

September 2011

Geoff Huston

Transitional Uncertainties

The telecommunications industry has been around for quite some time. Whether you take as a starting date the first efforts with the wired telegraph in the 1830's, or the telephone in the 1870's, this industry has been around for quite a long time. During this periods it has made huge achievements, and there is no doubt that the impacts of this industry have changed our lives in many ways. Indeed, this industry has a rich history of achievement. It is literally amazing that this industry has managed to preserve dial tone on telephone handsets while completely changing the underlying network and switching fabric of the telephone system numerous times.

But not all of this history is a story of unqualified success. The telephone industry's visions of broadband data networking represented some pretty poor technology choices. I would not describe ISDN as an unqualified success, and the efforts to construct a useful broadband data infrastructure using ATM as the ubiquitous substrate was not exactly a high point for this industry. And many of the recent successes in this industry have come as a complete surprise. Nobody in the industry expected that SMS would be used at all, let alone become so popular. The SMS user interface on traditional numeric keypads was about as unfriendly as one can imagine and the limited message length was incredibly limiting. Why would anyone want to use it? Yet they did, in their millions!

SMS was not the only surprising success for this industry. The Internet itself came as a complete surprise to the traditional telco industry of the time. Indeed, the success of the Internet probably came as somewhat of a shock to the early developers of the Internet as well, given that the mantra of the early days of the Internet was that this was an "experiment" undertaken within the academic and research sector that would be useful to seed ideas for subsequent broadband data networking within the far larger telecommunications industry. It may have been a wistful thought, but it was never the firm intention that this particular exercise in network research was to inherit the load of the entire global telecommunications infrastructure. It just happened that way, and it was surprising to everyone.

So the telecommunications industry gets things wrong just as easily as it can get things right. In many ways the size of the industry is no indicator of its ability to make astute technology choices. Being large, and commanding vast resources in terms of workforce and capital does not necessarily help in making the right decisions. Some argue that it increases the probability of getting it wrong!

This then brings me onto the obvious question: How will this industry manage the transition from IPv4 to IPv6? Will it be successful? Or will our children assign this exercise to the top shelf of the industry's record of failures?

When I've posed this question to my colleagues, I've received a mixed set of reactions, but the most common reaction is one that simply asks why we need to worry about this question at all. Surely, they say, IPv6 is simply inevitable, and this is just a matter of sitting back and watching it happen. The exhaustion of the entire IPv4 address pool is inevitable, and the increased industry pressures from IPv4 address scarcity, including escalating pricing for IPv4 addresses in an address after-market all serve as drivers for IPv6 adoption. The higher the IPv4 price goes, the greater the incentive for IP service providers to seek relief by migrating their customers into IPv6. All of this process of transition is just inevitable, and there is simply no need to worry.

Its all very reassuring, but I'm not sure that it is convincing. This month I'd like to look at this proposition in a bit more detail and look at whether IPv6 is an inevitable outcome of IPv4 exhaustion, or whether we should worry, even just a little, about what is going on with this transition.

Technology Transition

When we look back at previous technology transitions in our industry, they all look just so logical. The early telephone networks used a dedicated pair of conductors to carry a single conversation. Scaling the telephone networks of the late nineteenth century was a matter more of the logistics of putting more and more copper wires up on poles in the cities, and the scarcity factors were centered around the limits to the number of discrete wires that could be hung from single poles than any other factors. The key technological innovation was the concept of replacing a physical electrical circuit with a virtual circuit. A number of these virtual circuits could be placed on a single conductor pair through frequency division multiplexing, in exactly the same way that multiple radio stations coexist in the spectrum space. The switching elements between trunks was augmented with electrical oscillators and now a single conductor could carry hundreds of telephone conversations simultaneously. The virtual circuit model was further refined in response to more sophisticated electronic capability, and the second generation of virtual circuitry was constructed using a conversion of the analogue conversation to a digital data stream, and replacing the frequency division multiplexing technique with time division multiplexing. Again, in retrospect it all looks so logical and so inevitable.

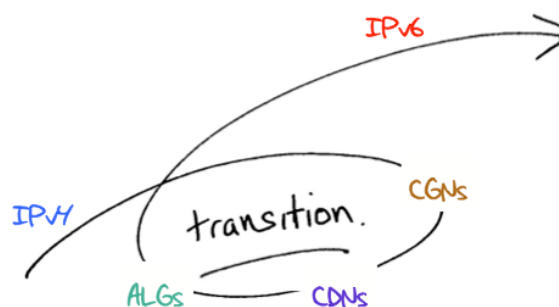
Even the transition from circuit to packet switching looks like an inevitable progression in retrospect. Instead of the network providing a constant time base for the transmitted data, the data unit itself carries sufficient information to allow each switching point to direct the data element towards its destination, and the sequence and timing information is pushed deeper into the data frame to become part of the end-to-end information stream, rather than trying to force the network into preserving timing and sequencing. These changes allow networks to be constructed in ways that are both faster and cheaper. And a combination of a technology shift that produces both financial and service opportunities is often irresistible.

What about the transition from IPv4 to IPv6? Is this also an inevitable transition? There is no doubt that the designers of IPv6 certainly envisaged this as an inevitable transition. But there are some challenges here.

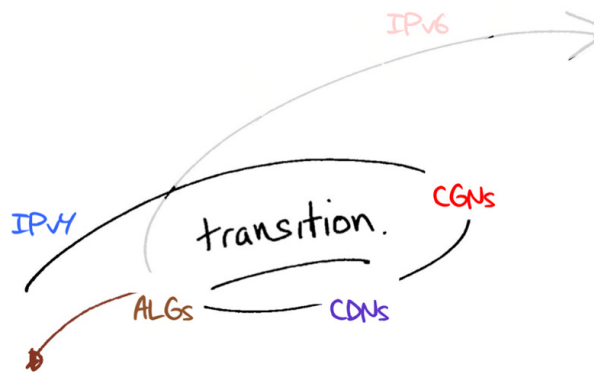
The first is that the transition does not provide for backwards compatibility. A host cannot switch from using IPv4 to IPv6 and still communicate with all the hosts still using IPv4. So the transition has an essential "dual stack" phase where, during the transition, hosts operate with both protocol stacks concurrently, using the IPv6 protocol stack to speak to other IPv6 hosts and the IPv4 protocol stack to speak to other IPv4 hosts. This lack of backwards compatibility in IPv6 makes the transition slightly more complex, but not prohibitively so. What it does mean is that applications and host operating systems need to be aware of IPv6 and explicitly have capabilities to use IPv6. It's not a seamless augmentation at the application interface level.

There's an additional challenge here that is formidable, and one that was largely unforeseen when IPv6 was being designed. At the time there was the general impression that the telecommunications industry behaved prudently, and given the warnings of the prospect of exhaustion of the IPv4 address space, industry actors, being prudent and risk averse, would embark on the transition to IPv6 well in advance of IPv4 address exhaustion. And one or two did. But everyone else did not. And now we have the challenge of trying to undertake this dual stack transition while one stack is critically short of further address space. This factor radically alters the dynamics of the transition. In order to make the IPv4 part of the transition work for the requisite number of additional years it will be necessary to deploy additional "middleware" in the network, and head in a different direction architecturally.

The most obvious shift is probably going to be one of deployment of Carrier Grade NATs (CGNs) in access networks. This will allow a single public IPv4 address to be shared across multiple end clients. The longer the transition takes the more likely that this alone will not be sufficient, and we may expect to see a push to re-architect content into Content Distribution Networks that have points of presence in major access networks. It is also possible that network providers may resort to Application Level Gateways (ALGs) and managed services in an effort to further contain the level of IPv4 address and port consumption by user services.



The risk here is that after making this additional capital investment in network infrastructure, the network service provider is then highly motivated to protect the value of this investment. What lengths will network service providers be prepared to go to in order to protect this investment in transitional services? And if these transitional services generate higher revenues for the network service provider than basic commodity packet transit services, to what extent is the network service provider then motivated to lock itself into this "transitional" service model for an extended period? Would this imply that rather than being a transitory state we see these changes to the network lasting for an indefinite period.



If one sector of the industry finds that this transitional model of providing services sufficiently attractive, is it possible that it could have sufficient market influence such the entire service provider industry collectively locks into this "transitional" model as an enduring service model? If this was to eventuate the internet would be driven in an entirely different direction than IPv6!

Challenges to Transition

Is it possible to manage this level of risk in the transition, and thereby ensure that this industry is not distracted by attempting to optimize what was intended to be a transitory phase of network transition. What can be done to ensure that this industry maintains collective focus on an IPv6 network as the objective of this exercise?

To try to assess the feasibility of such an approach it is useful to try and understand the collective pressures that are working against such an outcome.

There is no plan!

The first is that this is a deregulated and highly competitive environment. There is no single industry and no single "either/or" decision, but instead there are many different parts to this industry and each actor has their own perspective on the situation. Over the course of the transition it is likely that many different approaches will be explored by a diverse collection of industry actors. Any visible collective outcome of this process is not necessarily the outcome of a particular plan, but more likely to be the outcome of the interplay of various market pressures.

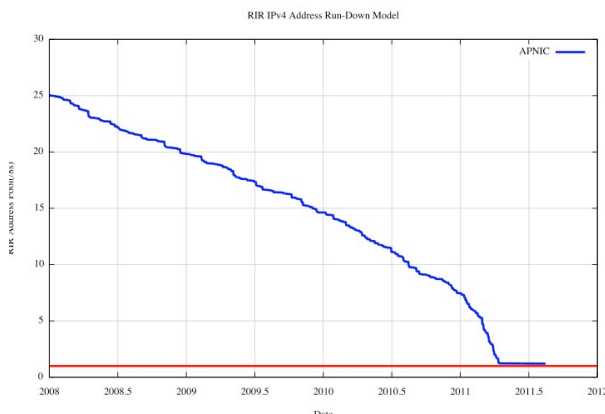
If the hope here is that there is some overall plan of a coordinated response to IPv4 exhaustion and the associated transition to IPv6, then it is useful to bear in mind that there really is no plan at all, and all we have instead is the interplay of market pressures and their outcome.

We are not experiencing the same pressures at the same time!

The second issue is related to the regional structure of the RIRs and the timelines of IPv4 address exhaustion.

In the Asia Pacific region APNIC ceased its IPv4 allocation function on the 19th April 2011, at which point the registry had 16,777,216 addresses left in its local pool. APNIC has now switched to its "last /8 allocation policy", which allows each requestor

access to only one allocation of up to 1024 addresses and no more. In effect APNIC has "run out".

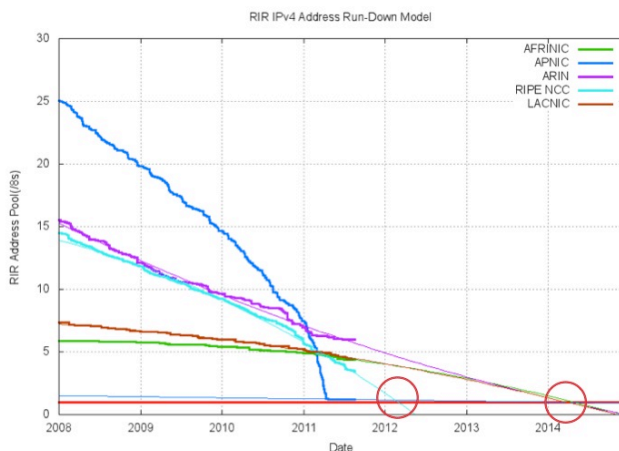


APNIC's IPv4 Address Allocations - 2008 to the present

The state of the other RIRs is not quite the same. The following table shows the amount of available addresses in all the RIRs, using units of a "/8" (or 16,777,216 individual addresses).

RIR	Projected Exhaustion Date	Remaining Addresses in RIR Pool (/8s)
APNIC	19-Apr-2011	1.2044
RIPENCC	29-Feb-2012	3.4683
LACNIC	19-Mar-2014	4.4370
AFRINIC	04-May-2014	4.3840
ARIN	26-May-2014	5.9743

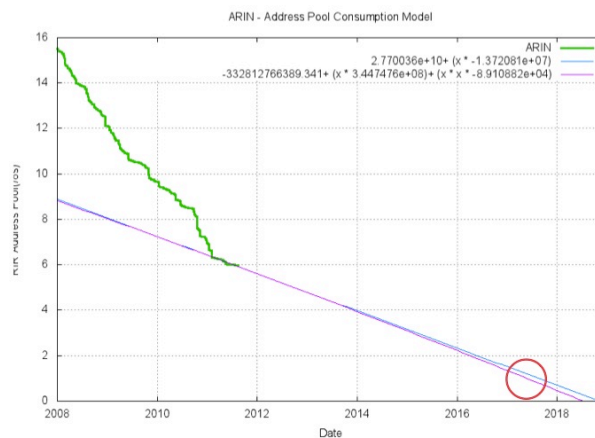
These predictions are based on a weighted least squares best fit of an $O(2)$ polynomial function over the past 3 years of allocations.



RIRs' IPv4 Address Allocations - 2008 to the present

In the North American region, the local regional registry, ARIN, changed its general IPv4 address allocation policy in February 2011, coinciding with the exhaustion of the

IANA-managed pool of addresses. Surprisingly, when this recent address consumption data is used to generate a prediction model for addresses, it appears that ARIN will not reach its final /8 until sometime in 2017. However I am somewhat wary of such a prediction, given the significant uncertainties that it encompasses.



Modified ARIN Exhaustion Model – 2011 – present

The situation we face today is that APNIC and the Asia Pacific region is already operating under circumstances where addresses are not abundant. However, other regions are still able to distribute IPv4 addresses. For other regions the concept of IPv4 exhaustion is a "future" problem rather than a "here and now" problem, and there is still a pervasive air of denial in many parts of the industry in those regions of the world where address exhaustion has not occurred yet. I'm reminded of the observation that: *"It's not happening unless it's happening to me!"*

We are not united with a common purpose!

From these two pressures comes the third consideration, namely the variation in the transition process.

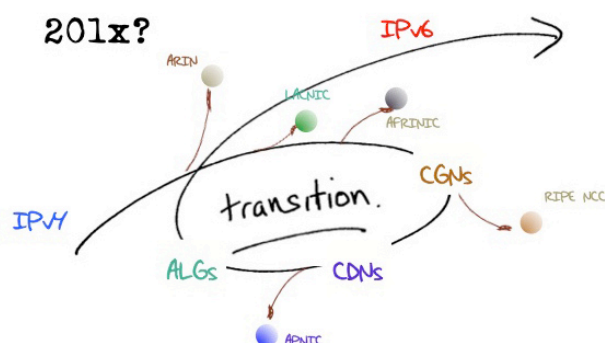
The challenge we face is to sustain the IPv4 half of the dual stack environment in the face of continuing escalation of demand for addresses. For many years the conventional networking environment has included the use of a NAT device at the interface between the network and the user. Increased pressure on addresses is now forcing network service providers to place a second level of NAT inside their network as part of the network infrastructure.

This process of transition is expected to take many years, I have heard commentary to suggest that five years is unrealistically short, and we should expect a transition that may take a decade or longer. But will CGNs last for a further decade of network growth during this extended transition? The next step after CGNs is to break apart the end-to-end network model and start to erect connectivity "barriers" or "walled gardens". The tools to do this include re-homing a copy of certain content "inside" the network as a next step, then, as a further step in address 'compression', using application level gateways rather than address level IP header translators.

The current approach to IPv4 exhaustion will see different regions experiencing different IPv4 scarcity pressures at any point in time. In the Asia Pacific Region the momentum to deploy CGNs as the first response to IPv4 address scarcity is already visible. However, other regions are not experiencing the same pressure at this time. If

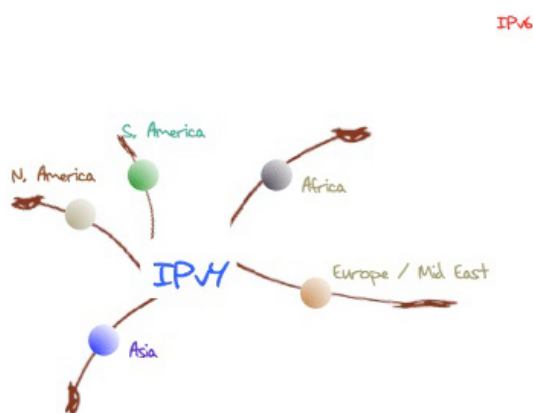
one were to project this further forward by 18 months to 2013 the European region would also have exhausted its pool of IPv4 addresses, but the other three regions may well be operating in a mode that is still able to meet regional demands for IPv4 addresses. It is highly likely that at that time the different regions will be experiencing very different market pressures for the provision of Internet services, due to differing transitional pressures from IPv4 exhaustion.

The consequent question is: What's the level of risk that the differing environments of transition lead to significantly different outcomes in each region as the process of transition takes of a different momentum in different regions? And if this eventuates will we still have a single coherent Internet as a common asset, or will we find that market forces interact in unpredictable ways that create different outcomes in each region?



What of the plan to ultimately converge to an IPv6 network? It may be useful to remember the myth of the long term plan. Are we still as firmly committed to the long term plans we formulated 5 or 10 years ago? Or have we found that our plans are continually modified and refined over time, and there is actually little left of the original plan. So will we be as firmly committed to the transition to IPv6 in five years time? Or will we manage to lose the plot and head into different directions because of the different spread of pressures on service providers in each of the regions. We will forget about the intention to preserve the concept of a single global network in amidst the difficulties of this disparate transition?

20xx?



On Maintaining the Momentum for IPv6

Can we help the Internet during this transition, and try to ensure that the Internet remains a single coherent network with some essential architectural attributes of end-to-end clarity? Or, if we want to aim a little lower, can we at least minimize the potential for disastrous long term damage to this phenomenally productive and valuable networking environment that the Internet has enabled?

I don't know the answer to those questions, but I would like to offer a small number of thoughts that I have had when thinking about this topic.

If we want a single working Internet at the end of all of this, then we need to keep an eye on the larger picture of network evolution during transition. We need to find ways for self interest and local interest to converge with what is in our common interest. Without that convergence we will see a form of market failure where the common interest of a single global network, and the value that such a service can generate, being lost to network divergence through the exercise of differing local market pressures. I'm not sure that I understand how to ensure that self interest aligns with common interest in every circumstance, but what would be good to avoid is building a network that imposes major barriers and inefficiencies all in the name of address conservation in IPv4, and then citing the investment in this additional infrastructure as grounds for not progressing with the transition to IPv6

Secondly, IPv4 addresses should be used in working networks and not hoarded. Its probably a natural reaction to impending scarcity to hoard a resource, but its not necessarily a good reaction. Hoarding behaviour exacerbates the scarcity of the resource in both its intensity and duration. This generalization is also true in the specific context of IPv4 exhaustion and transition. The scarcity of IPv4 addresses creates market uncertainty and market pain in the form of a reduced revenue outlook across the transition period. Extending this scarcity through hoarding and other forms of withholding addresses from use in the network acts to prolong the market pain and increase the unpredictability of the entire transition process.

And finally, we need to keep the transition as quick as possible. A rapid transition represents the best chance of achieving an IPv6 network as an outcome of this process. The more time we spend investing time, money and effort in deploying IPv4 address extension mechanisms, the higher the risk that we will lose track of the temporary nature of transition and the higher the possibility that we'll get stuck with the wrong Internet at the end! If we are truly committed to achieving a single and coherent IPv6 Internet then perhaps its necessary to act now to compress the timelines for transition, not extend them!

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