V6OPS Working Group
Internet-Draft
Intended status: Informational
Expires: April 25, 2013

P. Matthews
Alcatel-Lucent

October 22, 2012

Design Guidelines for IPv6 Networks
draft-matthews-v6ops-design-guidelines-01

Abstract

This document presents advice on the design choices that arise when designing IPv6 networks (both dual-stack and IPv6-only). The intended audience is someone designing an IPv6 network who is knowledgeable about best current practices around IPv4 network design, and wishes to learn the corresponding practices for IPv6.

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction

This document presents advice on the design choices that arise when designing IPv6 networks (both dual-stack and IPv6-only). The intended audience is someone designing an IPv6 network who is knowledgeable about best current practices around IPv4 network design, and wishes to learn the corresponding practices for IPv6.

The focus of the document is on design choices where there are differences between IPv4 and IPv6, either in the range of possible alternatives (e.g. the extra possibilities introduced by link-local addresses in IPv6) or the recommended alternative. The document presents the alternatives and discusses the pros and cons in detail. Where consensus currently exists around the best practice, this is documented; otherwise the document simply summarizes the current state of the discussion. Thus this document serves to both to document the reasoning behind best current practices for IPv6, and to allow a designer to make an intelligent choice where no such consensus exists.

This document does not present advice on strategies for adding IPv6 to a network, nor does it discuss transition mechanisms. For advice in these areas, see [RFC6180] for general advice, [I-D.ietf-v6ops-wireline-incremental-ipv6] for wireline service providers, [RFC6342] for mobile network providers, [RFC5963] for exchange point operators, [I-D.ietf-v6ops-icp-guidance] for content providers, and both [RFC4852] and [I-D.ietf-v6ops-enterprise-incremental-ipv6] for enterprises. Nor does the document cover the ins and outs of creating an IPv6 addressing plan; for advice in this area, see [RFC5375].

The current version of this document focuses on unicast network design only. It does not cover multicast,, nor supporting infrastructure such as DNS. This may change in future versions.

The current version is still work in progress, and it is expected that the presentation and discussion of additional design choices will be added as the document matures.

2. Design Choices

This section consists of a list of specific design choices a network designer faces when designing an IPv6-only or dual-stack network, along with guidance and advice to the designer when making a choice.

2.1. Mix IPv4 and IPv6 on the Same Link?

Should IPv4 and IPv6 traffic be logically separated on a link? That is:

a. Mix IPv4 and IPv6 traffic on the same layer 2 connection, OR

b. Separate IPv4 and IPv6 by using separate physical or logical links (e.g., two physical links or two VLANs on the same link)?

Option (a) implies a single layer 3 interface at each end with both IPv4 and IPv6 addresses; while option (b) implies two layer 3 interfaces, one for IPv4 addresses and one with IPv6 addresses.

The advantages of option (a) include:

- Requires only half as many layer 3 interfaces as option (b), thus providing better scaling;

- May require fewer physical ports, thus saving money;

- Can make the QoS implementation much easier (for example, rate-limiting the combined IPv4 and IPv6 traffic to or from a customer);

- Provides better support for the expected future of increasing IPv6 traffic and decreasing IPv4 traffic;

- And is generally conceptually simpler.

For these reasons, there is a pretty strong consensus in the operator community that option (a) is the preferred way to go.

However, there can be times when option (b) is the pragmatic choice. Most commonly, option (b) is used to work around limitations in network equipment. One big example is the generally poor level of support today for individual statistics on IPv4 traffic vs IPv6 traffic when option (a) is used. Other, device-specific, limitations exist as well. It is expected that these limitations will go away as support for IPv6 matures, making option (b) less and less attractive until the day that IPv4 is finally turned off.

Most networks today use option (a) wherever possible.

2.2. Links with Only Link-Local Addresses?

Should the link:

a. Use only link-local addresses ("unnumbered"), OR

b. Have global or unique-local addresses assigned in addition to link-locals?

There are two advantages of unnumbered links. The first advantage is ease of configuration. In a network with a large number of unnumbered links, the operator can just enable an IGP on each router, without going through the tedious process of assigning and tracking the addresses for each link. The second advantage is security. Since link-local addresses are unroutable, the associated interfaces cannot be attacked from an off-link device. This implies less effort around maintaining security ACLs.

Countering this advantage are various disadvantages to unnumbered links in IPv6:

- It is not possible to ping an interface that has only a link-local address from a device that is not directly attached to the link. Thus, to troubleshoot, one must typically log into a device that is directly attached to the device in question, and execute the ping from there.

- A traceroute passing over the unnumbered link will return the loopback or system address of the router, rather than the address of the interface itself.

- On some devices, by default the link-layer address of the interface is derived from the MAC address assigned to interface. When this is done, swapping out the interface hardware (e.g. interface card) will cause the link-layer address to change. In some cases (peering config, ACLs, etc) this may require additional changes. However, many devices allow the link-layer address of an interface to be explicitly configured, which avoids this issue.

- The practice of naming router interfaces using DNS names is difficult-to-impossible when using LLAs only.

- It is not possible to identify the interface or link (in a database, email, etc) by just giving its address.

For more discussion on the pros and cons, see [I-D.ietf-opsec-lla-only].

Today, most operators use numbered links (option b).

2.3. Link-Local Next-Hop in a Static Route?

What form of next-hop address should one use in a static route?

a. Use the far-end's link-local address as the next-hop address, OR

b. Use the far-end's GUA/ULA address as the next-hop address?

Recall that the IPv6 specs for OSPF [RFC5340] and ISIS [RFC5308] dictate that they always use link-locals for next-hop addresses. For static routes, [RFC4861] section 8 says:

> A router MUST be able to determine the link-local address for each of its neighboring routers in order to ensure that the target address in a Redirect message identifies the neighbor router by its link-local address. For static routing, this requirement implies that the next-hop router's address should be specified using the link-local address of the router.

This implies that using a GUA or ULA as the next hop will prevent a router from sending Redirect messages for packets that "hit" this static route. All this argues for using a link-local as the next-hop address in a static route.

However, there are two cases where using a link-local address as the next-hop clearly does not work. One is when the static route is an indirect (or multi-hop) static route. The second is when the static route is redistributed into another routing protocol. In these cases, the above text from RFC 4861 notwithstanding, either a GUA or ULA must be used.

Furthermore, many network operators are concerned about the dependency of the default link-local address on an underlying MAC address, as described in the previous section.

Today most operators use GUAs as next-hop addresses.

2.4. Separate or Combined eBGP Sessions?

For a dual-stack peering connection where eBGP is used as the routing protocol, then one can either:

a. Use one BGP session to carry both IPv4 and IPv6 routes, OR

b. Use two BGP sessions, a session over IPv4 carrying IPv4 routes and a session over IPv6 carrying IPv6 routes.

The main advantage of (a) is a reduction in the number of BGP

sessions compared with (b).

However, there are three main concerns with option (a). First, on most existing implementations, adding or removing an address family to an established BGP session will cause the router to tear down and re-establish the session. Thus adding the IPv6 family to an existing session carrying just IPv4 routes will disrupt the session, and the eventual removal of IPv4 from the dual IPv4/IPv6 session will also disrupt the session. This disruption problem will persist until something similar to [I-D.ietf-idr-dynamic-cap] is widely deployed. Second, there is the question of which protocol to use to carry the dual IPv4/IPv6 session: over IPv4 or over IPv6? Carrying it over IPv4 makes sense initially from a stability and troubleshooting perspective, but will eventually seem out-of-date. Third, carrying (for example) IPv6 routes over IPv4 means that route information is transported over a different transport plane than the data packets themselves. If the IPv6 data plane was to fail, then IPv6 routes would still be exchanged, but any IPv6 traffic resulting from these routes would be dropped.

Given these disadvantages, option (b) is the better choice in most situations, and this is the choice selected in most networks today.

2.5. eBGP Endpoints: Global or Link-Local Addresses?

When running eBGP over IPv6, there are two options for the addresses to use at each end of the eBGP session (or more properly, the underlying TCP session):

a. Use link-local addresses for the eBGP session, OR

b. Use global addresses for the eBGP session.

Note that the choice here is the addresses to use for the eBGP sessions, and not whether the link itself has global (or unique-local) addresses. In particular, it is quite possible for the eBGP session to use link-local addresses even when the link has global addresses.

The big attraction for option (a) is security: an eBGP session using link-local addresses is impossible to attack from a device that is off-link. This provides very strong protection against TCP RST and similar attacks. Though there are other ways to get an equivalent level of security (e.g. GTSM [RFC5082], MD5 [RFC5925], or ACLs), these other ways require additional configuration which can be forgotten or potentially mis-configured.

However, there are a number of small disadvantages to using link-

local addresses:

- Using link-local addresses only works for single-hop eBGP sessions; it does not work for multi-hop sessions.

- One must use "next-hop self" at both endpoints, otherwise redistributing routes learned via eBGP into iBGP will not work. (Some products enable "next-hop self" in this situation automatically).

- Operators and their tools are used to referring to eBGP sessions by address only, something that is not possible with link-local addresses.

- If one is configuring parallel eBGP sessions for IPv4 and IPv6 routes, then using link-local addresses for the IPv6 session introduces an extra difference between the two sessions which could otherwise be avoided.

- On some products, an eBGP session using a link-local address is more complex to configure than a session that use a global address.

- Finally, a strict interpretation of RFC 2545 can be seen as forbidding running eBGP between link-local addresses, as RFC 2545 requires the BGP next-hop field to contain at least a global address.

For these reasons, most operators today choose to have their eBGP sessions use global addresses.

3. General Observations

There are two themes that run though many of the design choices in this document. This section presents some general discussion on these two themes.

3.1. Use of Link-Local Addresses

The proper use of link-local addresses is a common theme in the IPv6 network design choices. Link-layer addresses are, of course, always present in an IPv6 network, but current network design practice mostly ignores them, despite efforts such as [I-D.ietf-opsec-lla-only].

There are three main reasons for this current practice:

- Network operators are concerned about the volitility of link-local addresses based on MAC addresses, despite the fact that this concern can be overcome by manually-configuring link-local addresses;

- It is impossible to ping a link-local address from a device that is not on the same subnet. This is a troubleshooting disadvantage, though it can also be viewed as a security advantage.

- Most operators are currently running networks that carry both IPv4 and IPv6 traffic, and wish to harmonize their IPv4 and IPv6 design and operational practices where possible.

3.2. Separation of IPv4 and IPv6

Currently, most operators are running or planning to run networks that carry both IPv4 and IPv6 traffic. Hence the question: To what degree should IPv4 and IPv6 be kept separate? As can be seen above, this breaks into two sub-questions: To what degree should IPv4 and IPv6 traffic be kept separate, and to what degree should IPv4 and IPv6 routing information be kept separate?

The general consensus around the first question is that IPv4 and IPv6 traffic should generally be mixed together. This recommendation is driven by the operational simplicity of mixing the traffic, plus the general observation that the service being offered to the end user is Internet connectivity and most users do not know or care about the differences between IPv4 and IPv6. Thus it is very desirable to mix IPv4 and IPv6 on the same link to the end user. On other links, separation is possible but more operationally complex, though it does occasionally allow the operator to work around limitations on network devices. The situation here is roughly comparable to IP and MPLS traffic: many networks mix the two traffic types on the same links without issues.

By contrast, there is more of an argument for carrying IPv6 routing information over IPv6 transport, while leaving IPv4 routing information on IPv4 transport. By doing this, one gets fate-sharing between the control and data plane for each IP protocol version: if the data plane fails for some reason, then often the control plane will too.

4. IANA Considerations

This document makes no requests of IANA.

5. Security Considerations

(TBD)

6. Acknowledgements

Many, many people in the V6OPS working group provided comments and suggestions that made their way into this document. A partial list includes: Rajiv Asati, Fred Baker, Michael Behringer, Marc Blanchet, Ron Bonica, Randy Bush, Cameron Byrne, Brian Carpenter, Tim Chown, Lorenzo Colitti, Gert Doering, Bill Fenner, Kedar K Gaonkar, Chris Grundemann, Steinar Haug, Ray Hunter, Joel Jaeggli, KK, Victor Kuarsingh, Alexandru Petrescu, Mark Smith, Jean-Francois Tremblay, Tina Tsou, Dan York, and Xuxiaohu. There are probably others which are not listed here, likely because they made a helpful comment at the mic during a WG session and I didn't catch the name.

I would also like to thank Pradeep Jain and Alastair Johnson for helpful comments on a very preliminary version of this document.

7. History

Version -01

Many, many changes from version -00, too many to document individually. Most of these changes are due to the many helpful comments and suggestions received by email or at the mic during the lengthy discussion at IETF 84 in Vancouver.

Version -00

Initial, very preliminary, version.

8. Informative References

[I-D.ietf-idr-dynamic-cap]
Ramachandra, S. and E. Chen, "Dynamic Capability for BGP-4", draft-ietf-idr-dynamic-cap-14 (work in progress), December 2011.

[I-D.ietf-opsec-lla-only]
Behringer, M. and E. Vyncke, "Using Only Link-Local Addressing Inside an IPv6 Network", draft-ietf-opsec-lla-only-01 (work in progress), September 2012.

[I-D.ietf-v6ops-enterprise-incremental-ipv6]
Chittimaneni, K., Chown, T., Howard, L., Kuarsingh, V., Pouffary, Y., and E. Vyncke, "Enterprise IPv6 Deployment Guidelines", draft-ietf-v6ops-enterprise-incremental-ipv6-01 (work in progress), September 2012.

[I-D.ietf-v6ops-icp-guidance]
Carpenter, B. and S. Jiang, "IPv6 Guidance for Internet Content and Application Service Providers", draft-ietf-v6ops-icp-guidance-04 (work in progress), September 2012.

[I-D.ietf-v6ops-wireline-incremental-ipv6]
Kuarsingh, V. and L. Howard, "Wireline Incremental IPv6", draft-ietf-v6ops-wireline-incremental-ipv6-06 (work in progress), September 2012.

[RFC4852] Bound, J., Pouffary, Y., Klynsma, S., Chown, T., and D. Green, "IPv6 Enterprise Network Analysis - IP Layer 3 Focus", RFC 4852, April 2007.

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

[RFC5082] Gill, V., Heasley, J., Meyer, D., Savola, P., and C. Pignataro, "The Generalized TTL Security Mechanism (GTSM)", RFC 5082, October 2007.

[RFC5308] Hopps, C., "Routing IPv6 with IS-IS", RFC 5308, October 2008.

[RFC5340] Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF for IPv6", RFC 5340, July 2008.

[RFC5375] Van de Velde, G., Popoviciu, C., Chown, T., Bonness, O., and C. Hahn, "IPv6 Unicast Address Assignment Considerations", RFC 5375, December 2008.

[RFC5925] Touch, J., Mankin, A., and R. Bonica, "The TCP Authentication Option", RFC 5925, June 2010.

[RFC5963] Gagliano, R., "IPv6 Deployment in Internet Exchange Points (IXPs)", RFC 5963, August 2010.

[RFC6180] Arkko, J. and F. Baker, "Guidelines for Using IPv6 Transition Mechanisms during IPv6 Deployment", RFC 6180,

May 2011.

[RFC6342] Koodli, R., "Mobile Networks Considerations for IPv6 Deployment", RFC 6342, August 2011.

Author's Address

Philip Matthews
Alcatel-Lucent
600 March Road
Ottawa, Ontario K2K 2E6
Canada

Phone: +1 613-784-3139
Email: philip_matthews@magma.ca