September 2009

Geoff Huston

# Is the Transition to IPv6 a "Market Failure?"

This is an extended commentary on a presentation to the 1st Workshop on Internet Economics, hosted by CAIDA in September 2009. http://www.caida.org/workshops/wie/0909/

*The Fine Print: I am not a economist in terms of my professional qualifications. Worse still, I think I fit in to the category of amateur economic dilettante! So most of what I offer here I do so tentatively, as it probably needs a little more rigor and precision in basic economic terms than I am able to provide!*

At the outset I should say that here I would like to restrict my view to the **transition** from the IPv4 Internet to the IPv6 Internet, and, in particular, to examine the topic of the appropriate market structure that lies behind the dual stack transition strategy, and the manner in which the Internet can transition from the universal use of IPv4 as the underlying datagram protocol to the universal use of IPv6. I should also stress that this is not an examination of IPv6 itself and its utility, but is limited to an examination of the mechanisms of transition from the IPv4 Internet to the IPv6 Internet, using an approach that attempts to take a perspective of the fundamentals of markets and the associated market models that attempt to bring suppliers and consumers into some form of equilibration.

My motivation for looking at this topic was most recently prompted by a report on a conference on IPv6 held in New Zealand in August 2009:

IPv6: NZ Gov't will lead by example not by regulation
Monday, 24 August 2009

A New Zealand IPv6 Steering Group, convened by the Ministry of Economic Development, last week staged a 'Hui' (conference) on IPv6. Addressing the conference, acting minister for communications and information technology, Nathan Guy, said: "The rationale for these Hui is to convince you as key business decision-makers that there is a business case to be made for IPv6 adoption...We need to have IPv6 widely adopted if we are to be able to take full advantage of the mobile Internet and of the 'Internet of things' involving such things as intelligent houses and intelligent transport systems. A key driver for IPv6 is to make possible new services and business opportunities on a large scale, such as networked sensors for industrial or home automation services."

He added : "the minister for communications and information technology does not believe that regulatory intervention is appropriate. Adoption of IPv6 needs to be lead by the private sector. The private sector must recognize that adopting IPv6 is in their own best interests to protect their investment in online capabilities into the future. Issues of advantages and disadvantages, costs, risks, timing, methodology etc, have to be for each enterprise to assess for itself. There is no single best answer for all."

http://www.itwire.com/content/view/27166/127/

What I found noteworthy in this report is the position being advocated at the level of telecommunications industry public policy in New Zealand, namely that the transition to IPv6 is activity that should be directed entirely by the private sector. It appears to assume that the combination of the factors of the imminent exhaustion of the unallocated pool of IPv4 addresses and the conventional dynamics of an open competitive marketplace in the domestic IP service industry will be sufficient to propel the New Zealand ISP sector to adopt IPv6. Presumably the situation in New Zealand is little different from most other markets, so that this view of one national market could be generalised to be an observation about the entire Internet sector.

The question I would like to pose here is: Is this an appropriate view of the transition to IPv6?

An alternative view is that this transition to IPv6 has already stalled over the past decade, and we should be prepared to view the current situation as an instance of a *market failure* in economic terms, where the transition will require the impetus of some form of response associated with the distribution of a *public good*, and that conventional market dynamics are in and of themselves incapable of sustaining such a transition.

***Market Failure***: a market failure exists when the production or use of goods and services by the market is not efficient. That is, there exists another outcome where market participants' total gains from the new outcome outweigh their losses (even if some participants lose under the new arrangement). Market failures can be viewed as scenarios where individuals' pursuit of pure self-interest leads to results that are not efficient – that can be improved upon from the societal point-of-view. The first known use of the term by economists was in 1958, but the concept has been traced back to the Victorian philosopher Henry Sidgwick. Market failures are often associated with non-competitive markets, externalities or public goods. The existence of a market failure is often used as a justification for government intervention in a particular market.

Wikipedia: http://en.wikipedia.org/wiki/Market_Failure

***Public Good***: In economics, a public good is a good that is non-rivalled and non-excludable. This means, respectively, that consumption of the good by one individual does not reduce availability of the good for consumption by others; and that no one can be effectively excluded from using the good.[1] In the real world, there may be no such thing as an absolutely non-rivalled and non-excludable good; but economists think that some goods approximate the concept closely enough for the analysis to be economically useful. … Non-rivalness and non-excludability may cause problems for the production of such goods. Specifically, some economists have argued that they may lead to instances of market failure, where uncoordinated markets driven by parties working in their own self interest are unable to provide these goods in desired quantities. These issues are known as public goods problems, and there is a good deal of debate and literature on how to measure their significance to an economy, and to identify the best remedies. These debates can become important to political arguments about the role of markets in the economy.

Wikipedia: http://en.wikipedia.org/wiki/Public_goods

## 1. The Background to Dual Stack Transition

One way of looking at this topic is to take the admittedly simplistic view that IPv6 is just IPv4 with larger address headers, and while this view ignores a significant body of adjustments that have been made in IPv6, to a first level of approximation this is a reasonable view. While it is possible to map the smaller (32 bit) IPv4 address pool into the larger (128 bit) IPv6 address pool, the reverse is simply not possible. So if an IPv4 host wants to send a packet to an IPv6 host the challenge is to somehow represent the 128 bits of IPv6 address in the 32 bits reserved for the destination address in the IPv4 header. This is a practical impossibility in any general fashion, and while small subsets of IPv6 address space can be 1:1 mapped into equally small subsets of IPv4 space, the larger problem is not solvable in this fashion. The consequence from this limitation is that even though IPv6 was designed to be a successor protocol to IPv4, its not a *backward compatible* technology, nor is IPv4 a *forward compatible* technology in this respect.

**Backward Compatibility**: In technology, for example in telecommunications and computing, a device or technology is said to be backwards (or downwards) compatible if it allows input generated by older devices. A standard, for example a data format or a communication protocol, is said to allow backward compatibility, if products designed for the new standard can receive, read, view or play older standards or formats.

This should not be confused with forward compatibility, which implies that old devices allow (or are expected to allow) data formats generated by new (or future) devices, perhaps without supporting all new features. A standard supports forward compatibility if older product versions can receive, read, view or play the new standard.

For example, the introduction of FM stereo transmission, or colour television, allowed backward compatibility since new receivers could receive monophonic or black-and-white signals generated by old transmitters. It also allowed forward compatibility, since old (and new) monophonic FM radio receivers and black-and-white TV sets still could receive a signal from a new transmitter.

Wikipedia: http://en.wikipedia.org/wiki/Backwards_compatibility

An instance of the IPv6 protocol stack cannot receive an IPv4 packet header and make any sense of it. Similarly, an IPv6 protocol stack cannot generate a form of the IPv6 packet header that would make sense to an IPv4 protocol instance. They are in effect entirely different and incompatible protocols, and while the two protocols can coexist on the same network, they simply cannot interoperate from one host to another. The reason for this is that while most of the IPv4 header fields can be mapped into either the equivalent IPv6 field or some form of extension header in one fashion or another, the problem comes when attempting to map the source and destination address values between the two protocols. A host using only an IPv4 stack cannot directly communicate with a host using an IPv6 stack.

Maybe it could've been different, and perhaps in retrospect perhaps we should've looked harder at alternative approaches when working on IPv6 that deliberately eased the burden of transition. But such speculation is not all that helpful given that the agenda before us now is pretty clear: just how can we undertake this transition from IPv4 to IPv6?

Given that we can't bind the IPv4 and IPv6 worlds together in a seamless fashion at the IP packet header level, then the starting point of the transition is that the underlying network infrastructure needs to support two distinct protocol families: IPv4 and IPv6.

Consequently, for the past 12 years or so what we've seen is that most of the plans for this transition to an IPv6 network pass through an intermediate stage of progressive dual protocol stack deployment where IPv4 and IPv6 are both deployed. The general idea is that initially new deployments would be both "conventional" IPv4 and also IPv6, with IPv6 network support being provided possibly via tunnelling across IPv4 infrastructure.

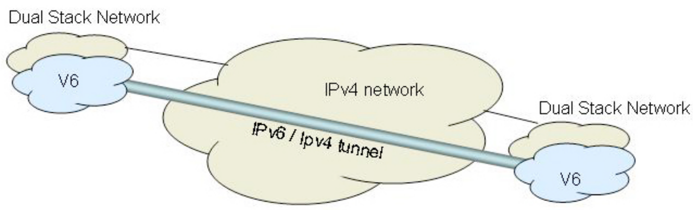


*Figure 1 - Initial Phase of the Dual Stack Infrastructrure*

Progressive stages of the transition would see these IPv6 "islands" interconnect directly, and also see the commencement of the older IPv4-only networks converting to a dual stack environment.

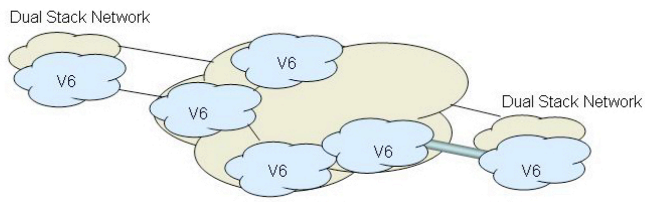


*Figure 2 - Subsequent Phase of the Dual Stack Infrastructrure*

Presumably we'd know that we were in the final stages of this transition when there are new network deployments using only IPv6 and the dual stack hosts make no further use of IPv4 to reach other hosts, at which stage the continued need for IPv4 would've finished.

This broad plan raises more questions than it answers, particularly relating to the nature of the drivers that would see existing deployments of IPv4 infrastructure convert to a dual stack environment, and to the drivers that would see the culmination of the transition and termination of the dual stack environment. There is no particular timetable here, and no particular means of coordination of activity, so that knowing when we've completely finished with IPv4 may not be an obvious call.

The dual stack deployment effort is predicated on the constraint that IPv4 hosts simply cannot talk directly with IPv6 hosts. So instead we have the dual stack negotiated behaviour that involves behaviours on the part of the application as well as the host's protocol stack. The approach is that the application first attempts to expose the protocol capabilities of the other end by asking the DNS for both IPv6 and IPv4 addresses that are associated with the other end's DNS name. Assuming that the DNS returns both address forms, then the application initially attempts to start up a session with one protocol (typically IPv6), and if this fails, then it falls back to the other protocol.

So, with help from various forms of tunnelling support to bridge over any protocol-specific transport continuity gaps, the basic idea of this transition is that we enter into an extended period where hosts need to use IPv4 to talk to IPv4 hosts and IPv6 to talk to IPv6 hosts. As long as hosts first try the IPv6 handshake, then, so goes the line of reasoning, we should see the overall traffic mix in this dual stack environment tend to move towards IPv6 as the legacy IPv4 infrastructure migrates to IPv6, and the dual stack nature of the deployment means that dual stack hosts can always fall back to IPv4 to speak to IPv4 legacy infrastructure.

As transition plans go, it's a reasonable plan, particularly considering that IPv6 is not a backwards-compatible protocol. In other words while IPv6 is a *substitute* for IPv4, the lack of backward compatibility imply that in an existing IPv4 environment IPv6 could be regarded as an *imperfect substitute* for IPv4. Another view is that the protocol are not direct substitutes until there is comprehensive deployment of the support infrastructure for IPv6, at which point, and only at that point, the two protocols are sufficiently similar in capability that they then could be regarded as *perfect substitutes*.

> ***Substitute Goods***: In economics, one kind of good (or service) is said to be a substitute good for another kind in so far as the two kinds of goods can be consumed or used in place of one another in at least some of their possible uses. [*in this sense IPv6 is a substitute for IPv4*]. Perfect substitutes may alternatively be characterized as goods having a constant marginal rate of substitution. Imperfect substitutes exhibit variable marginal rates of substitution along the consumer indifference curve.
>
> Wikipedia: http://en.wikipedia.org/wiki/Substitute_good

However there is one salient aspect of this dual stack transition that is significant. This approach to transition implies that for the duration of the transitional environment both IPv4 and IPv6 support is required in hosts, across the network, in the routing system, in switches and routers, and in all forms of infrastructure services such as the DNS, firewalls, security, management and rating systems.

This is not a very satisfying situation. This overall process of dual stack transition demands that, ultimately, the IPv4 legacy environment must migrate to support IPv6, while the characteristic of many forms of legacy environments is that they are strongly resistant to change. This process also requires new deployments to be equipped with legacy capabilities for as long as there are legacy IPv4 environments that have communications requirements. It appears that within such a transitional process we would not have moved very far from our current addiction to IPv4 addresses while at the same time attempting to push IPv4 legacy systems into upgrading for IPv6 support.

This form of the dual stack world does not alter the inherent demand for IPv4 addresses, as every communicating host that wants to maximizes its communications realm now needs the ability to communicate in both IPv6 and IPv4. An IPv6-only host can only communicate with other IPv6 hosts, while a dual stacked host can communicate with any other party.

The corollary of this dual transition plan is that, ideally, it needs to be complete before we exhaust the pool of IPv4 addresses that we will need to support the continual expansion of the Internet.

## 2. The Transition Process

An idealised model of the dynamics of this transition is shown in figure 3. The green line in that figure shows the total size of the network, with a continual growth factors. The blue line shows the depletion of the unallocated IPv4 address pool. Given that the growth of the Internet will need to be sustained using IPv4 addresses during the transition phase, the IPv4 address pool will deplete in size in direct relationship to the increase in the size of the network during this transition. The red line shows the total size of the IPv6 network.

Given that the major motivation of the development of IPv6 was to avoid the situation of depletion of available IPv4 addresses, the assumption behind the transition plan was that the uptake of the IPv6 network would need to progress at a faster rate than the growth of the network itself, so that before the time that the IPv4 address pool was fully depleted the takeup of IPv6 would be universal, or close enough. At that stage there would be no further requirement for IPv4 addresses and the transition would be complete.

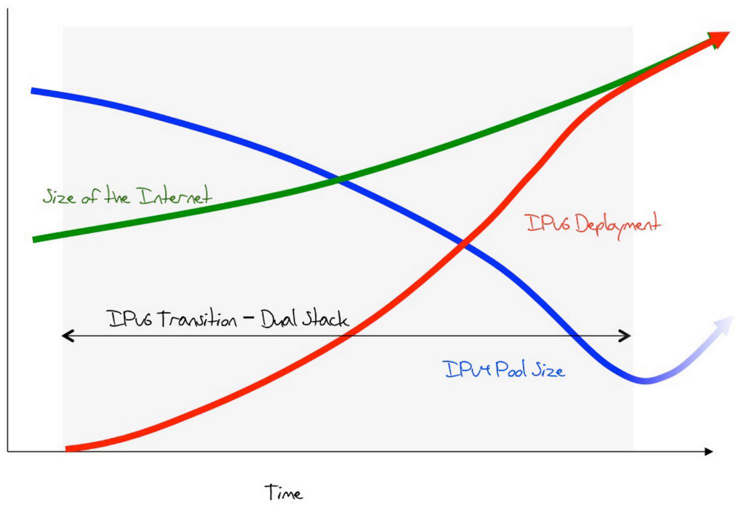


*Figure 3 - The Dual Stack Transition Process*

How well we are doing with this plan?

Current estimates (http://ipv4.potaroo.net) indicate that the IANA will exhaust its IPv4 address pool in September 2011, and the first RIR will fully exhaust all its pool of addresses by June 2012. Given the level of uncertainty in the predictive process, a prudent consideration would be that the entire IPv6 transition should be well and truly completed by July 2011 if we want to avoid the dual problems of driving the dual stack transition and attempting to expand the Internet without any residual pool of available addresses. This implies that the end date in the above transition model is mid-2011. So we have a good idea of when the transition process should complete, assuming that we want to avoid IPv4 address depletion.

The next task is to put the current date into the process. How far are we with the level of IPv6 deployment? There are a number of ways we can measure this (these have been discussed at http://www.potaroo.net/ispcol/2008-04/ipv6.html). In a small scale controlled measurement of IPv6 end-to-end deployment, the APNIC web server (http://www.apnic.net) has gathered data on the number of unique IPv4 and IPv6 addresses that have accessed the site each day since 2004. On the assumption that end hosts will prefer IPv6 over IPv4 as long they are equipped to handle IPv6 and as log as some form of IPv6 transport is available, the long term data of the ratio of

IPv6 to IPv4 access provides a reasonable window on the true level of permeation of IPv6 deployment across the population of end user sites. This data is shown in Figure 4. (The increase in the ratio in September 2008 though to early 2009 was due to a prominent listing of this web site on a Chinese academic and research network's IPv6 links page.)

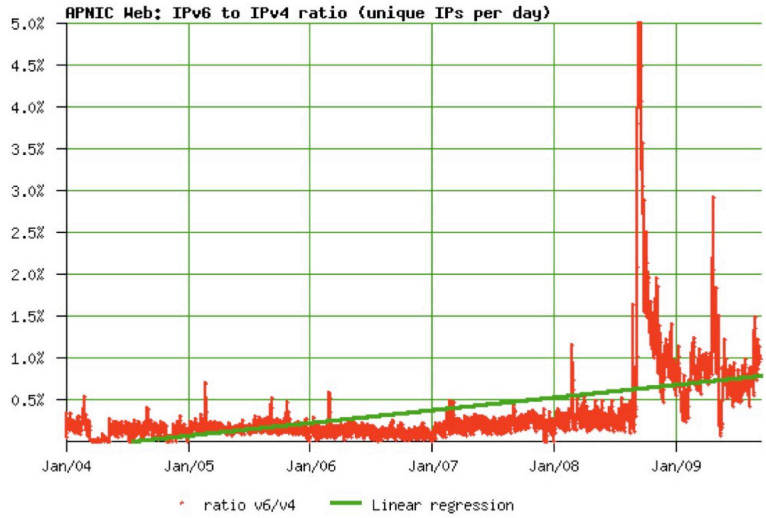


*Figure 4 - Relative IPv6 Deployment*

The overall 5 year trend in figure 4 shows that the level of end host IPv6 capability is still under 1% of the total set of end hosts that access this site.

Assuming that this is a mirror of a more general measurement of the IPv6 permeation levels, then we can put a timeline under the data presented in the earlier IPv6 transition plan.

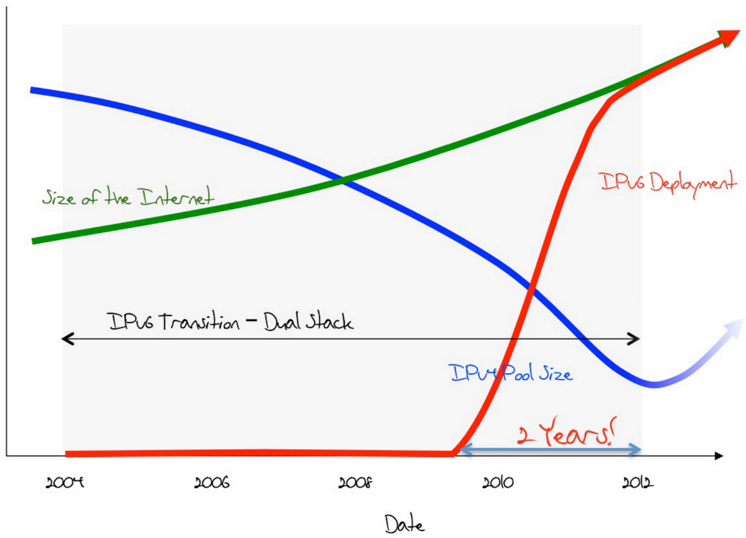


*Figure 5 - The Revised Dual Stack Transition Process*

The implication of this additional data on the originally envisaged process of dual stack transition is that if we wish to undertake this transition prior to the exhaustion of the IPv4 unallocated address pool, then a conservative estimate is that we have less than two years to lift the relative penetration of IPv6 by more than a hundredfold, from some 1% of the Internet of today to complete deployment, or effectively 100% of the Internet of 2011.

This appears to be a rather daunting objective. The Internet appears to embrace some 1.6 billion users (based on data published by the ITU-T). There are at least a billion end hosts (and likely to be far more), hundreds of millions of routers, and a comparable number of firewalls and other middleware devices, and a cumulative total of billions of lines of configuration codes and filters that are spread across hundreds of millions of ancillary support systems. We've been very busy over the past couple of decades and the IPv4 Internet is now widely dispersed across an impressive count of various forms of devices and active network elements.

The task, and associated cost, is that all these devices must be altered in some form or fashion to support IPv6. While some of these conversion tasks can be automated so that the cost of conversion can be minimized, other tasks require significant capital and operational costs, and also require an extended time to complete. The general tenor of industry comment on this transition timetable is that while it may have been feasible to complete this transition prior to IPv4 address exhaustion if the industry had commenced with this effort in the late 90's, this is no longer a feasible objective given our current situation. We are now incapable of orchestrating a comprehensive transition to IPv6 within the time available as determined by the anticipated time remaining for the unallocated pool of IPv4 addresses.

The implication is that the most likely outcome is that the efforts to undertake transition will be taking place at the same time as the industry will have to face the challenges, and costs, of operating the IPv4 Internet without sufficient addresses to perform unique numbering of the end points even within the current widely deployed model of edge Network Address Translation (NAT) deployment.

**Edge NAT deployment**: The current architecture used by most ISPs has managed to *externalize* the costs of address pool management to the customer. Rather than allocating the customer a pool of addresses and allow the customer to number their devices from the service provider's address pool, with the consequent issues and costs of address pool management, address portability and provider lock-in, the adopted solution has been for the provider to number the customer interface with a single address and have the customer use a private address pool for their devices, The customer owns and operates the NAT device that maps between this local private address pool and the service provider's public address. A consequence of this framework is that the service provider has managed to externalize the costs of address scarcity, transferring this cost directly to the customer.

Complete exhaustion of the unallocated IPv4 address space implies that service providers may need to deploy NAT devices within their own networks if they wish to service an expended set of IPv4 customers. The deployment of such Carrier Grade NATs (CGNs) will impose higher costs of the service provider without improving the service provided to the customer. Indeed the case can be made that the CGN deteriorates the service provided to the customer, so in a heterogeneous competitive environment where the address scarcity is not uniformly experienced, any service provided that attempted to pass on the costs of CGN deployment to its customer base would find itself at a short term competitive disadvantage to its competition though higher prices and lower service quality.

It appears that we can't seem to be able to undertake this transition in a manner that completely avoids IPv4 address exhaustion, and the issues with IPv4 address exhaustion expose a set of issues relating to a shift in the cost base for ISPs that are not necessarily exposed to the incumbent customer base, but are a significant factor for new customers and new products. This unequal exposure to the cost of IPv4 address exhaustion does not help in motivating the incumbent base of the network to underwrite their part of the costs of a dual stack transition to IPv6.

The question posed here is: In the light of these considerations, is the transition to IPv6 a "natural" outcome for a deregulated competitive industry? Does the impact of exhaustion of the unallocated IPv4 address pool provide further impetus for this transition, or, does it remove some of the impetus for transition through the creation of potentially mutually exclusive choices for an ISP's expenditure in IPv6 transition or CGN deployment? Will this industry require some form of external intervention to create an environment where the transition can proceed in an efficient manner?

As the original quotation at the start of this article illustrates, there is still some level of conviction in public policy circles that the open market is capable of reaching an IPv6 outcome without any need for external intervention in the transition process. It may be useful to understand the some of the factors that lie behind such convictions in the power and effectiveness of open competitive markets in the telecommunications sector, particularly when one considered that during the majority of he period between 1920 and 1990 this particular market was generally characterised by the use of public monopoly enterprises of one form or another. One possible reason for this recent conviction on the part of the regulatory and public policy bodies in the effectiveness of open competitive markets for telecommunications services is to observe the period of the Internet's own genesis. After all, if one is prepared to accept that the Internet itself is the outcome of an open market process of competitive interests working to ultimately benefit the consumer with novel innovation and efficient pricing and service delivery, surely this transition to IPv6 is another equally valid example where an open competitive market will create the appropriate outcome. So lets look at how the Internet arrived to see if there are similarities here.

## 3. The "Transition" to IPv4

The original deployment of IPv4 has often been cited as an excellent example of the beneficial outcomes of an open deregulated market in action. Indeed not only did this original deployment of the Internet happen without direct external intervention, and particularly without direct regulatory impost, it could be argued that at the time it happened in spite of the public policy regime at the time. In the late 80's there was a visible general level of governmental support for adoption of the OSI suite of protocols instead of the Internet, and the proliferation of various GOSIP programs of the late 80's and early 90's is illustrative of this. So why did the deregulated industry undertake the transition to IPv4? And are the factors that were at play in this previous transition relevant to the issues related to the transition from IPv4 to IPv6?

At the time IPv4 included a number of major technical innovations. The first of these was the use of datagram packet switching, in comparison to virtual circuit switching which formed the foundation of the then-prevalent data services. This form of stateless packet switching enabled access to lower costs through the transfer of functionality (and cost) from the network to the edge systems. This allowed a single data circuit to be efficiently shared across a community of up to thousands of IP end users. The related technical innovations in IP included a sufficiently large address field to accommodate a collection of users that was commensurate with the population of public network users, and the unreliable datagram model of switching control that pushed the responsibility for end-to-end reliability also out from the network and onto the edge systems. At the same time the provisioning methodology used by the telephone operating companies used a methodology of massive up-front provisioning of capacity, and then ameliorating this potential oversupply by careful rationing of availability of data services through extensive use of price premiums. In other words there was a significant pool of available infrastructure that could be accessed given a use model that achieved a certain threshold of exploitative value. The Internet's packet-switched architecture represented such an efficient user of the existing available carriage infrastructure.

In economic terms this Internet-induced shift can be modelled as a supply curve shift, where the cost of the provision of upper level data services was drastically reduced thorough the use of IP placed on top of the basic bit carriage services. If the supply side was able to respond to this opportunity then this lower price would result in an increase in quantity of service at the equilibrium point of the new price. At the same time the extremely simple and flexible model of service delivery in IP allowed for a entirely new set of applications for data networking. This was first seen in the rise of electronic mail, then with file sharing and very quickly to services that were built upon the emerging common substrate of the worldwide web. This dramatic expansion of the service portfolio increased the user's perceived value of the service, which can be modelled in the demand schedule as a demand curve shift within the framework of the same basic economic model.

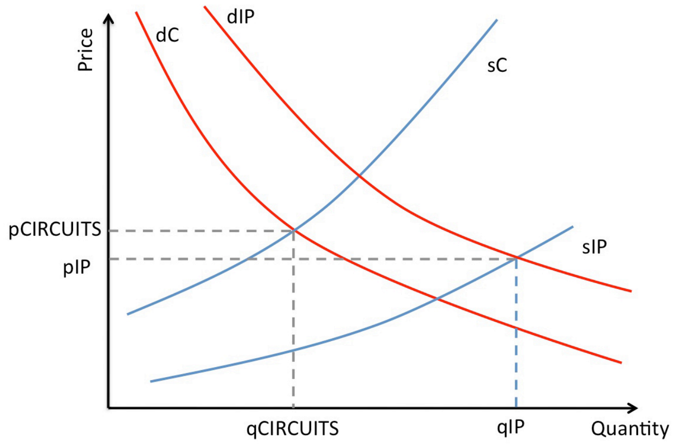


*Figure 6 - Modelling the Demand Schedule change from Circuits to IP Networks in the 90's*

The combination of these two factors in the supply and demand schedules worked together in terms of reduced price and increased user perception of value, shown in Figure 6.

Progressive deregulation of the telecommunications industry exposed this industry to the private capital investment markets, and a new class of entrepreneur that was seeking strategic weaknesses in the incumbents' service portfolio that could be readily exploited. IP provided such an opportunity, as the very significant differential between the efficiency of circuit-based and packet-based networks was of such a magnitude that it was readily possible to purchase a circuit, add IP switching elements and resell the service as IP to a large community of customers.

The initial growth of the Internet was fuelled by this entrepreneur sector, using highly agile, high risk private investment, or *venture capital*.

**Venture Capital**: Venture capital (also known as VC or Venture) is a type of private equity capital typically provided to early-stage, high-potential, growth companies in the interest of generating a return through an eventual realization event such as an IPO or trade sale of the company. Venture capital investments are generally made as cash in exchange for shares in the invested company. It is typical for venture capital investors to identify and back companies in high technology industries such as biotechnology and ICT (information and communication technology). A core skill within VC is the ability to identify novel technologies that have the potential to generate high commercial returns at an early stage. By definition, VCs also take a role in managing entrepreneurial companies at an early stage, thus adding skills as well as capital (thereby differentiating VC from buy out private equity which typically invest in companies with proven revenue), and thereby potentially realizing much higher rates of returns.

http://en.wikipedia.org/wiki/Venture_capital

This is not a long-term stable situation, particularly in the telecommunications services sector which is characterised by significant economies of scale. The increasing intensity of competition in a marketplace occupied by a set of small scale providers lead to a wave of consolidation in the nascent ISP industry, where the smaller players merged into larger enterprises that were then able to use their resultant greater volume of activity to achieve greater economics of scale.

It was only a matter of time before the larger telephone operators were able to overcome the short term issues of cannibalisation of existing data service products and directly address the strategic weakness in their service portfolio that had been exposed by the Internet. The latter part of the 1990's saw the larger players undertake the deployment of capital and the creation of new support infrastructure and process to support the large scale deployment of Internet services, and at the same time buy out many of the initial Internet service providers. The overall picture of the industry across the 90's and early 00's is indicated in Figure 7. This picture is not dissimilar to the early ears of the telephone industry in the late 19th century, or that of the early years of the auto industry, the television industry or the computer industry. The basic technical innovation that lay at the heart of the market initially motivated a wave of entrepreneurs who rushed products to market in order to obtain an early market lead. This stimulated greater levels of demand, which created further impetus for additional entrants in provision of goods and services. The pressures of competition then started to impose a need for efficient production, which lead to a need to improve production efficiency through economies of scale. This motivated a wave of consolidation in the supply side of the industry, leading to the consolidation of the market into a far smaller number of large scale providers.

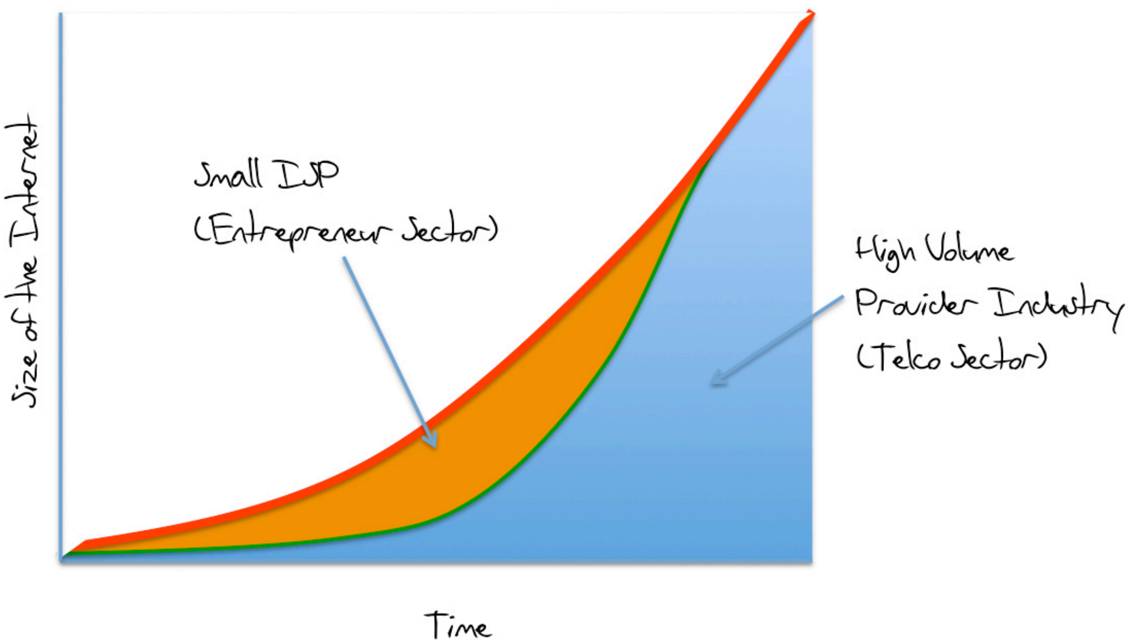


*Figure 7 - The Rise of the IPv4 Internet*

I have recently heard a claim from one ex-telephone company public affairs spokesman who claimed that the only way to successfully provide innovative services in new markets was through vertical integration and bundling of services.

The experience of the Internet points entirely in the opposite direction, where the presence of a wide diversity of highly specialised service providers and the related activity of disintegration and unbundling of all the components of the service lead to intense competition in both service efficiency and innovation in every part of the service portfolio. This, in turn, lead to a wave of innovative consumer products that left the incumbent telephone service operators looking distinctly anachronistic and out of touch with consumers. The side effect of this wave of entrepreneurial activity in the early days of the internet had a very notable outcome in the creation of a distinct market in the provision of specialised network services. This market is populated today by flagship enterprises such as EBay, Amazon and Google, as well as thousands of other smaller enterprises that continually push at the innovative edge. These providers do not operate an access network and have no direct relationship with the provision of Internet access services to end users, and collectively they represent a significant step in the unbundling of communications services, a process which previously had been successfully resisted by the incumbent telephone operators.

Despite the entry of these telephone operators into the internet services market, the Internet has managed to remain unbundled, allowing a large and diverse set of independent service providers to operate in an overlay across the substrate of the Internet.

## 4. The Transition to IPv6 as a Market Failure

The rather obvious question is: If it happened for IPv4, then why won't it happen for the transition to IPv6? Will we once again see the a deregulated industry embark upon this transition as a "natural" outcome of the conventional operation of an open, competitive, deregulated market?

The one major difference here is that we are not comparing the costs of datagram-based packet switching architectures to those of virtual circuits, but instead we are comparing two remarkably similar versions of IP. These versions are so similar in fact that the behaviour of the upper-layer transport protocols are essentially identical - TCP and UDP are both unaltered.

In comparing IPv4 to IPv6, the cost of the network service is identical. Adopting IPv6 does not provide any difference in the costs of the network's operation. Similarly, the functionality of the network services are so similar as to be effectively identical. There are some subtle differences in packet fragmentation and some subtle differences in auto-configuration, but these are extremely subtle changes and in general have no visible impact on the upper level services and on the resultant architecture of applications. The only significant difference here lies in the address space, where the span of devices that can possibly be addressed in an IPv6 network is vastly greater than that of IPv4.

In summary, there is no consumer-visible difference in the services provided by IPv6 that would lead to a change in demand, and nor is there a difference on the cost of production of services that would lead to a change in supply. And because IPv6 is not backward compatible with IPv4 the two goods are not *perfect substitutes*. Instead, there is a cost associated with the provision of IPv6 related to the cost of the provisioning and operating the network infrastructure with IPv6 support.

Normally this would lead to a situation where such a transition would fail, given that the two services are essentially identical and there is a non-negligible cost of transition from the incumbent protocol to the other protocol.

It has often been observed that this transition would've been a lot easier had the industry commenced the transition in 1999, while the network was a lot smaller than it is today and while the projection of completing the transition was well before the prospect of IPv4 address depletion became an imminent event. The obvious question here is that if we are allowed to assume that all the market players have been well informed at all times, then is the fact that this transition has not commenced when it would've been far easier and cheaper to undertake an indication that the transition is one that will not be undertaken at all? Is an inference from leaving this transition until its too late a clear signal that it will not happen at all within the constraints of the prevailing environment?

Another way to phrase this question is to inquire whether the transition to IPv6 is an instance of a *market failure*?

As noted earlier, a *market failure* is characterized by an inefficient production environment, where individual self interest does not provide sufficient incentive to undertake efficient production. This is often associated with a *public good*, and the associated question is whether this transition is an activity that can be characterised as a *public good*? On the face of it a strong case can be made that this transition has non-rivalrous and non-exclusionary properties. An individual's actions in supporting this transition does not lead to any other individual being unable to do so, and noone is excluded from the benefit of networking services by virtue of any individual's transition. In other words, if an ISP provides an IPv6 service in additional to an IPv4 service in support of a dual stack transition, then given that there are no services that are accessible using IPv6 exclusively there are no additional benefits or services than can be provided by this IPv6 service. The IPv4 users of the network can access all the clients of this dual stack service, so the additional expenditure on the part of the ISP has not created any additional value for the consumer and cannot be translated into an increase in the value of the service or the size of the service.

If this is the case, then it would appear that leaving this transition for competitive market forces to resolve appears to be an ineffectual stance if the transition is considered to be in the common or *public interest.* Conventional market dynamics in a deregulated market will not lead to this transition being undertaken.

**Public Interest**: The public interest refers to the "common well-being" or "general welfare." The public interest is central to policy debates, politics, democracy and the nature of government itself. While nearly everyone claims that aiding the common well-being or general welfare is positive, there is little, if any, consensus on what exactly constitutes the public interest. Rather than as an absolute, the public interest is often defined relative to the concept of a private or individual interest. It is possible for acts in the public interest to be bad for given individuals and vice versa.

Wikipedia: http://en.wikipedia.org/wiki/Public_interest

## 5. Possible Remedies

If this transition is considered to be necessary in terms of the public interest, then perhaps we need to look a little more widely for possible remedies.

One approach is to allow the complete exhaustion of IPv4 addresses to occur. Would this provide sufficient motivation for the industry to undertake this transition?

The major differentiator between IPv6 and IPv4 is that of future risk associated with IPv4 address exhaustion. Following exhaustion service providers will incur increased costs associated with continued expansion of the customer base through increased costs incurred through the combination of the increased cost of acquiring addresses (a scarcity premium) and the increased costs in making efficient use of these addresses through the use of service provider-provisioned NATs (an efficient exploitation cost). New customers and products that are deployed may have impaired functionality as a result of the use of additional NATs that are deployed in the data path. While the installed base of the network would not be disadvantaged, new customers would have a lower level of functionality and their service providers would experience higher costs in an IPv4 network that had to make more efficient use of the pool of addresses though multiple levels of NAT deployment. (see figure 8)

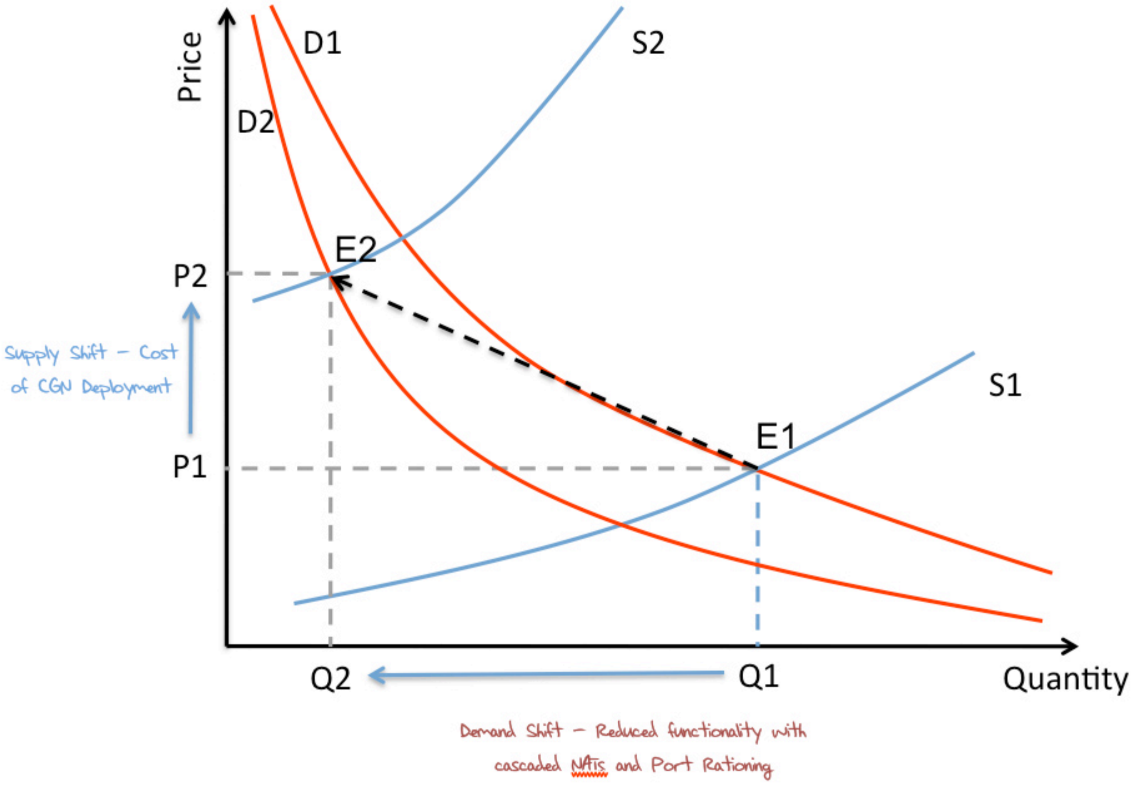


*Figure 8 - The Effect of IPv4 Address Exhaustion on the Demand Schedule*

However it must be noted that the entire Internet does not experience this shift in cost and functionality. This alteration to the demand schedule replates principally to the population of new Internet users and the deployment of new products. So the question is whether this shift of cost and functionality for new consumers and new product deployments would provide sufficient impetus for all providers to embark on the transition to IPv6 for all their customers and most on with a comprehensive deployment of IPv6 alongside IPv4? Interestingly, this does not change the demand schedule for new users, and the shift indicated in Figure 8 still applies, as a Dual Stack deployment implies a continuing need for IPv4 for such users. For existing customers the functionality and service remains constant, so there is no increase in the perceived benefit of the service, but the additional cost in the deployment of IPv6 in parallel to the IPv4 infrastructure would need to be borne by the service provider. If this cost were to be passed on the consumers this could result in a decrease in demand from that service provider (Figure 9).

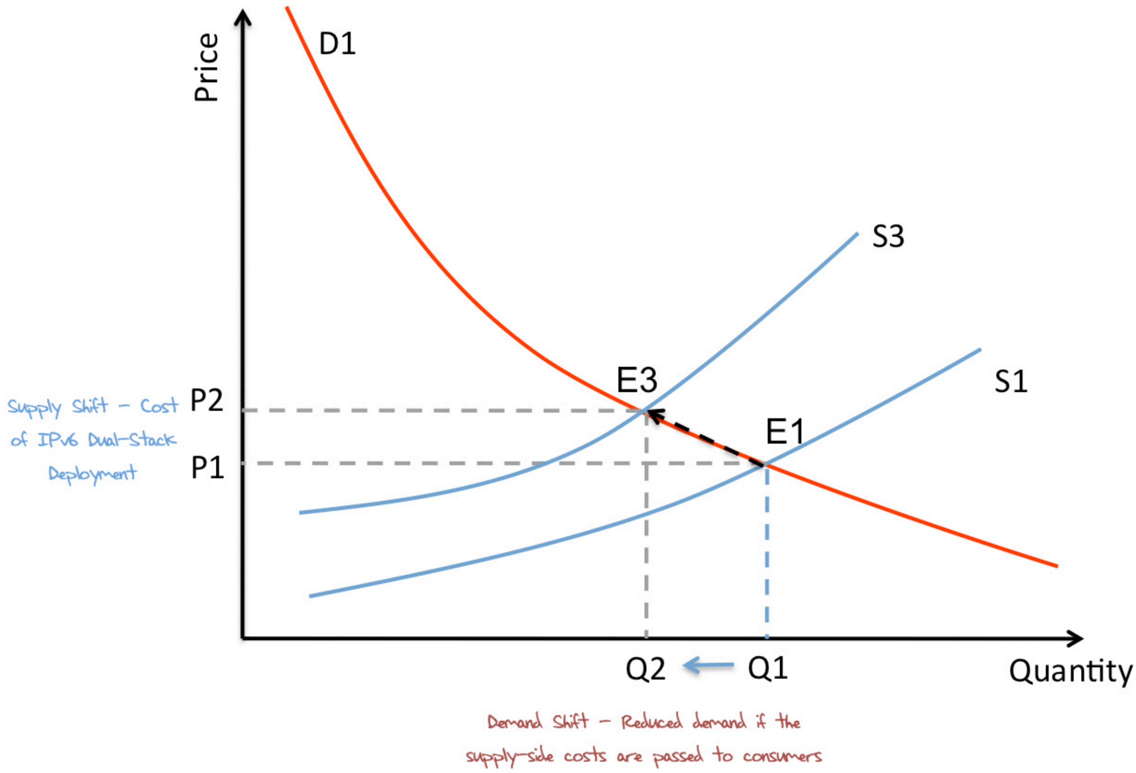


*Figure 9 - The Demand Schedule Shift with Dual Stack Deployment*

It appears that a service provider that elects to commence a dual-stack deployment of IPv6 is putting itself at a short term competitive disadvantage with respect to a service provider that elects to remain with an IPv4 service portfolio, even though the necessity of this move in the longer term is a common constraint.

This appears to be a more general case of a form of market failure where short term competitive interests appear to be contrary to longer term common interests.

What remedies exist if indeed this situation is an instance of a market failure? The range of possible measures include the usual suspects of :

- **regulatory impost**, where a regulatory constraint is placed on the licence holders that IPv6 services are to be provided by a given deadline (as has happened with digital television in many regulatory sectors). This regulatory constraint acts a form of a assurance contract, where all providers are in effect bound to produce a particular solution).

- **government purchase contracts**, where the government sector collectively require the provision in IPv6 in all their service contracts. This is a form of a *coasian solution* where a group of potential beneficiaries pool together their willingness to pay for the public good. We have seen this approach in the past with the Government OSI Profiles (GOSIP) of the late 1980's when the approach proved ineffectual, and there is no assurance that such collective actions have sufficient mass and momentum to create a broader sustainable market.

- **subsidies and incentives**, where the production of the good is subsidised in some fashion by public funds, This can be in the form of direct payments to service providers, or in the form of vouchers to consumers which can be redeemed only in exchange for the supply of a specified service. Other potential measures include the use of taxation incentives related to infrastructure investment, where the investment in a certain class of infrastructure or in a certain sector can be provided with advantaged taxation treatment.

- **public provision of the service**, where the service is provided by a publically-owned enterprise. The funding for such an enterprise can be provided by government-backed investment bonds, or directly from public revenues, and operating losses are underwritten by the public purse. This measure was used for most national telephone service providers for a significant part of the twentieth century, so it is not exactly a completely foreign concept for this industry.

## 6. Where now?

The one undeniable assertion that can be made about the current situation is that where we are today is an entirely unintended scenario. The dual stack transition was meant to be not only largely underway, but reaching its conclusion by now if we were to avoid the imposition of additional costs associated with IPv4 address exhaustion.

Given that this is now unavoidable, the next question to ask is whether this industry will undertake this transition at all if left to its own devices. It may well be the case that this is another instance of a public good with common interest implications, in that the problem of undertaking the transition is simultaneously a problem that is in every consumer's common interest as a desirable outcome, yet lies in no individual consumer's sphere of immediate self interest.

If this is indeed the case, then we may be facing a situation of a *market failure* with this transition, in which case we may need to contemplate some for of additional impetus in the form of a response to stimulating a viable and efficient market for what is at present a public good of the transition to IPv6.

## Disclaimer

The above views do not necessarily represent the views or positions of the Asia Pacific Network Information Centre, nor those of the Internet Society.

## About the Author

GEOFF HUSTON is the Chief Scientist at APNIC, the Regional Internet Registry serving the Asia Pacific region. he graduated from the Australian National University with a B.Sc, and M.Sc. in Computer Science. 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.

http://www.potaroo.net