Behave
Internet-Draft
Updates: 2765, 2766
(if approved)
Intended status: Standards Track
Expires: March 21, 2009

X. Li
C. Bao
CERNET Center/ Tsinghua University
F. Baker
K. Yin
Cisco Systems

September 17, 2008

TOC

IVI Update to SIIT and NAT-PT
draft-baker-behave-ivi-01

Status of this Memo

Abstract

This note proposes an address and service architecture designed to facilitate transition from an IPv4 Internet to an IPv6 Internet. This service contains three parts: A DNS Application Layer Gateway, a stateful Network Address Translator that enables IPv6 clients to initiate connections to IPv4 servers and peers, and a stateless Network Address Translator that enables IPv4 and IPv6 systems to interoperate freely.
It is couched as an update to RFCs 2765 and 2766. This is because the stateless service is essentially the SIIT with a different address format, and because the DNS Application Layer Gateway and the stateful translator have significant similarities to NAT-PT. There are, however,

important differences from NAT-PT, responsive to the issues raised in RFC 4966.

Table of Contents

1. Introduction

TOC

This note documents the prototype being used for translation between the IPv4 CERNET and the IPv6 CNGI-CERNET2 networks. This uses the algorithms of SIIT (Nordmark, E., “Stateless IP/ICMP Translation Algorithm (SIIT),” February 2000.) [RFC2765] with a modified address format, and a modified version of NAT-PT (Tsirtsis, G. and P. Srisuresh, “Network Address Translation - Protocol Translation (NAT-PT),” February 2000.) [RFC2766]. In general, we recommend the use of native communication and dual stack deployment. However, in several scenarios, the temporary use of translation can simplify service deployment. Hence, we describe a translation function.

It should be understood that protocol translation in any form is not a viable long term solution for IPv6 deployment; it has value during a certain part of the adoption curve, but will become less relevant at later points in the adoption curve. The objective of any transition strategy, of which IVI is an example, is to facilitate transition, not to enter a phase of heightened operational and capital expenditure running two networks in parallel only to stay there. When IPv6 is widely deployed and economic conditions support the move, we expect service providers to withdraw IPv4 service.

The objectives of the translation function are to enable systems that are unable to communicate with each other due to routing, implementation, or parameter differences to communicate. Almost any translation function will connect IPv6 systems with IPv4 systems or systems in an IPv4 network. The difficulty is that this gives no incentive to administrations to move their servers and peers from the IPv4 domain to the IPv6 domain. Noting that dual stack implementations such as recommended in [RFC4213] (Nordmark, E. and R. Gilligan, “Basic Transition Mechanisms for IPv6 Hosts and Routers,” October 2005.) are not being widely deployed by operators, the IVI model is designed to facilitate placing servers and peers in the IPv6 domain, achieving native IPv6 connectivity without giving up IPv4 accessibility.

More specifically, the objectives are several:

*As with any network, IPv4 systems connected by an IPv4 network can talk among themselves and IPv6 systems connected by an IPv6 network can talk among themselves. The first objective is to preserve this and its scaling characteristics.

*If one or both domains are IPv4+IPv6 but there exist systems with only one architecture, we presume that IPv4 and IPv6 routing crosses the gateway or a parallel router, and the systems are able to communicate directly.

*We want to enable systems that have no IPv6 address to access servers and peers with IPv4-derived IPv6 addresses (IVI addresses) in the IPv6 domain. This requires translation similar to that described in SIIT (Nordmark, E., “Stateless IP/ICMP Translation Algorithm (SIIT),” February 2000.) [RFC2765]. This operation is stateless.

*We want to enable systems that have no IPv4 address to access servers and peers in the IPv4 domain. For systems with IPv4-derived IPv6 addresses (IVI addresses), this is solved by the SIIT extension described in this document, given the appropriate AAAA record by IVI DNS ALG. This operation is also stateless. Other systems with non-IVI IPv6 addresses require some form of stateful translation. This has similarities to the mechanisms described in NAT-PT (Tsirtsis, G. and P. Srisuresh, “Network Address Translation - Protocol Translation (NAT-PT),”

February 2000.) [RFC2766]. We wish to do this with a minimum of maintained state.

Some have questioned the need for IPv4 access to IPv6-only servers and peers, noting that in the Internet of 2008 there is no market requirement for such access and any server or peer will require accessibility from an IPv4 network. The issue is that this presumes a certain point in the adoption curve; at another point in the adoption curve, one hopes that there will be few takers for IPv4-only service. In between, before IPv6 service for a server or peer becomes a requirement, IPv6-only service for a server or peer must be feasible (it must be conceivable that a server or peer with an IPv6 address will be useful). We argue that it is easier for IPv6 service for a server or peer to become feasible if it is possible to configure it with an IPv4-derived IPv6 address than if it must also have IPv4 service. In the long term, we believe that translation is not a service that service providers will normally use, but is a helpful and perhaps necessary step in transitioning to an IPv6 world.

2. The IVI model

TOC

The Name "IVI" contracts "IV<->VI"; we are describing a translation connection between systems using IPv4 or IPv6 that cannot communicate using either IPv4 or IPv6. In any normal case where native communication is possible between two systems, we argue that it is preferable.

2.1. IVI Network Model and communication objectives

TOC

An IVI Network, as shown in Figure 1 (IVI Network Model), consists of two or more network domains connected by one or more IVI gateways. One of those networks either routes IPv4 but not IPv6, or contains some hosts that only implement IPv4. The other network either routes IPv6 but not IPv4, or contains some hosts that only implement IPv6. Both networks contain clients, servers, and peers. It would be advisable and preferable to implement a dual stack architecture in both domains, but either due to address scarcity or the process involved in IPv6 turn-up, that is not practical at the moment.

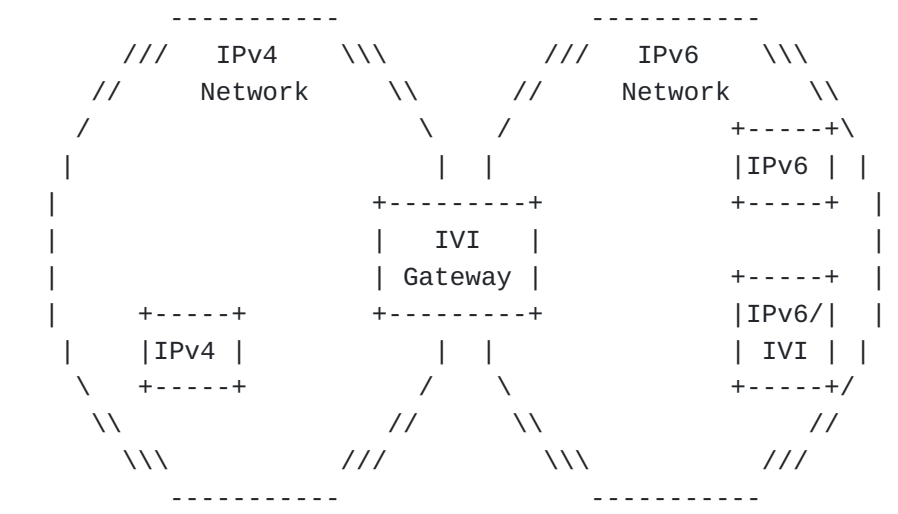

Figure 1: IVI Network Model

Clearly, there are issues in IP addressing, and routing, DNS, and the specifics of translation.

2.2. IVI Address Format

TOC

The IVI Address is an IPv4 address embedded in an IPv6 address and predictable by the gateway and systems on either side. The selection of the LIR prefix, including its length and absolute value, is at the option of the network administration; it is not fixed. Figure 2 (Example IVI Address Format) shows one possible model. It enables the IPv6 domain to assign the equivalent of IPv4 /24 prefixes to IPv6 LANs (/64).

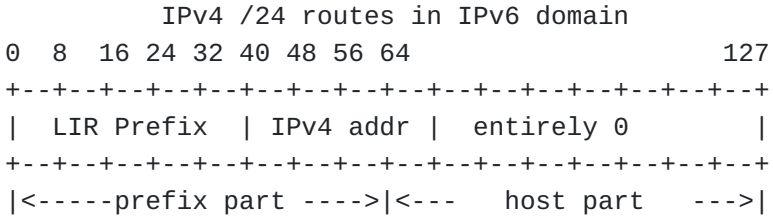

Figure 2: Example IVI Address Format

In the IPv4 domain, this represents a prefix no longer than /24. In the IPv6 domain, the "default route" advertising the entire IPv4 address space is the LIR /40 prefix. More specific prefixes up to /64 may be advertised as needed, or host (/128) routes.

The objective here is to enable the network administration to be in control of the impact of the tradeoff on its routing.
The need to change the address format used by SIIT bears repetition, although it has come up in other discussions. [RFC4291] (Hinden, R. and S. Deering, “IP Version 6 Addressing Architecture,” February 2006.) deprecated the address format with the brusque comment that "current IPv6 transition mechanisms no longer use these addresses." The reason that they were not widely deployed was that they gave network operators little control in routing, or ways to ensure that route redistribution worked correctly. A prefix that lets the LIR specify the upper bits gives the operator the flexibility to identify the IVI gateway advertising the prefix and better control the distribution of routes.

2.3. Routing in IVI networks

TOC

The IVI Gateway may be a general purpose router; in that mode, it operates like any other router. However, it also advertises one or more prefixes into both the IPv4 and the IPv6 domain, and when a datagram is directed to an address within the translation prefix(es), it translates the datagram.
As a Network Address Translator, the IVI Gateway offers one or both of two services: stateless translation of addresses conforming to Figure 2 (Example IVI Address Format) to and from IPv4 addresses, and stateful translation between IPv6 addressing and a combination of an IPv4 address and transport source port as is done in normal NATs.
In IPv4, the IVI gateway advertises the IPv4 prefix being used for stateless IVI address translation; for example, if an IPv4 /20 is being used as a set of /24 prefixes in the IPv6 domain, it would advertise a /20 into the IPv4 domain. If the IVI gateway is offering stateful translation, it may also advertise the addresses or prefix being used for that service unless another router handles this.
In IPv6, the IVI gateway advertises a “default route for global IPv4” - in the example given in Figure 2 (Example IVI Address Format), it would normally advertise the /40 LIR prefix. If that is inappropriate - there are multiple non-overlapping IPv4 domains or other concerns apply - it would advertise "more-specific" prefixes as appropriate.
In the IPv6 domain, the routers or hosts that have been assigned IVI prefixes or addresses subsidiary to the IVI prefix for a service advertise the IVI /64s corresponding to those IPv4 /24s.
Clearly, there may be multiple non-overlapping IPv4 domains, multiple non-overlapping IPv6 domains, and there may be multiple IVI gateways. These are handled in a manner consistent with normal routing practice in the Internet.

As shown in Figure 3 (IVI Reachability example), routing is slightly more complex in an IVI service, but follows simple routing concepts. In this example,

*IPv4 interfaces can open a session to any IVI address (e.g. 4Host1 -> IVI1),

*IPv4 interfaces cannot open sessions to non-IVI IPv6 addresses (e.g. 4Host1 X-> 6Host1),

*IPv6 IVI interfaces can open a session to any IPv4 interface, statelessly (e.g. IVI1 -> 4Host1),

*Non-IVI IPv6 hosts can open sessions to IPv4 interfaces, statefully (e.g. 6Host1 -> 4Host1),

*Any two IPv4 hosts can open a session to either each other using native routing (e.g. 4Host1 -> 4Host2, 4Host2 -> 4Host1),

*Any two IPv6 hosts can open a session to either each other using native routing, even using the IVI addresses (e.g. 6Host1 -> IVI1, IVI1 -> 6Host1, 6Host1 -> 6Host2, IVI1 -> IVI2).

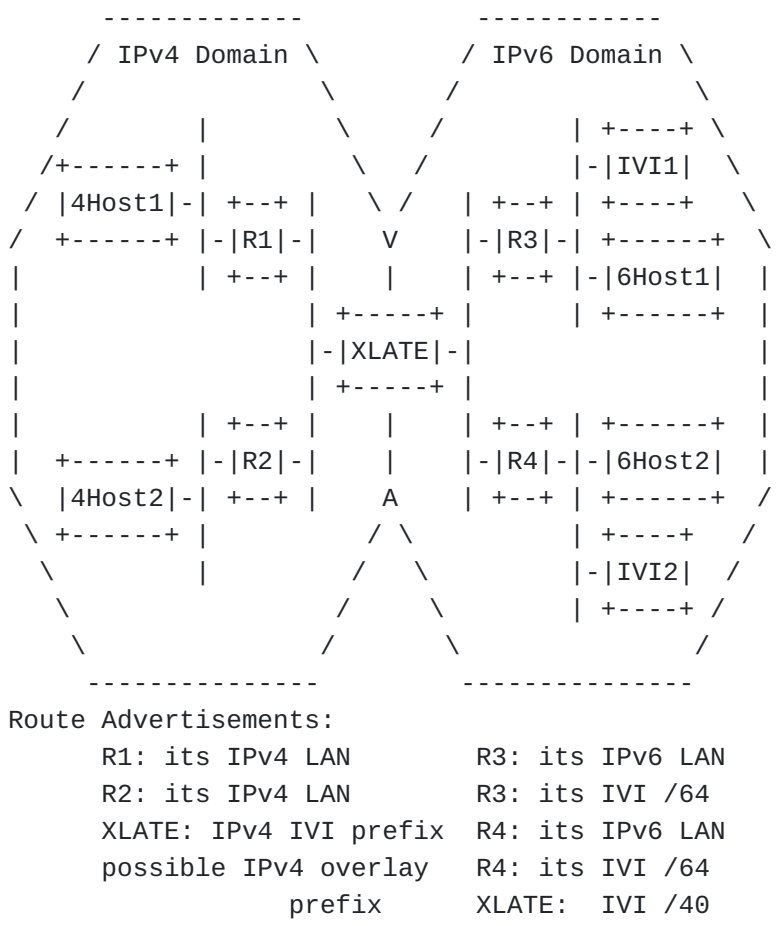

Figure 3: IVI Reachability example

2.4. DNS service in IVI networks

TOC

Rather than using the DNS Application Layer Gateway described in [RFC2766] (Tsirtsis, G. and P. Srisuresh, “Network Address Translation - Protocol Translation (NAT-PT),” February 2000.) as specified, the IVI DNS ALG is a one-way translation of A and MX records to AAAA records with a predictable address. The DNS server may be in the gateway or in a separate system related to it.

As illustrated in Figure 4 (Normal DNS Service), in the IPv4 domain, the DNS server holds and advertises A records for systems with IPv4 addresses and for systems (servers or peers) that have IVI addresses. These are generally pre-populated, if only via Dynamic DNS. The IPv4 network cannot distinguish them from other A records or from other IPv4 addresses, so this works without host changes.

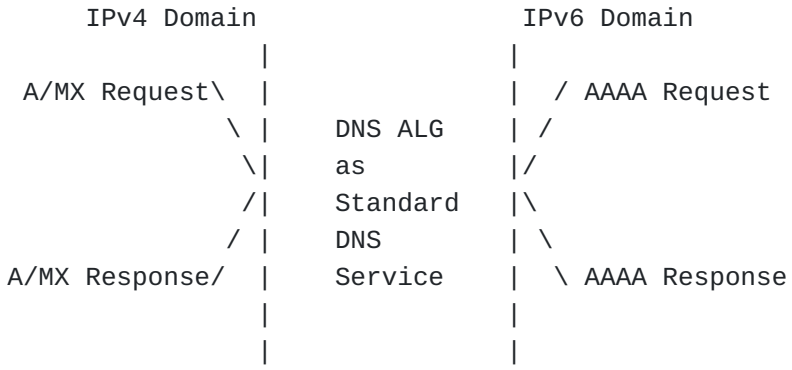

Figure 4: Normal DNS Service

Also as illustrated in Figure 4 (Normal DNS Service), in the IPv6 domain, the DNS server holds and advertises AAAA records in the usual fashion for systems with general IPv6 addresses.
As illustrated in Figure 5 (DNS Record Translation Service), in the IPv6 domain, when the DNS ALG receives a request for a AAAA record for which it has nothing to reply, or for which normal DNS processing receives a failure, it obtains an A or MX record from its own database or another server, manufactures a corresponding AAAA record using an IVI address, and returns that. The IPv6 network cannot distinguish between these and other AAAA records, or between these and any other address. Routing takes traffic through the gateway without host changes.

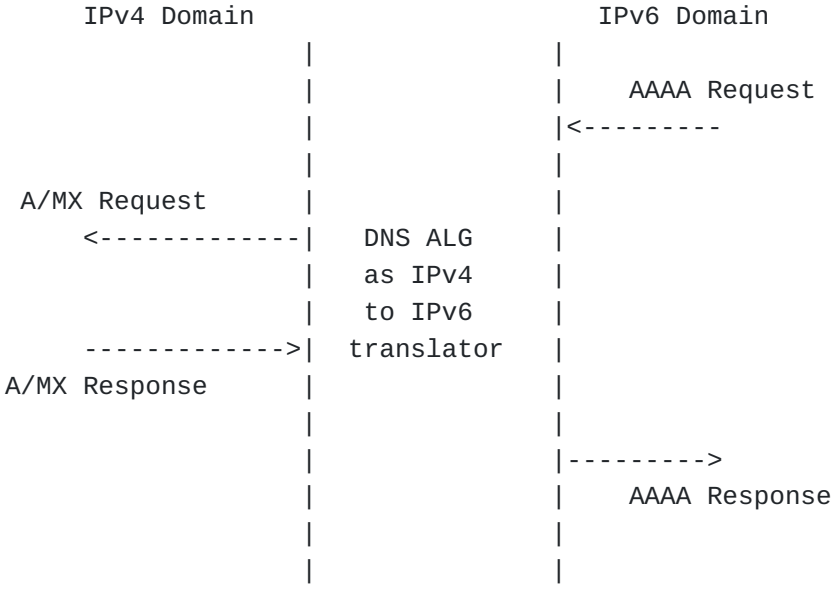

Figure 5: DNS Record Translation Service

To avoid conflicts, the DNS server should have access to all AAAA records advertised in the IPv6 domain. Otherwise, it may not know when to create AAAA records from A or MX records.
An issue arises with Dynamic DNS, in that the IVI host will only know to post its IPv6 addresses. The DNS service should store the AAAA records as generated by the host, and also generate an A record for each IVI address posted. Whether the IVI address itself is stored as a AAAA record as well as in an A record is at the implementer's option; if it is not stored it will be created from the A record if needed.

2.5. Host operation in IVI networks

TOC

Host behavior is unchanged by this specification. However, the local administration might want to configure host [RFC3484] (Draves, R., "Default Address Selection for Internet Protocol version 6 (IPv6)," February 2003.) address selection tables to optimize session behavior.

2.5.1. Interaction of IVI Addresses with RFC3484 Address Selection

TOC

[RFC3484] (Draves, R., "Default Address Selection for Internet Protocol version 6 (IPv6)," February 2003.) could be summarized as saying that IPv6 systems should select source and destination addresses that are as similar as possible. "Similarity" is defined in terms of prefix length. Each remote address is compared to each local address, and the remote address is considered to be most similar to the local address with the longest string of equivalent prefix bits. The specification recommends that sessions between the two systems should prefer the address pair with the longest "similar" prefix.
For example, if Alice has the addresses

- *2001:db8:1234:1::A and
- *2001:db8:5678::A,

and Bob has the following addresses

- *2001:db8:1234:2::B and
- *2001:db8:5ABC:B,

2001:db8:1234:1::A is more similar to 2001:db8:1234:2::B (the first 48 bits are the same as opposed to only the first 33) and 2001:db8:5678::A is more similar to 2001:db8:5ABC:B (the first 36 bits are the same as opposed to the first 33). When Alice and Bob communicate, the default

address policy selects the address pair in 2001:db8:1234::/48 over 2001:db8:5000::/36 because it has a longer "similar" prefix.
IPv4-only systems connect to IPv6 systems having IVI addresses through the gateway, and lack a means to initiate a connection to other IPv6 systems. Since IPv4 addresses appear in the IPv6 domain as IVI addresses, [RFC3484] (Draves, R., “Default Address Selection for Internet Protocol version 6 (IPv6),” February 2003.) will guide IPv6-only systems with IVI addresses to connect from their IVI address when communicating with IPv4-only systems, as they are the "most similar" addresses to those of their IPv4 counterparts. This is important, because it promotes stateless translation operation.
IVI systems may also find the IVI address pair "most similar" when communicating with other systems with IVI addresses. This is acceptable, as to the IPv6 domain they are simply IPv6 addresses and will communicate directly.
In general, a system with both an IPv4 address and an IPv6 address can connect to a similar system using either technology. There need be no preference order, and if one is chosen that is a local matter.

2.5.2. Interaction of IPv4 and IVI addresses on the same host

TOC

Systems that have both native IPv4 and translated IVI addresses require attention to the configuration of the address choice mechanism described in [RFC3484] (Draves, R., “Default Address Selection for Internet Protocol version 6 (IPv6),” February 2003.). In such a case, the redundancy suggests different uses for those addresses and the possibility that IPv4 reachability has been fragmented.
For example, consider a host with a private IPv4 address and an IVI address attempting to open a session with an IPv4 system with a public address. Apart from actually successfully opening a session, the addresses give no clue to actual reachability; the remote host might be reachable via IPv4, or that might be a private network disconnected from the Internet. If the remote host is reachable, there is likely to be a NAT between the host and that system, making the point moot. Similarly, the remote host might be reachable via IVI, but it might not. It might be reachable via both simultaneously, and it might not be reachable at all.
In general, native operation should be preferred to translated operation, but the specifics of the environment may guide this choice otherwise. As such, if an application is unable to open a session using one address, it should try another, and the local administration may consider configuring the [RFC3484] (Draves, R., “Default Address Selection for Internet Protocol version 6 (IPv6),” February 2003.) tables to manage the case.

TOC

2.6. Operation of the IVI Gateway

The IVI Gateway has two modes, depending on the address of the IPv6 system using its services. These are the Stateless Mode, used to connect between IPv6-only systems with IVI addresses and IPv4 systems, and the Stateful Mode, used to connect other (non-IVI) IPv6-only systems with IPv4 systems. IPv6 routing should not take traffic between IPv6 systems in the same IPv6 domain through the gateway, as it will follow more specific routes.

In either mode, the gateway is subject to the usual ills of Network Address Translation. Protocols that exchange IP addresses should in general not be exchanged across an IVI gateway, as the addresses are not necessarily translatable or meaningful after translation. Also, IPsec AH is compromised, so end-to-end privacy and authentication issues should be dealt with in another way such as IPsec ESP.

In general, native (IPv6<->IPv6 or IPv4<->IPv4) communications are preferable to any form of translation, and stateless translation is preferable to stateful translation. In the first case, this derives from the End-to-End principle discussed in [Saltzer] (Saltzer, JH., Reed, DP., and DD. Clark, “End-to-end arguments in system design,” Nov 1984.) - the utility of the network to the applications that use it is generally maximized by staying out of their way. In the latter case, this is due to the Simplicity Principle discussed in [RFC3439] (Bush, R. and D. Meyer, “Some Internet Architectural Guidelines and Philosophy,” December 2002.); given an easy and a hard way to do something, and given equivalence of outcome, the easy way is generally better for all concerned. Stateful and Stateless operation both enable communication at the cost of a header exchange. Stateful operation requires supporting dynamically-created per-flow tables in the gateway while stateless operation transforms datagrams algorithmically without per-flow state.

2.6.1. Stateless (1:1) Operation

TOC

In the stateless mode, the IVI gateway translates datagrams exchanged between IPv4 systems and IPv6 systems that have an IVI address. The translation is as described in SIIT (Nordmark, E., “Stateless IP/ICMP Translation Algorithm (SIIT),” February 2000.) [RFC2765], with the exception that the address format is as described in Section 2.2 (IVI Address Format) rather than the IPv4 Compatible Address described in section 2.1 of that document and deprecated in [RFC4291] (Hinden, R. and S. Deering, “IP Version 6 Addressing Architecture,” February 2006.). This includes the necessary correction of transport layer checksums.

This is referred to as 'stateless', because the transformation between IPv4 and IPv6 communication is entirely algorithmic and requires no long-term state in either the hosts or the gateway.

2.6.2. Stateful (1:n) Operation

TOC

In the stateful mode, the IVI gateway operates as a standard Network Address Translator, but between IPv4 and IPv6 domains. This is similar in many respects to the translation carried out in NAT-PT (Tsirtsis, G. and P. Srisuresh, “Network Address Translation - Protocol Translation (NAT-PT),” February 2000.) [RFC2766]. This includes the necessary correction of transport layer checksums.
IPv4 addresses and port numbers are mapped to IPv6 addresses in a stateful manner, much as is done in IPv4-IPv4 network address translation. The difference is that it is unidirectional; while the source port in an IPv6-> IPv4 translation may have to be changed to provide adequate flow identification, there is no necessity to change the source port in the IPv4->IPv6 direction.

3. Transition plan

TOC

Merriam-Webster defines a "transition" as "passage from one state, stage, subject, or place to another". Any transition plan that it doesn't describe how one can expect to transition from an IPv4 to an IPv6 network using it is incomplete. Coexistence is a necessary part, and is likely to last for a period of time measured in the durations of contracts. But if the increased operations and capital expenditures implied in a state of IPv4+IPv6 coexistence doesn't ultimately lead to the reduced expenditure state of a single network, it has not solved the problem it was intended to address.
In the IVI model, the network is presumed to traverse four relatively stable states. These are:

*IPv4-only Network

*IPv4+IPv6 Dual Stack Network

*IPv6+IPv4-accessible Network

*IPv6 Network

3.1. IPv4-only Network

TOC

The Internet, by and large, runs on IPv4 today. There are experimental uses of IPv6 and infrastructure uses of supporting internetwork protocols like MPLS and ATM, but end-to-end the protocol is IPv4.

3.2. IPv4+IPv6 Dual Stack Network

TOC

[RFC4213] (Nordmark, E. and R. Gilligan, “Basic Transition Mechanisms for IPv6 Hosts and Routers,” October 2005.) recommends the deployment of a dual stack architecture. The reason is straightforward: if while we can map IPv4 and IPv6 addresses 1:1 we aggressively deploy IPv6, we have two opportunities. First, should there be a problem (and there are always problems), user connectivity can be supported using IPv4 while the IPv6 issues are sorted out. Second, at the point where the availability of IPv4 addresses becomes a serious issue, IPv6 connectivity will be widespread, meaning that one can progress to the next phase rather than scrambling for business continuity.
We presume that service providers and enterprise networks can deploy IPv6 in parallel with IPv4, enabling current hosts (which are mostly if not all IPv4+IPv6 capable) to communicate with either architecture.

3.3. IPv6+IPv4-accessible Network

TOC

The problem with Section 3.2 (IPv4+IPv6 Dual Stack Network) is that, although people have had warning, they have chosen to not make use of it. Hence, we are likely to see an interval in the near future during which large numbers of IPv4 addresses are not available to extend services and IPv6 is not readily available as a deployed and purchasable service.
In such a case, a service provider has two main choices: obtain what IPv4 addresses can be obtained at whatever cost they may be available and extend his IPv4 service lifetime for a limited time period, or obtain those addresses and use them in a strategic manner to encourage movement to IPv6.
The IVI model suggests that remaining available IPv4 addresses could be mapped to IPv6 addresses in such a manner than both IPv4 and IPv6 systems can access servers and peers using them. A subscriber might be given an IPv6 /56 or /48 prefix for native use and a smaller IPv4 /30 or /24 prefix for translated use for servers and peers, giving him an IPv6-only network whose servers and peers are available using IPv4 via translation. Since the vast majority of systems operate as clients or as peer-to-peer application peers, this would in fact work.

3.4. IPv6 Network

TOC

At some point, enough systems have IPv6 addresses that it no longer makes economic sense to support the two networks in parallel. At this point, one can expect customers to no longer purchase IPv4 or IVI connectivity, IPv4 and IVI services to become economically uninteresting, and a global IPv6-only network to emerge.

4. Future extensions of the IVI Model

TOC

If the IPv6 hosts can be modified, the IVI model can have a stateless (1:n)operation, which can support both IPv6 initiated communication as well as IPv4 initiated communication.
For the operation and management concerns, the IVI model has ICMP extension, which can be used in the traceroute or similar cases.
The IVI model can also support the use of multicast between IPv4 and IPv6.
These extensions will be addressed in other documents.

5. Reflections on RFC 4966

TOC

RFCs [RFC4291] (Hinden, R. and S. Deering, “IP Version 6 Addressing Architecture,” February 2006.) and [RFC4966] (Aoun, C. and E. Davies, “Reasons to Move the Network Address Translator - Protocol Translator (NAT-PT) to Historic Status,” July 2007.) commented on the SIIT and NAT-PT proposals, and called for an improved proposal. This note is in response to that request. It would be of value to summarize the responses to the issues.
The DNS Application layer Gateway used by IVI, described in Section 2.4 (DNS service in IVI networks), is essentially that of DNS64 (Bagnulo, M., Matthews, P., and I. Beijnum, “NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers,” March 2009.) [I-D.bagnulo-behave-nat64] apart from the exact format of the address. NAT-PT has a problem with its DNS Application Layer gateway in that the overlay relationship between IPv4 addresses and ports and IPv6 addresses is dynamic, created when the RR is requested. There are complexities, scoping, and timing issues with that approach. The DNS Application Layer Gateway used by IVI could be configured as a completely static mapping, although that would not scale. The form of the relationship between A/MX and AAAA records is algorithmic and fixed, and the translation is done on the fly without saved state. So the DNS Application Layer Gateway is simpler and more predictable.
The stateless data plane translation algorithm, described in Section 2.6.1 (Stateless (1:1) Operation), is essentially that of SIIT (Nordmark, E., “Stateless IP/ICMP Translation Algorithm (SIIT),” February 2000.) [RFC2765] apart from the address format. As noted, [RFC4291] (Hinden, R. and S. Deering, “IP Version 6 Addressing Architecture,” February 2006.) deprecated the address format with the brusque comment that "current IPv6 transition mechanisms no longer use these addresses." The reason that they were not widely deployed was that they gave network operators little control in routing, or ways to ensure that route redistribution worked correctly. A prefix that lets the LIR specify the upper bits gives the operator the flexibility to

identify the IVI gateway advertising the prefix and better control the distribution of routes. In addition, moving the IPv4-mapped portion of the IVI address into the upper 64 bits of the address enables and encourages the network operator to permit IPv4-capable systems to be in various places topologically while retaining an address format familiar from other IPv6 addresses. The stateless mode applies to every session initiated from the IPv4 side of the gateway, as IVI does not provide DNS reachability to non-IVI addresses. It also applies to sessions initiated from IVI addresses on the IPv6 side of the gateway.
The stateful data plane algorithm, described in Section 2.6.2 (Stateful (1:n) Operation), is similar to that performed in normal IPv4/IPv4 NATs. Traffic from a non-IVI IPv6 address is overlaid on an IPv4 address and disambiguated by the source port number, and traffic to that IPv4 address using that port number as its destination is mapped back to the IPv6 address and port. This is essentially the "Address+Port" model proposed in [APNAT] (Maennel, O., Bush, R., Cittadini, L., and S. Bellovin, “A Better Approach than Carrier-Grade-NAT,” Aug 2008.).
These three algorithms have two years of operational experience behind them in the CERNET/CNGI-CERNET2 network. In a world of "rough consensus and running code", these are running code.
Many of the issues raised in [RFC4966] (Aoun, C. and E. Davies, “Reasons to Move the Network Address Translator - Protocol Translator (NAT-PT) to Historic Status,” July 2007.) are in fact not specific to NAT-PT, but are true of NATs in general and in particular NATs that translate between Internet Protocols with different address formats. In that sense, the document is not so much a complaint about NAT-PT as it is a complaint about translation. IVI's stateful mode suffers many of the same limitations, as would any translator. IVI's stateless mode suffers from less of them due to its stateless nature, but is nonetheless a translation and therefore breaks some of the end-to-end semantics of the Internet.

6. IANA Considerations

TOC

This memo adds no new IANA considerations.
Note to RFC Editor: This section will have served its purpose if it correctly tells IANA that no new assignments or registries are required, or if those assignments or registries are created during the RFC publication process. From the author's perspective, it may therefore be removed upon publication as an RFC at the RFC Editor's discretion.

TOC

7. Security Considerations

Three error cases are apparent: DNS errors, IPsec issues, and application address errors.
As noted in Figure 4 (Normal DNS Service), the errors that happen in NAT-PT implementations can happen in an IVI network as well. These mostly relate to the propagation of DNS records outside their domain of applicability.
As noted in Section 2.6 (Operation of the IVI Gateway), the side-effects of Network Address Translation between IPv4 and IPv4 apply when translating between IPv4 and IPv6. IPsec AH, whose checksum covers the IP header, fails when the header is changed. IPsec ESP, either directly on IP or over UDP are usable across NATs and presumably across translators as long as the IPv4 address is not in the certificate. Protocols that exchange IP addresses should not normally be used across a translator, as the addresses are generally not applicable on the far side. Such protocols should be filtered, or permitted but used with care.
[APNAT] (Maennel, O., Bush, R., Cittadini, L., and S. Bellovin, "A Better Approach than Carrier-Grade-NAT," Aug 2008.) raises a variety of issues with Carrier Grade Network Address Translators; those issues apply to the stateful mode of IVI, and in fact to any NAT. The stateless mode mitigates most of the issues raised there, however. If anything, this is the reason that we recommend dual stack deployment of IPv4 and IPv6 where possible in the near term, and target general IPv6 deployment in the medium term, as opposed to remaining in a dual address space environment forever.

8. Acknowledgements

TOC

Dan Wing and Tony Hain helped with the review of the document.

9. References

TOC

9.1. Normative References

TOC

[RFC2765] Nordmark, E., "Stateless IP/ICMP Translation Algorithm (SIIT)," RFC 2765, February 2000 (TXT).

[RFC2766] Tsirtsis, G. and P. Srisuresh, "Network Address Translation - Protocol Translation (NAT-PT)," RFC 2766, February 2000 (TXT).

9.2. Informative References

TOC

[APNAT] Maennel, O., Bush, R., Cittadini, L., and S. Bellovin, "A Better Approach than Carrier-Grade-NAT," Aug 2008.

[I-D.bagnulo-behave-nat64] Bagnulo, M., Matthews, P., and I. Beijnum, "NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers," draft-bagnulo-behave-nat64-03 (work in progress), March 2009 (TXT).

[RFC3439] Bush, R. and D. Meyer, "Some Internet Architectural Guidelines and Philosophy," RFC 3439, December 2002 (TXT).

[RFC3484] Draves, R., "Default Address Selection for Internet Protocol version 6 (IPv6)," RFC 3484, February 2003 (TXT).

[RFC4213] Nordmark, E. and R. Gilligan, "Basic Transition Mechanisms for IPv6 Hosts and Routers," RFC 4213, October 2005 (TXT).

[RFC4291] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture," RFC 4291, February 2006 (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).

[Saltzer] Saltzer, JH., Reed, DP., and DD. Clark, "End-to-end arguments in system design," ACM Transactions on Computer Systems (TOCS) v.2 n.4, p277-288, Nov 1984.

Authors' Addresses

TOC

Xing Li
CERNET Center/Tsinghua University
Room 225, Main Building, Tsinghua University
Beijing, 100084
China
Phone: +86 62785983
Email: xing@cernet.edu.cn

Congxiao Bao
CERNET Center/Tsinghua University
Room 225, Main Building, Tsinghua University
Beijing, 100084
China
Phone: +86 62785983
Email: congxiao@cernet.edu.cn

Fred Baker
Cisco Systems

Santa Barbara, California 93117
USA
Phone: +1 408 526 4257
Email: fred@cisco.com

Kevin Yin
Cisco Systems
No. 2 Jianguomenwai Ave, Chaoyang District
Beijing, 100022
China
Phone: +86 10 8515 5094
Email: kyin@cisco.com

Full Copyright Statement

TOC

Intellectual Property