Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: January 16, 2011

R. Gagliano
Cisco Systems

July 15, 2010

IPv6 Deployment in Internet Exchange Points (IXPs)
draft-ietf-v6ops-v6inixp-09.txt

Abstract

This document provides guidance on IPv6 deployment in Internet Exchange Points (IXP). It includes information regarding the switch fabric configuration, the addressing plan and general organizational tasks that need to be performed. IXPs are mainly a layer 2 infrastructure and in many cases the best recommendations suggest that the IPv6 data, control and management plane should not be handled differently than in IPv4.

Status of this Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on January 16, 2011.

Table of Contents

1. Introduction

Most Internet Exchange Points (IXP) work at the Layer 2 level, making the adoption of IPv6 an easy task. However, IXPs normally implement additional services such as statistics, route servers, looking glasses and broadcast control that may be impacted by the implementation of IPv6. This document clarifies the impact of IPv6 on a new or an existing IXP. The document assumes an Ethernet switch fabric, although other layer 2 configurations can be deployed.

2. Switch Fabric Configuration

An Ethernet based IXP switch fabric implements IPv6 over Ethernet as described in [RFC2464] . Therefore, the switching of IPv6 traffic happens in the same way as in IPv4. However, some management functions (such as switch management, SNMP [RFC3411] support or flow analysis exportation) may require IPv6 as underlying layer and this should be assessed by the IXP operator.

There are two common configurations of IXP switch ports to support IPv6:

1. dual-stack LAN (Local Area Network): when both IPv4 and IPv6 traffic share a common LAN. No extra configuration is required in the switch.

2. independent VLAN (Virtual Local Area Network)[IEEE.P802-1Q.1998]: when an IXP logically separates IPv4 and IPv6 traffic in different VLANs.

In both configurations, IPv6 and IPv4 traffic can either share a common physical port or use independent physical ports. The use of independent ports can be more costly in both capital expenses (as new ports are needed) and operational expenses.

When using the same physical port for both IPv4 and IPv6 traffic, some changes may be needed at the participants' interfaces configurations. If the IXP implements the "dual-stack configuration", IXP's participants will configure dual-stack interfaces. On the other hand, if the IXP implements the "independent VLAN configuration", IXP participants are required to pass one additional VLAN tag across the interconnection. In this case, if the IXP did not originally use VLAN tagging, VLAN tagging should be established and previously configured LAN may continue untagged as a "native VLAN" or be transitioned to a tagged VLAN. The "independent VLAN" configuration provides a logical separation of IPv4 and IPv6 traffic, simplifying separate statistical analysis for

IPv4 and IPv6 traffic. Conversely, the "dual-stack" configuration (when performing separate statistical analysis for IPv4 and IPv6 traffic) would require the use of flows techniques such as IPFIX [RFC5101] to classify traffic based on the different ether-types (0x0800 for IPv4, 0x0806 for ARP and 0x86DD for IPv6).

The only technical requirement for IPv6 referring link MTUs is that they need to be greater than or equal to 1280 octets [RFC2460]. The MTU size for every LAN in an IXP should be well known by all its participants.

3. Addressing Plan

Regional Internet Registries (RIRs) have specific address policies to assign Provider Independent (PI) IPv6 address to IXPs. Those allocations are usually /48 or shorter prefixes [RIR_IXP_POLICIES]. Depending on the country and region of operation, address assignments may be made by NIRs (National Internet Registries). Unique Local IPv6 Unicast Addresses ([RFC4193]) are normally not used in an IXP LAN as global reverse DNS resolution and whois services are required.

IXPs will normally use manual address configuration. The manual configuration of IPv6 addresses allows IXP participants to replace network interfaces with no need to reconfigure Border Gateway Protocol (BGP) sessions information and it also facilitates management tasks. The IPv6 Addressing Architecture [RFC4291] requires that interface identifiers are 64 bits in size for prefixes not starting with binary 000, resulting in a maximum prefix length of /64. Longer prefix lengths up to /127 have been used operationally. If prefix lengths longer than 64 bits are chosen, the implications described in [RFC3627] need to be considered. A /48 prefix allows the addressing of 65536 /64 LANs.

When selecting the use of static Interface Identifiers (IIDs), there are different options on how to fill its 64 bits (or 16 hexadecimal characters). A non-exhaustive list of possible IID selection mechanisms is the following:

1. Some IXPs like to include the participants' ASN number decimal encoding inside each IPv6 address. The ASN decimal number is used as the BCD (binary code decimal) encoding of the upper part of the IID such as shown in this example:

   * IXP LAN prefix: 2001:db8::/64

   * ASN: 64496

* IPv6 Address: 2001:db8:0000:0000:0000:0006:4496:0001/64 or its equivalent representation 2001:db8::6:4496:1/64

In this example we are right justifying the participant's ASN number from the 112nd bit. Remember that 32 bits ASNs require a maximum of 10 characters. With this example, up to 2^16 IPv6 addresses can be configured per ASN.

2. Although BCD encoding is more "human-readable", some IXPs prefer to use the hexadecimal encoding of the ASNs number as the upper part of the IID as follow:

* IXP LAN prefix: 2001:db8::/64

* ASN: 64496 (DEC) or fbf0 (HEX)

* IPv6 Address: 2001:db8:0000:0000:0000:0000:fbf0:0001/64 or its equivalent representation 2001:db8::fbf0:1/64

In this case a maximum of 8 characters will be needed to represent 32 bits ASNs.

3. A third scheme for statically assigning IPv6 addresses on an IXP LAN could be to relate some portions of a participant's IPv6 address to its IPv4 address. In the following example, the last four decimals of the IPv4 address are copied to the last hexadecimals of the IPv6 address, using the decimal number as the BCD encoding for the last three characters of the IID such as in the following example:

* IXP LAN prefix: 2001:db8::/64

* IPv4 Address: 192.0.2.123/23

* IPv6 Address: 2001:db8:2::123/64

4. A fourth approach might be based on the IXPs ID for that participant.

IPv6 prefixes for IXP LANs are typically publicly well known and taken from dedicated IPv6 blocks for IXP assignments reserved for this purpose by the different RIRs. These blocks are usually only meant for addressing the exchange fabric, and may be filtered out by DFZ (Default Free Zone) operators. When considering the routing of the IXP LANs two options are identified:

o IXPs may decide that LANs should not to be globally routed in order to limit the possible origins of a Denial of Service (DoS)

attack to its participants' AS boundaries. In this configuration participants may route these prefixes inside their networks (e. g. using BGP no-export communities or routing the IXP LANs within the participants' IGP) to perform fault management. Using this configuration, the monitoring of the IXP LANs from outside of its participants' AS boundaries is not possible.

- o IXP may decide that LANs should (attempt to be) be globally routed. In this case, IXP LANs monitoring from outside its participants' AS boundaries may be possible but the IXP LANs will be vulnerable to DoS from outside of those boundaries.

Additionally, possible IXP external services (such as DNS, web pages, FTP servers) need to be globally routed. These should be addressed from separate address blocks, either from upstream providers' address space, or separate independent assignments. Strict prefix length filtering could be a reason for requesting more than one /48 assignment from a RIR (i.e. requesting one /48 assignment for the IXPs LANs that may not be globally routed and a different, non-IXP /48 assignment for the IXP external services that will be globally routed).

4. Multicast IPv6

There are two elements that need to be evaluated when studying IPv6 multicast in an IXP: multicast support for neighbor discovery and multicast peering.

4.1. Multicast Support and Monitoring for Neighbor Discovery at an IXP

IXPs typically control broadcast traffic across the switching fabric in order to avoid broadcast storms by only allowing limited ARP [RFC0826] traffic for address resolution. In IPv6 there is not broadcast support but IXPs may intend to control multicast traffic in each LAN instead. ICMPv6 Neighbor Discovery [RFC4861] implements the following necessary functions in an IXP switching fabric: Address Resolution, Neighbor Unreachability Detection and Duplicate Address Detection. In order to perform these functions, Neighbor Solicitation and Neighbor Advertisement packets are exchanged using the link-local all-nodes multicast address (ff02::1) and/or solicited-node multicast addresses (ff02:0:0:0:0:1:ff00:0000 to ff02:0:0:0:0:1:ffff:ffff). As described in [RFC4861] routers will initialize its interfaces by joining its solicited-node multicast addresses using either Multicast Listener Discovery (MLD) [RFC2710] or MLDv2 [RFC3810] . MLD messages may be sent to the corresponding group address ff02::2 (MLD) or ff02::16 (MLDv2). Depending on the addressing plan selected by the IXP, each solicited-node multicast

group may be shared by a sub-set of participants' conditioned by how the last three octets of the addresses are selected. In Section 3 example 1, only participants with ASNs with the same two last digits are going to share the same solicited-node multicast group.

Similarly to the ARP policy an IXP may limit multicast traffic across the switching fabric in order to only allow ICMPv6 Neighbor Solicitation, Neighbor Advertisement and MLD messages. Configuring default routes in an IXP LAN without an agreement between the parties is normally against IXP policies. ICMPv6 Router Advertisement packets should neither be issued nor accepted by routers connected to the IXP. Where possible, the IXP operator should block link-local RA packets using IPv6 RA-GUARD [I-D.ietf-v6ops-ra-guard] . If this is not possible, the IXP operator should monitor the exchange for rogue Router Advertisement packets as described in [I-D.ietf-v6ops-rogue-ra] .

4.2. IPv6 Multicast traffic exchange at an IXP

For IPv6 Multicast traffic exchange, an IXP may decide to use either the same LAN being used for unicast IPv6 traffic exchange, the same LAN being used for IPv4 Multicast traffic exchange or a dedicated LAN for IPv6 Multicast traffic exchange. The reason for having a dedicated LAN for multicast is to prevent unwanted multicast traffic to reach participants that do not have multicast support. Protocol Independent Multicast (PIM) [RFC4601] messages will be sent to the link-local IPv6 'ALL-PIM-ROUTERS' multicast group ff02::d in the selected LAN and should be allowed. Implementing IPv6 PIM snooping will allow only the participants associated to a particular group to receive its multicast traffic. BGP reachability information for IPv6 multicast address-family (SAFI=2) is normally exchanged using MP-BGP [RFC4760] and is used for Reverse Path Forwarding (RPF) lookups performed by the IPv6 PIM. If a dedicated LAN is configured for Multicast IPv6 traffic exchange, reachability information for IPv6 Multicast address family should be carried in new BGP sessions. ICMPv6 Neighbor Discovery should be allowed in the Multicast IPv6 LAN as described in the previous paragraph.

5. Reverse DNS

The inclusion of PTR records for all addresses assigned to participants in the IXP reverse zone under "ip6.arpa" facilitates troubleshooting, particularly when using tools such as traceroute. If reverse DNS is configured, DNS servers should be reachable over IPv6 transport for complete IPv6 support.

6. Route-Server

IXPs may offer a Route-Server service, either for Multi-Lateral Peering Agreements (MLPA) service, looking glass service or route-collection service. IPv6 support needs to be added to the BGP speaking router. The equipment should be able to transport IPv6 traffic and to support Multi-protocol BGP (MP-BGP) extensions for IPv6 address family ([RFC2545] and [RFC4760]).

A good practice is that all BGP sessions used to exchange IPv6 network information are configured using IPv6 data transport. This configuration style ensures that both network reachability information and generic packet data transport use the same transport plane. Because of the size of the IPv6 space, limiting the maximum number of IPv6 prefixes in every session should be studied.

External services should be available for external IPv6 access, either by an IPv6 enabled web page or an IPv6 enabled console interface.

7. External and Internal support

Some external services that need to have IPv6 support are traffic graphics, DNS, FTP, Web, Route Server and Looking Glass. Other external services such as NTP servers, or SIP Gateways need to be evaluated as well. In general, each service that is currently accessed through IPv4 or that handle IPv4 addresses should be evaluated for IPv6 support.

Internal services are also important when considering IPv6 adoption at an IXP. Such services may not deal with IPv6 traffic but may handle IPv6 addresses; that is the case of provisioning systems, logging tools and statistics analysis tools. Databases and tools should be evaluated for IPv6 support.

8. IXP Policies and IPv6

IXP Policies and contracts should be revised as any mention of IP should be clarified if it refers to IPv4, IPv6 or both.

Policies for IPv6 traffic monitoring and filtering may be in place as described in Section Section 4 .

9. IANA Considerations

This memo includes no request to IANA.

10. Security Considerations

This memo includes references to procedures for monitoring and/or avoiding particular ICMPv6 traffic at IXPs' LANs. None of these procedures prevent Ethernet loops caused by mischief in the LAN. The document also mentions how to limit IPv6 DoS attacks to the IXP switch fabric by not globally announce the IXP LANs prefix.

11. Acknowledgements

The author would like to thank the contributions from Alain Aina, Bernard Tuy, Stig Venaas, Martin Levy, Nick Hilliard, Martin Pels, Bill Woodcock, Carlos Friacas, Arien Vijn, Fernando Gont and Louis Lee.

12. Informative References

[I-D.ietf-v6ops-ra-guard]
Levy-Abegnoli, E., Velde, G., Popoviciu, C., and J. Mohacsi, "IPv6 RA-Guard", draft-ietf-v6ops-ra-guard-05 (work in progress), May 2010.

[I-D.ietf-v6ops-rogue-ra]
Chown, T. and S. Venaas, "Rogue IPv6 Router Advertisement Problem Statement", draft-ietf-v6ops-rogue-ra-00 (work in progress), May 2009.

[IEEE.P802-1Q.1998]
Institute of Electrical and Electronics Engineers, "Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks", IEEE Draft P802.1Q, March 1998.

[RFC0826] Plummer, D., "Ethernet Address Resolution Protocol: Or converting network protocol addresses to 48.bit Ethernet address for transmission on Ethernet hardware", STD 37, RFC 826, November 1982.

[RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.

[RFC2464] Crawford, M., "Transmission of IPv6 Packets over Ethernet

Networks", RFC 2464, December 1998.

[RFC2545] Marques, P. and F. Dupont, "Use of BGP-4 Multiprotocol Extensions for IPv6 Inter-Domain Routing", RFC 2545, March 1999.

[RFC2710] Deering, S., Fenner, W., and B. Haberman, "Multicast Listener Discovery (MLD) for IPv6", RFC 2710, October 1999.

[RFC3411] Harrington, D., Presuhn, R., and B. Wijnen, "An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks", STD 62, RFC 3411, December 2002.

[RFC3627] Savola, P., "Use of /127 Prefix Length Between Routers Considered Harmful", RFC 3627, September 2003.

[RFC3810] Vida, R. and L. Costa, "Multicast Listener Discovery Version 2 (MLDv2) for IPv6", RFC 3810, June 2004.

[RFC4193] Hinden, R. and B. Haberman, "Unique Local IPv6 Unicast Addresses", RFC 4193, October 2005.

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

[RFC4601] Fenner, B., Handley, M., Holbrook, H., and I. Kouvelas, "Protocol Independent Multicast - Sparse Mode (PIM-SM): Protocol Specification (Revised)", RFC 4601, August 2006.

[RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760, January 2007.

[RFC4861] Narten, T., Nordmark, E., Simpson, W., and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, September 2007.

[RFC5101] Claise, B., "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of IP Traffic Flow Information", RFC 5101, January 2008.

[RIR_IXP_POLICIES]
Numbers Resource Organization (NRO)., "RIRs Allocations Policies for IXP. NRO Comparison matrix", 2009, <http://www.nro.net/documents/comp-pol.html#3-4-2>.

Author's Address

Roque Gagliano
Cisco Systems
Avenue des Uttins 5
Rolle, 1180
Switzerland

Email: rogaglia@cisco.com