

V6OPS Working Group
Internet-Draft
Intended status: Informational
Expires: March 8, 2015

P. Matthews
Alcatel-Lucent
V. Kuarsingh
Dyn
September 4, 2014

Design Choices for IPv6 Networks
draft-ietf-v6ops-design-choices-02

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

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 March 8, 2015.

Copyright Notice

Copyright (c) 2014 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in [Section 4.e](#) of

the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
2. Design Choices	3
2.1. Links	3
2.1.1. Mix IPv4 and IPv6 on the Same Link?	3
2.1.2. Links with Only Link-Local Addresses?	4
2.2. Static Routes	6
2.2.1. Link-Local Next-Hop in a Static Route?	6
2.3. IGPs	6
2.3.1. IGP Choice	6
2.4. BGP	8
2.4.1. BGP Sessions for Unlabeled Routes	10
2.4.2. BGP sessions for Labeled or VPN Routes	11
2.4.3. eBGP Endpoints: Global or Link-Local Addresses?	11
3. General Observations	12
3.1. Use of Link-Local Addresses	12
3.2. Separation of IPv4 and IPv6	13
4. IANA Considerations	14
5. Security Considerations	14
6. Acknowledgements	14
7. Informative References	14
Authors' Addresses	16

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, [RFC6782] for wireline service providers, [RFC6342] for mobile network providers, [RFC5963] for exchange point operators, [RFC6883] 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].

This document focuses on unicast network design only. It does not cover multicast, nor supporting infrastructure such as DNS.

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. Links

2.1.1. 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:

- o Requires only half as many layer 3 interfaces as option (b), thus providing better scaling;
- o May require fewer physical ports, thus saving money;

- o Can make the QoS implementation much easier (for example, rate-limiting the combined IPv4 and IPv6 traffic to or from a customer);
- o Works well in practice, as any increase in IPv6 traffic is usually counter-balanced by a corresponding decrease in IPv4 traffic to or from the same host (ignoring the common pattern of an overall increase in Internet usage);
- o 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. Most networks today use option (a) wherever possible.

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.

2.1.2. 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:

- o 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.

- o 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.
- o In cases of parallel point to point links it is difficult to determine which of the parallel links was taken when attempting to troubleshoot unless one sends packets directly between the two attached link-locals on the specific interfaces. Since many network problems behave differently for traffic to/from a router than for traffic through the router(s) in question, this can pose a significant hurdle to some troubleshooting scenarios.
- o 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.
- o The practice of naming router interfaces using DNS names is difficult and not recommended when using link-locals only. More generally, it is not recommended to put link-local addresses into DNS; see [[RFC4472](#)].
- o It is not often not possible to identify the interface or link (in a database, email, etc) by giving just its address without also specifying the link in some manner.

It should be noted that it is quite possible for the same link-local address to be assigned to multiple interfaces. This can happen because the MAC address is duplicated (due to manufacturing process defaults or the use of virtualization), because a device deliberately re-uses automatically-assigned link-local addresses on different links, or because an operator manually assigns the same easy-to-type link-local address to multiple interfaces. All these are allowed in IPv6 as long as the addresses are used on different links.

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

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

2.2. Static Routes

2.2.1. 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.

2.3. IGPs

2.3.1. IGP Choice

One of the main decisions for an IPv6 implementor is the choice of IGP (Interior Gateway Protocol) within the network. The primary choices are the IETF protocols of RIP [RFC2080], OSPF [RFC2328] [RFC5340] and IS-IS [RFC5120] [RFC5308], though some operators may

consider non-IETF protocols. Here we limit our discussion to the pros and cons of OSPF vs. IS-IS.

Considering just OSPF vs. IS-IS, the discussion in this section revolves around the options in the table below:

Option	IGP for IPv4	IGP for IPv6	Known to work well	Hard separation	Similar configuration possible
a	IS-IS	IS-IS	YES	-	YES
b	IS-IS	OSPFv3	-	YES	-
c	OSPFv2	IS-IS	YES	YES	-
d	OSPFv2	OSPFv3	YES	YES	YES
e	OSPFv3	IS-IS	-	YES	-
f	OSPFv3	OSPFv3	-	-	YES

Three of the options above are marked as "Known to work well". These options have seen significant deployments and are generally considered to be good choices. The other options represent valid choices, but have not seen widespread use, so it is hard to offer comments on how well they work. In particular, options (e) and (f) use OSPFv3 to route IPv4 [[RFC5838](#)], which is still rather new and untested.

A number of options are marked "Gives hard separation". These options use a different IGP for IPv4 vs IPv6. With these options, a problem with routing IPv6 is unlikely to affect IPv4 or visa-versa.

Three options are marked "Similar configuration possible". This means it is possible (but not required) to use very similar IGP configuration for IPv4 and IPv6: for example, the same area boundaries, area numbering, link costing, etc. If you are happy with your IPv4 IGP design, then this will likely be a consideration. By contrast, the options that uses IS-IS for one IP version and OSPF for

the other version will require quite different configuration, and will also require the operations staff to become familiar with the difference between the two protocols.

With option (a), there is an additional choice of whether to run IS-IS in single-topology mode (where IPv4 and IPv6 share a single topology and a single set of link costs[RFC5308]) or multi-topology mode (where IPv4 and IPv6 have separate topologies and potentially different link costs[RFC5120]). A big problem with single-topology mode is that it cannot easily accommodate devices that support IPv4-only or IPv6-only. Thus, today there is general agreement that multi-topology is the right choice as this gives the greatest flexibility in network design.

It should be noted that a number of ISPs have run OSPF as their IPv4 IGP for quite a few years, but have selected IS-IS as their IPv6 IGP. However, there are very few (none?) that have made the reverse choice. This is, in part, because routers generally support more nodes in an IS-IS area than in the corresponding OSPF area, and because IS-IS is seen as more secure because it runs at layer 2.

2.4. BGP

The discussion in this section revolves around the following table.

Route Family	Transport	Comments
Unlabeled IPv4	IPv4	Works well
Unlabeled IPv4	IPv6	Next-hop issues
Unlabeled IPv6	IPv4	Next-hop issues
Unlabeled IPv6	IPv6	Works well
Labeled IPv4	IPv4	Works well
Labeled IPv4	IPv6	Next-hop issues
Labeled IPv6	IPv4	(6PE) Works well
Labeled IPv6	IPv6	???
VPN IPv4	IPv4	Works well
VPN IPv4	IPv6	Next-hop issues
VPN IPv6	IPv4	(6VPE) Works well
VPN IPv6	IPv6	???

The first column lists various route families, where "unlabeled" means SAFI 1, "labeled" means SAFI 4, and "VPN" means SAFI 128. The second column lists the protocol used to transport the BGP session, frequently specified by giving either an IPv4 or IPv6 address in the "neighbor" statement.

The third column comments on the combination in the first two columns:

- o For combinations marked "Works well", these combinations are widely supported and are generally recommended.
- o For combinations marked "Next-hop issues", these combinations are less-widely supported and when supported, often have next-hop

issues. That is, the next-hop address is typically a v4-mapped IPv6 address, which is based on some IPv4 address on the sending router. This v4-mapped IPv6 address is often not reachable by default using IPv6 routing. One common solution to this problem is to use routing policy to change the next-hop to a different IPv6 address.

- o For combinations marked as "???", it is believed that these combinations will not be supported until MPLS over IPv6 becomes available. [Need to Confirm].

Also, it is important to note that changing the set of address families being carried over a BGP session requires the BGP session to be reset (unless something like [[I-D.ietf-idr-dynamic-cap](#)] or [[I-D.ietf-idr-bgp-multisession](#)] is in use). This is generally more of an issue with eBGP sessions than iBGP sessions: for iBGP sessions it is common practice for a router to have two iBGP sessions, one to each member of a route reflector pair, and so one can change the set of address families on first one session and then the other.

The following subsections discuss specific scenarios in more detail.

2.4.1. BGP Sessions for Unlabeled Routes

Unlabeled routes are commonly carried on eBGP sessions, as well as on iBGP sessions in networks where Internet traffic is carried unlabeled across the network. In these scenarios, operators today most commonly use two BGP sessions: one session is transported over IPv4 and carries the unlabeled IPv4 routes, while the second session is transported over IPv6 and carries the unlabeled IPv6 routes.

There are several reasons for this choice:

- o It gives a clean separation between IPv4 and IPv6.
- o This avoids the next-hop problem described in note 1 above.
- o The status of the routes follows the status of the underlying transport. If, for example, the IPv6 data path between the two BGP speakers fails, then the IPv6 session between the two speakers will fail and the IPv6 routes will be withdrawn, which will allow the traffic to be re-routed elsewhere. By contrast, if the IPv6 routes were transported over IPv4, then the failure of the IPv6 data path might leave a working IPv4 data path, so the BGP session would remain up and the IPv6 routes would not be withdrawn, and thus the IPv6 traffic would be sent into a black hole.

- o It avoids resetting the BGP session when adding IPv6 to an existing session, or when removing IPv4 from an existing session.

2.4.2. BGP sessions for Labeled or VPN Routes

In these scenarios, it is most common today to carry both the IPv4 and IPv6 routes over sessions transported over IPv4. This can be done with either: (a) one session carrying both route families, or (b) two sessions, one for each family.

Using a single session is usually appropriate for an iBGP session going to a route reflector handling both route families. Using a single session here usually means that the BGP session will reset when changing the set of address families, but as noted above, this is usually not a problem when redundant route reflectors are involved.

In eBGP situations, two sessions are usually more appropriate.

2.4.3. 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:

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

- o One must use "next-hop self" at both endpoints, otherwise re-advertising routes learned via eBGP into iBGP will not work. (Some products enable "next-hop self" in this situation automatically).
- o Operators and their tools are used to referring to eBGP sessions by address only, something that is not possible with link-local addresses.
- o If one is configuring parallel eBGP sessions for IPv4 and IPv6 routes, then using link-local addresses for the IPv6 session introduces extra operational differences between the two sessions which could otherwise be avoided.
- o On some products, an eBGP session using a link-local address is more complex to configure than a session that use a global address.
- o If hardware or other issues cause one to move the cable to a different local interface, then reconfiguration is required at both ends: at the local end because the interface has changed (and with link-local addresses, the interface must always be specified along with the address), and at the remote end because the link-local address has likely changed. (Contrast this with using global addresses, where less re-configuration is required at the local end, and no reconfiguration is required at the remote end).
- o 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 through 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:

- o Network operators are concerned about the volatility of link-local addresses based on MAC addresses, despite the fact that this concern can be overcome by manually-configuring link-local addresses;
- o 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.
- o 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, KK Chittimaneni, Tim Chown, Lorenzo Colitti, Gert Doering, Bill Fenner, Kedar K Gaonkar, Chris Grundemann, Steinar Haug, Ray Hunter, Joel Jaeggli, Victor Kuarsingh, Ivan Pepelnjak, Alexandru Petrescu, Rob Shakir, Mark Smith, Jean-Francois Tremblay, Tina Tsou, Dan York, and Xuxiaohu.

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

7. Informative References

[I-D.ietf-idr-bgp-multisession]

Scudder, J., Appanna, C., and I. Varlashkin, "Multisession BGP", [draft-ietf-idr-bgp-multisession-07](#) (work in progress), September 2012.

[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-10](#) (work in progress), July 2014.

[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-ipv6-06](#) (work in progress), July 2014.

- [RFC2080] Malkin, G. and R. Minnear, "RIPng for IPv6", [RFC 2080](#), January 1997.
- [RFC2328] Moy, J., "OSPF Version 2", STD 54, [RFC 2328](#), April 1998.
- [RFC4472] Durand, A., Ihren, J., and P. Savola, "Operational Considerations and Issues with IPv6 DNS", [RFC 4472](#), April 2006.
- [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.
- [RFC5120] Przygienda, T., Shen, N., and N. Sheth, "M-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs)", [RFC 5120](#), February 2008.
- [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.
- [RFC5838] Lindem, A., Mirtorabi, S., Roy, A., Barnes, M., and R. Aggarwal, "Support of Address Families in OSPFv3", [RFC 5838](#), April 2010.
- [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.
- [RFC6782] Kuarsingh, V. and L. Howard, "Wireline Incremental IPv6", [RFC 6782](#), November 2012.
- [RFC6883] Carpenter, B. and S. Jiang, "IPv6 Guidance for Internet Content Providers and Application Service Providers", [RFC 6883](#), March 2013.

Authors' Addresses

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

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

Victor Kuarsingh
Dyn
150 Dow Street
Manchester, NH 03101
USA

Email: victor@jvknet.com