Internet Engineering Task Force	A. Durand	
Internet-Draft	Comcast	
Intended status: Informational	July 09, 2008	
Expires: January 10, 2009		

TOC

Dual-stack lite broadband deployments post IPv4 exhaustion
draft-durand-dual-stack-lite-00

Status of this Memo

Abstract

The common thinking for the last 10+ years has been that the transition to IPv6 will be based on the dual stack model and that most things would be converted this way before we ran out of IPv4.
It has not happened. The IANA free pool of IPv4 addresses will be depleted soon, way before any significant IPv6 deployment will have occurred.
This document revisits the dual-stack model and introduces the dual-stack lite technology aimed at better aligning the cost and the benefits of deploying IPv6. It will provide the necessary bridge between the two protocols, offering an evolution path of the Internet post IANA IPv4 depletion.

Table of Contents

---

1. Introduction

TOC

This memo will present views on deployments post IPv4 exhaustion and some of the necessary technologies to achieve it. The views expressed

are the author personal opinions and in no way imply that Comcast is going to deploy the technologies described here.

1.1. Requirements language

TOC

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 (Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels," March 1997.) [RFC2119].

1.2. Terminology

TOC

This document makes a distinction between a dual-stack capable and a dual-stack provisioned device. The former is a device that has code that implements both IPv4 and IPv6, from the network layer to the applications. The later is a similar device that has been provisioned with both an IPv4 and an IPv6 address on its interface(s). This document will also further refine this notion by distinguishing between interfaces provisioned directly by the service provider from those provisioned by the customer.

1.3. IPv4 exhaustion coming sooner than expected

TOC

Global public IPv4 addresses coming from the IANA free pool are running out faster than many predicted a few years ago. The current model shows that exhaustion could happen as early as 2010 or 2011. See http://ipv4.potaroo.net for more details. Those projection are based on the assumption that tomorrow is going to be very similar to today, ie looking at recent address consumption figures is a good indicator of future consumption patterns. This of course, does not take into account any new large scale deployment of IP technology or any human reaction when facing an upcoming shortage.
The prediction of the exact date of exhaustion of the IANA free pool is outside the scope of this document, however one conclusion must be drawn from that study: there will be in the near future a point where new global public IPv4 addresses will not be available in large enough quantity thus any new broadband deployment may have to consider the option of not provisioning any (global) IPv4 addresses to the WAN facing interface of edge devices. However, the classic IPv6 deployment model

known as "dual stack provisioning" can be a non starter in such environments.

2. Handling the legacy

TOC

2.1. Legacy edges of the Internet for broadband customers

TOC

Broadband home customers have a mix and match of IP enable devices. The most recent operating systems (eg Windows Vista or MacOS-X) can operate in an IPv6-only environment, however most of the legacy one can't. It has been reported, for example, that windows XP cannot process DNS requests over IPv6 transport. Expecting broadband customers to massively upgrade their software (and in most cases the corresponding hardware) to deploy IPv6 is a very tall order.

2.2. Content and Services available on the Internet

TOC

IPv6 deployment has been very long to take off, so the current situation is that almost none of the contents and services available on the Internet are accessible over IPv6. This will probably change in the future, but for now, one has to make the assumption that most of the traffic generated by (and to) broadband customers will be sent to (and originated by) IPv4 nodes.

2.3. Additional impact on new broadband deployment

TOC

Even when considering new, green field, broadband deployments, such as always-on 4G, service providers have to face the same situation as described above, that is, contents and services available on the Internet are, today, generally accessible only over IPv4 and not IPv6. This makes adoption of IPv6 for green field deployment difficult. Solutions like NAT-PT, now deprecated, do not provide, as of today, a satisfying and scalable answer.

TOC

2.4. Burden on service providers

As a conclusion, broadband service providers may be faced with the situation where they have IPv4 customers willing to communicate with IPv4 servers on the Internet but may not have any IPv4 addresses left to assign to them... However, without some form of backward compatibility with IPv4, the cost and the benefits of deploying IPv6 are not a aligned, making incremental IPv6 deployment very difficult.

2.5. The dual-stack lite model: IPv4 address sharing on top of IPv6-only provisioning

TOC

The core idea behind dual-stack lite is to move from a deployment model where a globally unique IPv4 address is provisoned per customer and shared among several devices within that customer premise to a deployment model where that globally unique IPv4 address is shared among many customers. Instead of relying on a cascade of NATs, the dual-stack lite model is build on IPv4 over IPv6 tunnels to cross the network to reach a carrier-grade IPv4-IPv4 NAT. As such, it simplify the management of the service provider network and provide the customer the benefit of having only one layer of NAT. The additional benefit of this model is to gradually introduce IPv6 in the Internet by making it virtually backward compatible with IPv4.

2.6. Domain of application

TOC

The dual-stack lite deployment model has been designed with broadband networks in mind. It is certainly applicable to other domains although the author does not make any specific claim of suitability.

---

3. Expectations for dual-stack lite deployment

TOC

3.1. Expectations for home gateway based scenarios

TOC

This section mainly address home style networks characterized by the presence of a home gateway.
Legacy, unmodified, IPv4-only devices inside the home network are expected to keep using RFC1918 address space, a-la 192.168.0.0/16 and

should be able to access the IPv4 Internet in a similar way they do it today through a home gateway IPv4 NAT.
Unmodified IPv6 capable devices are expected to be able to reach directly the IPv6 Internet, without going through any translation. It is expected that most IPv6 capable devices will also be IPv4 capable and will simply be configured with an IPv4 RFC1918 style address within the home network and access the IPv4 Internet the same way as the legacy IPv4-only devices within the home.
IPv6-only devices that do not include code for an IPv4 stack are outside of the scope of this document.
It is expected that the home gateway is either software upgradable, replaceable or provided by the service provider as part of a new contract. Outside of early IPv6 deployments done prior to IPv4 exhaustion using some form of tunnel, this is pretty much a requirement to deploy any IPv6 service to the home. It is expected that this home gateway will be a dual stack capable device that would only be provisioned with IPv6 on its WAN side. IPv4 and IPv6 are expected to be locally provisonned on any LAN interfaces of such devices. IPv4 addresses on such interfaces are expected to be RFC1918. The key point here is that the service provider will not provision any IPv4 addresses for those home gateway devices.

3.2. Expectations for devices directly connected to the broadband service provider network

TOC

Under this deployment model, devices directly connected to the broadband service provider network without the presence of a home gateway will have to be dual stack capable devices. The service provider facing interface(s) of such device will only be provisioned with IPv6. IPv4 may or may not be provisioned locally on other interfaces of such devices. Similarly to the above section, the key point here is that the service provider will not provision any IPv4 addresses for those directly connected devices.
It is expected that directly connected devices will implement code to support the dual-stack lite functionality. The minimum support required is an IPv4 over IPv6 tunnel.
IPv4-only devices and IPv6-only devices are specifically left out of scope for this document. It is expected that most modern device directly connected to the service provider network would not have memory constraints to implement both stack.

TOC

3.3. Application expectations

Most applications that today work transparently through an IPv4 home gateway NAT should keep working the same way. However, it is not expected that applications that requires specific port assignment or port mapping from the NAT box will keep working. Details and recommendations for application behavior are outside the scope of this document and should be discussed in the behave working group.

3.4. Service provider network expectations TOC

The dual-stack lite deployment model is based on the notion that IPv4 addresses will be shared by several customers. This implies that the NAT functionality will move from the home gateway to a device hosted within the service provider network. It is important to observe that this functionality does not have to be performed deep in the core of the network and that it might be better implemented close to the aggregation point of customer traffic.

---

4. Dual-stack lite TOC

4.1. Dual-stack lite interface TOC

A dual-stack lite interface on a dual-stack capable device is modeled as a point to point IPv4 over IPv6 tunnel. Its configuration requires that the device is provisioned with IPv6 but does not require it to be provisioned with a global IPv4 by the service provider.
Any locally unique IPv4 address can be configure on the local side of the dual-stack lite tunnel. It is recommended that dual-stack lite implementations use simply 0.0.0.1.
Note: A future version of this draft may request IANA to reserve an IPv4 address for this usage.
The tunnel end point of a dual-stack lite interface is the IPv6 address of a dual-stack lite carrier-grade NAT within the service provider network.
A dual-stack lite interface is not required to maintain any state beside the IPv6 address of the remote tunnel end point and the local IPv4 address assigned to the local tunnel end point.

4.2. Dual-stack lite device

TOC

A dual-stack lite device is a dual-stack capable device implementing a dual-stack lite interface. In the absence of better routing information, a dual-stack lite device will configure a static IPv4 default route over the dual-stack lite interface.

4.3. Dual-stack lite home router

TOC

A dual-stack lite home router is a dual-stack capable home router implementing a dual-stack lite interface layered on top of its WAN interface. In the absence of better routing information, a dual-stack lite home router will configure a static IPv4 default route over the dual-stack lite interface.
Note: a dual-stack lite home router SHOULD NOT perform any IPv4 address translation. It should simply act as a router and pass packets from the LAN to the dual-stack lite interface and back without changing any address. The dual-stack lite router will have to take into account the lowered MTU of the tunnel and possibly perform IPv4 fragmentation.

4.4. Discovery of the dual-stack lite carrier-grade NAT device

TOC

The IPv6 address of a dual-stack lite carrier-grade NAT device can be configured on a dual-stack lite interface using a variety of way, ranging from out-of-band mechanism, manual configuration, a to-be-defined DHCPv6 option or a to-be-defined IPv6 router advertisement. It is expected that over time all the above methods and maybe more will be defined. The requirements and specifications of such methods are out of scope for this document.

4.5. Dual-stack lite carrier-grade NAT

TOC

A dual-stack lite carrier grade NAT is a special IPv4 to IPv4 NAT deployed within the service provider network. It is reachable by customers via a series of point to point IPv4 over IPv6 tunnels.
A dual-stack lite carrier-grade NAT uses a combination of the IPv6 source address of the tunnel and the inner IPv4 source address to establish and maintain the IPv4 NAT mapping table.
A dual-stack lite carrier-grade NAT does not have to perform any IPv6-IPv6, IPv6-IPv4 or IPv4-IPv6 NAT.

A dual-stack lite carrier-grade NAT should implement a full-cone NAT with hair-pinning (symmetric NAT may break applications using several simultaneous connections). It will have to implement the ALGs to support the classic applications. However, manual port forwarding or UPnP may or may not be supported.

4.6. Carrier-grade NAT considerations

TOC

Because IPv4 addresses will be share among customers and potentially a large address space reduction factor may be applied, in average, only a limited number of TCP or UDP port numbers will be available per customer. This means that applications opening a very large number of TCP ports may have a harder time to work. For example, it has been reported that a very well know web site was using AJAX techniques and was opening up to 69 TCP ports per web page... If we make the hypothesis of an address space reduction of a factor 100 (one IPv4 address per 100 customers), and 65k ports per IPv4 addresses available, that makes a total of 650 ports available simultaneously to be shared among the various devices behind the dual-stack lite tunnel end-point.

---

5. Multicast considerations

TOC

This document only describes unicast IPv4 as IPv4 Multicast is not widely deployed in broadband networks. Some multicast IPv4 considerations might to be discussed as well in a future revision of this document.

---

6. Comparison with an architecture with multiple-layers of NAT

TOC

An alternative architecture could consist on layering multiple levels of IPv4-IPv4 NAT toward the edges of the service provider network. Such architecture has a key advantage: it would work with any existing IPv4 home gateway. However, it would have a number of drawbacks:

- Each NAT device in the path will have its own application level gateways, increasing the odds of failure or miss-configuration.

- The larger private IPv4 address space available today is Net 10.0.0.0/8. It can in theory accommodate for about 16 million addresses, in practice, with an 80% allocation efficiency, it can address about 13 million devices. This may not be enough for

existing and future large scale deployments, thus multiple overlapping copies of Net 10 might have to be used simultaneously. This in itself create more complexity:

- If multiple copies of Net 10 are in use, network troubleshooting gets more difficult. One first need to figure out in which Net 10 realm the customer is before sending a ping to a home gateway to test it. This means that provisioning systems need to be modified to include this information.

- Multiple overlapping copies of Net 10 often intersect in some places with routers and firewalls. The configuration of such devices need to carefully take into accounts the overlapping address space. Debugging problems created by miss-configuration can be time consuming.

* Both legacy customers with global IPv4 addresses and new customers with private IPv4 addresses may be connected to the same aggregation router. That router will have to make the decision to send packets directly to the Internet or via a translator based on the source address of those packets, which is known as source-based routing. Although not impossible to implement, this adds complexity to the management of the network.

None of the issues described above are show stoppers, but put together, they seriously increase the complexity of the management of the network.

---

7. Comparison with NAT-PT (or its potential replacements)

TOC

NAT-PT [RFC2766] (Tsirtsis, G. and P. Srisuresh, “Network Address Translation - Protocol Translation (NAT-PT),” February 2000.) deals with the translation from IPv6 to IPv4 and vice versa. As such, it would not help solving the problem of offering IPv4 service to legacy IPv4 hosts. NAT-PT is targeted at green field IPv6 deployments, allowing them to access services and content on the IPv4 Internet. In that sense, NAT-PT could be in competition with dual-stack lite for green field deployment of new devices directly connected to the broadband service provider network.
In this situation, NAT-PT has the advantage of enabling to remove entirely the IPv4 stack on edge devices. This may be critical on sensor type devices with a very small memory footprint. However, it is unclear if such devices really need to have access to the whole global IPv4 Internet in the first place or if they only need to communicate with servers that can be made IPv6 enable.
In the more general case, dual-stack lite has several advantages over NAT-PT:

*Dual-stack lite does not require any hack to the DNS. In other words, there is no need to synthesize fake AAAA records to represent IPv4 addresses. This make DNSsec works more reliably.

*Because of the DNS ALG hack, NAT-PT places restriction on the topology, in most cases requiring that all exiting traffic go through a single point of contention. Because there is no DNS ALG with dual-stack lite and because each dual-stack lite device can be directed independently to a different dual-stack lite NAT, the dual-stack lite architecture can scale better.

*ALG sometimes need to manipulate literal IP address in the payload of packets. In the case of an IPv4-IPv4 NAT, this is a simple 32 bit field replacement. In the case of IPv6-IPv4 (or IPv4-IPv6) NAT, a 128 bit field need to be substituted by a 32 bit field (or vice versa). This makes NAT-PT ALG more complex than dual-stack lite NAT ALG.

For more detail on NAT-PT related issues, see [RFC4966] (Aoun, C. and E. Davies, “Reasons to Move the Network Address Translator - Protocol Translator (NAT-PT) to Historic Status,” July 2007.).

---

8. Comparison with DSTM

TOC

DSTM [I-D.bound-dstm-exp] (Bound, J., “Dual Stack IPv6 Dominant Transition Mechanism (DSTM),” October 2005.) was adressing IPv6 backward compatibility with IPv4 by providing a temporary IPv4 address to dual-stack nodes. Connectivity was also provided by the way of IPv4 over IPv6 tunnels. However, DSTM was relying on nodes acquiring and releasing IPv4 addresses on a need to have basis. It is the author opinion that such mechanism may not provide the necessary savings in address space for large scale broadband deployments.

---

9. Acknowledgements

TOC

The author would like to acknowledge the role of Mark Townsley for his input on the overall architecture of this technology by pointing this work in the direction of [I-D.droms-softwires-snat] (Droms, R. and B. Haberman, “Softwires Network Address Translation (SNAT),” July 2008.). Also to be acknowledged are the many discussions with a number of people including Shin Miyakawa, Katsuyasu Toyama, Akihide Hiura, Takashi Uematsu, Tetsutaro Hara, Yasunori Matsubayashi, Ichiro Mizukoshi. The

auhor would also like to thank David Ward, Jari Arkko, Thomas Narten and Geoff Huston for their constructive feedback.

10. IANA Considerations

TOC

This draft does not request any IANA action.

11. Security Considerations

TOC

Security issues associated with NAT have long been documented. See [RFC2663] (Srisuresh, P. and M. Holdrege, "IP Network Address Translator (NAT) Terminology and Considerations," August 1999.) and [RFC2993] (Hain, T., "Architectural Implications of NAT," November 2000.). However, moving the NAT functionality from the home gateway to the core of the service provider network and sharing IPv4 addresses among customers create additional requirements when logging data for abuse treatment. With any architecture including a carrier-grade NAT, IPv4 addresses and a timestamps are no longer sufficient to identify a particular broadband customer. Additional information like TCP port numbers will be be required for that purpose.

12. References

TOC

12.1. Normative references

TOC

[RFC2119]	Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels," BCP 14, RFC 2119, March 1997 (TXT, HTML, XML).

12.2. Informative references

TOC

[I-D.bound-dstm-exp]	Bound, J., "Dual Stack IPv6 Dominant Transition Mechanism (DSTM)," draft-bound-dstm-exp-04 (work in progress), October 2005 (TXT).
[I-D.droms-softwires-snat]	Droms, R. and B. Haberman, "Softwires Network Address Translation (SNAT)," draft-droms-softwires-snat-01 (work in progress), July 2008 (TXT).

[RFC2663]	Srisuresh, P. and M. Holdrege, “IP Network Address Translator (NAT) Terminology and Considerations,” RFC 2663, August 1999 (TXT).
[RFC2766]	Tsirtsis, G. and P. Srisuresh, “Network Address Translation - Protocol Translation (NAT-PT),” RFC 2766, February 2000 (TXT).
[RFC2993]	Hain, T., “Architectural Implications of NAT,” RFC 2993, November 2000 (TXT).
[RFC4966]	Aoun, C. and E. Davies, “Reasons to Move the Network Address Translator - Protocol Translator (NAT-PT) to Historic Status,” RFC 4966, July 2007 (TXT).

Author's Address

TOC

	Alain Durand
	Comcast
	1500 Market st
	Philadelphia, PA 19102
	USA
Email:	alain_durand@cable.comcast.com

Full Copyright Statement

TOC

Intellectual Property