Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: June 18, 2012

M. Townsley
O. Troan
cisco
December 16, 2011

Basic Requirements for Customer Edge Routers - multihoming and transition
draft-townsley-troan-ipv6-ce-transitioning-02

Abstract

This document specifies general IPv6 multihoming and specific 6rd transitioning requirements for an IPv6 Customer Edge (CE) router. It also provides an illustrative implementation model for IPv4 multihoming in order to support shared IPv4 transition mechanisms that utilize IPv6 as a transport for IPv4.

Table of Contents

1. Introduction

The CE requirements specified in RFC 6204 are based on a fundamental assumption that a CE router has a single active WAN interface for forwarding IPv4 and IPv6 traffic towards an ISP. The operation of IPv6 via 6rd, IPv6 via Native, IPv4 via DS-lite and IPv4 via Native together, forces us to reconsider this basic assumption.

There are three possible steady-state combinations of "native" and "virtual" (tunneled) dual-stack service models (an IPv6-only service model, while imperative for finalizing the transition to IPv6, is currently out of scope in [RFC6204] and [I-D.ietf-v6ops-6204bis]) that do not break the fundamental assumption of having no more than one WAN interface per IP version.

1. One Native IPv4 and IPv6 interface (Classic Dual-Stack)

2. One Native IPv4 and one Virtual IPv6 interface (6rd, Softwires Hub and Spoke via L2TP, TSP, etc)

3. One Virtual IPv4 and one Native IPv6 interface (DS-Lite, 4rd, etc)

For (1), IPv4 and IPv6 each share a single WAN interface, so there is no problem when enabling one vs. the other.

For (2), when enabling tunneled IPv6 on an existing IPv4-only network there is no significant change in the basic model as each IP version still has its own distinct single WAN interface. Multihoming issues arise when enabling native IPv6 alongside tunneled IPv6 (needed for "6rd sunsetting") as IPv6 may be enabled on two distinct interfaces at the same time.

For (3) there similarly is no problem when enabling tunneled IPv4 on an existing IPv6-only network, and we understand that there are greenfield deployments just like this happening. Multihoming issues arise when enabling tunneled IPv4 on a network that has native IPv4 available at the same time.

The multihoming model described in this document assumes the operator or user can configure any WAN interface type at any time. The CE follows forwarding rules defined in this document in order to ensure packets make it out the right interface on WAN egress, and liberally accepts packets on WAN ingress. This is "classic multihoming" and should work for any order of planned incremental transition steps, as well as failover and/or transient situations.

While this authors of this document believe that the forward-oriented

model, there is another approach which we identify as "Configuration-oriented" and describe briefly in Appendix A.

1.1. Requirements Language

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 [RFC2119].

2. Terminology

SRIB	A Source Address Routing Information Base containing an entry per delegated prefix. Each entry points to one or more Destination Address Routing Tables (DRIB).
DRIB	A Destination Address Routing Information Base used for destination address longest matching lookups. Each entry points to one or more next-hops.
NPIB	Network Address and Port Translation (NAPT) Information Base used binding "flows" to an egress WAN interface.
NPIB entry	The address and port mapping state [RFC4787] at the NAT necessary for network address and port translation.

3. IPv6 MultiPrefix Multihoming

A multihomed, multiprefix, IPv6 CE router has multiple WAN interfaces connecting it to one or more Service Providers. The interfaces may be "real" or "virtual" in the case of tunneling technology such as 6rd [RFC5969]. The CE router receives one or more delegated prefixes, each associated with one or more WAN interfaces. The CE router has a single SRIB, and one DRIB associated with each WAN Interface.

WAN interfaces are used to send Ingress traffic from the Internet to the End-User, and Egress traffic from the End-User network to the Internet. Ingress traffic may be received on any active interface at any time. Egress traffic follows a set of rules within the CE in order to choose the proper WAN interface. This is important not only in order to choose the best path, but also because the networks that the CE are connected to typically employ source address verification

mechanisms.

Packets arriving at the CE have an IPv6 source address chosen by the host [RFC3484]. The SRIB contains an entry for each delegated prefix with a pointer to one or more DRIBs. A longest matching lookup based upon the source address of each arriving packet is performed within the SRIB to determine the DRIB(s). The egress WAN interface to use for sending a packet is then chosen by performing a longest matching lookup within the resulting DRIB(s).

3.1. Multihoming requirements

MH-1: An IPv6 CE router MUST create a separate DRIB for each WAN interface (real or virtual) and installs a route for the associated delegated prefix, default route and more specific routes.

MH-2: An IPv6 CE router MUST create an SRIB containing entries for associated delegated prefixes. Each entry points to one or more DRIBs. An entry points to multiple DRIBs only in the case where an identical delegated prefix is associated with multiple WAN interfaces.

MH-3: When forwarding a packet from a LAN interface, the CE router MUST do a longest matching lookup based on the packet's Source Address in the SRIB. A Destination Address lookup is then performed in the corresponding DRIB or DRIBs. When there are multiple equal matches, the route with the lowest cost is chosen.

3.2. 6rd Sunsetting Requirements

6RDS-1: Multihoming as defined in section Section 3 MUST be supported, allowing 6rd and native packets to be sent and received as long as 6rd configuration is provided by the ISP.

6RDS-2: By default, the 6rd virtual interface MUST be assigned a higher routing cost than a native IPv6 interface.

6RDS-3: The IPv6 CE router MUST support that 6rd and native IPv6 delegated prefixes are identical or different, and operate as defined in the multihoming section.

4. IPv4 NAT Multihoming

This section describes a general implementation model used to illustrate CE IPv4 multihoming, alongside specifics for using the model to support IPv6 transition technologies aimed at delivering IPv4 within an IPv6 transport (e.g., DS-Lite).

A multihomed IPv4 CE router has multiple "physical" or "virtual" WAN interfaces connecting it to one or more Service Providers. WAN interfaces are used to send Ingress traffic from the Internet to the End-User, and Egress traffic from the End-User network to the Internet. Each WAN interface may be configured with a single public IPv4 address, private IPv4 address [RFC1918], a shared IPv4 address [A+P], or no IPv4 address at all.

An IPv4 CE router WAN interface is often configured in the same manner as a single host, i.e., with a single 32-bit IPv4 address. A CE router NAPT function, in turn, allows more than one device within the end-user network to appear as a single host with a single IPv4 address facing the ISP network. The CE router NAPT function is responsible for rewriting IPv4 headers with the address assigned to the associated WAN interface with the expectation that return traffic will be sent back to that address.

IPv4 WAN interfaces which are not configured with an IPv4 address by the ISP (e.g., DS-Lite, L2-aware NAT), bypass the translation function of the NAPT when forwarding traffic. In this case, the ISP's centralized NAPT function is responsible for rewriting IPv4 headers with a source address which ensures return traffic will reach the proper NAPT binding within the ISP.

4.1. IPv4 Multihoming Data Structures

The CE router has a single NAPT Information Base (NPIB) which consists of Dynamic and Configured entries. A WAN interface provisioned with an IPv4 address has an associated NAPT function which examines packet flow and programs the NPIB with Dynamic entries as they are identified. For example, an active TCP session on the CE correlates to a single Dynamic entry within the NPIB. A Dynamic entry in the NPIB table unambiguously identifies packets that are associated with it when an NPIB Lookup is performed.

Configured NPIB entries allow for rules to be specified which direct traffic in a specific manner. Unlike Dynamic entries, Configured entries allow for "wildcards", and may be end-user configured or configured by a protocol such as NAT-PMP, UPnP, PCP, etc. Packets directed by configured entries may or may not instantiate Dynamic NPIB entries for specific flows.

Finally, the CE router also has a single IPv4 Routing Information Base (RIB), containing IPv4 routes learned from active LAN and WAN interfaces. The RIB is only consulted for packets that fail to match an NPIB entry. The RIB supports a classic IPv4 routing function, including use of the longest matching algorithm and preferences that may be assigned to each interface. A RIB lookup ultimately resolves to an output interface which, in turn may have an NAPT function enabled. If the output interface has an NAPT function enabled, it is responsible for programming the NPIB with a new Dynamic entry and translating the packet before sending.

4.2. IPv4 Packet flow

Ingress (from the Internet to the End-User) traffic may be received on any active interface at any time. Egress (from the End-User to the Internet) traffic follows a set of rules within the CE in order to choose the proper WAN interface and associated NAPT function on that interface if present.

Egress packets arriving at the CE trigger a lookup in the NPIB. A successful match on a Dynamic entry in the NPIB provides all information necessary to translate and send a packet out the proper interface (e.g., no additional lookup in the RIB or elsewhere is required). Packets which do not have a matching NPIB entry but match a Configured entry are treated similarly. Packets which fail the NPIB lookup entirely are sent to the RIB.

The RIB performs a classic IPv4 longest-matching routing lookup based on the destination address of the IPv4 packet. If more than one interface is selected (which will be the case when more than one active WAN interface programs a default route in the RIB), then packets are sent via the interface with the highest configured preference. If the preference is the same, packets may be load-balanced. After selecting the proper output interface for the packet, the packet is either sent immediately or translated and sent if a NAPT function is enabled. If NAPT function is enabled on that interface, it is responsible for programming the NPIB with a new Dynamic entry and translating the packet before sending.

4.3. Example RIB Policy for IPv4 to IPv6 Transition

Interface preferences in the RIB allow policy definition for choosing one type of interface over another. Well-defined defaults which encourage transition to IPv6, less use of NAPT, and more distributed state may be defined. The policy may be represented as a simple table which may be altered by the operator or user without any change to the CE packet forwarding implementation itself.

In this example, an interface with a higher preference value is preferred over an interface with a lower preference value. Weighted values are assigned according to the following basic principles for IPv4 interface selection:

1. IPv6 transport is preferred over any other.

2. Less address translation occurrences is preferred over more. [RFC5864][I-D.donley-nat444-impacts]

3. The closer the state is to the edge, the better.[RFC1958]

In the case of DS-Lite and Native IPv4 configuration being present at the same time, DS-Lite would be preferred as it uses IPv6 transport and Native IPv4 does not. When transitioning an active dual-stack network to DS-Lite, this means that when the DS-Lite IPv4 interface is made active, traffic that does not match an active entry in the NPIB table would be directed over the DS-Lite tunnel. As entries in the NPIB table naturally time out, or if the Native interface is deactivated, the CGN within the DS-Lite AFTR takes over the NAPT state of the CE router. In the event the DS-Lite tunnel fails, the Native IPv4 interface and local NAPT will naturally begin taking over.

The above principles may be applied to other methods of IPv4 transport as well. The following table indicates a basic ordering (least to most preferred) for some of the known IPv4 extension and IPv6 transition mechanisms under development today.

Mechanism	#1 (100)	#2 (10)	#3 (1)	Total
CGN	1	1	1	111
SD-NAT	1	1	2	112
Native IPv4	1	2	2	122
MAP-T	2	1	2	212
DS-lite	2	2	1	221
MAP-E	2	2	2	222

Table 1: IPv4 Preference Table

5. Security Considerations

6. Acknowledgements

7. IANA Considerations

This memo includes no request to IANA.

8. References

8.1. Normative References

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

[RFC2131] Droms, R., "Dynamic Host Configuration Protocol", RFC 2131, March 1997.

[RFC5969] Townsley, W. and O. Troan, "IPv6 Rapid Deployment on IPv4 Infrastructures (6rd) -- Protocol Specification", RFC 5969, August 2010.

8.2. Informative References

[I-D.donley-nat444-impacts] Donley, C., Howard, L., Kuarsingh, V., Berg, J., and U. Colorado, "Assessing the Impact of Carrier-Grade NAT on Network Applications", draft-donley-nat444-impacts-03 (work in progress), November 2011.

[I-D.ietf-v6ops-6204bis] Singh, H., Beebee, W., Donley, C., Stark, B., and O. Troan, "Basic Requirements for IPv6 Customer Edge Routers", draft-ietf-v6ops-6204bis-04 (work in progress), December 2011.

[RFC1918] Rekhter, Y., Moskowitz, R., Karrenberg, D., Groot, G., and E. Lear, "Address Allocation for Private Internets", BCP 5, RFC 1918, February 1996.

[RFC1958] Carpenter, B., "Architectural Principles of the Internet", RFC 1958, June 1996.

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

[RFC4787] Audet, F. and C. Jennings, "Network Address Translation (NAT) Behavioral Requirements for Unicast UDP", BCP 127, RFC 4787, January 2007.

[RFC5864] Allbery, R., "DNS SRV Resource Records for AFS", RFC 5864,

April 2010.

[RFC6204] Singh, H., Beebee, W., Donley, C., Stark, B., and O. Troan, "Basic Requirements for IPv6 Customer Edge Routers", RFC 6204, April 2011.

Appendix A. Configuration-oriented multihoming

This method attempts to actively avoid multihoming by forcing only a single configured WAN egress interface to be active at any time for a given IP version. For this to work, one of the following assumptions must hold.

a. From the perspective of the CE router, the network supports only one type of interface for a given IP version, or

b. The CE router is configured in advance of any IP configuration to support only one type of interface for a given IP version, or

c. The CE router goes through an ordered set of configuration attempts in series, each requiring a timeout before moving to the next. Transition-oriented changes after steady-state is reached will require "reboot" to go through the ordered process from scratch.

d. The CE router chooses one type of interface and shuts down all others based on a predetermined priority when more than one interface with the same IP version is configured. This allows parallel configuration attempts and changes after reaching steady-state, but requires the CE router and network to manage a "flash cut" from one configured interface to the other and may be prone to tricky race-conditions.

Authors' Addresses

Mark Townsley
cisco

Email: mark@townsley.net

Ole Troan
cisco

Email: ot@cisco.com