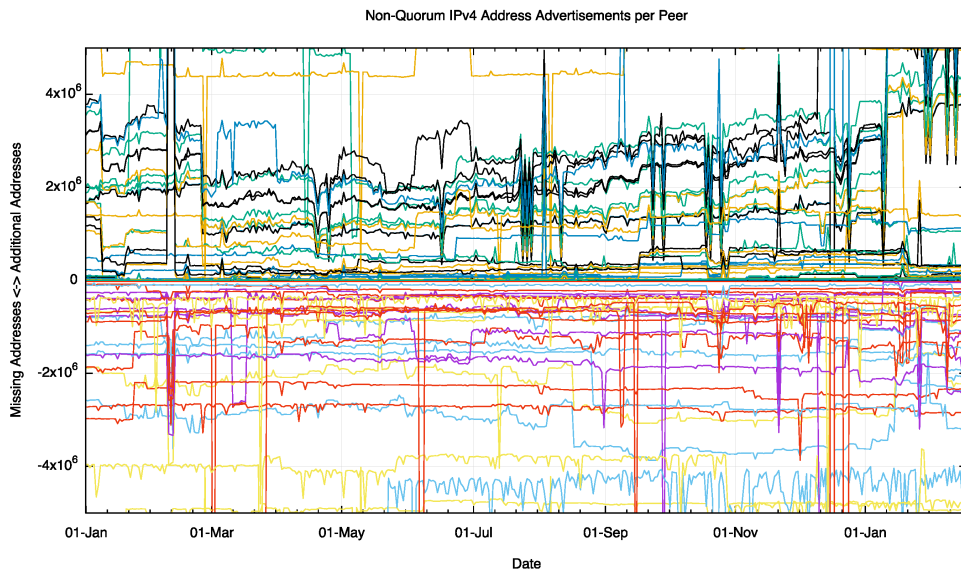

Figure 5 – Non Quorum IPv4 Address Advertisements per peer for 2015
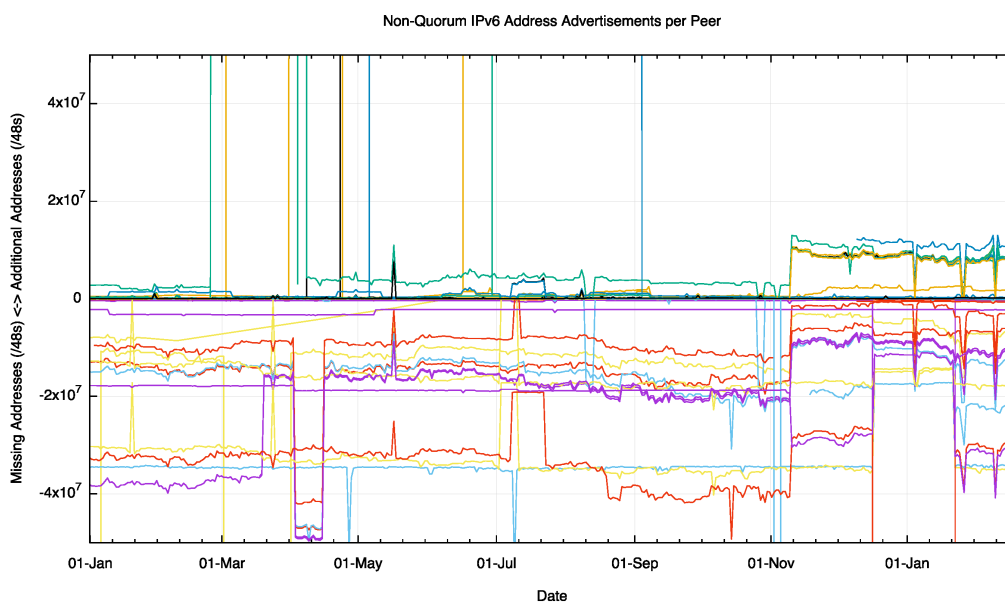

Figure 6 – Non Quorum IPv4 Address Advertisements per peer for 2015

There is a similar story for the IPv6 network, where individual peers can see differences in the address span from the common core set (Figure 6).

How serious is this issue?

The basic question is perhaps unanswerable in any precise manner. This is the question: "To what extent are attempts by one Internet user to contact anoher user, or some service, prevented by a structrual fragmentation of the interconnectivity of the Internet?"

Or, to put the question in the opposite sense, why isn't all of the Internet fully connected?

Surely this is a case where individual motivations coincide with the common good. Each connected network is best served by being reachable from the entirety of all other connected networks, and network is potentially detrimentally affected when there are networks that cannot reach them. This is also a symmetric desire, in that the same applies to the set of networks that can be reached by this connecting network. The theoretical value of the connection is maximised when the network can reach, and be reached by, all other connected networks.

In practice, however, it is not possible to purchase a service that guarantees such universal reachability. Service providers strive to fulfil such expectations on the part of their customers, but universal connectivity falls into the category of 'best effort' as distinct from "service garantee".

Why is this?

Universal interconnection is not a requirement imposed by any regulatory fiat, nor by any deliberate arrangement between network operators. Interconnection is its own market, and the outcomes can be viewed as market-based outcomes. Each individual service provider network operates in a domain or "peering" and "tiering", and if you add the customer routes, the result is the accumulated route set. A network that has a different collection of customers, peers and upstreams may well see a slightly different set of reachable addresses.

One view is that is a surprising outcome that the connectivity on the Internet is as stable and as comprehensive as it is, given that this is market driven activity of peering and tiering is one that comes without any particular guarantees of the "right" outcome.

But then again, perhaps it's not as surprising as that. Another, admittedly cynical, view is that its all about what one would loosely call an "informal cartel" of the tier 1 providers that are at the core of connectivity. As long as each connecting network takes the effort to ensure that their routes are advertised to at least one tier one router via one or more customer / provider relationships then some level of basic connectivity is an outcome. After that basic connectivity is achieved, then peering is there to minimise the cost and/or improve the service for selected routes. Within this perspective Internet-wide connectivity is defined almost completely by the ability to have one's routes passed into the tier one provider cartel. This group of peered interconnected networks essentially define what it is to be connected in the Internet. So another view of Internet connectivity is that this is not a distributed open market for connectivity, and instead we have a self-perpetuating routing monopoly at its 'core'!

But the core tier 1 networks are not the entire Internet – there are routes that are used that are not necessarily passed into the core cartel, and these routes may well be part of the set of routes that have some level of variable reachability. The result is what we see in the figures above. There is considerable variability in terms of who sees which addresses and this impacts on who can connect to whom.

Is the Internet fully interconnected?

No.

Around its edges there is a grey zone of connective asymmetry where you might be able to send a packet to me, but that does not mean that you get to see my response!

This is a summary of a longer report on connectivity in today's Internet. The full report can be found at http://bit.ly/1TzGLL4.

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 building the Internet within the Australian academic and research sector in the early 1990's. 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 and chaired a number of IETF Working Groups. He has worked as an Internet researcher, as an ISP systems architect and a network operator at various times.

www.potaroo.net

Disclaimer

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