Internet Engineering Task Force	A. Durand	
Internet-Draft	Comcast	
Intended status: Informational	R. Droms	
Expires: May 7, 2009	Cisco	
	B. Haberman	
	JHU APL	
	J. Woodyatt	
	Apple	
	November 03, 2008	

TOC

Dual-stack lite broadband deployments post IPv4 exhaustion
draft-durand-softwire-dual-stack-lite-01

Status of this Memo

Abstract

The common thinking for more than 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, well 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 costs and benefits of deploying IPv6. Dual-stack lite 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 document presents views on IP deployments after the exhaustion of IPv4 addresses and some of the necessary technologies to achieve it. The views expressed are the authors' personal opinions and in no way imply that Comcast plans to deploy or that Cisco will implement 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, i.e., 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

The dual-stack lite technology is intended for maintaining connectivity to legacy IPv4 devices and networks after the exhaustion of the IPv4 address space while service provider networks make a transition to IPv6-only deployments. This section describes some of the specific legacy scenarios addressed by dual-stack lite.

2.1. Legacy edges of the Internet for broadband customers

TOC

Broadband home customers have a mix and match of IP enabled devices. The most recent operating systems (e.g., Windows Vista, Mac OS X and various versions of Linux) can operate in an IPv6-only environment; however most of the legacy devices can't. Windows XP, for example, 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.

TOC

2.2. Content and Services available on the Internet

IPv6 deployment has taken a very long time to take off, so the current situation is that almost none of the content and services available on the Internet are accessible over IPv6. This situation 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, content 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.

2.4. Burden on service providers

TOC

As a conclusion, broadband service providers may be faced with the situation where they have IPv4 customers who need to communicate with IPv4 servers on the Internet but may not have any IPv4 addresses left to assign to those customers. A service providers may also be in a situation where it wants to deploy IPv6 in its core network, avoiding the use of scarce IPv4 addresses. However, without some form of backward compatibility with IPv4, the cost and the benefits of deploying IPv6 are not aligned, making incremental IPv6 deployment very difficult.

---

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 provisioned 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

The core ideas behind dual-stack lite are:

- Move from a deployment model where a globally unique IPv4 address is provisioned 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

- Provide transport of IPv4 traffic to customers over a core network that uses only IPv6

Instead of relying on a cascade of NATs or NAT-PT, the dual-stack lite model is built on IPv4 over IPv6 tunnels to cross the network to reach a carrier-grade IPv4-IPv4 NAT. As such, it simplifies the management of the service provider network by using only IPv6 and provides 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.

TOC

4.1. Domain of application

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

4.2. 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 configured on the subscriber network end of the dual-stack lite tunnel. In the case of dual-stack lite in which the tunnel endpoint is in a host Section 5.2 (Host based architecture), it is recommended that dual-stack lite implementations use the well known value a.b.c.d (to be defined by IANA) as the IPv4 host side of the tunnel and a.b.c.d+1 (TBD by IANA) as the address of the IPv4 default gateway, with a netmask to cover a /30 network.
Note: because of this static configuration using well known values, there is no need to run a DHCPv4 client on a Dual-stack lite interface.
The service provider network 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.

4.3. 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.

4.4. 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. The dual-stack lite home router can use the IPv4 address a.b.c.d (TBD by IANA) to source its own IPv4 packets, embedded into the IPv6 tunnel. If the dual-stack lite home router need

to configure a router pointing to an IPv4 default router, it can use the value a.b.c.d+1 (TBD by IANA) for that purpose with a prefix It also configure a.b.c.d+1 (TBD by IANA), with a netmask to cover a /30 network.
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.

4.5. Dual-stack lite router

TOC

The concept of a dual-stack lite home router can be extended to any IPv4 router serving as a gateway between a leaf IP domain and the rest of the Internet.

4.6. 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 methods, ranging from an out-of-band mechanism, manual configuration, a to-be-defined DHCPv6 option [I-D.dhankins-softwire-tunnel-option] (Hankins, D. and T. Mrugalski, “Dynamic Host Configuration Protocol (DHCPv6) Option for Dual-Stack Lite,” November 2009.) or a to-be-defined IPv6 router advertisement. It is expected that over time some or all the above methods, as well as others, will be defined. The requirements and specifications of such methods are out of scope for this document.

4.7. 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 can use the IPv4 address a.b.c.d+1 (TBD by IANA) in the IPv4 ICMP packets it will originate toward a dual-stack lite client to enable meaningful ping and traceroute results.

5. Example architectures

TOC

The underlying technology behind dual-stack lite is the combination of two well-known technologies: NAT and tunneling. This combination can be best described using the terminology developed in the softwire working group as Softwire NAT, or SNAT.
Two architectures can be deployed for dual-stack lite. One is targeting the legacy installed base of IPv4 only hosts (and dual-stack capable hosts) sitting behind a gateway. The second is targeting dual-stack capable hosts initiating the tunnel themselves.

5.1. Router-based architecture

TOC

This architecture is targeted at residential broadband deployments but can be adapted easily to other types of deployment where the installed base of IPv4-only device is important.
As illustrated in Figure 1 (SNAT gateway-based architecture), this dual-stack lite deployment model consists of three components: the dual-stack lite home router, the dual-stack lite carrier-grade NAT and a softwire between the softwire initiator (SI) in the dual-stack lite home router and the softwire concentrator (SC) in the dual-stack lite carrier-grade NAT. The dual-stack lite carrier-grade NAT performs IPv4-IPv4 NAT translations to multiplex multiple subscribers through a single global IPv4 address. Overlapping address spaces used by subscribers are disambiguated through the identification of tunnel endpoints.

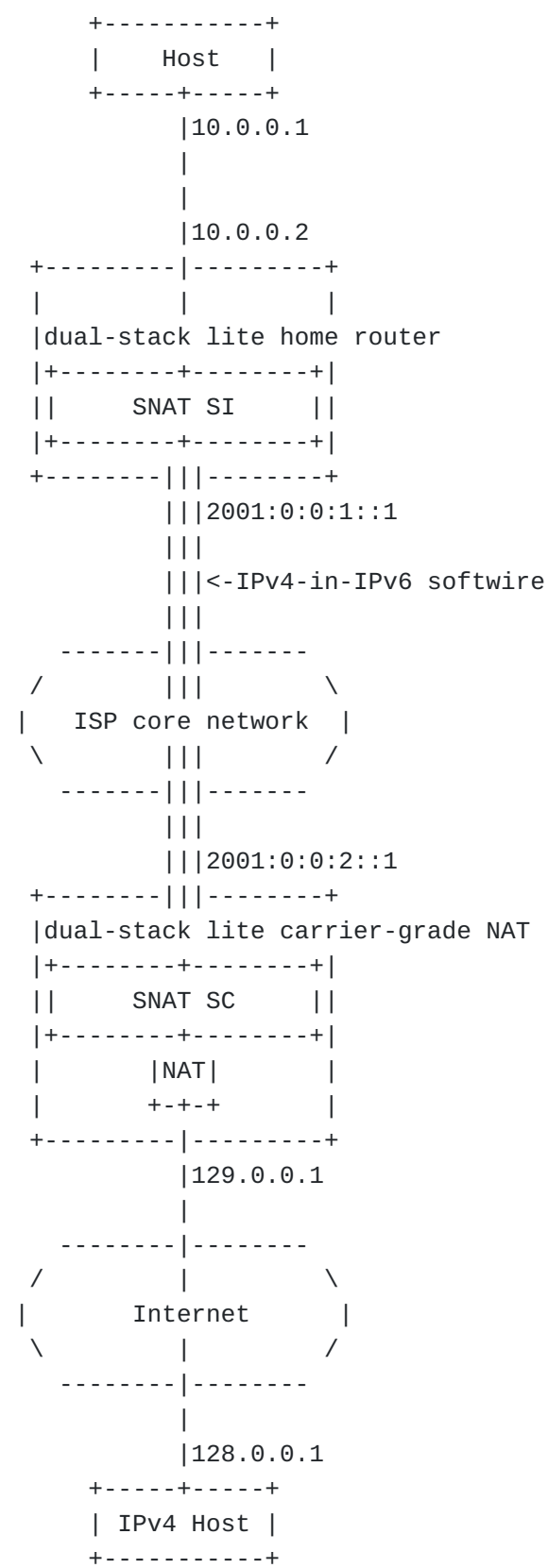
Host
10.0.0.1
10.0.0.2
dual-stack lite home router
SNAT SI
2001:0:0:1::1
<-IPv4-in-IPv6 softwire
ISP core network
2001:0:0:2::1
dual-stack lite carrier-grade NAT
SNAT SC
NAT
129.0.0.1
Internet
128.0.0.1
IPv4 Host

Figure 1: SNAT gateway-based architecture

Notes:

*The dual-stack lite home router is not required to be on the same link as the host

*The dual-stack lite home router could be replaced by a dual-stack lite router in the service provider network

The resulting solution accepts an IPv4 datagram that is translated into an IPv4-in-IPv6 softwire datagram for transmission across the softwire. At the corresponding endpoint, the IPv4 datagram is decapsulated, and the translated IPv4 address is inserted based on a translation from the softwire.

5.1.1. Example message flow

TOC

In the example shown in Figure 2 (Outbound Datagram), the translation tables in the dual-stack lite carrier-grade NAT is configured to forward between IP/TCP (10.0.0.1/10000) and IP/TCP (129.0.0.1/5000). That is, a datagram received by the dual-stack lite home router from the host at address 10.0.0.1, using TCP DST port 10000 will be translated a datagram with IP SRC address 129.0.0.1 and TCP SRC port 5000 in the Internet.

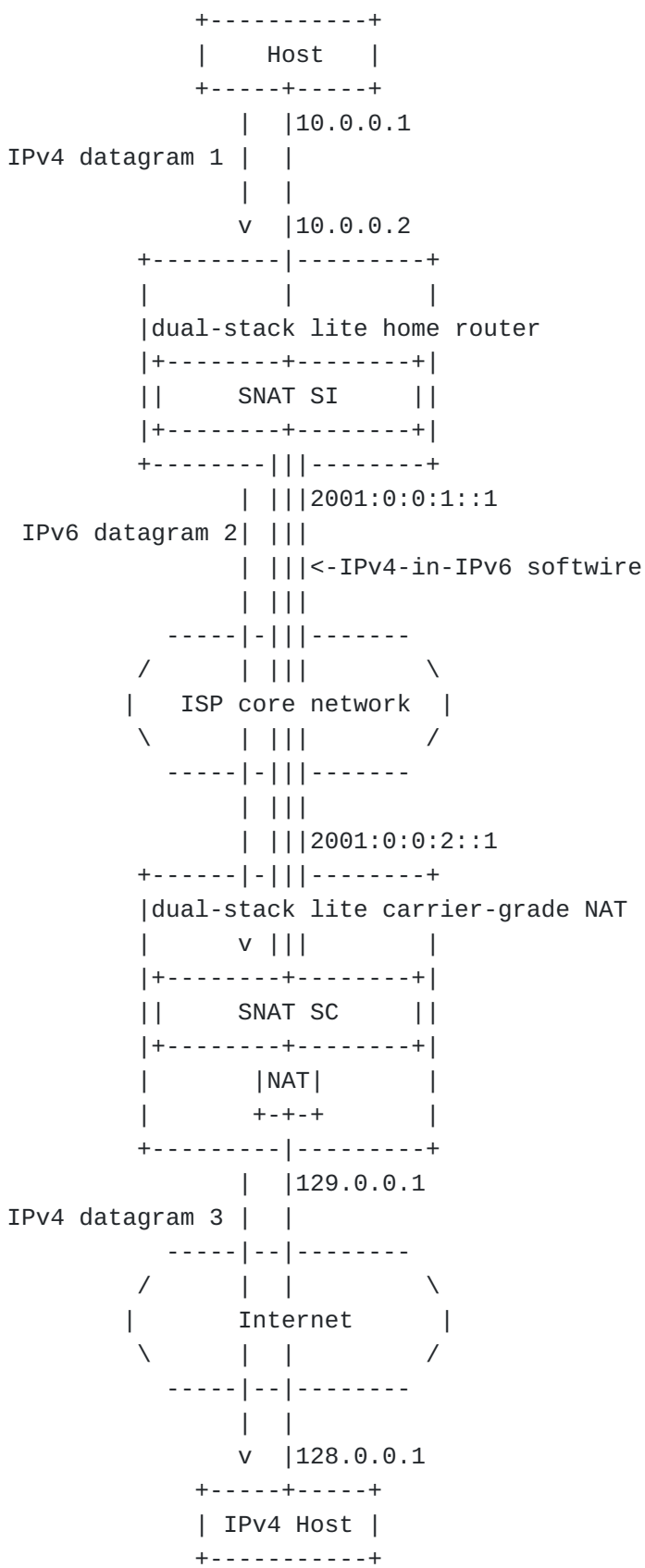
Host
10.0.0.1
IPv4 datagram 1
10.0.0.2
dual-stack lite home router
SNAT SI
2001:0:0:1::1
IPv6 datagram 2
<-IPv4-in-IPv6 softwire
ISP core network
2001:0:0:2::1
dual-stack lite carrier-grade NAT
SNAT SC
NAT
129.0.0.1
IPv4 datagram 3
Internet
128.0.0.1
IPv4 Host

Figure 2: Outbound Datagram

Datagram	Header field	Contents
IPv4 datagram 1	IPv4 Dst	128.0.0.1
	IPv4 Src	10.0.0.1
	TCP Dst	80
	TCP Src	10000
---	---	---
IPv6 Datagram 2	IPv6 Dst	2001:0:0:2::1
	IPv6 Src	2001:0:0:1::1
	IPv4 Dst	128.0.0.1
	IPv4 Src	10.0.0.1
	TCP Dst	80
	TCP Src	10000
---	---	---
IPv4 datagram 3	IPv4 Dst	128.0.0.1
	IPv4 Src	129.0.0.1
	TCP Dst	80
	TCP Src	5000

Datagram header contents

When datagram 1 is received by the dual-stack lite home router, the SI function encapsulates the datagram in datagram 2 and forwards it to the dual-stack lite carrier-grade NAT over the softwire.
When it receives datagram 2, the SC in the dual-stack lite carrier-grade NAT hands the IPv4 datagram to the NAT, which determines from its translation table that the datagram received on Softwire_1 with TCP SRC port 10000 should be translated to datagram 3 with IP SRC address 129.0.0.1 and TCP SRC port 5000.
Figure 3 (Inbound Datagram) shows an inbound message received at the dual-stack lite carrier-grade NAT. When the NAT function in the dual-stack lite carrier-grade NAT receives datagram 1, it looks up the IP/TCP DST in its translation table. In the example in Figure 3, the NAT translates the TCP DST port to 10000, sets the IP DST address to 10.0.0.1 and hands the datagram to the SC for transmission over Softwire_1. The SI in the dual-stack lite home router decapsulates IPv4 datagram from the inbound softwire datagram, and forwards it to the host.

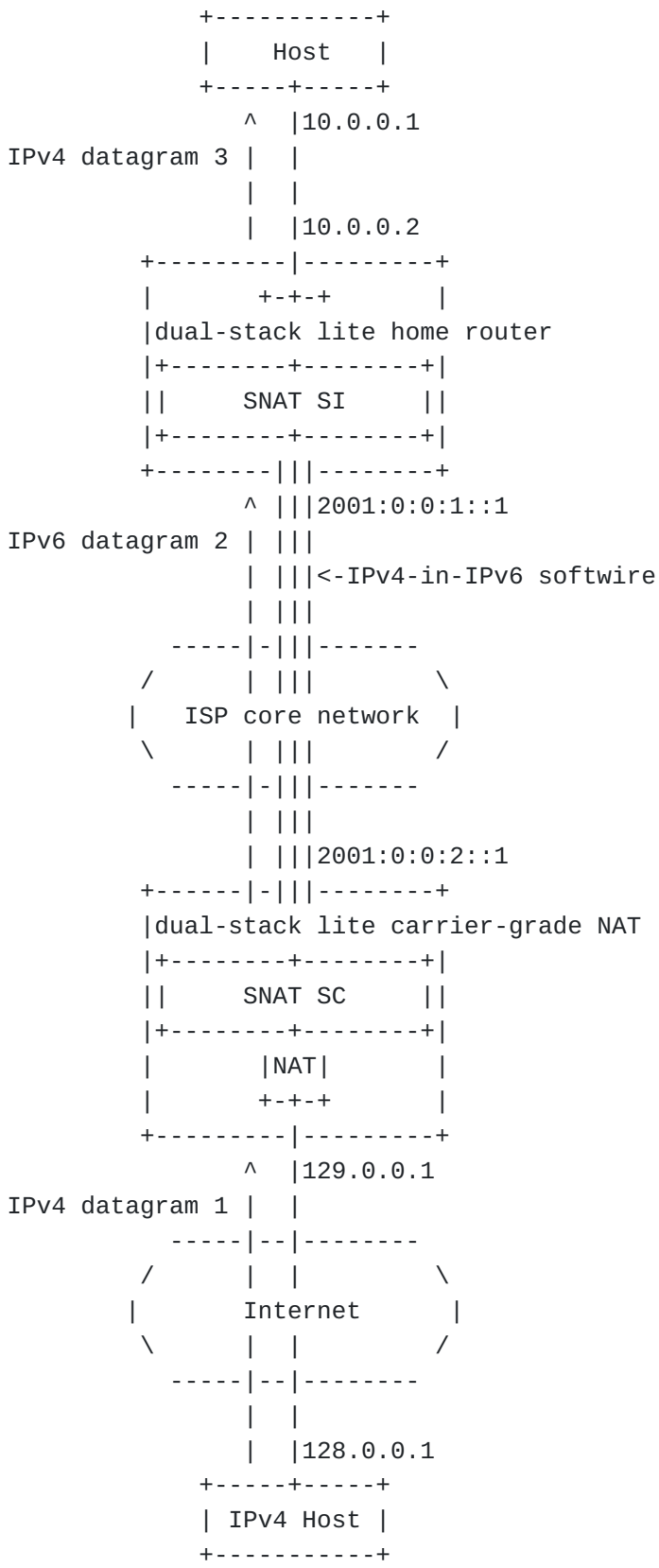

Figure 3: Inbound Datagram

Datagram	Header field	Contents
IPv4 datagram 1	IPv4 Dst	129.0.0.1
	IPv4 Src	128.0.0.1
	TCP Dst	5000
	TCP Src	80
---	---	---
IPv6 Datagram 2	IPv6 Dst	2001:0:0:1::1
	IPv6 Src	2001:0:0:2::1
	IPv4 Dst	10.0.0.1
	IP Src	128.0.0.1
	TCP Dst	10000
	TCP Src	80
---	---	---
IPv4 datagram 3	IPv4 Dst	10.0.0.1
	IPv4 Src	128.0.0.1
	TCP Dst	10000
	TCP Src	80

Datagram header contents

5.1.2. Translation details

TOC

The dual-stack lite carrier-grade NAT has a NAT that translates between softwire/port pairs and IPv4-address/port pairs. The same translation is applied to IPv4 datagrams received on the device's external interface and from the softwire endpoint in the device.
In Figure 2 (Outbound Datagram), the translator network interface in the dual-stack lite carrier-grade NAT is on the Internet, and the softwire interface connects to the dual-stack lite home router. The dual-stack lite carrier-grade NAT translator is configured as follows:

> **Network interface:** Translate IPv4 destination address and TCP destination port to the softwire identifier and TCP destination port

Softwire interface:

> Translate softwire identifier and TCP source port to IPv4 source address and TCP source port

Here is how the translation in Figure 3 (Inbound Datagram) works:

*Datagram 1 is received on the dual-stack lite carrier-grade NAT translator network interface. The translator looks up the IPv4-address/port pair in its translator table, rewrites the IPv4 destination address to 10.0.0.1 and the TCP source port to 10000, and hands the datagram to the SE to be forwarded over the softwire.

*The IPv4 datagram is received on the dual-stack lite home router SI. The SI function extracts the IPv4 datagram and the dual-stack lite home router forwards datagram 3 to the host.

Softwire/IPv4/Port	IPv4/Port
Softwire_1/10.0.0.1/TCP 10000	129.0.0.1/TCP 5000

dual-stack lite carrier-grade NAT translation table

5.2. Host based architecture

TOC

This architecture is targeted at new, large scale deployments of dual-stack capable devices implementing a dual-stack lite interface.
As illustrated in Figure 4 (SNAT host-based architecture), this dual-stack lite deployment model consists of three components: the dual-stack lite host, the dual-stack lite carrier-grade NAT and a softwire between the softwire initiator (SI) in the host and the softwire concentrator (SC) in the dual-stack lite carrier-grade NAT. The dual-stack lite host is assumed to have IPv6 service and can exchange IPv6 traffic with the dual-stack lite carrier-grade NAT.
The dual-stack lite carrier-grade NAT performs IPv4-IPv4 NAT translations to multiplex multiple subscribers through a single global IPv4 address. Overlapping IPv4 address spaces used by the dual-stack lite hosts are disambiguated through the identification of tunnel endpoints.
In this situation, the dual-stack lite host configures the well known IPv4 address a.b.c.d (TBD by IANA) on it dual-stack lite interface

acting as the SI. It also configure a.b.c.d+1 (TBD by IANA) as the address of its default gateway, with a netmask to cover a /30 network.

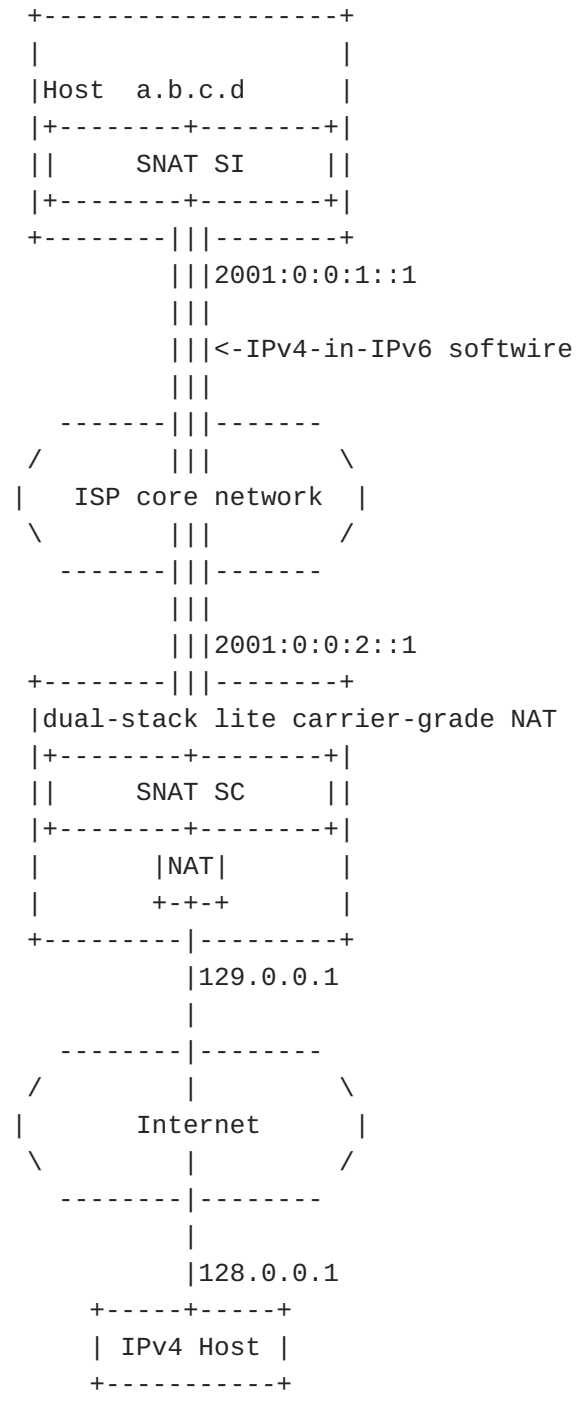

Figure 4: SNAT host-based architecture

The resulting solution accepts an IPv4 datagram that is translated into an IPv4-in-IPv6 softwire datagram for transmission across the softwire. At the corresponding endpoint, the IPv4 datagram is decapsulated, and the translated IPv4 address is inserted based on a translation from the softwire.

5.2.1. Example message flow

TOC

In the example shown in Figure 5 (Outbound Datagram), the translation tables in the dual-stack lite carrier-grade NAT is configured to forward between IP/TCP (a.b.c.d/10000) and IP/TCP (129.0.0.1/5000). That is, a datagram received from the host at address a.b.c.d, using TCP DST port 10000 will be translated a datagram with IP SRC address 129.0.0.1 and TCP SRC port 5000 in the Internet.

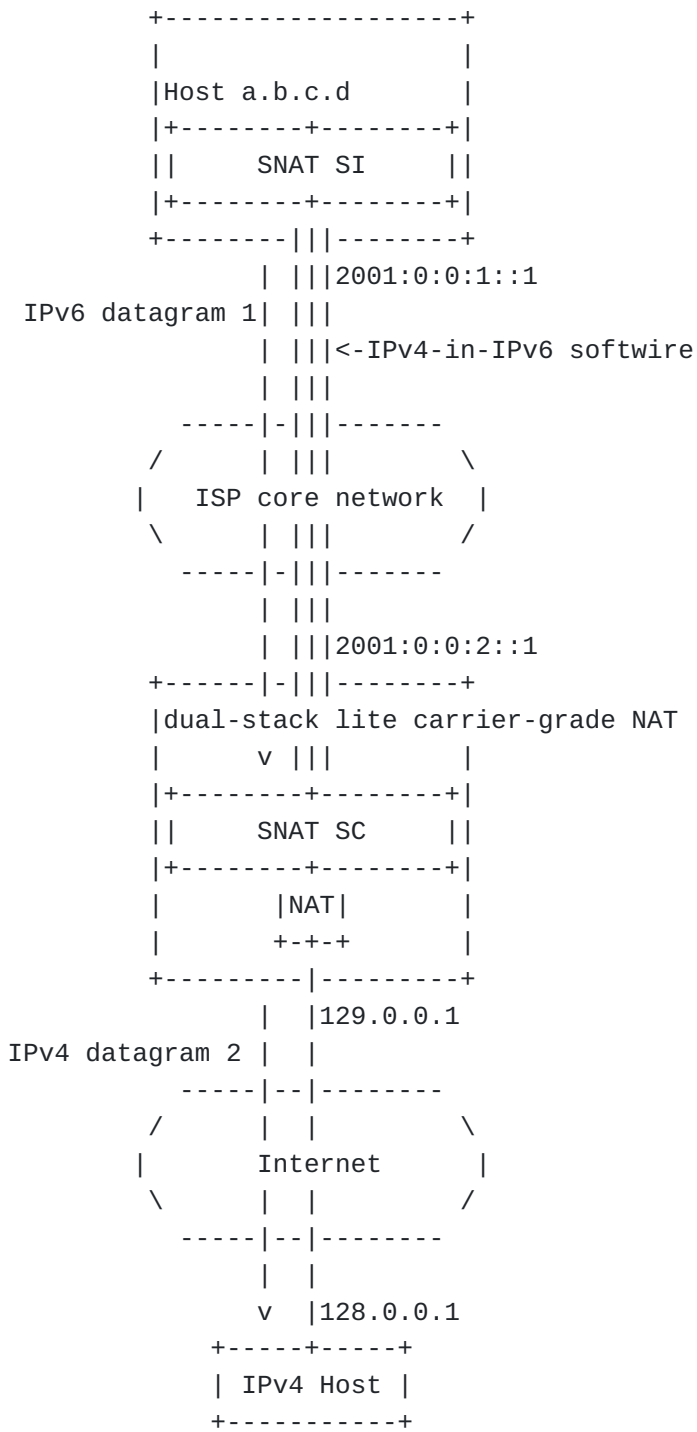

Figure 5: Outbound Datagram

Datagram	Header field	Contents
IPv6 Datagram 1	IPv6 Dst	2001:0:0:2::1
	IPv6 Src	2001:0:0:1::1
	IPv4 Dst	128.0.0.1
	IPv4 Src	a.b.c.d
	TCP Dst	80
	TCP Src	10000
---	---	---
IPv4 datagram 2	IPv4 Dst	128.0.0.1
	IPv4 Src	129.0.0.1
	TCP Dst	80
	TCP Src	5000

Datagram header contents

When sending an IPv4 packet, the dual-stack lite host encapsulates it in datagram 1 and forwards it to the dual-stack lite carrier-grade NAT over the softwire.
When it receives datagram 1, the SC in the dual-stack lite carrier-grade NAT hands the IPv4 datagram to the NAT, which determines from its translation table that the datagram received on Softwire_1 with TCP SRC port 10000 should be translated to datagram 3 with IP SRC address 129.0.0.1 and TCP SRC port 5000.
Figure 6 (Inbound Datagram) shows an inbound message received at the dual-stack lite carrier-grade NAT. When the NAT function in the dual-stack lite carrier-grade NAT receives datagram 1, it looks up the IP/TCP DST in its translation table. In the example in Figure 3, the NAT translates the TCP DST port to 10000, sets the IP DST address to a.b.c.d and hands the datagram to the SC for transmission over Softwire_1. The SI in the dual-stack lite home router decapsulates IPv4 datagram from the inbound softwire datagram, and forwards it to the host.

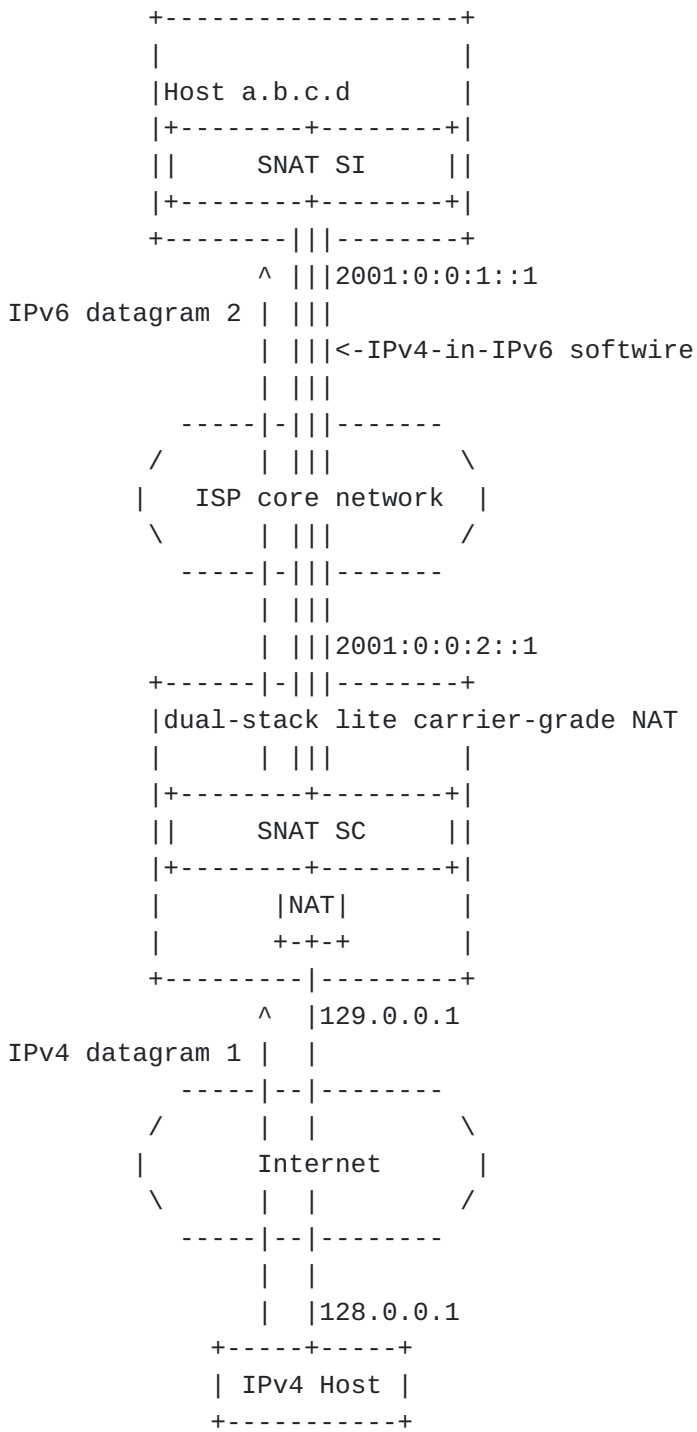

Figure 6: Inbound Datagram

Datagram	Header field	Contents
IPv4 datagram 1	IPv4 Dst	129.0.0.1
	IPv4 Src	128.0.0.1
	TCP Dst	5000
	TCP Src	80
---	---	---
IPv6 Datagram 2	IPv6 Dst	2001:0:0:1::1
	IPv6 Src	2001:0:0:2::1
	IPv4 Dst	a.b.c.d
	IP Src	128.0.0.1
	TCP Dst	10000
	TCP Src	80

Datagram header contents

5.2.2. Translation details

TOC

The translations happening in the dual-stack lite carrier-grade NAT are the same as in the previous examples. The well known IPv4 address a.b.c.d used by all the hosts are disambiguated by the IPv6 source address of the softwire.

---

6. Encapsulations

TOC

In its simplest deployment model, dual-stack lite only requires IPv4 in IPv6 encapsulation. In more complex scenario where a site gateway would play the role of the softwire initiator, more complex encapsulation might be desired. Thus dual-stack lite hosts, dual-stack lite home gateway and dual-stack lite NAT devices must at minimum implement IPv4 in IPv6 encapsulation. On top of that, dual-stack lite NAT devices should be able to support other encapsulation, like L2TPv2/v3, GRE, MPLS,...

TOC

7. Carrier-grade NAT considerations

A dual-stack lite carrier-grade NAT SHOULD implement behavior conforming to the best current practice, currently documented in [RFC4787] (Audet, F. and C. Jennings, “Network Address Translation (NAT) Behavioral Requirements for Unicast UDP,” January 2007.), [I-D.ietf-behave-tcp] (Guha, S., Biswas, K., Ford, B., Sivakumar, S., and P. Srisuresh, “NAT Behavioral Requirements for TCP,” September 2008.) and [I-D.ietf-behave-nat-icmp] (Srisuresh, P., Ford, B., Sivakumar, S., and S. Guha, “NAT Behavioral Requirements for ICMP protocol,” January 2009.). Other requirements for carrier-grade NATs can be found in [I-D.nishitani-cgn] (Yamagata, I., Nishitani, T., Miyakawa, S., Nakagawa, A., and H. Ashida, “Common requirements for IP address sharing schemes,” March 2010.). DISCUSSION: those requirements need to be harmonized.

7.1. Per customer port allocation

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 N 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 N=650 ports available simultaneously to be shared among the various devices behind the dual-stack lite tunnel end-point.
There is an important operational difference if those N ports are pre-allocated in a cookie-cutter fashion versus allocated on demand by incoming connections. This is a difference between an average of N ports and a maximum of N ports. Note that an hybrid system can be design where a limited number of ports are pre-allocated to customers for special applications and the others are reserved for dynamic allocations.

7.2. ALG

TOC

Concerns have been expressed that any type of carrier-grade NAT could stifle innovation by making it harder to deploy ALGs, and, as such, it would be better to architect the NAT at the edges of the network rather than in the core. Such an edge based architecture (SAM [I-D.despres-sam] (Despres, R., “Scalable Multihoming across IPv6

Local-Address Routing Zones Global-Prefix/Local-Address Stateless Address Mapping (SAM)," July 2009.), A+P [I-D.ymbk-aplusp] (Bush, R., "The A+P Approach to the IPv4 Address Shortage," October 2009.)) would require significantly more complexity to be placed either in the end host or in the home gateway than simply having an interface that encapsulate IPv4 in IPv6. This complexity may be an heavy price to pay considering that most recent applications have been developped to work without any ALG support at all, as illustrated by the trend to now use HTTP as transport... More over, as was stated during the Montreal IETF interim meeting, what really matters is not so much the placement of the NAT itself, but the control over the ALG. For that reason, the carrier-grade NAT SHOULD avoid performing any ALG on unknown protocols and provide a method for edge devices to learn about the external binding (IPv4 address+port) that will be used by the carrier-grade NAT to translate the packets and then perform any necessary ALG.
DISCUSSION: A carrier-grade NAT MAY implement ALGs supporting all the classic applications, e.g. FTP, RTSP/RTP, IPsec and PPTP VPN pass-through, etc.
Manual port forwarding or UPnP IGD may or may not be supported.

7.3. On-demand port reservation

TOC

A port mapping protocol might be developed to run between a dual-stack lite host (or a dual-stack lite router) and the dual-stack lite carrier-grade NAT to reserve at connection time a binding with an external IPv4 address and a port number, and for use by privately addressed hosts to determine the which public address the NAT will pair with it. In a dual-stack lite router, such a protocol could serve as a proxy for UPnP IGD (UPnP Forum, "Universal Plug and Play Internet Gateway Device Standardized Gateway Device Protocol," September 2006.) [UPnP-IGD] or NAT-PMP (Cheshire, S., "NAT Port Mapping Protocol (NAT-PMP)," April 2008.) [I-D.cheshire-nat-pmp].

7.4. Pre-allocating ports

TOC

An alternate mechanism would be to use a DHCPv4 option to request the allocation of a (small) block of port number on a shared IPv4 address. Such a mechanism is described in [I-D.bajko-v6ops-port-restricted-ipaddr-assign] (Bajko, G. and T. Savolainen, "Port Restricted IP Address Assignment," November 2008.)
An even simpler mechanism would be to enable a user to pre-register a limited number of external mappings on the service provider web site.

---

8. Future work

TOC

The items described bellow could be included in a future version of this document or be the object of a separate document.

8.1. 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.

8.2. 3rd party carrier-grade NAT

TOC

The dual-stack lite architecture can be easily extended to support 3rd party carrier-grade NATs. The dual-stack lite interface just need to be pointed to the IPv6 address of that 3rd party carrier-grade NAT instead of the IPv6 address of the service provider carrier-grade NAT. Implementation of dual-stack lite should enable users to override the mechanism used for automatic discovery of the carrier grade NAT and, for example, manually enter the DNS name of the selected carrier-grade NAT.

8.3. Interface initialization

TOC

The initialization sequence of each interface of a dual-stack lite node need to be analyzed and heuristics need to be defined to determined if each interface operates in IPv4 mode, IPv6 mode, dual-stack mode or dual-stack lite mode. The absence/presence of the DHCPv6 option discussed above in requests/responses could be a trigger to decide in which mode to operate.

---

9. 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.

---

10. 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.).

---

11. Comparison with DSTM

TOC

DSTM [I-D.bound-dstm-exp] (Bound, J., “Dual Stack IPv6 Dominant Transition Mechanism (DSTM),” October 2005.) was addressing 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 authors' opinion that such mechanism may not provide the necessary savings in address space for large scale broadband deployments.

TOC

12. Acknowledgements

The authors 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.). Note that this document results from a merging of [I-D.durand-dual-stack-lite] (Durand, A., “Dual-stack lite broadband deployments post IPv4 exhaustion,” July 2008.) and [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 author would also like to thank David Ward, Jari Arkko, Thomas Narten and Geoff Huston for their constructive feedback. A special thank you goes to Dave Thaler for his review and comments.

---

13. IANA Considerations

TOC

This draft request IANA to allocate a well know IPv4 a.b.c.d/30 network prefix. The IPv4 address a.b.c.d is reserved for sourcing IPv4 packets inside on IPv6 tunnel. The IPv4 address a.b.c.d+1 is reserved as the IPv4 address of the default router for such dual-stack lite hosts.

---

14. 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 where an IPv4 address does not uniquely represent an end host, 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.
Similarly, some attack mitigation techniques put an IPv4 address in a "penalty box" for a period of time ifan abnormal behavior is observed. Such techniques may need to be revisited as they would impact more than just one user (presumably the offender) at a time.

15. References

TOC

15.1. 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).

15.2. Informative references

TOC

[I-D.bajko-v6ops-port-restricted-ipaddr-assign]	Bajko, G. and T. Savolainen, “Port Restricted IP Address Assignment,” draft-bajko-v6ops-port-restricted-ipaddr-assign-02 (work in progress), November 2008 (TXT).
[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.cheshire-nat-pmp]	Cheshire, S., “NAT Port Mapping Protocol (NAT-PMP),” draft-cheshire-nat-pmp-03 (work in progress), April 2008 (TXT).
[I-D.despres-sam]	Despres, R., “Scalable Multihoming across IPv6 Local-Address Routing Zones Global-Prefix/Local-Address Stateless Address Mapping (SAM),” draft-despres-sam-03 (work in progress), July 2009 (TXT).
[I-D.dhankins-softwire-tunnel-option]	Hankins, D. and T. Mrugalski, “Dynamic Host Configuration Protocol (DHCPv6) Option for Dual-Stack Lite,” draft-dhankins-softwire-tunnel-option-05 (work in progress), November 2009 (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).
[I-D.durand-dual-stack-lite]	Durand, A., “Dual-stack lite broadband deployments post IPv4 exhaustion,” draft-durand-dual-stack-lite-00 (work in progress), July 2008 (TXT).
[I-D.ietf-behave-nat-icmp]	Srisuresh, P., Ford, B., Sivakumar, S., and S. Guha, “NAT Behavioral Requirements for ICMP protocol,” draft-ietf-behave-nat-icmp-12 (work in progress), January 2009 (TXT).
[I-D.ietf-behave-tcp]	Guha, S., Biswas, K., Ford, B., Sivakumar, S., and P. Srisuresh, “NAT Behavioral Requirements for TCP,” draft-ietf-behave-tcp-08 (work in progress), September 2008 (TXT).
[I-D.nishitani-cgn]	Yamagata, I., Nishitani, T., Miyakawa, S., Nakagawa, A., and H. Ashida, “Common requirements for IP address sharing schemes,” draft-nishitani-cgn-04 (work in progress), March 2010 (TXT).
[I-D.ymbk-aplusp]	Bush, R., “The A+P Approach to the IPv4 Address Shortage,” draft-ymbk-aplusp-05 (work in progress), October 2009 (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).
[RFC4787]	Audet, F. and C. Jennings, “Network Address Translation (NAT) Behavioral Requirements for Unicast UDP,” BCP 127, RFC 4787, January 2007 (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).
[UPnP-IGD]	UPnP Forum, “Universal Plug and Play Internet Gateway Device Standardized Gateway Device Protocol,” September 2006.

Authors' Addresses

TOC

	Alain Durand
	Comcast
	1500 Market st
	Philadelphia, PA 19102
	USA
Email:	alain_durand@cable.comcast.com
	Ralph Droms
	Cisco
	1414 Massachusetts Avenue
	Boxborough, MA 01714
	US
Phone:	+1 978.936.1674
Email:	rdroms@cisco.com
	Brian Haberman
	Johns Hopkins University Applied Physics Lab
	11100 Johns Hopkins Road
	Laurel, MD 20723-6099
	US
Phone:	+1 443 778 1319
Email:	brian@innovationslab.net
	James Woodyatt

	Apple Inc.
	1 Infinite Loop
	Cupertino, CA 95014
	US
Email:	jhw@apple.com

Full Copyright Statement

TOC

Intellectual Property