Are we there yet?

The saga of the IPv6 transition continues to surprise us all. RFC 2460, the first complete effort at a specification of the IPv6 protocol was published in December 1998, more than twenty years ago. The entire point of IPv6 was to specify a successor protocol to IPv4 due to the prospect of running out of IPv4 addresses. Yet we ran out of IPv4 addresses more than a decade ago. This transition to IPv6 has been going on for 20 years now, and if there was any urgency that was instilled in the effort by the prospect of IPv4 address exhaustion then we've been living with exhaustion for a decade now. So perhaps it's time to ask the question: How much longer is this transition going to take?

This was the question that was put to a panel at the recent ARIN 49 meeting, and, predictably, there was no clear consensus as to what the answer might be. Here I'd like to explore this question here in a little more detail.

A little bit of History

In 1991 it was clear that IP had a problem. It was still a tiny Internet at the time, but the growth patterns were exponential, doubling in size every 12 months. We were stressing out the pool of Class B IPv4 addresses and in the absence of any corrective measures this address pool would be fully depleted in 1994. We were also placing pressure on the routing system and the deployed routers in 1992 only had enough memory to support a further 12 to 18 months of routing growth. The combination of these routing and addressing pressures was collectively addressed in the IETF at the time under the umbrella of the ROAD effort (RFC 1380).

There was a collection of short, medium and longer term responses that were adopted to address the problem. In the short term we dispensed with the class-based address plan and instead adopted the variably sized address prefix model. Routing protocols, including BGP, were quickly modified to support these classless address prefixes. Variably-sized address prefixes added additional burdens to the address allocation process, and in the medium term we adopted the Regional Internet Registry structure to allow each region to resource their address allocation and registry functions. This increased specificity of address allocations and adequate resources that permitted a more diligent application of relatively conservative address allocation practices permitted a significant increase in address utilization efficiency. At this time the concept of "address sharing" using Network Address Translation (NATs) also gained traction in the ISP word. Not only did this dramatically simplify the processes in ISPs, NATs also played a major role in reducing the pressures on address consumption.

While these measures pushed a two-year imminent crisis into a manageable decade-long scenario, they were not considered to be a stable long-term response. This response really needed to extend the 32-bt address field used in IPv4.

There was no way that any such change would be backward compatible with the installed base of IPv4 systems. As a result, there were a few divergent schools of thought as to what to do. One approach was to jump streams and switch over to use the Connectionless Transport profile of the OSI protocol suite and adopt OSI NSAP addresses along the way. Another was to change as little as possible in IP except

the size of the address fields. And there were a number of ideas being thrown about in the area of proposing significant changes to the IP model.

By 1994 the IETF had managed to settle on the minimal change approach, which was IPv6. The address field was expanded to 128 bits, a Flow ID field was introduced, fragmentation behaviour was altered and pushed into an optional header and ARP was replaced with multicast.

The bottom line was that IPv6 did not offer any new functionality that was not already present in IPv4. It did not introduce any significant changes to the operation of IP. It was just IP, with larger addresses.

Transition

While the design of IPv6 consumed a lot of attention at the time, the concept of transition of the network from IPv4 to IPv6 did not.

Given the runaway adoption of IPv4, there was a naive expectation that IPv6 would similarly just take off, and there was no need to give this much thought. In the first phase, we would expect to see applications, hosts and networks adding support for IPv6 in addition to IPv4, transforming the internet into a dual stack environment. In the second phase we could then phase out support for IPv4.

There were a number of problems with this plan! Perhaps the most serious of these was a resource allocation problem. The Internet was growing extremely quickly, and most of our effort was devoted to keeping pace with demand. More users, more capacity, larger servers, more content and services, more responsive services, more security, better defence. All of these shared a common theme: scale. So we could either concentrate our resources on meeting the incessant demands of scaling, or we could work on IPv6 deployment. The short and medium-term measures we had already taken had addressed the immediacy of the problems of address depletion, so in terms of priority, scaling was a for more important priority for the industry than IPv6 transition.

The scaling problem accelerated by a whole new order of magnitude in the mid 2000's with the introduction of the iPhone and its brethren. All of a sudden this was not just a scale problem of households and enterprises, but it transformed to a scale problem of individuals and mobility. The entire reason why IPv6 was a necessity was coming into fruition, but we were just not ready to deploy IPv6 in response. So we increased our consumption of the remaining pools of IPv4 addresses and we supported the first wave of large scale mobile services with IPv4. Dual stack was not even an option in the mobile world at the time. The rather bizarre economics of financing 3G infrastructure meant that dual stack infrastructure in a 3G platform was impractical.

At the same time the decentralised nature of the Internet was hampering IPv6 transition efforts. What point was there in developing application support for IPv6 services if no host had integrated IPv6 into its network stack? What point was there in adding IPv6 to a host networking stack if no ISP was providing IPv6 support? And what point was there in an ISP in deploying IPv6 if no hosts and no applications would make use of it? So nothing happens.

The first to try and break this impasse of mutual dependence was the operating system folk, and fully functional IPv6 stacks were added to the various flavours of Linux, Windows and MAC OS, as well as in the mobile host stacks of iOS and Android.

But even this was not enough to allow a transition to achieve critical momentum. It could be argued that this situation actually made the IPv6 situation worse and set back the transition by some years. The problem was that with IPv6-emnabled hosts there was some desire to use IPv6. However, these hosts were isolated "islands" of IPv6 sitting in an ocean of IPv4. Then concentration of the transition effort then fixated on various tunnelling methods to tunnel IPv6 packets through the IPv4 networks. While this can be performed in a manual manner when you have control over both tunnel endpoints, this was not that useful an approach. What we wanted was an automated tunnelling mechanism that took care of

all these details. The first such approach that gathered some momentum was 6to4. The first problem with 6to4 was that is required public IPv4 addresses, so it could not provide services to IPv6 hosts that were behind a NAT. The more critical problem was that firewalls had no idea who to handle 6to4 packets, and the default action when in doubt is to deny access. So 6to4 showed a 20% - 20% failure rate, which made it all but unusable. The NAT issue was also a problem, so a second auto-tunnel mechanism was devised that performed NAT sensing and traversal. This mechanism was even worse in terms of failure rates, and some 40% of Teredo connection attempts were observed to fail. (https://www.potaroo.net/presentations/2011-03-31-dualstack.pdf).

Not only were these initial transition tools extremely poor performers, as they were so unreliable, but even when they worked the connection was both fragile and slower than IPv4. The result was perhaps predictable, even if unfair. It was not just the transition mechanisms that were viewed with disfavour, but IPv6 itself also attracted opprobrium.

Up until 2011 IPv6 was largely ignored as a result. A small number of service providers tried to deploy IPv6, but in each case they found themselves with a unique set of challenges that they and their vendors, had to solve, and without a rich set of content and services on IPv6 the value of the entire exercise was highly dubious! So, nothing happened.

Movement at last!

It wasn't until the central IPv4 address pool managed by the IANA was depleted at the start of 2011, and the first RIR ran down on its general allocation pool in April of that year, that the ISP industry started to pay some more focussed attention to this transition.

At first these were faltering steps, but over the past decade the transition has gathered momentum. A plot of the estimated share of IPv6 users who can access IPv6 services is shown in Figure 1.
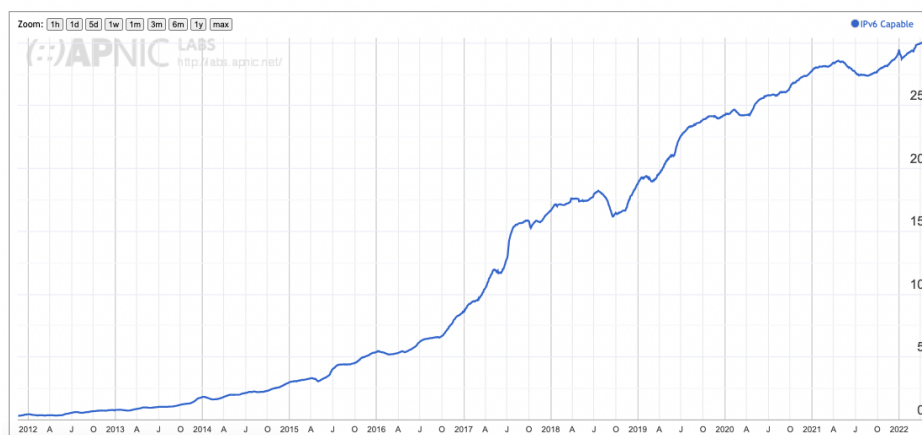

Figure 1 – IPv6 Uptake as a % of all Internet Users – 2012 to the present – from https://stats.labs.apnic.net/ipv6/XA

Across the 2-year period 2016 – 2017 the IPv6 numbers tripled from 5% at the start of 2016 to 17% at the end of 2017. Much of this was due to the rapid deployment of IPv6 in mobile networks in India at that time. In the following 3-year period this number rose from 17% to 30%. Much of this second growth phase can be attributed to the deployment of IPv6 in China. However, its not just a story about deployments India and China. Other economies have seen a steady and sustained growth in IPv6 capability across the entire 5-year period, such as Mexico, Brazil and the United States If we are taking a proportion of IPv6 deployment across users as our metric, then at some 30% of all users, then it's reasonable to conclude that transition is well and truly underway. Obviously this 30% is not uniformly distributed. When we take national economies into account, we also see a varied picture of the current state of this transition (Figure 2).
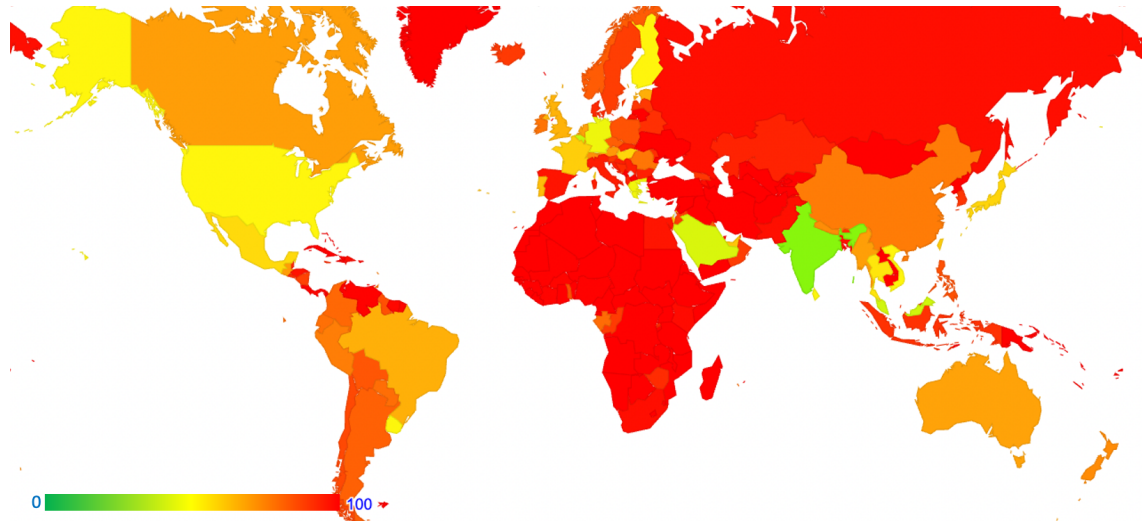
Figure 2 – IPv6 Uptake in April 2022- from https://stats.labs.apnic.net/ipv6

IPv6 transition is underway in a relatively small subset of national communities. Just 32 national economies have IPv6 adoption rates above the global average of 30%. The level of IPv6 adoption appears to be highest in South Asia, North America, and Western Europe, and the lowest adoption rates are seen in Africa and the Pacific Oceania. However, such comparisons at a national level can be misleading, as they place China (population 1.4B) on equal terms with Pitcairn Island (population 50). It we look at the 10 largest national populations of IPv6 users we get a somewhat different view (Table 1).

Economy	V6-Capable Users (est.)
India	455M
China	205M
USA	123M
Brazil	57M
Japan	49M
Mexico	39M
Germany	36M
Vietnam	26M
UK	23M
France	22M

Table 1 – IPv6 National IPv6 User Population – from https://www.potaroo.net/bgp/iso3166/v6dcc.html

This data substantiates the observation that IPv6 adoption in large-scale consumer networks is well underway in many parts of the world.

How much longer?

Now that we are in the middle of this transition, then next question is how much longer is this transition is going to last?

This seems like a simple question, but it does need a little more elucidation. What is the "end point" when we can declare the transition to be over? When will this transition be "complete"? Is it the time when there is no more IPv4-based traffic on the internet? Or is it the time when there is no requirement for IPv4 in public services on the Internet? Or do we mean the point when IPv6-only services are viable? Or perhaps we should look at the market for IPv4 addresses and define the endpoint of this transition at the time when the price of IPv4 addresses completely collapses? Perhaps we can take a more pragmatic position here and rather than looking for completion as the point when the Internet is completely bereft of all use of IPv4 addresses we can define completion as the point when use of IPv4 is no longer necessary. This would imply that when a service provider can operate a viable Internet service using only IPv6 and having no supported IPv4 access mechanisms, then we would've completed this transition.

What does this imply? Certainly, the ISP needs to provide IPv6. But as well all the connected edge networks and the hosts in these networks need to support IPv6. After all, the ISP has no IPv4 services at this point of completion of the transition. It also implies that all the services used by the clients of this ISP must be accessible over IPv6. Yes, this includes all the popular cloud services and cloud platforms, all the content streamers and all the content distribution platforms. It also includes specialised platforms such as Slack, Xero, Atlassian and similar.

We are certainly not there today, and not likely to reach this point in the next 2 or 3 years. The data published at the World IPv6 Launch site (https://www.worldipv6launch.org/measurements/) reports that only some 30% of the Alexa Top 1000 web sites are reachable over IPv6, and clearly a lot of service platforms have work to do, and this will take more time. In the ARIN 49 Panel the prognostications as to how long "more time" might be there were various views that tended towards numbers from at least a further decade to a further quarter century of transition.

Why is this transition taking so long?

To state the entirely obvious, it's taking so long because there is no common sense of urgency. While some actors have made the necessary changes to deploy dual stack services and infrastructure others do not feel that such actions are a current priority. And for as long as deferring any such action does not have unacceptable business consequences, then deferral just keeps on happening.

Part of the issue here is that there is no major competitive advantage to be realised by adopting IPv6. It provides no unique capability over IPv4 that would result in greater efficiencies, lower costs, or unique service profiles. The initial phase of operating a dual stack environment represents a higher cost and support overhead and little in the way of offset benefit.

The case for IPv6 was originally based on risk aversion. Address sharing in the form of NATs was seen as an unacceptable measure, and when the pool of available IPv4 address was depleted, it was thought that further growth of the Internet would be impossible, as we would comprehensively eschew the use of NATs. The timely adoption of IPv6 was seen as an imperative measure to avoid such a situation.

However, NATs were not an ISP issue at first. ISPs effectively outsourced NATs to CPE vendors and thereby pushed the entire issue to end users and the application space. Application designers were faced with the simple reality that either their application operated seamlessly in the presence of NATs, or it just did not work. Applications moved away from end-to-end connection models and instead adopted server/client service models. The next step was taken with the deployment of mobile networks, where there was no CPE at the client end of the connection. ISPs pulled the NAT into their own infrastructure and the use of so-called "Carrier Grade NATs" (CGNs) become commonplace. All of this could've been avoided had we completed the transition prior to the mass uptake of mobile services. However as this was happening during the transition every mobile service provider had to provide an answer to provide IPv4 services to their customer base, irrespective of whether they had already deployed IPv6 or not. CGNs on one form or other was a mandatory requirement for all. It was IPv6 that was the optional extra.

It certainly seems that the pressures to embark on a dual stack service are felt in different parts of the Internet at different levels of intensity. Some operators have felt compelled to deploy IPv6 with some urgency to alleviate the pressures on an otherwise insufficient pool of public IPv4 addresses to support their CGNs, while other operators evidently do not feel any particular current pressure to deploy and are willing to wait for the "right" moment.

Signalling Common Needs

It's useful to remember that the Internet is not a single entity. There are many component networks, including consumer retail service networks, enterprise networks, Internet of Things support networks, and so on.

There is a broad diversity of providers in the Internet ecosystem. There are IP access operators, IP carriage providers, platform providers, chip manufacturers, application providers, content platforms and so on. These individual actors normally communicate their respective needs to with each other through market signals, and market signals are normally expressed as price. Goods and services under demand experience price pressure, while goods and services no longer needed experience a price slump.

However, for many years we withheld IP addresses from such market-based distribution mechanisms. There were many reasons behind this, including the consideration of addresses as a public good and the desire to exercise conservation principles over the consumption of addresses in order to achieve some form of equity of access to addresses ("fairness") as well as countering potential disruptions of address distribution through market distortions.

The result was that signalling impending scarcity of IPv4 was not performed by conventional pricing mechanisms. It was only when the address transfer market was added to the environment in the 2010's that pricing information was available as a form of market signalling.

The record of transferred IPv4 address prices over time is shown in Figure 3. If price movement is a signal of a change in market sentiment, then the period of 2014 to the start of 2018 shows no substantial change in common sentiment. The relatively slow pace of price increase shows no significant level of concern over looming scarcity of supply of IPv4 addresses. Even the price increase in 2018 was offset by the steady, if not slightly declining, price point across 2019 and 2020. The situation changed radically over 2021, and the price doubled over this period.

The escalating price for IPv4 addresses signals that at present demand is exceeding supply.

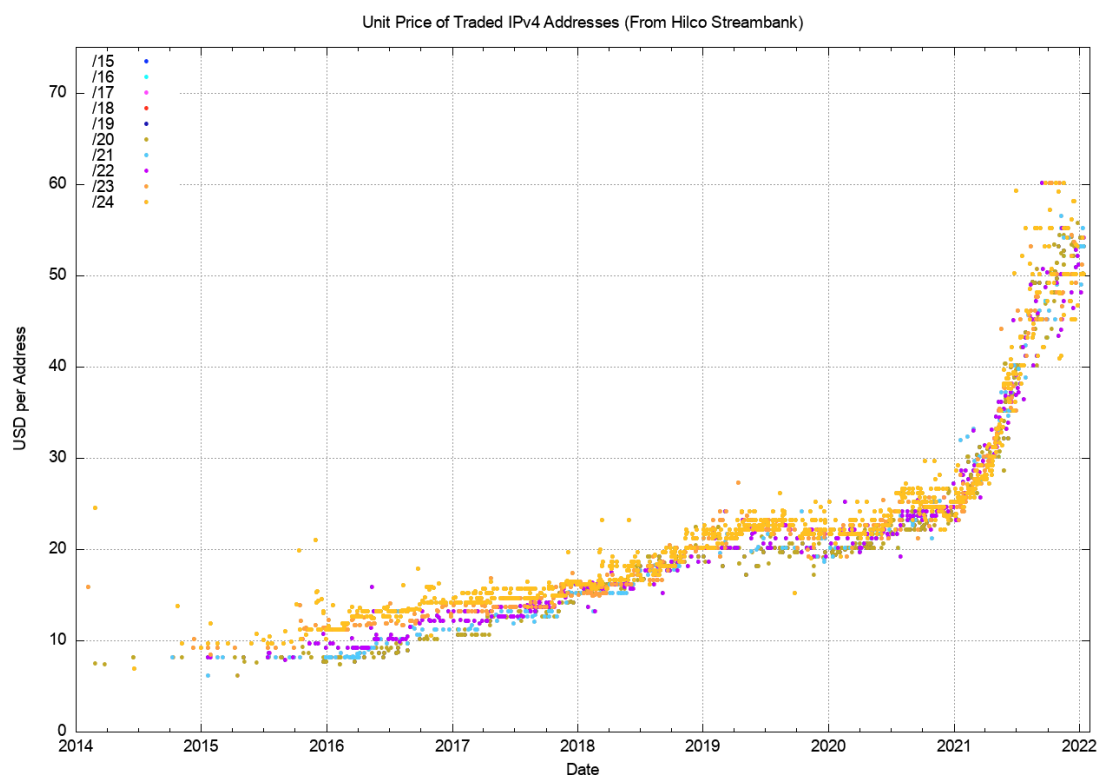Markets are not all that good in predicting the future, but they are useful in informing us where we are.

Figure 3 – IPv4 Address Prices 2014 – 2022 – from Hilco Streambank

What does the current address price escalation tell us? It appears to be saying that there is no common consensus that the transition is coming to completion, and that demand for IPv4 address will continue to outpace supply for some time yet. But as to how long this situation will continue, market pricing information really has little to say.

What would we expect to see when the transition to IPv6 is coming to an end? Presumably at that point there is no further demand for IPv4, the market position would flip from demand driven to oversupply. Given that the transition is coming to an end in this scenario, then there would be no prospect of any resurgence of demand. This would cause the market to collapse.

Is transition going to take forever?
If we can't predict when this transition will be effectively over, can we at least form an opinion on whether this transition will ever be completed at all?

It strikes me that there is a human temptation to regard the current situation as some form of imposed default. Now that we are some 25 years into this IPv6 transition, and the Internet appears to be functional for users and the services that they access, then we find it natural to believe that this situation can continue for some time further, or even indefinitely.

But whether or not you might conclude that this situation is tenable indefinitely depends on the model of Internet growth you are prepared to believe. If the growth phase of the Internet is effectively over and we are now looking at a saturated market then yes, we are in a tenable situation where we are right now. But in this current situation there is little further ability to absorb more growth given that new entrants would still require some form of access to IPv4 addresses, and this situation will become less and less tenable as the growth pressures continue. What this implies is that the established markets, including the mass-market residential services in many parts of the world, are not necessarily under any major pressure to expand their platform and secure more IPv4 addresses. But that's by no means the entirety of the network environment. The areas of growth appear to be in the cloud service space and the IOT space, as well as the continued expansion of conventional retail Internet access markets in Africa, and parts pf Asia. The current estimates of the Internet's user population is some 4.2 billion out of a total world population of 7.8 billion. There is still some growth that the Internet has to encompass.

Is transition to IPv6 the only option?
This conversation about transition assumes that IPv6 is the only available option, and if this is the case then the continual expansion of the scope of digital systems into a world of embedded objects tends to make the case that IPv6-only services are inevitable sooner or later. Perhaps we should be asking if this view of a very limited set of options, namely one, is indeed an accurate one.

We have explored a number of different network models over the years, and one that appears to offer a potentially viable alternative here are name-based networks. In its original concept of name-based networking we thought about replacing the address fields in a packet header with service identifier names and route and forward packets based on these names. When viewed from a sufficiently large distance endpoint identification based on integer assignments and endpoint identification based on some encoded form of character string are largely isomorphic. The point is that name-based networking is not a terribly different concept, and the essential different is an even larger token space with a sparse usage pattern.

However, what we have built in the heavily NATted world of IPv4 is something a little different. When we resolve a service name in the DNS, the DNS attempts to provide us with an address that allows us to access the service. The address is not necessarily unique to that particular service name, as resolution queries for different DNS service names may also return the same IP address. The address is not fixed in time, in that the same DNS resolution query made at a different time may result in different addresses. And the address is not universal, as different users making the same DNS resolution query may receive different addresses.

This potential ambiguity in the mapping of service names to IP addresses is resolved at the transport and application level. When a connection is made to the address, then the client is required to name the service that it wants to connect to. This explicit service identification is part of the Client Hello exchange in the TLS protocol, part of the HTTP initial exchange, and part of many other service protocols. The intent of this level of indirection is to allow a server to host many services and allow a service to be hosted

on many servers. The result is that IP addresses are simply not uniquely associated with a services or network endpoints. Today's network is in fact a version of a named network, and the role of IP addresses is ephemeral session-level tokens that allow the network to distinguish between concurrent packet flows and little else.

As we continue to explore this digital space it is a little presumptuous to assume that the architectures that we devised in the 1970's about packet switched networking are the only ones that can work, and that we will never come up with different architectures. Today's combinations of NATs, anycast systems, content distribution networks and increased application-level functionality all point to another transition which is quietly underway on the Internet already. It's a transition that no longer relies on the IP layer as the universal adaption protocol between diverse network media and diverse applications but pushes the critical common dependence further up the protocol stack into the application layer.

Are we there yet with IPv6?

Well, clearly no.

But we are closing in. If the end point is being able to provide a IPv6-only service to customers that meets their requirements, then ongoing efforts of IPv6 adoption in the content and service platform space are having some impact on when we might reach this point. I have not seen published data on the provisioning rules of the number of pooled IPv4 addresses used in CGNs vs. the size of the user base, but with the combination of increased use of IPv6 in the service platforms and the use of IPv6 preference rules in applications I strongly suspect that this ratio of required IPv4 pool size against the customer base is declining over time.

That leads to the second observation that it's likely that we will not reach the end of this transition with a bang, but with a whimper. It's not that all ISPs will shut down their CGNs and quite IPv4 on any particular date. A more likely scenario is that the size of the required pools of IPv4 addresses to service the ISP's client base will continue to decline. At some point it makes no business sense to continue to devote resources to continue to operate these CGN services within the ISP's own infrastructure. Customers with a requirement for IPv4 services would need to find a different ISP that still offers IPv4 or switch over to an external IPv4 service and use some form of VPN tunnel to access it. This will likely not occur everywhere all at one, but in a piecemeal fashion over an extended period of time.

So, when will this transition end?

I still just don't know!

The recording of the Panel Session ARIN 49 on the IPv6 transition can be found at https://www.youtube.com/watch?v=jsgFCiTB15Y&t=17654s

Disclaimer

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

Author

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

www.potaroo.net