

V6OPS Working Group
Internet-Draft
Updates: 3316 (if approved)
Intended status: Informational
Expires: April 8, 2013

D. Binet
M. Boucadair
France Telecom
A. Vizdal
Deutsche Telekom AG
C. Byrne
T-Mobile
G. Chen
China Mobile
October 5, 2012

Internet Protocol Version 6 (IPv6) for Cellular Hosts
draft-binnet-v6ops-cellular-host-reqs-rfc3316update-03

Abstract

This document lists a set of IPv6-related requirements to be supported by cellular hosts.

This document updates RFC3316.

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].

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 April 8, 2013.

Copyright Notice

Copyright (c) 2012 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	3
1.1. Scope	3
2. Connectivity Requirements	4
2.1. WiFi Connectivity	7
3. Advanced Requirements	7
4. Cellular Devices with LAN Capabilities	8
5. APIs & Applications	10
6. Security Considerations	10
7. IANA Considerations	11
8. Acknowledgements	11
9. References	11
9.1. Normative References	11
9.2. Informative References	13
Authors' Addresses	14

1. Introduction

[RFC3316] lists a set of features to be supported by cellular hosts to connect to 3GPP cellular networks. Since the publication of that document, new functions have been specified within the 3GPP and the IETF whilst others have been updated. Moreover, in the light of recent IPv6 production deployments, additional features to facilitate IPv6-only deployments while accessing IPv4-only service are to be considered.

This document updates [RFC3316] with additional new functionalities which cellular hosts need to support.

While [RFC3316] considered only GPRS and UMTS networks; this document also considers EPS (Evolved Packet System).

A detailed overview of IPv6 support in 3GPP architectures is provided in [RFC6459].

This document makes use of the terms defined in [RFC6459].

PREFIX64 denotes an IPv6 prefix used to build IPv4-converted IPv6 addresses [RFC6052].

1.1. Scope

Various types of nodes can be connected to 3GPP networks requiring specific functions. Indeed, a 3GPP network can be used to connect user equipment such as a mobile telephone, a CPE or a machine-to-machine (M2M) device. Because of this diversity of terminals, it is necessary to define a set of IPv6 functionalities valid for any node directly connecting to a 3GPP network. This document describes these functionalities.

This document is structured to initially provide the generic IPv6 requirements which are valid for all nodes, whatever their function or service (e.g., SIP [RFC3261]) capability. The document also contains, dedicated sections covering specific functionalities the specific device types must support (e.g., smartphones, devices providing some LAN functions (mobile CPE or broadband dongles)).

M2M devices profile is not considered in the first version of this document.

The requirements listed below are valid for both 3GPP GPRS and 3GPP EPS access. For EPS, "PDN type" terminology is used instead of "PDP context".

2. Connectivity Requirements

REQ#1: The cellular host MUST support the IPv6 addressing architecture described in ([RFC4291]). For address representation, [RFC5952] MUST be supported.

REQ#2: The cellular host MUST support both IPv6 and IPv4v6 PDP Contexts.

This allows each operator to select their own strategy regarding IPv6 introduction. Both IPv6 and IPv4v6 PDP contexts MUST be supported in addition to the IPv4 PDP context. IPv4, IPv6 or IPv4v6 PDP-Context request acceptance depends on the mobile network configuration.

REQ#3: The cellular host MUST comply with the behavior defined in [TS.23060] [TS.23401] [TS.24008] for requesting a PDP context type. In particular, the cellular host MUST request an IPv6 PDP context if the cellular host is IPv6-only and requesting an IPv4v6 PDP context if the cellular host is dual stack or when the cellular host is not aware of connectivity types requested by devices connected to it (e.g., cellular host with LAN capabilities):

- * If the requested IPv4v6 PDP context is not supported by the network, but IPv4 and IPv6 PDP types are allowed, then the cellular host will be configured with an IPv4 address and/or an IPv6 prefix by the network. It MAY initiate another PDP request in addition to the one already activated for a given APN.
- * If the requested PDP type and subscription data allows only one IP address family (IPv4 or IPv6), the cellular host MUST NOT request a second PDP context to the same APN for the other IP address family.

REQ#4: The cellular host MUST support the PCO (Protocol Configuration Options) [TS.24008] to retrieve the IPv6 address(es) of the Recursive DNS server(s).

In-band signaling is a convenient method to inform the cellular host about various services, including DNS server information. It does not require any specific protocol to be supported and it is already deployed in IPv4 cellular networks to convey such DNS information.

REQ#5: The cellular host MUST support IPv6 aware Traffic Flow Templates (TFT) [TS.24008].

Traffic Flow Templates are employing a Packet Filter to couple an IP traffic with a PDP-Context. Thus a dedicated PDP-Context and radio resources can be provided by the mobile network for certain IP traffic.

REQ#6: The cellular host MUST support ICMPv6 ([RFC4443]).

The base protocol MUST be fully implemented by every IPv6 node as indicated in Section 2 of [RFC4443].

REQ#7: The device MUST support the Neighbor Discovery Protocol ([RFC4861] and [RFC5942]).

In particular, MTU communication via Router Advertisement SHOULD be supported since many 3GPP networks do not have a standard MTU setting due to inconsistencies in GTP [RFC3314] mobility tunnel infrastructure deployments.

REQ#8: The cellular host MUST support IPv6 Stateless Address Autoconfiguration ([RFC4862]) apart from the exceptions noted in [TS.23060] (3G) and [TS.23401] (LTE):

Stateless mode is the only way to configure a cellular host. The GGSN must allocate a prefix that is unique within its scope to each primary PDP context.

The cellular host MUST use the interface identifier sent in PDP Context Accept message to configure its link local address. The cellular host may use a different Interface Identifiers to configure its global addresses.

REQ#9: The cellular host SHOULD support Router Advertisement Options [RFC6106] for DNS configuration.

The support of this function allows for a consistent method of informing cellular hosts about DNS recursive servers across various types of access networks. The cellular host SHOULD support RA-based DNS information discovery.

REQ#10: The cellular host SHOULD embed a DHCPv6 client [RFC3736].

If [RFC6106] is not supported, the cellular host SHOULD retrieve DNS information using stateless DHCPv6 [RFC3736].

If the cellular host receives the DNS information in several channels for the same interface, the following preference order MUST be followed:

1. PCO
2. RA
3. DHCPv6

REQ#11: The cellular host SHOULD support a method to locally construct IPv4-embedded IPv6 addresses [RFC6052]. A method to learn PREFIX64 SHOULD be supported by the cellular host.

This solves the issue when applications use IPv4 referrals on IPv6-only access networks.

The cellular host SHOULD implement the method specified in [I-D.ietf-behave-nat64-discovery-heuristic] to retrieve the PREFIX64.

REQ#12: The cellular host SHOULD implement the Customer Side Translator function (CLAT, [I-D.ietf-v6ops-464xlat]) function which is compliant with [RFC6052][RFC6145][RFC6146].

CLAT function in the cellular host allows for IPv4-only application and IPv4-referrals to work on an IPv6-only PDP. CLAT function requires a NAT64 capability [RFC6146] in the core network.

REQ#13: The cellular device SHOULD embed a DNS64 function [RFC6147].

Local DNS64 functionality allows for compatibility with DNSSEC. Means to configure or discover a PREFIX64 is also required on the cellular device.

REQ#14: The cellular host SHOULD support PCP [I-D.ietf-pcp-base].

The support of PCP is seen as a driver to save battery consumption exacerbated by keepalive messages. PCP also gives the possibility of enabling incoming connections to the user. Indeed, because several stateful devices may be deployed in mobile networks (e.g., NAT and/or Firewalls), PCP can be used by the cellular host to control network based NAT and Firewall functions which will reduce per-application signaling and save battery consumption.

REQ#15: The cellular host SHOULD support means to prefer native IPv6 connection over NAT64 devices or NAT44 when the cellular host gets dual stack connectivity.

Cellular hosts SHOULD follow the procedure specified in [RFC6724] for source address selection.

Some potential issues are discussed in [I-D.ietf-mif-happy-eyeballs-extension] for MIFed devices.

REQ#16: The cellular host SHOULD support the procedure defined in [RFC6555].

REQ#17: The cellular host SHOULD NOT perform Duplicate Address Detection (DAD) for these Global IPv6 addresses (as the GGSN or PDN-GW must not configure any IPv6 addresses using the prefix allocated to the cellular host).

REQ#18: The cellular device MAY embed a BIH function [RFC6535] facilitating the communication between an IPv4 application and an IPv6 server.

2.1. WiFi Connectivity

It is increasingly common for cellular hosts have a Wi-Fi interface in addition to their cellular interface. These hosts are likely to be connected to private or public hotspots. Below are listed some generic requirements:

REQ#19: IPv6 MUST be supported on the Wi-Fi interface.

REQ#20: DHCPv6 client SHOULD be supported on Wi-Fi interface ([RFC3736]).

REQ#21: Wi-Fi interface SHOULD support Router Advertisement Options for DNS configuration ([RFC6106]).

3. Advanced Requirements

REQ#22: The cellular host MUST support Path MTU discovery ([RFC1981]). If the MTU used by cellular hosts is larger than 1280 bytes, they can rely on Path MTU discovery function to discover the real path MTU.

REQ#23: The cellular host SHOULD support the Privacy Extensions for Stateless Address Autoconfiguration in IPv6 ([RFC4941]).

The activation of privacy extension makes it more difficult to track a host over time when compared to using a permanent interface identifier. [RFC4941] does not require any DAD mechanism to be activated as the GGSN (or PDN-GW) MUST NOT configure any global address based on the prefix allocated to the cellular host.

REQ#24: The cellular host SHOULD support ROHC for IPv6 ([RFC5795]).

Bandwidth in mobile environments must be optimized as much as possible. ROHC provides a solution to reduce bandwidth consumption and to reduce the impact of having bigger packet headers in IPv6 compared to IPv4.

REQ#25: The cellular host SHOULD support IPv6 Router Advertisement Flags Options ([RFC5175]).

Some flags are used by the GGSN (or PDN-GW) to inform cellular hosts about the autoconfiguration process.

REQ#26: The cellular host SHOULD support Router Advertisement extension for communicating default router preferences and more-specific routes as described in [RFC4191].

This function can be used for instance for traffic offload.

4. Cellular Devices with LAN Capabilities

This section focuses on cellular devices (e.g., CPE, smartphones or dongles with tethering features) which provide IP connectivity to other devices connected to them. In such case, all connected devices are sharing the same GPRS, UMTS or EPS connection. In addition to the generic requirements listed in Section 2, these cellular devices have to meet the requirements listed below.

Prefix delegation which allows to allocate a shorter prefix to a cellular host is only available since 3GPP Release 10. For deployments requiring to share the same /64 prefix, the cellular device has to support a mechanism to enable sharing a /64 prefix between the 3GPP interface towards the GGSN (WAN interface) and the LAN interfaces.

[NOTE: Update the text with a pointer to I-D.byrne-v6ops-64share once adopted in v6ops.]

REQ#27: The cellular device MUST support Prefix Delegation capabilities [RFC3633] and MUST support Prefix Exclude Option for DHCPv6-based Prefix Delegation as defined in [RFC6603]. Particularly, it MUST behave as a Requesting Router.

Cellular networks are more and more perceived as an alternative to fixed networks for home IP-based services delivery; especially with the advent of smartphones and 3GPP data dongles. There is a need for an efficient mechanism to assign shorter prefix than /64 to cellular hosts so that each LAN segment can get its own /64 prefix and multilink subnet issues to be avoided.

In case a prefix is delegated to a cellular host using DHCPv6, the cellular device will be configured with two prefixes:

- (1) one for 3GPP link allocated using SLAAC mechanism and

- (2) another one delegated for LANs acquired during Prefix Delegation operation.

Note that the 3GPP network architecture requires both the WAN and the Delegated Prefix to be aggregatable, so the subscriber can be identified using a single prefix.

Without the Prefix Exclude Option, the delegating router (GGSN/PDN-GW) will have to ensure [RFC3633] compliancy (e.g., halving the Delegated prefix and assigning the WAN prefix out of the 1st half and the prefix to be delegated to the terminal from the 2nd half).

REQ#28: The cellular device MUST be compliant with the CPE requirements specified in [RFC6204].

REQ#29: The cellular device SHOULD support the Customer Side Translator (CLAT) [I-D.ietf-v6ops-464xlat].

Various IP devices are likely to be connected to cellular device, acting as a CPE. Some of these devices can be dual-stack, others are IPv6-only or IPv4-only. IPv6-only connectivity for cellular device

does not allow IPv4-only sessions to be established for hosts connected on the LAN segment of cellular devices.

In order to allow IPv4 sessions establishment initiated from devices located on LAN segment side and target IPv4 nodes, a solution consists in integrating the CLAT function in the cellular device. As elaborated in Section 2, the CLAT function allows also IPv4 applications to continue running over an IPv6-only host.

REQ#30: If a RA MTU is advertised from the 3GPP network, the cellular device SHOULD relay that upstream MTU information to the downstream attached LAN devices in RA.

Since 3GPP networks extensively use IP-in-IP/UDP GTP tunnels, the effective MTU is frequently effectively reduced to 1440 bytes. While a host may generate packets with an MTU of 1500 bytes, this results in undesirable fragmentation of the GTP IP/UDP packets.

Receiving and relaying RA MTU values facilitates a more harmonious functioning of the mobile core network where end nodes transmit packets that do not exceed the MTU size of the mobile network's GTP tunnels.

5. APIs & Applications

REQ#31: Name resolution libraries MUST support both IPv4 and IPv6.

In particular, the cellular host MUST support [RFC3596].

REQ#32: Applications MUST be independent of the underlying IP address family.

This means applications must be IP version agnostic.

REQ#33: Applications using URIs MUST follow [RFC3986]. For example, SIP applications MUST follow the correction defined in [RFC5954].

6. Security Considerations

The security considerations identified in [RFC3316] are to be taken into account.

REQ#34: If the cellular device provides LAN features, it SHOULD be compliant with the security requirements specified in [RFC6092].

7. IANA Considerations

This document does not require any action from IANA.

8. Acknowledgements

Many thanks to H. Soliman, H. Singh, L. Colliti, T. Lemon, B. Sarikaya, J. Korhonen and M. Mawatari for the discussion in the v6ops mailing list.

9. References

9.1. Normative References

- [RFC1981] McCann, J., Deering, S., and J. Mogul, "Path MTU Discovery for IP version 6", RFC 1981, August 1996.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC3261] Rosenberg, J., Schulzrinne, H., Camarillo, G., Johnston, A., Peterson, J., Sparks, R., Handley, M., and E. Schooler, "SIP: Session Initiation Protocol", RFC 3261, June 2002.
- [RFC3596] Thomson, S., Huitema, C., Ksinant, V., and M. Souissi, "DNS Extensions to Support IP Version 6", RFC 3596, October 2003.
- [RFC3633] Troan, O. and R. Droms, "IPv6 Prefix Options for Dynamic Host Configuration Protocol (DHCP) version 6", RFC 3633, December 2003.
- [RFC3736] Droms, R., "Stateless Dynamic Host Configuration Protocol (DHCP) Service for IPv6", RFC 3736, April 2004.
- [RFC3986] Berners-Lee, T., Fielding, R., and L. Masinter, "Uniform Resource Identifier (URI): Generic Syntax", STD 66, RFC 3986, January 2005.
- [RFC4191] Draves, R. and D. Thaler, "Default Router Preferences and

More-Specific Routes", RFC 4191, November 2005.

- [RFC4291] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 4291, February 2006.
- [RFC4443] Conta, A., Deering, S., and M. Gupta, "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", RFC 4443, March 2006.
- [RFC4861] Narten, T., Nordmark, E., Simpson, W., and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, September 2007.
- [RFC4862] Thomson, S., Narten, T., and T. Jinmei, "IPv6 Stateless Address Autoconfiguration", RFC 4862, September 2007.
- [RFC4941] Narten, T., Draves, R., and S. Krishnan, "Privacy Extensions for Stateless Address Autoconfiguration in IPv6", RFC 4941, September 2007.
- [RFC5175] Haberman, B. and R. Hinden, "IPv6 Router Advertisement Flags Option", RFC 5175, March 2008.
- [RFC5795] Sandlund, K., Pelletier, G., and L-E. Jonsson, "The RObust Header Compression (ROHC) Framework", RFC 5795, March 2010.
- [RFC5942] Singh, H., Beebee, W., and E. Nordmark, "IPv6 Subnet Model: The Relationship between Links and Subnet Prefixes", RFC 5942, July 2010.
- [RFC5952] Kawamura, S. and M. Kawashima, "A Recommendation for IPv6 Address Text Representation", RFC 5952, August 2010.
- [RFC5954] Gurbani, V., Carpenter, B., and B. Tate, "Essential Correction for IPv6 ABNF and URI Comparison in RFC 3261", RFC 5954, August 2010.
- [RFC6052] Bao, C., Huitema, C., Bagnulo, M., Boucadair, M., and X. Li, "IPv6 Addressing of IPv4/IPv6 Translators", RFC 6052, October 2010.
- [RFC6106] Jeong, J., Park, S., Beloeil, L., and S. Madanapalli, "IPv6 Router Advertisement Options for DNS Configuration", RFC 6106, November 2010.
- [RFC6145] Li, X., Bao, C., and F. Baker, "IP/ICMP Translation Algorithm", RFC 6145, April 2011.

- [RFC6146] Bagnulo, M., Matthews, P., and I. van Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", RFC 6146, April 2011.
- [RFC6147] Bagnulo, M., Sullivan, A., Matthews, P., and I. van Beijnum, "DNS64: DNS Extensions for Network Address Translation from IPv6 Clients to IPv4 Servers", RFC 6147, April 2011.
- [RFC6535] Huang, B., Deng, H., and T. Savolainen, "Dual-Stack Hosts Using "Bump-in-the-Host" (BIH)", RFC 6535, February 2012.
- [RFC6555] Wing, D. and A. Yourtchenko, "Happy Eyeballs: Success with Dual-Stack Hosts", RFC 6555, April 2012.
- [RFC6603] Korhonen, J., Savolainen, T., Krishnan, S., and O. Troan, "Prefix Exclude Option for DHCPv6-based Prefix Delegation", RFC 6603, May 2012.
- [RFC6724] Thaler, D., Draves, R., Matsumoto, A., and T. Chown, "Default Address Selection for Internet Protocol Version 6 (IPv6)", RFC 6724, September 2012.

9.2. Informative References

- [I-D.ietf-behave-nat64-discovery-heuristic]
Savolainen, T., Korhonen, J., and D. Wing, "Discovery of IPv6 Prefix Used for IPv6 Address Synthesis", draft-ietf-behave-nat64-discovery-heuristic-11 (work in progress), July 2012.
- [I-D.ietf-mif-happy-eyeballs-extension]
Chen, G., Williams, C., Wing, D., and A. Yourtchenko, "Happy Eyeballs Extension for Multiple Interfaces", draft-ietf-mif-happy-eyeballs-extension-00 (work in progress), July 2012.
- [I-D.ietf-pcp-base]
Wing, D., Cheshire, S., Boucadair, M., Penno, R., and P. Selkirk, "Port Control Protocol (PCP)", draft-ietf-pcp-base-28 (work in progress), October 2012.
- [I-D.ietf-v6ops-464xlat]
Mawatari, M., Kawashima, M., and C. Byrne, "464XLAT: Combination of Stateful and Stateless Translation", draft-ietf-v6ops-464xlat-08 (work in progress), September 2012.

- [RFC3314] Wasserman, M., "Recommendations for IPv6 in Third Generation Partnership Project (3GPP) Standards", RFC 3314, September 2002.
- [RFC3316] Arkko, J., Kuijpers, G., Soliman, H., Loughney, J., and J. Wiljakka, "Internet Protocol Version 6 (IPv6) for Some Second and Third Generation Cellular Hosts", RFC 3316, April 2003.
- [RFC6092] Woodyatt, J., "Recommended Simple Security Capabilities in Customer Premises Equipment (CPE) for Providing Residential IPv6 Internet Service", RFC 6092, January 2011.
- [RFC6204] Singh, H., Beebee, W., Donley, C., Stark, B., and O. Troan, "Basic Requirements for IPv6 Customer Edge Routers", RFC 6204, April 2011.
- [RFC6459] Korhonen, J., Soininen, J., Patil, B., Savolainen, T., Bajko, G., and K. Iisakkila, "IPv6 in 3rd Generation Partnership Project (3GPP) Evolved Packet System (EPS)", RFC 6459, January 2012.
- [TS.23060] 3GPP, "General Packet Radio Service (GPRS); Service description; Stage 2", September 2011.
- [TS.23401] 3GPP, "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access", September 2011.
- [TS.24008] 3GPP, "Mobile radio interface Layer 3 specification; Core network protocols; Stage 3", June 2011.

Authors' Addresses

David Binet
France Telecom
Rennes,
France

Email: david.binet@orange.com

Mohamed Boucadair
France Telecom
Rennes, 35000
France

Email: mohamed.boucadair@orange.com

Ales Vizdal
Deutsche Telekom AG

Phone:
Email: ales.vizdal@t-mobile.cz
URI:

Cameron Byrne
T-Mobile
USA

Phone:
Email: Cameron.Byrne@T-Mobile.com

Gang Chen
China Mobile

Email: phdgang@gmail.com

V6OPS Working Group
Internet-Draft
Intended Status: Informational
Expires: April 12, 2013

C. Byrne
T-Mobile USA
D. Drown
October 9, 2012

Sharing /64 3GPP Mobile Interface Subnet to a LAN
draft-byrne-v6ops-64share-03

Abstract

This document describes a known and implemented method of sharing a /64 IPv6 subnet from a User Equipment 3GPP radio interface to a tethered LAN.

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 April 12, 2013.

Copyright and License Notice

Copyright (c) 2012 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	3
2. The Challenge of Providing IPv6 Addresses to a 3GPP Tethered LAN	3
3. Method for Sharing the 3GPP Interface /64 to the Tethered LAN .	3
4. Security Considerations	4
5. IANA Considerations	4
6. Acknowledgments	4
7. Informative References	4

1. Introduction

3GPP mobile cellular networks such as GSM, UMTS, and LTE have architectural support for IPv6 [RFC6459], but only 3GPP Release-10 and onwards of the 3GPP specification supports DHCPv6 [RFC3633] for delegating IPv6 addresses to a tethered LAN. To facilitate the use of IPv6 in a tethered LAN prior to deployment of DHCPv6 in a 3GPP network, this document describes how the 3GPP User Equipment (UE) interface assigned /64 subnet may be shared from the 3GPP interface to a tethered LAN. This is achieved by specifying the UE 3GPP interface as an IPv6 /128 subnet taken from the 3GPP interface's network assigned /64 subnet. Then, assign the same address to the tethered LAN interface with the full /64 subnet. The /64 tethered LAN subnet will then be advertised to the tethered LAN via Router Advertisements (RA) [RFC4861].

The end result is that all UE interfaces have link-local IPv6 addresses, the UE's 3GPP interface has a /128 address from the 3GPP network assigned /64, and the same address that is assigned to the 3GPP interface is assigned to the tethered LAN interface with a /64 subnet and advertised to the LAN via RA. This approach only impacts the UE configuration and does not require any changes to the 3GPP network.

2. The Challenge of Providing IPv6 Addresses to a 3GPP Tethered LAN

As described in [RFC6459], 3GPP networks assign a /64 subnet to the UE with RA. IPv6 prefix delegation is a part of 3GPP Release-10 and is not covered by any earlier releases. Neighbor Discovery Proxy (ND Proxy) [RFC4389] functionality has been suggested as an option for sharing the assigned /64 from the 3GPP interface to the LAN, but ND Proxy is an experimental protocol and has some limitations with loop-avoidance.

DHCPv6 is the best way to assign subnets to tethered LANs. The method described in this document should only be applied when deploying DHCPv6 is not achievable in the 3GPP network.

3. Method for Sharing the 3GPP Interface /64 to the Tethered LAN

As [RFC6459] describes, the 3GPP network assigned /64 is completely dedicated to the UE and the gateway does not consume any of the /64 addresses. Communication between the UE and the gateway is only done using link-local addresses and the link is point-to-point. This allows for the UE to use the 3GPP network assigned /64 to assign itself a /128 subnet address to the 3GPP radio interface for consistent network reachability and the same address with a /64

subnet to the tethered LAN interface. The tethered LAN interface may then advertise the /64 subnet to the LAN with RA.

For example, if the 3GPP network assigns to the UE via RA the subnet 2001:db8:ac10:f002::/64, the UE may choose the address for its 3GPP interface to be 2001:db8:ac10:f002:1234:4567::9/128. When tethering a LAN, the UE may then assign that same address to its LAN interface with a /64 subnet, such as 2001:db8:ac10:f002:1234:4567::9/64. The UE may then advertise the 2001:db8:ac10:f002::/64 subnet to the tethered LAN using RA. Since the UE only consumes one address from the 3GPP network assigned /64 for both the 3GPP interface and the LAN interface, there is no address conflict potential. On the LAN, the /64 subnet is announced via RA and the interface address is defended with Duplicate Address Detection (DAD) [RFC4862]. Since the 3GPP interface is a point-to-point link and the gateway does not consume an address from the network assigned /64, there is no chance of address conflict on the 3GPP interface for the /64.

The UE should be compliant with the relevant requirements in [I-D.binet-v6ops-cellular-host-reqs-rfc3316update].

4. Security Considerations

Security considerations identified in [I-D.binet-v6ops-cellular-host-reqs-rfc3316update] are to be taken into account.

5. IANA Considerations

This document does not require any action from IANA.

6. Acknowledgments

Many thanks for review and discussion from Masanobu Kawashima, Teemu Savolainen, Mikael Abrahamsson, Eric Vyncke, and Ales Vizdal.

7. Informative References

[I-D.binet-v6ops-cellular-host-reqs-rfc3316update] Binet, D., Boucadair, M., A. Vizdal, C. Byrne, "Internet Protocol Version 6 (IPv6) for Cellular Hosts", draft-binet-v6ops-cellular-host-reqs-rfc3316update-03 (work in progress), October 2012.

[RFC3633] Troan, O. and R. Droms, "IPv6 Prefix Options for Dynamic Host Configuration Protocol (DHCP) version 6", RFC 3633, December 2003.

[RFC4389] Thaler, D., Talwar, M., and C. Patel, "Neighbor Discovery

Proxies (ND Proxy)", RFC 4389, April 2006.

- [RFC4861] Narten, T., Nordmark, E., Simpson, W., and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, September 2007.
- [RFC4862] Thomson, S., Narten, T., and T. Jinmei, "IPv6 Stateless Address Autoconfiguration", RFC 4862, September 2007.
- [RFC6459] Korhonen, J., Ed., Soininen, J., Patil, B., Savolainen, T., Bajko, G., and K. Iisakkila, "IPv6 in 3rd Generation Partnership Project (3GPP) Evolved Packet System (EPS)", RFC 6459, January 2012.

Authors' Addresses

Cameron Byrne
T-Mobile USA
Bellevue, Washington, USA

EMail: Cameron.Byrne@T-Mobile.com

Dan Drown
Email: Dan@Drown.org

Network Working Group
Internet-Draft
Intended status: Informational
Expires: September 29, 2013

F. L. Templin, Ed.
Boeing Research & Technology
March 28, 2013

Operational Considerations for Tunnel Fragmentation and Reassembly
draft-generic-v6ops-tunmtu-13.txt

Abstract

The Maximum Transmission Unit (MTU) for popular IP-in-IP tunnels is currently recommended to be set to 1500 (or less) minus the length of the encapsulation headers when static MTU determination is used. This requires the tunnel ingress to either fragment any IP packet larger than the MTU or drop the packet and return an ICMP Packet Too Big (PTB) message. Concerns for operational issues with Path MTU Discovery (PMTUD) point to the possibility of MTU-related black holes when a packet is dropped due to an MTU restriction. The current "Internet cell size" is effectively 1500 bytes (i.e., the minimum MTU configured by the vast majority of links in the Internet) and should therefore also be the minimum MTU assigned to tunnels, but this has proven to be problematic in common operational practice. This document therefore discusses operational considerations for tunnel fragmentation and reassembly necessary to accommodate this Internet cell size.

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 September 29, 2013.

Copyright Notice

Copyright (c) 2013 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. Tunnel Fragmentation and Reassembly	3
3. Jumbo Packet Accommodation	5
4. Common Tunneling Mechanisms	5
5. IANA Considerations	5
6. Security Considerations	5
7. Acknowledgments	5
8. References	6
8.1. Normative References	6
8.2. Informative References	6
Author's Address	7

1. Introduction

The Maximum Transmission Unit (MTU) for popular IP-in-IP tunnels is currently recommended to be set to 1500 (or less) minus the length of the encapsulation headers when static MTU determination is used. This requires the tunnel ingress to either fragment any IP packet larger than the MTU or drop the packet and return an ICMP Packet Too Big (PTB) message [RFC0791][RFC2460]. Concerns for operational issues with Path MTU Discovery (PMTUD) [RFC1191][RFC1981] point to the possibility of MTU-related black holes when a packet is dropped due to an MTU restriction. The current "Internet cell size" is effectively 1500 bytes (i.e., the minimum MTU configured by the vast majority of links in the Internet) and should therefore also be the minimum MTU assigned to tunnels, but this has proven to be problematic in common operational practice.

[RFC4459] discusses "MTU and Fragmentation Issues with In-the-Network Tunneling" and provides a comprehensive study of the various techniques that could be applied to alleviate the issues, including:

1. Fragmenting all too big encapsulated packets to fit in the paths, and reassembling them at the tunnel endpoints.

2. Signal to all the sources whose traffic must be encapsulated, and is larger than fits, to send smaller packets, e.g., using PMTUD.
3. Ensure that in the specific environment, the encapsulated packets will fit in all the paths in the network, e.g., by using MTU bigger than 1500 in the backbone used for encapsulation.
4. Fragmenting the original too big packets so that their fragments will fit, even encapsulated, in the paths, and reassembling them at the destination nodes. Note that this approach is only available for IPv4 under certain assumptions.

After considerable effort by many individuals since the publication of [RFC4459], these four alternatives continue to cover the domain of potential solutions - all of which have drawbacks and/or impracticalities. In this document, we discuss further considerations within the framework of the only solution alternative that can be applied generically - namely, fragmentation and reassembly at the tunnel endpoints.

2. Tunnel Fragmentation and Reassembly

Pushing the tunnel MTU to 1500 bytes or beyond is met with the challenge that the addition of encapsulation headers would cause an inner IP packet that is 1500 bytes (or slightly smaller) to appear as a slightly larger than 1500 byte outer IP packet on the wire, where it may be too large to traverse the path in one piece. When an IP tunnel configures an MTU smaller than 1500 bytes, packets that are small enough to traverse earlier links in the path toward the final destination may be dropped at the tunnel ingress which then returns a PTB message to the original source. However, operational experience has shown that the PTB messages can be lost in the network [RFC2923], in which case the source does not receive notification of the loss.

It is therefore highly desirable that the tunnel configure an MTU of at least 1500 bytes even though encapsulation would cause some tunneled packets to be slightly larger than 1500 bytes. In that case, the tunnel ingress would need to make special adaptations to deliver packets that are no larger than 1500 bytes yet larger than can be accommodated in a single piece.

One possibility is to use IP fragmentation of the inner IP layer protocol before encapsulation so that inner packet fragments can be delivered via the tunnel without loss due to a size restriction and then reassembled at the final destination. This option removes the burden from the tunnel endpoints, but is only available for IPv4 packets (since IPv6 deprecates router fragmentation [RFC2460]), and is further only available when the IPv4 header sets the Don't Fragment (DF) bit in the IPv4 header to 0.

A second possibility is to use IP fragmentation of the outer IP layer protocol following encapsulation so that the outer packet fragments can be delivered via the tunnel without loss due to a size restriction and then reassembled at the tunnel egress. This option is available for tunnels over both IPv4 and IPv6, and indeed the tunnel ingress is permitted to use IPv6 fragmentation since it is acting as a "host" (i.e., and not a router) for the encapsulated packets it produces. While IPv6 fragmentation is assumed to be "safe at all speeds", IPv4 fragmentation can be dangerous at high data rates due to the possibility of Identification field wrapping while reassemblies are still active [RFC4963][RFC6864]. Also, if outer IP fragmentation were used the tunnel ingress has no assurance that the egress can reassemble packets larger than 1500 bytes, since the Minimum Reassembly Unit (MRU) is 1500 bytes for IPv6 [RFC2460] and only 576 bytes for IPv4 [RFC1122]. Finally, recent studies have shown that IPv6 fragments are sometimes dropped in the network due to middlebox misconfigurations [I-D.taylor-v6ops-fragdrop].

A third possibility for accommodating inner packets that are slightly too large is the use of "tunnel fragmentation" based on a mid-layer encapsulation that is inserted between the inner and outer IP headers. Tunnel fragmentation requires separate packet Identification and segmentation control bits in the mid-layer encapsulation that are distinct from those that appear in the inner and/or outer headers. As for outer fragmentation, the tunnel egress is responsible for reassembly. Tunnel fragmentation can be particularly useful for tunnels over IPv4, since the mid-layer encapsulation can include an extended Identification field that avoids the identification wrapping issue discussed above. However, tunnel fragmentation is not used in common widely-deployed tunneling mechanisms at the time of this writing. An example of tunnel fragmentation appears in SEAL [I-D.templin-intarea-seal].

Following any inner, tunnel or outer fragmentation, the ingress must allow the encapsulated packets or fragments to be further fragmented by a router on the path that configures a link with a too-small MTU. These fragments would be reassembled by the tunnel egress the same as if the fragmentation occurred within the tunnel ingress. This final form of fragmentation is undesirable and should be avoided if at all

possible through the application of fragmentation at the tunnel ingress. However, common widely-deployed tunneling mechanisms at the time of this writing make no such provisions.

3. Jumbo Packet Accommodation

In addition to failure to accommodate packets up to 1500 bytes in length, current tunneling solutions typically do not make provisions for delivering packets that are larger than 1500 bytes. As long as they are no larger than the underlying link used for tunneling, the tunnel ingress should admit such "jumbo" packets into the tunnel and allow them to either be delivered to the egress in one piece or be dropped with the possibility of a PTB message being returned. The original host will then be able to determine the correct packet sizes whether or not PTB messages are delivered if it is using [RFC4821]. However, this approach is not used in common widely-deployed tunneling mechanisms at the time of this writing.

4. Common Tunneling Mechanisms

The operational issues discussed in this document apply to existing IPv6-in-IPv4 transition mechanisms, including configured tunnels [RFC4213], 6to4 [RFC3056], Teredo [RFC4380], ISATAP [RFC5214], DSMIP [RFC5555], 6rd [RFC5969], etc.

The issues further apply to existing IP-in-IP tunneling mechanisms of all varieties, including GRE [RFC1701], IPv4-in-IPv4 [RFC2003], IPv6-in-IPv6 [RFC2473], IPv4-in-IPv6 [RFC6333], IPsec [RFC4301], etc.

5. IANA Considerations

There are no IANA considerations for this document.

6. Security Considerations

The security considerations for the various tunneling mechanisms apply also to this document.

7. Acknowledgments

This method was inspired through discussion on the IETF v6ops and NANOG mailing lists in the May/June 2012 timeframe.

8. References

8.1. Normative References

- [RFC0791] Postel, J., "Internet Protocol", STD 5, RFC 791, September 1981.
- [RFC2460] Deering, S.E. and R.M. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.
- [RFC4459] Savola, P., "MTU and Fragmentation Issues with In-the-Network Tunneling", RFC 4459, April 2006.

8.2. Informative References

- [I-D.taylor-v6ops-fragdrop]
Jaeggli, J., Colitti, L., Kumari, W., Vyncke, E., Kaeo, M., and T. Taylor, "Why Operators Filter Fragments and What It Implies", draft-taylor-v6ops-fragdrop-00 (work in progress), October 2012.
- [I-D.templin-intarea-seal]
Templin, F., "The Subnetwork Encapsulation and Adaptation Layer (SEAL)", draft-templin-intarea-seal-52 (work in progress), March 2013.
- [RFC1122] Braden, R., "Requirements for Internet Hosts - Communication Layers", STD 3, RFC 1122, October 1989.
- [RFC1191] Mogul, J. and S. Deering, "Path MTU discovery", RFC 1191, November 1990.
- [RFC1701] Hanks, S., Li, T., Farinacci, D., and P. Traina, "Generic Routing Encapsulation (GRE)", RFC 1701, October 1994.
- [RFC1981] McCann, J., Deering, S., and J. Mogul, "Path MTU Discovery for IP version 6", RFC 1981, August 1996.
- [RFC2003] Perkins, C., "IP Encapsulation within IP", RFC 2003, October 1996.
- [RFC2473] Conta, A. and S. Deering, "Generic Packet Tunneling in IPv6 Specification", RFC 2473, December 1998.
- [RFC2923] Lahey, K., "TCP Problems with Path MTU Discovery", RFC 2923, September 2000.

- [RFC3056] Carpenter, B. and K. Moore, "Connection of IPv6 Domains via IPv4 Clouds", RFC 3056, February 2001.
- [RFC4213] Nordmark, E. and R. Gilligan, "Basic Transition Mechanisms for IPv6 Hosts and Routers", RFC 4213, October 2005.
- [RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, December 2005.
- [RFC4380] Huitema, C., "Teredo: Tunneling IPv6 over UDP through Network Address Translations (NATs)", RFC 4380, February 2006.
- [RFC4821] Mathis, M. and J. Heffner, "Packetization Layer Path MTU Discovery", RFC 4821, March 2007.
- [RFC4963] Heffner, J., Mathis, M., and B. Chandler, "IPv4 Reassembly Errors at High Data Rates", RFC 4963, July 2007.
- [RFC5214] Templin, F., Gleeson, T., and D. Thaler, "Intra-Site Automatic Tunnel Addressing Protocol (ISATAP)", RFC 5214, March 2008.
- [RFC5555] Soliman, H., "Mobile IPv6 Support for Dual Stack Hosts and Routers", RFC 5555, June 2009.
- [RFC5969] Townsley, W. and O. Troan, "IPv6 Rapid Deployment on IPv4 Infrastructures (6rd) -- Protocol Specification", RFC 5969, August 2010.
- [RFC6333] Durand, A., Droms, R., Woodyatt, J., and Y. Lee, "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", RFC 6333, August 2011.
- [RFC6864] Touch, J., "Updated Specification of the IPv4 ID Field", RFC 6864, February 2013.

Author's Address

Fred L. Templin (editor)
Boeing Research & Technology
P.O. Box 3707
Seattle, WA 98124
USA

Email: fltemplin@acm.org

DHC Working Group
Internet-Draft
Intended status: Standards Track
Expires: March 15, 2015

S. Jiang
Huawei Technologies Co., Ltd
G. Chen
China Mobile
S. Krishnan
Ericsson
R. Asati
Cisco Systems, Inc.
September 11, 2014

Registering Self-generated IPv6 Addresses in DNS using DHCPv6
draft-ietf-dhc-addr-registration-07

Abstract

In networks that are centrally managed, self-generated addresses cause some traceability issues due to their decentralized nature. One of the most important issues in this regard is the inability to register such addresses in DNS. This document defines a mechanism to register self-generated and statically configured addresses in DNS through a DHCPv6 server.

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 15, 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. Terminology	3
3. Solution Overview	3
4. DHCPv6 ADDR-REGISTRATION-REQUEST Message	4
5. DHCPv6 Address Registration Procedure	5
5.1. DHCPv6 Address Registration Request	6
5.2. Registration Expiry and Refresh	6
5.3. Acknowledging Registration and Retransmission	6
6. Security Considerations	7
7. IANA Considerations	8
8. Acknowledgements	8
9. References	8
9.1. Normative References	8
9.2. Informative References	9
Authors' Addresses	9

1. Introduction

In several common network scenarios, IPv6 addresses are self-generated by the end-hosts by appending a self-generated interface identifier to a network-specified prefix. Examples of self-generated addresses include those created using IPv6 Stateless Address Configuration [RFC4862] , temporary addresses [RFC4941] and Cryptographically Generated Addresses (CGA) [RFC3972] etc. In several tightly controlled networks, hosts with self-generated addresses may face some limitations. One such limitation is related to the inability of nodes with self-generated addresses to register their IPv6-address-to-FQDN bindings in DNS. This is related to the fact that, in such networks, only certain nodes (e.g. The DHCPv6 server) are allowed to update these bindings in order to prevent end-hosts from registering arbitrary addresses for their FQDNs or associating their addresses with arbitrary domain names. The administrators may not want to distribute the address of authoritative name-server. Also, there is no way to propagate the address of authoritative name server by any protocols. It is preferred that the address registration server, which is under the same management with the authoritative name-server, to know the address of the authoritative name-server and make registration requests on behalf of clients. It is preferred by administrators to

establish and manage one trust relationship between a single DHCPv6 (address registration) server and the DNS authoritative name-server, rather than to distribute and manage trust relationships between many clients and the DNS authoritative name-server.

For nodes that obtain their addresses through DHCPv6, a solution has been specified in [RFC4704]. The solution works by including a Client FQDN option in the SOLICIT, REQUEST, RENEW or REBIND messages during the process of acquiring an address through DHCPv6. This document provides an analogous mechanism to register self-generated addresses in DNS.

A new ADDR-REGISTRATION-REQUEST DHCPv6 message type is defined to initiate the address registration request, and two new Status codes are defined to indicate registration errors on the server side.

2. Terminology

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 [RFC2119].

Certificate In this document, the term "Certificate" is all referred to public key certificate.

3. Solution Overview

After successfully assigning a self-generated IPv6 address on one of its interfaces, an end-host implementing this specification SHOULD send an ADDR-REGISTRATION-REQUEST message to a DHCPv6 address registration server. After receiving the address registration request, the DHCPv6 server registers the IPv6 address to FQDN binding towards a configured DNS server. An acknowledgement MUST be sent back to the end host to indicate whether or not the registration operation succeeded.

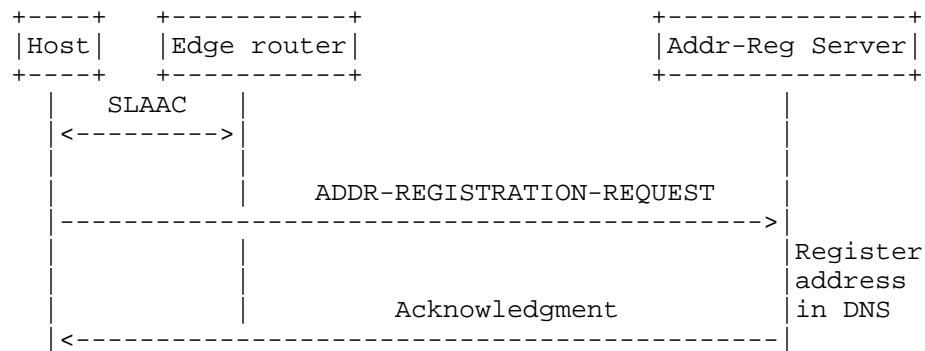


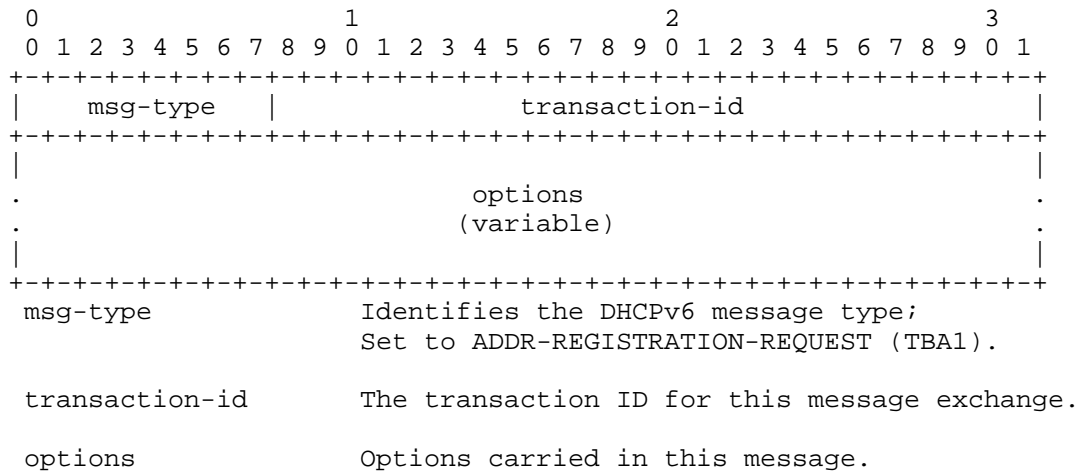
Figure 1: Address Registration Procedure

Furthermore, the registration server MAY apply certain filter/accept criteria for the address registration requests, particularly for the client chosen domain names.

It is RECOMMENDED to only set up one addressregistration server within an administration domain, although there may be multiple DHCPv6 servers. While using multiple address registration servers does potentially increase the load on DNS, because of how [RFC4703] and [RFC4704] work, this should NOT be an issue - the servers should work correctly in updating DNS (either adding or removing the entries). The broken part with multiple servers is the 'extension' of the registration. If there are two address registration servers and both receive the initial registration and (correctly) update DNS, the problem comes when the client extends this but one of the servers does not receive this extension. Then, the server that missed the extension removes the entry prematurely (i.e., when it expired originally).

4. DHCPv6 ADDR-REGISTRATION-REQUEST Message

The DHCPv6 client sends an ADDR-REGISTRATION-REQUEST message to a server to request an address to be registered in the DNS. The format of the ADDR-REGISTRATION-REQUEST message is described as follows:



DHCPv6 ADDR-REGISTRATION-REQUEST message

The ADDR-REGISTRATION-REQUEST message MUST NOT contain server-identifier option and MUST contain the IA Address option and the DHCPv6 FQDN option [RFC4704]. The ADDR-REGISTRATION-REQUEST message is dedicated for clients to initiate an address registration request toward an address registration server. Consequently, clients MUST NOT put any Option Request Option(s) in the ADDR-REGISTRATION-REQUEST message.

Clients MUST discard any received ADDR-REGISTRATION-REQUEST messages.

Servers MUST discard any ADDR-REGISTRATION-REQUEST messages that meet any of the following conditions:

- o the message does not include a Client Identifier option;
- o the message includes a Server Identifier option;
- o the message does not include at least one IA Address option;
- o the message does not include FQDN option (or include multiple FQDN options);
- o the message includes an Option Request Option.

5. DHCPv6 Address Registration Procedure

The DHCPv6 protocol is used as the address registration protocol when a DHCPv6 server performs the role of an address registration server. The DHCPv6 IA Address option [RFC3315] and the DHCPv6 FQDN option

[RFC4704] are adopted in order to fulfill the address registration interactions.

5.1. DHCPv6 Address Registration Request

The end-host sends a DHCPv6 ADDR-REGISTRATION-REQUEST message to the address registration server to the All_DHCP_Relay_Agents_and_Servers multicast address (ff02::1:2).

The end-host MUST include a Client Identifier option in the ADDR-REGISTRATION-REQUEST message to identify itself to the server. The DHCPv6 ADDR-REGISTRATION-REQUEST message MUST contain at least one IA Address option and exactly one FQDN option. The valid-lifetime field of the IA Address option MUST be set to the period for which the client would like to register the binding in DNS.

After receiving this ADDR-REGISTRATION-REQUEST message, the address registration server MUST register the binding between the provided FQDN and address(es) in DNS. If the DHCPv6 server does not support address registration function, it MUST silently drop the message.

5.2. Registration Expiry and Refresh

For every successful binding registration, the address registration server MUST record the IPv6-address-to-FQDN bindings and associated valid-lifetimes in its storage.

The address registration client MUST refresh the registration before it expires (i.e. before the valid-lifetime of the IA address elapses) by sending a new ADDR-REGISTRATION-REQUEST to the address registration server. If the address registration server does not receive such a refresh after the valid-lifetime has passed, it SHOULD remove the IPv6-address-to-FQDN bindings in DNS, also the local record.

It is RECOMMENDED that clients initiate a refresh at about 85% of the valid-lifetime. Because RAs may periodically 'reset' the valid-lifetime, the refresh timer MUST be independently maintained from the address valid-lifetime. Clients SHOULD set a refresh timer to 85% of the valid-lifetime when they complete a registration operation and only update this timer if 85% of any updated valid-lifetime would be sooner than the timer.

5.3. Acknowledging Registration and Retransmission

After an address registration server accepts an address registration request, it MUST send a Reply message as the response to the client. The acceptance reply only means that the server has taken

responsibility to registry for the client. It may not have actually completed the update yet. The server is responsible to register all the addresses in DNS. The server generates a Reply message and includes a Status Code option with value Success, a Server Identifier option with the server's DUID, and a Client Identifier option with the client's DUID.

If there is no reply received within some interval, the client SHOULD retransmits the message according to section 14 of [RFC3315], using the following parameters:

- o IRT ADDR_REG_TIMEOUT
- o MRT ADDR_REG_MAX_RT
- o MRC ADDR_REG_MAX_RC
- o MRD 0

The below presents a table of values used to describe the message transmission behavior of clients and servers:

Parameter	Default	Description
ADDR_REG_TIMEOUT	1 secs	Initial Addr Registration Request timeout
ADDR_REG_MAX_RT	60 secs	Max Addr Registration Request timeout value
ADDR_REG_MAX_RC	5	Max Request retry attempts

For each IA Address option in the ADDR-REGISTRATION-REQUEST message for which the server does not accept its associated registration request, the server adds an IA Address option with the associated IPv6 address, and includes a Status Code option with the value RegistrationDenied (TBA2) in the IA Address option. No other options are included in the IA Address option.

Upon receiving a RegistrationDenied error status code, the client MAY also resend the message following normal retransmission routines defined in [RFC3315] with above parameters. The client MUST wait out the retransmission time before retrying.

6. Security Considerations

An attacker may attempt to register large number of addresses in quick succession in order to overwhelm the address registration server. These attacks may be prevented generic DHCPv6 protection by using the AUTH option [RFC3315] or Secure DHCPv6 [I-D.ietf-dhc-sedhcpv6].

7. IANA Considerations

This document defines a new DHCPv6 message, the ADDR-REGISTRATION-REQUEST message (TBA1) described in Section 4, that requires an allocation out of the registry of Message Types defined at <http://www.iana.org/assignments/dhcpv6-parameters/>

Value	Description	Reference
TBA1	ADDR-REGISTRATION-REQUEST	this document

This document defines a new DHCPv6 Status code, the RegistrationDenied (TBA2) described in Section 5, that requires an allocation out of the registry of Status Codes defined at <http://www.iana.org/assignments/dhcpv6-parameters/>

Code	Name	Reference
TBA2	RegistrationDenied	this document

8. Acknowledgements

The authors would like to thank Ralph Droms, Ted Lemon, Bernie Volz, Sten Carlsen, Erik Kline, Lorenzo Colitti, Joel Jaeggli, Sten Carlsen, Mark Smith, Marcin Siodelski, Darpan Malhotra, Tomek Mrugalski, Jinmei Tatuya and other members of dhc and v6ops working groups for their valuable comments.

9. References

9.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997.
- [RFC2136] Vixie, P., Thomson, S., Rekhter, Y., and J. Bound, "Dynamic Updates in the Domain Name System (DNS UPDATE)", RFC 2136, April 1997.
- [RFC3315] Droms, R., Bound, J., Volz, B., Lemon, T., Perkins, C., and M. Carney, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", RFC 3315, July 2003.
- [RFC3972] Aura, T., "Cryptographically Generated Addresses (CGA)", RFC 3972, March 2005.

- [RFC4703] Stapp, M. and B. Volz, "Resolution of Fully Qualified Domain Name (FQDN) Conflicts among Dynamic Host Configuration Protocol (DHCP) Clients", RFC 4703, October 2006.
- [RFC4704] Volz, B., "The Dynamic Host Configuration Protocol for IPv6 (DHCPv6) Client Fully Qualified Domain Name (FQDN) Option", RFC 4704, October 2006.
- [RFC4862] Thomson, S., Narten, T., and T. Jinmei, "IPv6 Stateless Address Autoconfiguration", RFC 4862, September 2007.
- [RFC4941] Narten, T., Draves, R., and S. Krishnan, "Privacy Extensions for Stateless Address Autoconfiguration in IPv6", RFC 4941, September 2007.

9.2. Informative References

- [I-D.ietf-dhc-sedhcpv6]
Jiang, S., Shen, S., Zhang, D., and T. Jinmei, "Secure DHCPv6 with Public Key", draft-ietf-dhc-sedhcpv6-03 (work in progress), June 2014.

Authors' Addresses

Sheng Jiang
Huawei Technologies Co., Ltd
Q14, Huawei Campus
No.156 Beiqing Road
Hai-Dian District, Beijing 100095
P.R. China

Email: jiangsheng@huawei.com

Gang Chen
China Mobile
53A, Xibianmennei Ave., Xuanwu District, Beijing
P.R. China

Phone: 86-13910710674
Email: phdgang@gmail.com

Suresh Krishnan
Ericsson
8400 Decarie Blvd.
Town of Mount Royal, QC
Canada

Phone: +1 514 345 7900 x42871
Email: suresh.krishnan@ericsson.com

Rajiv Asati
Cisco Systems, Inc.
7025 Kit Creek road
Research Triangle Park, NC 27709-4987
USA

Email: rajiva@cisco.com

Network Working Group
Internet-Draft
Intended status: Informational
Expires: February 1, 2015

K. Chittimaneni
Dropbox Inc.
T. Chown
University of Southampton
L. Howard
Time Warner Cable
V. Kuarsingh
Dyn Inc
Y. Pouffary
Hewlett Packard
E. Vyncke
Cisco Systems
July 31, 2014

Enterprise IPv6 Deployment Guidelines
draft-ietf-v6ops-enterprise-incremental-ipv6-06

Abstract

Enterprise network administrators worldwide are in various stages of preparing for or deploying IPv6 into their networks. The administrators face different challenges than operators of Internet access providers, and have reasons for different priorities. The overall problem for many administrators will be to offer Internet-facing services over IPv6, while continuing to support IPv4, and while introducing IPv6 access within the enterprise IT network. The overall transition will take most networks from an IPv4-only environment to a dual stack network environment and eventually an IPv6-only operating mode. This document helps provide a framework for enterprise network architects or administrators who may be faced with many of these challenges as they consider their IPv6 support strategies.

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 February 1, 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	3
1.1. Enterprise Assumptions	4
1.2. IPv4-only Considerations	4
1.3. Reasons for a Phased Approach	5
2. Preparation and Assessment Phase	6
2.1. Program Planning	6
2.2. Inventory Phase	7
2.2.1. Network infrastructure readiness assessment	7
2.2.2. Applications readiness assessment	8
2.2.3. Importance of readiness validation and testing	8
2.3. Training	9
2.4. Security Policy	9
2.4.1. IPv6 is no more secure than IPv4	9
2.4.2. Similarities between IPv6 and IPv4 security	10
2.4.3. Specific Security Issues for IPv6	10
2.5. Routing	12
2.6. Address Plan	13
2.7. Tools Assessment	15
3. External Phase	16
3.1. Connectivity	16
3.2. Security	17
3.3. Monitoring	19
3.4. Servers and Applications	19
3.5. Network Prefix Translation for IPv6	20
4. Internal Phase	20
4.1. Security	21
4.2. Network Infrastructure	21
4.3. End user devices	22
4.4. Corporate Systems	23

5. IPv6-only	23
6. Considerations For Specific Enterprises	25
6.1. Content Delivery Networks	25
6.2. Data Center Virtualization	25
6.3. University Campus Networks	25
7. Security Considerations	27
8. Acknowledgements	27
9. IANA Considerations	27
10. Informative References	27
Authors' Addresses	32

1. Introduction

An Enterprise Network is defined in [RFC4057] as a network that has multiple internal links, one or more router connections to one or more Providers, and is actively managed by a network operations entity (the "administrator", whether a single person or department of administrators). Administrators generally support an internal network, consisting of users' workstations, personal computers, mobile devices, other computing devices and related peripherals, a server network, consisting of accounting and business application servers, and an external network, consisting of Internet-accessible services such as web servers, email servers, VPN systems, and customer applications. This document is intended as guidance for enterprise network architects and administrators in planning their IPv6 deployments.

The business reasons for spending time, effort, and money on IPv6 will be unique to each enterprise. The most common drivers are due to the fact that when Internet service providers, including mobile wireless carriers, run out of IPv4 addresses, they will provide native IPv6 and non-native IPv4. The non-native IPv4 service may be NAT64, NAT444, Dual-stack Lite, MAP-T, MAP-E, or other transition technologies. Compared to tunneled or translated service, native traffic typically performs better and more reliably than non-native. For example, for client networks trying to reach enterprise networks, the IPv6 experience will be better than the transitional IPv4 if the enterprise deploys IPv6 in its public-facing services. The native IPv6 network path should also be simpler to manage and, if necessary, troubleshoot. Further, enterprises doing business in growing parts of the world may find IPv6 growing faster there, where again potential new customers, employees and partners are using IPv6. It is thus in the enterprise's interests to deploy native IPv6, at the very least in its public-facing services, but ultimately across the majority or all of its scope.

The text in this document provides specific guidance for enterprise networks, and complements other related work in the IETF, including [I-D.ietf-v6ops-design-choices] and [RFC5375].

1.1. Enterprise Assumptions

For the purpose of this document, we assume:

- o The administrator is considering deploying IPv6 (but see Section 1.2 below).
- o The administrator has existing IPv4 networks and devices which will continue to operate and be supported.
- o The administrator will want to minimize the level of disruption to the users and services by minimizing number of technologies and functions that are needed to mediate any given application. In other words: provide native IP wherever possible.

Based on these assumptions, an administrator will want to use technologies which minimize the number of flows being tunnelled, translated or intercepted at any given time. The administrator will choose transition technologies or strategies which allow most traffic to be native, and will manage non-native traffic. This will allow the administrator to minimize the cost of IPv6 transition technologies, by containing the number and scale of transition systems.

Tunnels used for IPv6/IPv4 transition are expected as near/mid- term mechanisms, while IPv6 tunneling will be used for many long-term operational purposes such as security, routing control, mobility, multi-homing, traffic engineering, etc. We refer to the former class of tunnels as "transition tunnels"

1.2. IPv4-only Considerations

As described in [RFC6302] administrators should take certain steps even if they are not considering IPv6. Specifically, Internet-facing servers should log the source port number, timestamp (from a reliable source), and the transport protocol. This will allow investigation of malefactors behind address-sharing technologies such as NAT444, MAP, or Dual-stack Lite. Such logs should be protected for integrity, safeguarded for privacy and periodically purged within applicable regulations for log retention.

Other IPv6 considerations may impact ostensibly IPv4-only networks, e.g. [RFC6104] describes the rogue IPv6 RA problem, which may cause problems in IPv4-only networks where IPv6 is enabled in end systems

on that network. Further discussion of the security implications of IPv6 in IPv4-only networks can be found in [RFC7123]).

1.3. Reasons for a Phased Approach

Given the challenges of transitioning user workstations, corporate systems, and Internet-facing servers, a phased approach allows incremental deployment of IPv6, based on the administrator's own determination of priorities. This document outlines suggested phases: a Preparation and Assessment Phase, an Internal Phase, and an External Phase. The Preparation Phase is highly recommended to all administrators, as it will save errors and complexity in later phases. Each administrator must decide whether to begin with an External Phase (enabling IPv6 for Internet-facing systems, as recommended in [RFC5211]) or an Internal Phase (enabling IPv6 for internal interconnections first).

Each scenario is likely to be different to some extent, but we can highlight some considerations:

- o In many cases, customers outside the network will have IPv6 before the internal enterprise network. For these customers, IPv6 may well perform better, especially for certain applications, than translated or tunneled IPv4, so the administrator may want to prioritize the External Phase such that those customers have the simplest and most robust connectivity to the enterprise, or at least its external-facing elements.
- o Employees who access internal systems by VPN may find that their ISPs provide translated IPv4, which does not support the required VPN protocols. In these cases, the administrator may want to prioritize the External Phase, and any other remotely-accessible internal systems. It is worth noting that a number of emerging VPN solutions provide dual-stack connectivity; thus a VPN service may be useful for employees in IPv4-only access networks to access IPv6 resources in the enterprise network (much like many public tunnel broker services, but specifically for the enterprise). Some security considerations are described in [I-D.ietf-opsec-vpn-leakages].
- o Internet-facing servers cannot be managed over IPv6 unless the management systems are IPv6-capable. These might be Network Management Systems (NMS), monitoring systems, or just remote management desktops. Thus in some cases, the Internet-facing systems are dependent on IPv6-capable internal networks. However, dual-stack Internet-facing systems can still be managed over IPv4.

- o Virtual machines may enable a faster rollout once initial system deployment is complete. Management of VMs over IPv6 is still dependent on the management software supporting IPv6.
- o IPv6 is enabled by default on all modern operating systems, so it may be more urgent to manage and have visibility on the internal traffic. It is important to manage IPv6 for security purposes, even in an ostensibly IPv4-only network, as described in [RFC7123].
- o In many cases, the corporate accounting, payroll, human resource, and other internal systems may only need to be reachable from the internal network, so they may be a lower priority. As enterprises require their vendors to support IPv6, more internal applications will support IPv6 by default and it can be expected that eventually new applications will only support IPv6. The inventory, as described in Section 2.2, will help determine the systems' readiness, as well as the readiness of the supporting network elements and security, which will be a consideration in prioritization of these corporate systems.
- o Some large organizations (even when using private IPv4 addresses[RFC1918]) are facing IPv4 address exhaustion because of the internal network growth (for example the vast number of virtual machines) or because of the acquisition of other companies that often raise private IPv4 address overlapping issues.
- o IPv6 restores end to end transparency even for internal applications (of course security policies must still be enforced). When two organizations or networks merge [RFC6879], the unique addressing of IPv6 can make the merger much easier and faster. A merger may, therefore, prioritize IPv6 for the affected systems.

These considerations are in conflict; each administrator must prioritize according to their company's conditions. It is worth noting that the reasons given in one "Large Corporate User's View of IPng", described in [RFC1687], for reluctance to deploy have largely been satisfied or overcome in the intervening years.

2. Preparation and Assessment Phase

2.1. Program Planning

Since enabling IPv6 is a change to the most fundamental Internet Protocol, and since there are so many interdependencies, having a professional project manager organize the work is highly recommended. In addition, an executive sponsor should be involved in determining

the goals of enabling IPv6 (which will establish the order of the phases), and should receive regular updates.

It may be necessary to complete the Preparation Phase before determining whether to prioritize the Internal or External Phase, since needs and readiness assessments are part of that phase. For a large enterprise, it may take several iterations to really understand the level of effort required. Depending on the required schedule, it may be useful to roll IPv6 projects into other architectural upgrades--this can be an excellent way to improve the network and reduce costs. However, by increasing the scope of projects, the schedule is often affected. For instance, a major systems upgrade may take a year to complete, where just patching existing systems may take only a few months.

The deployment of IPv6 will not generally stop all other technology work. Once IPv6 has been identified as an important initiative, all projects, both new and in-progress, will need to be reviewed to ensure IPv6 support.

It is normal for assessments to continue in some areas while execution of the project begins in other areas. This is fine, as long as recommendations in other parts of this document are considered, especially regarding security (for instance, one should not deploy IPv6 on a system before security has been evaluated).

2.2. Inventory Phase

To comprehend the scope of the inventory phase we recommend dividing the problem space in two: network infrastructure readiness and applications readiness.

2.2.1. Network infrastructure readiness assessment

The goal of this assessment is to identify the level of IPv6 readiness of network equipment. This will identify the effort required to move to an infrastructure that supports IPv6 with the same functional service capabilities as the existing IPv4 network. This may also require a feature comparison and gap analysis between IPv4 and IPv6 functionality on the network equipment and software. IPv6 support will require testing; features often work differently in vendors' labs than production networks. Some devices and software will require IPv4 support for IPv6 to work.

The inventory will show which network devices are already capable, which devices can be made IPv6 ready with a code/firmware upgrade, and which devices will need to be replaced. The data collection consists of a network discovery to gain an understanding of the

topology and inventory network infrastructure equipment and code versions with information gathered from static files and IP address management, DNS and DHCP tools.

Since IPv6 might already be present in the environment, through default configurations or VPNs, an infrastructure assessment (at minimum) is essential to evaluate potential security risks.

2.2.2. Applications readiness assessment

Just like network equipment, application software needs to support IPv6. This includes OS, firmware, middleware and applications (including internally developed applications). Vendors will typically handle IPv6 enablement of off-the-shelf products, but often enterprises need to request this support from vendors. For internally developed applications it is the responsibility of the enterprise to enable them for IPv6. Analyzing how a given application communicates over the network will dictate the steps required to support IPv6. Applications should avoid instructions specific to a given IP address family. Any applications that use APIs, such as the C language, that expose the IP version specifically, need to be modified to also work with IPv6.

There are two ways to IPv6-enable applications. The first approach is to have separate logic for IPv4 and IPv6, thus leaving the IPv4 code path mainly untouched. This approach causes the least disruption to the existing IPv4 logic flow, but introduces more complexity, since the application now has to deal with two logic loops with complex race conditions and error recovery mechanisms between these two logic loops. The second approach is to create a combined IPv4/IPv6 logic, which ensures operation regardless of the IP version used on the network. Knowing whether a given implementation will use IPv4 or IPv6 in a given deployment is a matter of some art; see Source Address Selection [RFC6724] and Happy Eyeballs [RFC6555]. It is generally recommended that the application developer use industry IPv6-porting tools to locate the code that needs to be updated. Some discussion of IPv6 application porting issues can be found in [RFC4038].

2.2.3. Importance of readiness validation and testing

Lastly IPv6 introduces a completely new way of addressing endpoints, which can have ramifications at the network layer all the way up to the applications. So to minimize disruption during the transition phase we recommend complete functionality, scalability and security testing to understand how IPv6 impacts the services and networking infrastructure.

2.3. Training

Many organizations falter in IPv6 deployment because of a perceived training gap. Training is important for those who work with addresses regularly, as with anyone whose work is changing. Better knowledge of the reasons IPv6 is being deployed will help inform the assessment of who needs training, and what training they need.

2.4. Security Policy

It is obvious that IPv6 networks should be deployed in a secure way. The industry has learnt a lot about network security with IPv4, so, network operators should leverage this knowledge and expertise when deploying IPv6. IPv6 is not so different than IPv4: it is a connectionless network protocol using the same lower layer service and delivering the same service to the upper layer. Therefore, the security issues and mitigation techniques are mostly identical with same exceptions that are described further.

2.4.1. IPv6 is no more secure than IPv4

Some people believe that IPv6 is inherently more secure than IPv4 because it is new. Nothing can be more wrong. Indeed, being a new protocol means that bugs in the implementations have yet to be discovered and fixed and that few people have the operational security expertise needed to operate securely an IPv6 network. This lack of operational expertise is the biggest threat when deploying IPv6: the importance of training is to be stressed again.

One security myth is that thanks to its huge address space, a network cannot be scanned by enumerating all IPv6 address in a /64 LAN hence a malevolent person cannot find a victim. [RFC5157] describes some alternate techniques to find potential targets on a network, for example enumerating all DNS names in a zone. Additional advice in this area is also given in [I-D.ietf-opsec-ipv6-host-scanning].

Another security myth is that IPv6 is more secure because it mandates the use of IPsec everywhere. While the original IPv6 specifications may have implied this, [RFC6434] clearly states that IPsec support is not mandatory. Moreover, if all the intra-enterprise traffic is encrypted, both malefactors and security tools that rely on payload inspection (IPS, firewall, ACL, IPFIX ([RFC7011] and [RFC7012]), etc) will be thwarted. Therefore, IPsec is as useful in IPv6 as in IPv4 (for example to establish a VPN overlay over a non-trusted network or reserved for some specific applications).

The last security myth is that amplification attacks (such as [SMURF]) do not exist in IPv6 because there is no more broadcast.

Alas, this is not true as ICMP error (in some cases) or information messages can be generated by routers and hosts when forwarding or receiving a multicast message (see Section 2.4 of [RFC4443]). Therefore, the generation and the forwarding rate of ICMPv6 messages must be limited as in IPv4.

It should be noted that in a dual-stack network the security implementation for both IPv4 and IPv6 needs to be considered, in addition to security considerations related to the interaction of (and transition between) the two, while they coexist.

2.4.2. Similarities between IPv6 and IPv4 security

As mentioned earlier, IPv6 is quite similar to IPv4, therefore several attacks apply for both protocol families, including:

- o Application layer attacks: such as cross-site scripting or SQL injection
- o Rogue device: such as a rogue Wi-Fi Access Point
- o Flooding and all traffic-based denial of services (including the use of control plane policing for IPv6 traffic see [RFC6192])

A specific case of congruence is IPv6 Unique Local Addresses (ULAs) [RFC4193] and IPv4 private addressing [RFC1918], which do not provide any security by 'magic'. In both cases, the edge router must apply strict filters to block those private addresses from entering and, just as importantly, leaving the network. This filtering can be done by the enterprise or by the ISP, but the cautious administrator will prefer to do it in the enterprise.

IPv6 addresses can be spoofed as easily as IPv4 addresses and there are packets with bogon IPv6 addresses (see [CYMRU]). Anti-bogon filtering must be done in the data and routing planes. It can be done by the enterprise or by the ISP, or both, but again the cautious administrator will prefer to do it in the enterprise.

2.4.3. Specific Security Issues for IPv6

Even if IPv6 is similar to IPv4, there are some differences that create some IPv6-only vulnerabilities or issues. We give examples of such differences in this section.

Privacy extension addresses [RFC4941] are usually used to protect individual privacy by periodically changing the interface identifier part of the IPv6 address to avoid tracking a host by its otherwise always identical and unique MAC-based EUI-64. While this presents a

real advantage on the Internet, moderated by the fact that the prefix part remains the same, it complicates the task of following an audit trail when a security officer or network operator wants to trace back a log entry to a host in their network, because when the tracing is done the searched IPv6 address could have disappeared from the network. Therefore, the use of privacy extension addresses usually requires additional monitoring and logging of the binding of the IPv6 address to a data-link layer address (see also the monitoring section of [I-D.ietf-opsec-v6]). Some early enterprise deployments have taken the approach of using tools that harvest IP/MAC address mappings from switch and router devices to provide address accountability; this approach has been shown to work, though it can involve gathering significantly more address data than in equivalent IPv4 networks. An alternative is to try to prevent the use of privacy extension addresses by enforcing the use of DHCPv6, such that hosts only get addresses assigned by a DHCPv6 server. This can be done by configuring routers to set the M-bit in Router Advertisements, combined with all advertised prefixes being included without the A-bit set (to prevent the use of stateless auto-configuration). This technique of course requires that all hosts support stateful DHCPv6. It is important to note that not all operating systems exhibit the same behavior when processing RAs with the M-Bit set. The varying OS behavior is related to the lack of prescriptive definition around the A, M and O-bits within the ND protocol. [I-D.liu-bonica-dhcpv6-slaac-problem] provides a much more detailed analysis on the interaction of the M-Bit and DHCPv6.

Extension headers complicate the task of stateless packet filters such as ACLs. If ACLs are used to enforce a security policy, then the enterprise must verify whether its ACL (but also stateful firewalls) are able to process extension headers (this means understand them enough to parse them to find the upper layers payloads) and to block unwanted extension headers (e.g., to implement [RFC5095]). This topic is discussed further in [RFC7045].

Fragmentation is different in IPv6 because it is done only by source host and never during a forwarding operation. This means that ICMPv6 packet-too-big messages must be allowed to pass through the network and not be filtered [RFC4890]. Fragments can also be used to evade some security mechanisms such as RA-guard [RFC6105]. See also [RFC5722], and [RFC7113].

One of the biggest differences between IPv4 and IPv6 is the introduction of the Neighbor Discovery Protocol [RFC4861], which includes a variety of important IPv6 protocol functions, including those provided in IPv4 by ARP [RFC0826]. NDP runs over ICMPv6 (which as stated above means that security policies must allow some ICMPv6 messages to pass, as described in RFC 4890), but has the same lack of

security as, for example, ARP, in that there is no inherent message authentication. While Secure Neighbour Discovery (SeND) [RFC3971] and CGA [RFC3972] have been defined, they are not widely implemented). The threat model for Router Advertisements within the NDP suite is similar to that of DHCPv4 (and DHCPv6), in that a rogue host could be either a rogue router or a rogue DHCP server. An IPv4 network can be made more secure with the help of DHCPv4 snooping in edge switches, and likewise RA snooping can improve IPv6 network security (in IPv4-only networks as well). Thus enterprises using such techniques for IPv4 should use the equivalent techniques for IPv6, including RA-guard [RFC6105] and all work in progress from the SAVI WG, e.g. [RFC6959], which is similar to the protection given by dynamic ARP monitoring in IPv4. Other DoS vulnerabilities are related to NDP cache exhaustion, and mitigation techniques can be found in ([RFC6583]).

As stated previously, running a dual-stack network doubles the attack exposure as a malevolent person has now two attack vectors: IPv4 and IPv6. This simply means that all routers and hosts operating in a dual-stack environment with both protocol families enabled (even if by default) must have a congruent security policy for both protocol versions. For example, permit TCP ports 80 and 443 to all web servers and deny all other ports to the same servers must be implemented both for IPv4 and IPv6. It is thus important that the tools available to administrators readily support such behaviour.

2.5. Routing

An important design choice to be made is what IGP to use inside the network. A variety of IGPs (IS-IS, OSPFv3 and RIPng) support IPv6 today and picking one over the other is a design choice that will be dictated mostly by existing operational policies in an enterprise network. As mentioned earlier, it would be beneficial to maintain operational parity between IPv4 and IPv6 and therefore it might make sense to continue using the same protocol family that is being used for IPv4. For example, in a network using OSPFv2 for IPv4, it might make sense to use OSPFv3 for IPv6. It is important to note that although OSPFv3 is similar to OSPFv2, they are not the same. On the other hand, some organizations may chose to run different routing protocols for different IP versions. For example, one may chose to run OSPFv2 for IPv4 and IS-IS for IPv6. An important design question to consider here is whether to support one IGP or two different IGPs in the longer term. [I-D.ietf-v6ops-design-choices] presents advice on the design choices that arise when considering IGPs and discusses the advantages and disadvantages to different approaches in detail.

2.6. Address Plan

The most common problem encountered in IPv6 networking is in applying the same principles of conservation that are so important in IPv4. IPv6 addresses do not need to be assigned conservatively. In fact, a single larger allocation is considered more conservative than multiple non-contiguous small blocks, because a single block occupies only a single entry in a routing table. The advice in [RFC5375] is still sound, and is recommended to the reader. If considering ULAs, give careful thought to how well it is supported, especially in multiple address and multicast scenarios, and assess the strength of the requirement for ULA. [I-D.ietf-v6ops-ula-usage-recommendations] provides much more detailed analysis and recommendations on the usage of ULAs.

The enterprise administrator will want to evaluate whether the enterprise will request address space from a LIR (Local Internet Registry, such as an ISP), a RIR (Regional Internet Registry, such as AfriNIC, APNIC, ARIN, LACNIC, or RIPE-NCC) or a NIR (National Internet Registry, operated in some countries). The normal allocation is Provider Aggregatable (PA) address space from the enterprise's ISP, but use of PA space implies renumbering when changing provider. Instead, an enterprise may request Provider Independent (PI) space; this may involve an additional fee, but the enterprise may then be better able to be multihomed using that prefix, and will avoid a renumbering process when changing ISPs (though it should be noted that renumbering caused by outgrowing the space, merger, or other internal reason would still not be avoided with PI space).

The type of address selected (PI vs. PA) should be congruent with the routing needs of the enterprise. The selection of address type will determine if an operator will need to apply new routing techniques and may limit future flexibility. There is no right answer, but the needs of the external phase may affect what address type is selected.

Each network location or site will need a prefix assignment. Depending on the type of site/location, various prefix sizes may be used. In general, historical guidance suggests that each site should get at least a /48, as documented in RFC 5375 and [RFC6177]. In addition to allowing for simple planning, this can allow a site to use its prefix for local connectivity, should the need arise, and if the local ISP supports it.

When assigning addresses to end systems, the enterprise may use manually-configured addresses (common on servers) or SLAAC or DHCPv6 for client systems. Early IPv6 enterprise deployments have used SLAAC, both for its simplicity but also due to the time DHCPv6 has

taken to mature. However, DHCPv6 is now very mature, and thus workstations managed by an enterprise may use stateful DHCPv6 for addressing on corporate LAN segments. DHCPv6 allows for the additional configuration options often employed by enterprise administrators, and by using stateful DHCPv6, administrators correlating system logs know which system had which address at any given time. Such an accountability model is familiar from IPv4 management, though for DHCPv6 hosts are identified by DUID rather than MAC address. For equivalent accountability with SLAAC (and potentially privacy addresses), a monitoring system that harvests IP/MAC mappings from switch and router equipment could be used.

A common deployment consideration for any enterprise network is how to get host DNS records updated. Commonly, either the host will send DNS updates or the DHCP server will update records. If there is sufficient trust between the hosts and the DNS server, the hosts may update (and the enterprise may use SLAAC for addressing). Otherwise, the DHCPv6 server can be configured to update the DNS server. Note that an enterprise network with this more controlled environment will need to disable SLAAC on network segments and force end hosts to use DHCPv6 only.

In the data center or server room, assume a /64 per VLAN. This applies even if each individual system is on a separate VLAN. In a /48 assignment, typical for a site, there are then still 65,535 /64 blocks. Some administrators reserve a /64 but configure a small subnet, such as /112, /126, or /127, to prevent rogue devices from attaching and getting numbers; an alternative is to monitor traffic for surprising addresses or ND tables for new entries. Addresses are either configured manually on the server, or reserved on a DHCPv6 server, which may also synchronize forward and reverse DNS (though see [RFC6866] for considerations on static addressing). SLAAC is not recommended for servers, because of the need to synchronize RA timers with DNS TTLs so that the DNS entry expires at the same time as the address.

All user access networks should be a /64. Point-to-point links where Neighbor Discovery Protocol is not used may also utilize a /127 (see [RFC6164]).

Plan to aggregate at every layer of network hierarchy. There is no need for VLSM [RFC1817] in IPv6, and addressing plans based on conservation of addresses are short-sighted. Use of prefixes longer than /64 on network segments will break common IPv6 functions such as SLAAC[RFC4862]. Where multiple VLANs or other layer two domains converge, allow some room for expansion. Renumbering due to outgrowing the network plan is a nuisance, so allow room within it. Generally, plan to grow to about twice the current size that can be

accommodated; where rapid growth is planned, allow for twice that growth. Also, if DNS (or reverse DNS) authority may be delegated to others in the enterprise, assignments need to be on nibble boundaries (that is, on a multiple of 4 bits, such as /64, /60, /56, ..., /48, /44), to ensure that delegated zones align with assigned prefixes.

If using ULAs, it is important to note that AAAA and PTR records for ULA are not recommended to be installed in the global DNS. Similarly, reverse (address-to-name) queries for ULA must not be sent to name servers outside of the organization, due to the load that such queries would create for the authoritative name servers for the ip6.arpa zone. For more details please refer to section 4.4 of [RFC4193].

Enterprise networks more and more include virtual networks where a single physical node may host many virtualized addressable devices. It is imperative that the addressing plans assigned to these virtual networks and devices be consistent and non-overlapping with the addresses assigned to real networks and nodes. For example, a virtual network established within an isolated lab environment may at a later time become attached to the production enterprise network.

2.7. Tools Assessment

Enterprises will often have a number of operational tools and support systems which are used to provision, monitor, manage and diagnose the network and systems within their environment. These tools and systems will need to be assessed for compatibility with IPv6. The compatibility may be related to the addressing and connectivity of various devices as well as IPv6 awareness of the tools and processing logic.

The tools within the organization fall into two general categories, those which focus on managing the network, and those which are focused on managing systems and applications on the network. In either instance, the tools will run on platforms which may or may not be capable of operating in an IPv6 network. This lack in functionality may be related to Operating System version, or based on some hardware constraint. Those systems which are found to be incapable of utilizing an IPv6 connection, or which are dependent on an IPv4 stack, may need to be replaced or upgraded.

In addition to devices working on an IPv6 network natively, or via a transition tunnel, many tools and support systems may require additional software updates to be IPv6 aware, or even a hardware upgrade (usually for additional memory: IPv6 addresses are larger and for a while, IPv4 and IPv6 addresses will coexist in the tool). This awareness may include the ability to manage IPv6 elements and/or

applications in addition to the ability to store and utilize IPv6 addresses.

Considerations when assessing the tools and support systems may include the fact that IPv6 addresses are significantly larger than IPv4, requiring data stores to support the increased size. Such issues are among those discussed in [RFC5952]. Many organizations may also run dual-stack networks, therefore the tools need to not only support IPv6 operation, but may also need to support the monitoring, management and intersection with both IPv6 and IPv4 simultaneously. It is important to note that managing IPv6 is not just constrained to using large IPv6 addresses, but also that IPv6 interfaces and nodes are likely to use two or more addresses as part of normal operation. Updating management systems to deal with these additional nuances will likely consume time and considerable effort.

For networking systems, like node management systems, it is not always necessary to support local IPv6 addressing and connectivity. Operations such as SNMP MIB polling can occur over IPv4 transport while seeking responses related to IPv6 information. Where this may seem advantageous to some, it should be noted that without local IPv6 connectivity, the management system may not be able to perform all expected functions - such as reachability and service checks.

Organizations should be aware that changes to older IPv4-only SNMP MIB specifications have been made by the IETF related to legacy operation in [RFC2096] and [RFC2011]. Updated specifications are now available in [RFC4292] and [RFC4293] which modified the older MIB framework to be IP protocol agnostic, supporting both IPv4 and IPv6. Polling systems will need to be upgraded to support these updates as well as the end stations which are polled.

3. External Phase

The external phase for enterprise IPv6 adoption covers topics which deal with how an organization connects its infrastructure to the external world. These external connections may be toward the Internet at large, or to other networks. The external phase covers connectivity, security and monitoring of various elements and outward facing or accessible services.

3.1. Connectivity

The enterprise will need to work with one or more Service Providers to gain connectivity to the Internet or transport service infrastructure such as a BGP/MPLS IP VPN as described in [RFC4364] and [RFC4659]. One significant factor that will guide how an organization may need to communicate with the outside world will

involve the use of PI (Provider Independent) and/or PA (Provider Aggregatable) IPv6 space.

Enterprises should be aware that depending on which address type they selected (PI vs. PA) in their planning phase, they may need to implement new routing functions and/or behaviours to support their connectivity to the ISP. In the case of PI, the upstream ISP may offer options to route the prefix (typically a /48) on the enterprise's behalf and update the relevant routing databases. Otherwise, the enterprise may need to perform this task on their own and use BGP to inject the prefix into the global BGP system.

Note that the rules set by the RIRs for an enterprise acquiring PI address space have changed over time. For example, in the European region the RIPE-NCC no longer requires an enterprise to be multihomed to be eligible for an IPv6 PI allocation. Requests can be made directly or via a LIR. It is possible that the rules may change again, and may vary between RIRs.

When seeking IPv6 connectivity to a Service Provider, Native IPv6 connectivity is preferred since it provides the most robust and efficient form of connectivity. If native IPv6 connectivity is not possible due to technical or business limitations, the enterprise may utilize readily available transition tunnel IPv6 connectivity. There are IPv6 transit providers which provide robust tunnelled IPv6 connectivity which can operate over IPv4 networks. It is important to understand the transition tunnel mechanism used, and to consider that it will have higher latency than native IPv4 or IPv6, and may have other problems, e.g. related to MTUs.

It is important to evaluate MTU considerations when adding IPv6 to an existing IPv4 network. It is generally desirable to have the IPv6 and IPv4 MTU congruent to simplify operations (so the two address families behave similarly, that is, as expected). If the enterprise uses transition tunnels inside or externally for IPv6 connectivity, then modification of the MTU on hosts/routers may be needed as mid-stream fragmentation is no longer supported in IPv6. It is preferred that pMTUD is used to optimize the MTU, so erroneous filtering of the related ICMPv6 message types should be monitored. Adjusting the MTU may be the only option if undesirable upstream ICMPv6 filtering cannot be removed.

3.2. Security

The most important part of security for external IPv6 deployment is filtering and monitoring. Filtering can be done by stateless ACLs or a stateful firewall. The security policies must be consistent for IPv4 and IPv6 (else the attacker will use the less protected protocol

stack), except that certain ICMPv6 messages must be allowed through and to the filtering device (see [RFC4890]):

- o Packet Too Big: essential to allow Path MTU discovery to work
- o Parameter Problem
- o Time Exceeded

In addition, Neighbor Discovery Protocol messages (including Neighbor Solicitation, Router Advertisements, etc.) are required for local hosts.

It could also be safer to block all fragments where the transport layer header is not in the first fragment to avoid attacks as described in [RFC5722]. Some filtering devices allow this filtering. Ingress filters and firewalls should follow [RFC5095] in handling routing extension header type 0, dropping the packet and sending ICMPv6 Parameter Problem, unless Segments Left = 0 (in which case, ignore the header).

If an Intrusion Prevention System (IPS) is used for IPv4 traffic, then an IPS should also be used for IPv6 traffic. In general, make sure IPv6 security is at least as good as IPv4. This also includes all email content protection (anti-spam, content filtering, data leakage prevention, etc.).

The edge router must also implement anti-spoofing techniques based on [RFC2827] (also known as BCP 38).

In order to protect the networking devices, it is advised to implement control plane policing as per [RFC6192].

The potential NDP cache exhaustion attack (see [RFC6583]) can be mitigated by two techniques:

- o Good NDP implementation with memory utilization limits as well as rate-limiters and prioritization of requests.
- o Or, as the external deployment usually involves just a couple of exposed statically configured IPv6 addresses (virtual addresses of web, email, and DNS servers), then it is straightforward to build an ingress ACL allowing traffic for those addresses and denying traffic to any other addresses. This actually prevents the attack as a packet for a random destination will be dropped and will never trigger a neighbor resolution.

3.3. Monitoring

Monitoring the use of the Internet connectivity should be done for IPv6 as it is done for IPv4. This includes the use of IP Flow Information eXport (IPFIX) [RFC7012] to report abnormal traffic patterns (such as port scanning, SYN-flooding, related IP source addresses) from monitoring tools and evaluating data read from SNMP MIBs [RFC4293] (some of which also enable the detection of abnormal bandwidth utilization) and syslogs (finding server and system errors). Where Netflow is used, version 9 is required for IPv6 support. Monitoring systems should be able to examine IPv6 traffic, use IPv6 for connectivity, record IPv6 address, and any log parsing tools and reporting need to support IPv6. Some of this data can be sensitive (including personally identifiable information) and care in securing it should be taken, with periodic purges. Integrity protection on logs and sources of log data is also important to detect unusual behavior (misconfigurations or attacks). Logs may be used in investigations, which depend on trustworthy data sources (tamper resistant).

In addition, monitoring of external services (such as web sites) should be made address-specific, so that people are notified when either the IPv4 or IPv6 version of a site fails.

3.4. Servers and Applications

The path to the servers accessed from the Internet usually involves security devices (firewall, IPS), server load balancing (SLB) and real physical servers. The latter stage is also multi-tiered for scalability and security between presentation and data storage. The ideal transition is to enable native dual-stack on all devices; but as part of the phased approach, operators have used the following techniques with success:

- o Use a network device to apply NAT64 and basically translate an inbound TCP connection (or any other transport protocol) over IPv6 into a TCP connection over IPv4. This is the easiest to deploy as the path is mostly unchanged but it hides all IPv6 remote users behind a single IPv4 address which leads to several audit trail and security issues (see [RFC6302]).
- o Use the server load balancer which acts as an application proxy to do this translation. Compared to the NAT64, it has the potential benefit of going through the security devices as native IPv6 (so more audit and trace abilities) and is also able to insert a HTTP X-Forward-For header which contains the remote IPv6 address. The latter feature allows for logging, and rate-limiting on the real

servers based on the IPV6 address even if those servers run only IPv4.

In either of these cases, care should be taken to secure logs for privacy reasons, and to periodically purge them.

3.5. Network Prefix Translation for IPv6

Network Prefix Translation for IPv6, or NPTv6 as described in [RFC6296] provides a framework to utilize prefix ranges within the internal network which are separate (address-independent) from the assigned prefix from the upstream provider or registry. As mentioned above, while NPTv6 has potential use-cases in IPv6 networks, the implications of its deployment need to be fully understood, particularly where any applications might embed IPv6 addresses in their payloads.

Use of NPTv6 can be chosen independently from how addresses are assigned and routed within the internal network, how prefixes are routed towards the Internet, or whether PA or PI addresses are used.

4. Internal Phase

This phase deals with the delivery of IPv6 to the internal user-facing side of the IT infrastructure, which comprises various components such as network devices (routers, switches, etc.), end user devices and peripherals (workstations, printers, etc.), and internal corporate systems.

An important design paradigm to consider during this phase is "dual-stack when you can, tunnel when you must". Dual-stacking allows a more robust, production-quality IPv6 network than is typically facilitated by internal use of transition tunnels that are harder to troubleshoot and support, and that may introduce scalability and performance issues. Tunnels may of course still be used in production networks, but their use needs to be carefully considered, e.g. where the transition tunnel may be run through a security or filtering device. Tunnels do also provide a means to experiment with IPv6 and gain some operational experience with the protocol. [RFC4213] describes various transition mechanisms in more detail. [RFC6964] suggests operational guidance when using ISATAP tunnels [RFC5214], though we would recommend use of dual-stack wherever possible.

4.1. Security

IPv6 must be deployed in a secure way. This means that all existing IPv4 security policies must be extended to support IPv6; IPv6 security policies will be the IPv6 equivalent of the existing IPv4 ones (taking into account the difference for ICMPv6 [RFC4890]). As in IPv4, security policies for IPv6 will be enforced by firewalls, ACL, IPS, VPN, and so on.

Privacy extension addresses [RFC4941] raise a challenge for an audit trail as explained in section Section 2.4.3. The enterprise may choose to attempt to enforce use of DHCPv6, or deploy monitoring tools that harvest accountability data from switches and routers (thus making the assumption that devices may use any addresses inside the network).

One major issue is threats against Neighbor Discovery. This means, for example, that the internal network at the access layer (where hosts connect to the network over wired or wireless) should implement RA-guard [RFC6105] and the techniques being specified by SAVI WG [RFC6959]; see also Section 2.4.3 for more information.

4.2. Network Infrastructure

The typical enterprise network infrastructure comprises a combination of the following network elements - wired access switches, wireless access points, and routers (although it is fairly common to find hardware that collapses switching and routing functionality into a single device). Basic wired access switches and access points operate only at the physical and link layers, and don't really have any special IPv6 considerations other than being able to support IPv6 addresses themselves for management purposes. In many instances, these devices possess a lot more intelligence than simply switching packets. For example, some of these devices help assist with link layer security by incorporating features such as ARP inspection and DHCP Snooping, or they may help limit where multicast floods by using IGMP (or, in the case of IPv6, MLD) snooping.

Another important consideration in enterprise networks is first hop router redundancy. This directly ties into network reachability from an end host's point of view. IPv6 Neighbor Discovery (ND), [RFC4861], provides a node with the capability to maintain a list of available routers on the link, in order to be able to switch to a backup path should the primary be unreachable. By default, ND will detect a router failure in 38 seconds and cycle onto the next default router listed in its cache. While this feature provides a basic level of first hop router redundancy, most enterprise IPv4 networks are designed to fail over much faster. Although this delay can be

improved by adjusting the default timers, care must be taken to protect against transient failures and to account for increased traffic on the link. Another option to provide robust first hop redundancy is to use the Virtual Router Redundancy Protocol for IPv6 (VRRPv3), [RFC5798]. This protocol provides a much faster switchover to an alternate default router than default ND parameters. Using VRRPv3, a backup router can take over for a failed default router in around three seconds (using VRRPv3 default parameters). This is done without any interaction with the hosts and a minimum amount of VRRP traffic.

Last but not the least, one of the most important design choices to make while deploying IPv6 on the internal network is whether to use Stateless Automatic Address Configuration (SLAAC), [RFC4862], or Dynamic Host Configuration Protocol for IPv6 (DHCPv6), [RFC3315], or a combination thereof. Each option has advantages and disadvantages, and the choice will ultimately depend on the operational policies that guide each enterprise's network design. For example, if an enterprise is looking for ease of use, rapid deployment, and less administrative overhead, then SLAAC makes more sense for workstations. Manual or DHCPv6 assignments are still needed for servers, as described in the External Phase and Address Plan sections of this document. However, if the operational policies call for precise control over IP address assignment for auditing then DHCPv6 may be preferred. DHCPv6 also allows you to tie into DNS systems for host entry updates and gives you the ability to send other options and information to clients. It is worth noting that in general operation RAs are still needed in DHCPv6 networks, as there is no DHCPv6 Default Gateway option. Similarly, DHCPv6 is needed in RA networks for other configuration information, e.g. NTP servers or, in the absence of support for DNS resolvers in RAs [RFC6106], DNS resolver information.

4.3. End user devices

Most operating systems (OSes) that are loaded on workstations and laptops in a typical enterprise support IPv6 today. However, there are various out-of-the-box nuances that one should be mindful about. For example, the default behavior of OSes vary; some may have IPv6 turned off by default, some may only have certain features such as privacy extensions to IPv6 addresses (RFC 4941) turned off while others have IPv6 fully enabled. Further, even when IPv6 is enabled, the choice of which address is used may be subject to Source Address Selection (RFC 6724) and Happy Eyeballs (RFC 6555). Therefore, it is advised that enterprises investigate the default behavior of their installed OS base and account for it during the Inventory phases of their IPv6 preparations. Furthermore, some OSes may have some transition tunneling mechanisms turned on by default and in such

cases it is recommended to administratively shut down such interfaces unless required.

It is important to note that it is recommended that IPv6 be deployed at the network and system infrastructure level before it is rolled out to end user devices; ensure IPv6 is running and routed on the wire, and secure and correctly monitored, before exposing IPv6 to end users.

Smartphones and tablets are significant IPv6-capable platforms, depending on the support of the carrier's data network.

IPv6 support for peripherals varies. Much like servers, printers are generally configured with a static address (or DHCP reservation) so clients can discover them reliably.

4.4. Corporate Systems

No IPv6 deployment will be successful without ensuring that all the corporate systems that an enterprise uses as part of its IT infrastructure support IPv6. Examples of such systems include, but are not limited to, email, video conferencing, telephony (VoIP), DNS, RADIUS, etc. All these systems must have their own detailed IPv6 rollout plan in conjunction with the network IPv6 rollout. It is important to note that DNS is one of the main anchors in an enterprise deployment, since most end hosts decide whether or not to use IPv6 depending on the presence of IPv6 AAAA records in a reply to a DNS query. It is recommended that system administrators selectively turn on AAAA records for various systems as and when they are IPv6 enabled; care must be taken though to ensure all services running on that host name are IPv6-enabled before adding the AAAA record. Care with web proxies is advised; a mismatch in the level of IPv6 support between the client, proxy, and server can cause communication problems. All monitoring and reporting tools across the enterprise will need to be modified to support IPv6.

5. IPv6-only

Early IPv6 enterprise deployments have generally taken a dual-stack approach to enabling IPv6, i.e. the existing IPv4 services have not been turned off. Although IPv4 and IPv6 networks will coexist for a long time, the long term enterprise network roadmap should include steps to simplify engineering and operations by deprecating IPv4 from the dual-stack network. In some extreme cases, deploying dual-stack networks may not even be a viable option for very large enterprises due to the RFC 1918 address space not being large enough to support the network's growth. In such cases, deploying IPv6-only networks might be the only choice available to sustain network growth. In

other cases, there may be elements of an otherwise dual-stack network that may be run IPv6-only.

If nodes in the network don't need to talk to an IPv4-only node, then deploying IPv6-only networks should be straightforward. However, most nodes will need to communicate with some IPv4-only nodes; an IPv6-only node may therefore require a translation mechanism. As [RFC6144] points out, it is important to look at address translation as a transition strategy towards running an IPv6-only network.

There are various stateless and stateful IPv4/IPv6 translation methods available today that help IPv6 to IPv4 communication. RFC 6144 provides a framework for IPv4/IPv6 translation and describes in detail various scenarios in which such translation mechanisms could be used. [RFC6145] describes stateless address translation. In this mode, a specific IPv6 address range will represent IPv4 systems (IPv4-converted addresses), and the IPv6 systems have addresses (IPv4-translatable addresses) that can be algorithmically mapped to a subset of the service provider's IPv4 addresses. [RFC6146], NAT64, describes stateful address translation. As the name suggests, the translation state is maintained between IPv4 address/port pairs and IPv6 address/port pairs, enabling IPv6 systems to open sessions with IPv4 systems. [RFC6147], DNS64, describes a mechanism for synthesizing AAAA resource records (RRs) from A RRs. Together, RFCs 6146 and RFC 6147 provide a viable method for an IPv6-only client to initiate communications to an IPv4-only server. As described in the assumptions section, the administrator will usually want most traffic or flows to be native, and only translate as needed.

The address translation mechanisms for the stateless and stateful translations are defined in [RFC6052]. It is important to note that both of these mechanisms have limitations as to which protocols they support. For example, RFC 6146 only defines how stateful NAT64 translates unicast packets carrying TCP, UDP, and ICMP traffic only. The classic problems of IPv4 NAT also apply, e.g. handling IP literals in application payloads. The ultimate choice of which translation mechanism to choose will be dictated mostly by existing operational policies pertaining to application support, logging requirements, etc.

There is additional work being done in the area of address translation to enhance and/or optimize current mechanisms. For example, [I-D.xli-behave-divi] describes limitations with the current stateless translation, such as IPv4 address sharing and application layer gateway (ALG) problems, and presents the concept and implementation of dual-stateless IPv4/IPv6 translation (dIVI) to address those issues.

It is worth noting that for IPv6-only access networks that use technologies such as NAT64, the more content providers (and enterprises) that make their content available over IPv6, the less the requirement to apply NAT64 to traffic leaving the access network. This particular point is important for enterprises which may start their IPv6 deployment well into the global IPv6 transition. As time progresses, and given the current growth in availability of IPv6 content, IPv6-only operation using NAT64 to manage some flows will become less expensive to run versus the traditional NAT44 deployments since only IPv6 to IPv4 flows need translation. [RFC6883] provides guidance and suggestions for Internet Content Providers and Application Service Providers in this context.

Enterprises should also be aware that networks may be subject to future convergence with other networks (i.e. mergers, acquisitions, etc). An enterprise considering IPv6-only operation may need to be aware that additional transition technologies and/or connectivity strategies may be required depending on the level of IPv6 readiness and deployment in the merging networking.

6. Considerations For Specific Enterprises

6.1. Content Delivery Networks

Some guidance for Internet Content and Application Service Providers can be found in [RFC6883], which includes a dedicated section on Content Delivery Networks (CDNs). An enterprise that relies on a CDN to deliver a 'better' e-commerce experience needs to ensure that their CDN provider also supports IPv4/IPv6 traffic selection so that they can ensure 'best' access to the content. A CDN could enable external IPv6 content delivery even if the enterprise provides that content over IPv4.

6.2. Data Center Virtualization

IPv6 Data Center considerations are described in [I-D.ietf-v6ops-dc-ipv6].

6.3. University Campus Networks

A number of campus networks around the world have made some initial IPv6 deployment. This has been encouraged by their National Research and Education Network (NREN) backbones having made IPv6 available natively since the early 2000's. Universities are a natural place for IPv6 deployment to be considered at an early stage, perhaps compared to other enterprises, as they are involved by their very nature in research and education.

Campus networks can deploy IPv6 at their own pace; there is no need to deploy IPv6 across the entire enterprise from day one, rather specific projects can be identified for an initial deployment, that are both deep enough to give the university experience, but small enough to be a realistic first step. There are generally three areas in which such deployments are currently made.

In particular those initial areas commonly approached are:

- o External-facing services. Typically the campus web presence and commonly also external-facing DNS and MX services. This ensures early IPv6-only adopters elsewhere can access the campus services as simply and as robustly as possible.
- o Computer science department. This is where IPv6-related research and/or teaching is most likely to occur, and where many of the next generation of network engineers are studying, so enabling some or all of the campus computer science department network is a sensible first step.
- o The eduroam wireless network. Eduroam [I-D.wierenga-ietf-eduroam] is the de facto wireless roaming system for academic networks, and uses 802.1X-based authentication, which is agnostic to the IP version used (unlike web-redirection gateway systems). Making a campus' eduroam network dual-stack is a very viable early step.

The general IPv6 deployment model in a campus enterprise will still follow the general principles described in this document. While the above early stage projects are commonly followed, these still require the campus to acquire IPv6 connectivity and address space from their NREN (or other provider in some parts of the world), and to enable IPv6 on the wire on at least part of the core of the campus network. This implies a requirement to have an initial address plan, and to ensure appropriate monitoring and security measures are in place, as described elsewhere in this document.

Campuses which have deployed to date do not use ULAs, nor do they use NPTv6. In general, campuses have very stable PA-based address allocations from their NRENs (or their equivalent). However, campus enterprises may consider applying for IPv6 PI; some have already done so. The discussions earlier in this text about PA vs. PI still apply.

Finally, campuses may be more likely than many other enterprises to run multicast applications, such as IP TV or live lecture or seminar streaming, so may wish to consider support for specific IPv6 multicast functionality, e.g. Embedded-RP [RFC3956] in routers and MLDv1 and MLDv2 snooping in switches.

7. Security Considerations

This document has multiple security sections detailing how to securely deploy an IPv6 network within an enterprise network.

8. Acknowledgements

The authors would like to thank Robert Sparks, Steve Hanna, Tom Taylor, Brian Haberman, Stephen Farrell, Chris Grundemann, Ray Hunter, Kathleen Moriarty, Benoit Claise, Brian Carpenter, Tina Tsou, Christian Jaquet, and Fred Templin for their substantial comments and contributions.

9. IANA Considerations

There are no IANA considerations or implications that arise from this document.

10. Informative References

- [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.
- [RFC1687] Fleischman, E., "A Large Corporate User's View of IPng", RFC 1687, August 1994.
- [RFC1817] Rekhter, Y., "CIDR and Classful Routing", RFC 1817, August 1995.
- [RFC1918] Rekhter, Y., Moskowitz, R., Karrenberg, D., Groot, G., and E. Lear, "Address Allocation for Private Internets", BCP 5, RFC 1918, February 1996.
- [RFC2011] McCloghrie, K., "SNMPv2 Management Information Base for the Internet Protocol using SMIV2", RFC 2011, November 1996.
- [RFC2096] Baker, F., "IP Forwarding Table MIB", RFC 2096, January 1997.
- [RFC2827] Ferguson, P. and D. Senie, "Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing", BCP 38, RFC 2827, May 2000.

- [RFC3315] Droms, R., Bound, J., Volz, B., Lemon, T., Perkins, C., and M. Carney, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", RFC 3315, July 2003.
- [RFC3956] Savola, P. and B. Haberman, "Embedding the Rendezvous Point (RP) Address in an IPv6 Multicast Address", RFC 3956, November 2004.
- [RFC3971] Arkko, J., Kempf, J., Zill, B., and P. Nikander, "SEcure Neighbor Discovery (SEND)", RFC 3971, March 2005.
- [RFC3972] Aura, T., "Cryptographically Generated Addresses (CGA)", RFC 3972, March 2005.
- [RFC4038] Shin, M-K., Hong, Y-G., Hagino, J., Savola, P., and E. Castro, "Application Aspects of IPv6 Transition", RFC 4038, March 2005.
- [RFC4057] Bound, J., "IPv6 Enterprise Network Scenarios", RFC 4057, June 2005.
- [RFC4193] Hinden, R. and B. Haberman, "Unique Local IPv6 Unicast Addresses", RFC 4193, October 2005.
- [RFC4213] Nordmark, E. and R. Gilligan, "Basic Transition Mechanisms for IPv6 Hosts and Routers", RFC 4213, October 2005.
- [RFC4293] Routhier, S., "Management Information Base for the Internet Protocol (IP)", RFC 4293, April 2006.
- [RFC4292] Haberman, B., "IP Forwarding Table MIB", RFC 4292, April 2006.
- [RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.
- [RFC4443] Conta, A., Deering, S., and M. Gupta, "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", RFC 4443, March 2006.
- [RFC4659] De Clercq, J., Ooms, D., Carugi, M., and F. Le Faucheur, "BGP-MPLS IP Virtual Private Network (VPN) Extension for IPv6 VPN", RFC 4659, September 2006.
- [RFC4861] Narten, T., Nordmark, E., Simpson, W., and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, September 2007.

- [RFC4862] Thomson, S., Narten, T., and T. Jinmei, "IPv6 Stateless Address Autoconfiguration", RFC 4862, September 2007.
- [RFC4890] Davies, E. and J. Mohacsi, "Recommendations for Filtering ICMPv6 Messages in Firewalls", RFC 4890, May 2007.
- [RFC4941] Narten, T., Draves, R., and S. Krishnan, "Privacy Extensions for Stateless Address Autoconfiguration in IPv6", RFC 4941, September 2007.
- [RFC5095] Abley, J., Savola, P., and G. Neville-Neil, "Deprecation of Type 0 Routing Headers in IPv6", RFC 5095, December 2007.
- [RFC7012] Claise, B. and B. Trammell, "Information Model for IP Flow Information Export (IPFIX)", RFC 7012, September 2013.
- [RFC5157] Chown, T., "IPv6 Implications for Network Scanning", RFC 5157, March 2008.
- [RFC5211] Curran, J., "An Internet Transition Plan", RFC 5211, July 2008.
- [RFC5214] Templin, F., Gleeson, T., and D. Thaler, "Intra-Site Automatic Tunnel Addressing Protocol (ISATAP)", RFC 5214, March 2008.
- [RFC5375] Van de Velde, G., Popoviciu, C., Chown, T., Bonness, O., and C. Hahn, "IPv6 Unicast Address Assignment Considerations", RFC 5375, December 2008.
- [RFC5722] Krishnan, S., "Handling of Overlapping IPv6 Fragments", RFC 5722, December 2009.
- [RFC5798] Nadas, S., "Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6", RFC 5798, March 2010.
- [RFC5952] Kawamura, S. and M. Kawashima, "A Recommendation for IPv6 Address Text Representation", RFC 5952, August 2010.
- [RFC6052] Bao, C., Huitema, C., Bagnulo, M., Boucadair, M., and X. Li, "IPv6 Addressing of IPv4/IPv6 Translators", RFC 6052, October 2010.
- [RFC6104] Chown, T. and S. Venaas, "Rogue IPv6 Router Advertisement Problem Statement", RFC 6104, February 2011.

- [RFC6105] Levy-Abegnoli, E., Van de Velde, G., Popoviciu, C., and J. Mohacsi, "IPv6 Router Advertisement Guard", RFC 6105, February 2011.
- [RFC6106] Jeong, J., Park, S., Beloeil, L., and S. Madanapalli, "IPv6 Router Advertisement Options for DNS Configuration", RFC 6106, November 2010.
- [RFC6144] Baker, F., Li, X., Bao, C., and K. Yin, "Framework for IPv4/IPv6 Translation", RFC 6144, April 2011.
- [RFC6145] Li, X., Bao, C., and F. Baker, "IP/ICMP Translation Algorithm", RFC 6145, April 2011.
- [RFC6146] Bagnulo, M., Matthews, P., and I. van Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", RFC 6146, April 2011.
- [RFC6147] Bagnulo, M., Sullivan, A., Matthews, P., and I. van Beijnum, "DNS64: DNS Extensions for Network Address Translation from IPv6 Clients to IPv4 Servers", RFC 6147, April 2011.
- [RFC6177] Narten, T., Huston, G., and L. Roberts, "IPv6 Address Assignment to End Sites", BCP 157, RFC 6177, March 2011.
- [RFC6164] Kohno, M., Nitzan, B., Bush, R., Matsuzaki, Y., Colitti, L., and T. Narten, "Using 127-Bit IPv6 Prefixes on Inter-Router Links", RFC 6164, April 2011.
- [RFC6192] Dugal, D., Pignataro, C., and R. Dunn, "Protecting the Router Control Plane", RFC 6192, March 2011.
- [RFC6296] Wasserman, M. and F. Baker, "IPv6-to-IPv6 Network Prefix Translation", RFC 6296, June 2011.
- [RFC6302] Durand, A., Gashinsky, I., Lee, D., and S. Sheppard, "Logging Recommendations for Internet-Facing Servers", BCP 162, RFC 6302, June 2011.
- [RFC6434] Jankiewicz, E., Loughney, J., and T. Narten, "IPv6 Node Requirements", RFC 6434, December 2011.
- [RFC6555] Wing, D. and A. Yourtchenko, "Happy Eyeballs: Success with Dual-Stack Hosts", RFC 6555, April 2012.
- [RFC6583] Gashinsky, I., Jaeggli, J., and W. Kumari, "Operational Neighbor Discovery Problems", RFC 6583, March 2012.

- [RFC6724] Thaler, D., Draves, R., Matsumoto, A., and T. Chown, "Default Address Selection for Internet Protocol Version 6 (IPv6)", RFC 6724, September 2012.
- [RFC6866] Carpenter, B. and S. Jiang, "Problem Statement for Renumbering IPv6 Hosts with Static Addresses in Enterprise Networks", RFC 6866, February 2013.
- [RFC6879] Jiang, S., Liu, B., and B. Carpenter, "IPv6 Enterprise Network Renumbering Scenarios, Considerations, and Methods", RFC 6879, February 2013.
- [RFC6883] Carpenter, B. and S. Jiang, "IPv6 Guidance for Internet Content Providers and Application Service Providers", RFC 6883, March 2013.
- [RFC6959] McPherson, D., Baker, F., and J. Halpern, "Source Address Validation Improvement (SAVI) Threat Scope", RFC 6959, May 2013.
- [RFC6964] Templin, F., "Operational Guidance for IPv6 Deployment in IPv4 Sites Using the Intra-Site Automatic Tunnel Addressing Protocol (ISATAP)", RFC 6964, May 2013.
- [RFC7011] Claise, B., Trammell, B., and P. Aitken, "Specification of the IP Flow Information Export (IPFIX) Protocol for the Exchange of Flow Information", STD 77, RFC 7011, September 2013.
- [RFC7113] Gont, F., "Implementation Advice for IPv6 Router Advertisement Guard (RA-Guard)", RFC 7113, February 2014.
- [RFC7123] Gont, F. and W. Liu, "Security Implications of IPv6 on IPv4 Networks", RFC 7123, February 2014.
- [RFC7045] Carpenter, B. and S. Jiang, "Transmission and Processing of IPv6 Extension Headers", RFC7045, December 2013, <<http://tools.ietf.org/html/rfc7045>>.
- [I-D.xli-behave-divi]
 Bao, C., Li, X., Zhai, Y., and W. Shang, "dIVI: Dual-Stateless IPv4/IPv6 Translation", draft-xli-behave-divi-06 (work in progress), January 2014.
- [I-D.wierenga-ietf-eduroam]
 Wierenga, K., Winter, S., and T. Wolniewicz, "The eduroam architecture for network roaming", draft-wierenga-ietf-eduroam-03 (work in progress), February 2014.

- [I-D.ietf-v6ops-dc-ipv6]
Lopez, D., Chen, Z., Tsou, T., Zhou, C., and A. Servin,
"IPv6 Operational Guidelines for Datacenters", draft-ietf-
v6ops-dc-ipv6-01 (work in progress), February 2014.
- [I-D.ietf-v6ops-design-choices]
Matthews, P. and V. Kuarsingh, "Design Choices for IPv6
Networks", draft-ietf-v6ops-design-choices-01 (work in
progress), March 2014.
- [I-D.ietf-opsec-v6]
Chittimaneni, K., Kaeo, M., and E. Vyncke, "Operational
Security Considerations for IPv6 Networks", draft-ietf-
opsec-v6-04 (work in progress), October 2013.
- [I-D.ietf-opsec-ipv6-host-scanning]
Gont, F. and T. Chown, "Network Reconnaissance in IPv6
Networks", draft-ietf-opsec-ipv6-host-scanning-04 (work in
progress), June 2014.
- [I-D.liu-bonica-dhcpv6-slaac-problem]
Liu, B. and R. Bonica, "DHCPv6/SLAAC Address Configuration
Interaction Problem Statement", draft-liu-bonica-dhcpv6-
slaac-problem-02 (work in progress), September 2013.
- [I-D.ietf-v6ops-ula-usage-recommendations]
Liu, B. and S. Jiang, "Considerations of Using Unique
Local Addresses", draft-ietf-v6ops-ula-usage-
recommendations-03 (work in progress), July 2014.
- [I-D.ietf-opsec-vpn-leakages]
Gont, F., "Layer-3 Virtual Private Network (VPN) tunnel
traffic leakages in dual- stack hosts/networks", draft-
ietf-opsec-vpn-leakages-06 (work in progress), April 2014.
- [SMURF] "CERT Advisory CA-1998-01 Smurf IP Denial-of-Service
Attacks",
<<http://www.cert.org/advisories/CA-1998-01.html>>.
- [CYMRU] "THE BOGON REFERENCE",
<<http://www.team-cymru.org/Services/Bogons/>>.

Authors' Addresses

Kiran K. Chittimaneni
Dropbox Inc.
1600 Amphitheater Pkwy
Mountain View, California CA 94043
USA

Email: kk@dropbox.com

Tim Chown
University of Southampton
Highfield
Southampton, Hampshire SO17 1BJ
United Kingdom

Email: tjc@ecs.soton.ac.uk

Lee Howard
Time Warner Cable
13820 Sunrise Valley Drive
Herndon, VA 20171
US

Phone: +1 703 345 3513
Email: lee.howard@twcable.com

Victor Kuarsingh
Dyn Inc
150 Dow Street
Manchester, NH
US

Email: victor@jvknet.com

Yanick Pouffary
Hewlett Packard
950 Route Des Colles
Sophia-Antipolis 06901
France

Email: Yanick.Pouffary@hp.com

Eric Vyncke
Cisco Systems
De Kleetlaan 6a
Diegem 1831
Belgium

Phone: +32 2 778 4677
Email: evyncke@cisco.com

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: September 11, 2014

G. Chen
Z. Cao
China Mobile
C. Xie
China Telecom
D. Binet
France Telecom-Orange
March 10, 2014

NAT64 Deployment Options and Experience
draft-ietf-v6ops-nat64-experience-10

Abstract

This document summarizes NAT64 function deployment scenarios and operational experience. Both NAT64 Carrier Grade NAT (NAT64-CGN) and NAT64 server Front End (NAT64-FE) are considered in this document.

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 September 11, 2014.

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. Terminology	3
3. NAT64 Networking Experience	4
3.1. NAT64-CGN Consideration	4
3.1.1. NAT64-CGN Usages	4
3.1.2. DNS64 Deployment	4
3.1.3. NAT64 Placement	5
3.1.4. Co-existence of NAT64 and NAT44	5
3.2. NAT64-FE Consideration	6
4. High Availability	7
4.1. Redundancy Design	7
4.2. Load Balancing	9
5. Source Address Transparency	9
5.1. Traceability	9
5.2. Geo-location	10
6. Quality of Experience	11
6.1. Service Reachability	11
6.2. Resource Reservation	12
7. MTU Considerations	13
8. ULA Usages	14
9. Security Considerations	15
10. IANA Considerations	15
11. Acknowledgements	15
12. Additional Author List	16
13. References	16
13.1. Normative References	16
13.2. Informative References	18
Appendix A. Testing Results of Application Behavior	20
Authors' Addresses	21

1. Introduction

IPv6 is the only sustainable solution for numbering nodes on Internet due to the IPv4 depletion. Network operators have to deploy IPv6-only networks in order to meet the needs of the expanding internet without available IPv4 addresses.

Single-stack IPv6 network deployment can simplify networks provisioning, some justification was provided in 464xlat [RFC6877]. IPv6-only connectivity confers some benefits to mobile operators as an example. In the mobile context, IPv6-only usage enables the use of a single IPv6 Packet Data Protocol(PDP) context or Evolved Packet System (EPS) bearer on Long Term Evolution (LTE) networks. This

eliminates significant network costs caused by employing two PDP contexts in some cases, and the need for IPv4 addresses to be assigned to customers. In broadband networks overall, it can allow for the scaling of edge-network growth to be decoupled from IPv4 numbering limitations.

In transition scenarios, some existing networks are likely to be IPv4-only for quite a long time. IPv6 networks and hosts IPv6-only hosts will need to coexist with IPv4 numbered resources. Widespread dual-stack deployments have not materialized at the anticipated rate over the last 10 years, one possible conclusion being that legacy networks will not make the jump quickly. The Internet will include nodes that are dual-stack, nodes that remain IPv4-only, and nodes that can be deployed as IPv6-only nodes. A translation mechanism based on a NAT64[RFC6146] [RFC6145]function is likely to be a key element of Internet connectivity for IPv6-IPv4 interoperability.

[RFC6036] reports at least 30% of operators plan to run some kind of translator (presumably NAT64/DNS64). Advice on NAT64 deployment and operations are therefore of some importance. [RFC6586] documents the implications for IPv6 only networks. This document intends to be specific to NAT64 network planning.

2. Terminology

Regarding IPv4/IPv6 translation, [RFC6144] has described a framework for enabling networks to make interworking possible between IPv4 and IPv6 networks. This document has further categorized different NAT64 functions, locations and use-cases. The principle distinction of location is whether the NAT64 is located in a Carrier Grade NAT or server Front End. The terms of NAT-CGN/FE are understood to be a topological distinction indicating different features employed in a NAT64 deployment.

NAT64 Carrier Grade NAT (NAT64-CGN): A NAT64-CGN is placed in an ISP network. IPv6 enabled subscribers leverage the NAT64-CGN to access existing IPv4 internet services. The ISP as an administrative entity takes full control of the IPv6 side, but has limited or no control on the IPv4 internet side. NAT64-CGN deployments may have to consider the IPv4 Internet environment and services, and make appropriate configuration choices accordingly.

NAT64 server Front End (NAT64-FE): A NAT64-FE is generally a device with NAT64 functionality in a content provider or data center network. It could be for example a traffic load balancer or a firewall. The operator of the NAT64-FE has full control over the IPv4 network within the data center, but only limited influence or control over the external Internet IPv6 network.

3. NAT64 Networking Experience

3.1. NAT64-CGN Consideration

3.1.1. NAT64-CGN Usages

Fixed network operators and mobile operators may locate NAT64 translators in access networks or in mobile core networks. It can be built into various devices, including routers, gateways or firewalls in order to connect IPv6 users to the IPv4 Internet. With regard to the numbers of users and the shortage of public IPv4 addresses, stateful NAT64[RFC6146] is more suited to maximize sharing of public IPv4 addresses. The usage of stateless NAT64 can provide better transparency features [I-D.ietf-softwire-stateless-4v6-motivation], but has to be coordinated with A+P[RFC6346] processes as specified in [I-D.ietf-softwire-map-t] in order to address an IPv4 address shortage.

3.1.2. DNS64 Deployment

DNS64[RFC6147] is recommended for use in combination with stateful NAT64, and will likely be an essential part of an IPv6 single-stack network that couples to the IPv4 Internet. 464xlat[RFC6877] can enable access of IPv4 only applications or applications that call IPv4 literal addresses. Using DNS64 will help 464xlat to automatically discover NAT64 prefix through [RFC7050]. Berkeley Internet Name Daemon (BIND) software supports the function. It's important to note that DNS64 generates the synthetic AAAA reply when services only provide A records. Operators should not expect to access IPv4 parts of a dual-stack server using NAT64/DNS64. The traffic is forwarded on IPv6 paths if dual-stack servers are targeted. IPv6 traffic may be routed around rather than going through NAT64. Only the traffic going to IPv4-only service would traverse the NAT64 translator. In some sense, it encourages IPv6 usage and limits NAT translation compared to employing NAT44, where all traffic flows have to be translated. In some cases, NAT64-CGNs may serve double roles, i.e. as a translator and IPv6 forwarder. In mobile networks, NAT64 may be deployed as the default gateway serving all the IPv6 traffic. The traffic heading to a dual-stack server is only forwarded on the NAT64. Therefore, both IPv6 and IPv4 are suggested to be configured on the Internet faced interfaces of NAT64. We tested on Top100 websites (referring to [Alexa] statistics). 43% of websites are connected and forwarded on the NAT64 since those websites have both AAAA and A records. With expansion of IPv6 support, the translation process on NAT64 will likely become less-important over time. It should be noted the DNS64-DNSSEC Interaction[RFC6147] may impact validation of Resource Records retrieved from the the DNS64 process. In particular, DNSSEC

validation will fail when DNS64 synthesizes AAAA records where there is a DNS query with the "DNSSEC OK" (DO) bit set and the "Checking Disabled" (CD) bit set received.

3.1.3. NAT64 Placement

All connections to IPv4 services from IPv6-only clients must traverse the NAT64-CGN. It can be advantageous from the vantage-point of troubleshooting and traffic engineering to carry the IPv6 traffic natively for as long as possible within an access network and translate packets only at or near the network egress. NAT64 may be a feature of the Autonomous System (AS) border in fixed networks. It may be deployed in an IP node beyond the Gateway GPRS Support Node (GGSN) or Public Data Network- Gateway (PDN-GW) in mobile networks or directly as part of the gateway itself in some situations. This allows consistent attribution and traceability within the service provider network. It has been observed that the process of correlating log information is problematic from multiple-vendor's equipment due to inconsistent formats of log records. Placing NAT64 in a centralized location may reduce diversity of log format and simplify the network provisioning. Moreover, since NAT64 is only targeted at serving traffic flows from IPv6 to IPv4-only services, the user traffic volume should not be as high as in a NAT44 scenario, and therefore, the gateway's capacity in such location may be less of a concern or a hurdle to deployment. On the other-hand, placement in a centralized fashion would require more strict high availability (HA) design. It would also make geo-location based on IPv4 addresses rather inaccurate as is currently the case for NAT44 CGN already deployed in ISP networks. More considerations or workarounds on HA and traceability could be found at Section 4 and Section 5.

3.1.4. Co-existence of NAT64 and NAT44

NAT64 will likely co-exist with NAT44 in a dual-stack network where IPv4 private addresses are allocated to customers. The coexistence has already been observed in mobile networks, in which dual stack mobile phones normally initiate some dual-stack PDN/PDP Type[RFC6459] to query both IPv4/IPv6 address and IPv4 allocated addresses are very often private ones. [RFC6724] always prioritizes IPv6 connections regardless of whether the end-to-end path is native IPv6 or IPv6 translated to IPv4 via NAT64/DNS64. Conversely, Happy Eyeballs[RFC6555] will direct some IP flows across IPv4 paths. The selection of IPv4/IPv6 paths may depend on particular implementation choices or settings on a host-by-host basis, and may differ from an operator's deterministic scheme. Our tests verified that hosts may find themselves switching between IPv4 and IPv6 paths as they access identical service, but at different times [I-D.kaliwoda-sunset4-dual-ipv6-coexist]. Since the topology on each

path is potentially different, it may cause unstable user experience and some degradation of Quality of Experience (QoE) when falling back to the other protocol. It's also difficult for operators to find a solution to make a stable network with optimal resource utilization. In general, it's desirable to figure out the solution that will introduce IPv6/IPv4 translation service to IPv6-only hosts connecting to IPv4 servers while making sure dual-stack hosts to have at least one address family accessible via native service if possible. With the end-to-end native IPv6 environment available, hosts should be upgraded aggressively to migrate in favor of IPv6-only. There are ongoing efforts to detect host connectivity and propose a new DHCPv6 option[I-D.wing-dhc-dns-reconfigure] to convey appropriate configuration information to the hosts.

3.2. NAT64-FE Consideration

Some Internet Content Providers (ICPs) may locate NAT64 in front of an Internet Data Center (IDC), for example co-located with a load-balancing function. Load-balancers are employed to connect different IP family domains, and distribute workloads across multiple domains or internal servers. In some cases, IPv4 addresses exhaustion may not be a problem in some IDC's internal networks. IPv6 support for some applications may require some investments and workloads so IPv6 support may not be a priority. The use of NAT64 may be served to support widespread IPv6 adoption on the Internet while maintaining IPv4-only applications access.

Different strategy has been described in [RFC6883] referred to as "inside out" and "outside in". An IDC operator may implement the following practices in the NAT64-FE networking scenario.

- o Some ICPs who already have satisfactory operational experience might adopt single stack IPv6 operation in building data-center networks, servers and applications, as it allows new services delivery without having to integrate consideration of IPv4 NAT and address limitations of IPv4 networks. Stateless NAT64[RFC6145] can be used to provide services for IPv4-only enabled customers. [I-D.anderson-siit-dc] has provided further descriptions and guidelines.
- o ICPs who attempt to offer customers IPv6 support in their application farms at an early stage may likely run proxies load-balancers or translators, which are configured to handle incoming IPv6 flows and proxy them to IPv4 back-end systems. Many load balancers integrate proxy functionality. IPv4 addresses configured in the proxy may be multiplexed like a stateful NAT64 translator. A similar challenge exists once increasingly numerous users in IPv6 Internet access an IPv4 network. High loads on

load-balancers may be apt to cause additional latency, IPv4 pool exhaustion, etc. Therefore, this approach is only reasonable at an early stage. ICPs may employ dual-stack or IPv6 single stack in a further stage, since the native IPv6 is frequently more desirable than any of the transition solutions.

[RFC6144] recommends that AAAA records of load-balancers or application servers can be directly registered in the authoritative DNS servers. In this case, there is no need to deploy DNS64 name-servers. Those AAAA records can point to natively assigned IPv6 addresses or IPv4-converted IPv6 addresses[RFC6052]. Hosts are not aware of the NAT64 translator on communication path. For the testing purpose, operators could employ an independent sub domain e.g. ipv6exp.example.com to identify experimental ipv6 services to users. How to design the FQDN for the IPv6 service is out-of-scope of this document.

4. High Availability

4.1. Redundancy Design

High Availability (HA) is a major requirement for every service and network services. The deployment of redundancy mechanisms is an essential approach to avoid failure and significantly increase the network reliability. It's not only useful to stateful NAT64 cases, but also to stateless NAT64 gateways.

Three redundancy modes are mainly used: cold standby, warm standby and hot standby.

- o Cold standby HA devices do not replicate the NAT64 states from the primary equipment to the backup. Administrators switch on the backup NAT64 only if the primary NAT64 fails. As a result, all existing established sessions through a failed translator will be disconnected. The translated flows will need to be recreated by end-systems. Since the backup NAT64 is manually configured to switch over to active NAT64, it may have unpredictable impacts to the ongoing services.
- o Warm standby is a flavor of the cold standby mode. Backup NAT64 would keep running once the primary NAT64 is working. This makes warm standby less time consuming during the traffic failover. Virtual Router Redundancy Protocol (VRRP)[RFC5798] can be a solution to enable automatic handover in the warm standby. It was tested that the handover takes as maximum as 1 minute if the backup NAT64 needs to take over routing and re-construct the Binding Information Bases (BIBs) for 30 million sessions. In

deployment phase, operators could balance loads on distinct NAT64s devices. Those NAT64s make a warm backup of each other.

- o Hot standby must synchronize the BIBs between the primary NAT64 and backup. When the primary NAT64 fails, backup NAT64 would take over and maintain the state of all existing sessions. The internal hosts don't have to re-connect the external hosts. The handover time has been extremely reduced. Employing Bidirectional Forwarding Detection (BFD) [RFC5880] combined with VRRP, a delay of only 35ms for 30 million sessions handover was observed during testing. Under ideal conditions hotstandby deployments could guarantee the session continuity for every service. In order to timely transmit session states, operators may have to deploy extra transport links between primary NAT64 and distant backup. The scale of synchronization data instance is depending on the particular deployment. For example, If a NAT64-CGN is served for 200,000 users, the average amount of 800, 000 sessions per second is roughly estimated for new created and expired sessions. A physical 10Gbps transport link may have to be deployed for the sync data transmission considering the amount of sync sessions at the peak and capacity redundancy

In general, cold-standby and warm-standby is simpler and less resource intensive, but it requires clients to re-establish sessions when a fail-over occurs. Hot standby increases resource consumption in order to synchronize state, but potentially achieves seamless handover. For stateless NAT64 considerations are simple, because state synchronization is unnecessary. Regarding stateful NAT64, it may be useful to investigate performance tolerance of applications and the traffic characteristics in a particular network. Some testing results are shown in the Appendix A.

Our statistics in a mobile network shown that almost 91.21% of of traffic is accounted by http/https services. These services generally don't require session continuity. Hot-standby does not offer much benefit for those sessions on this point. In fixed networks, HTTP streaming, p2p and online games would be the major traffic beneficiaries of hot-standby replication[Cisco-VNI]. Consideration should be given to the importance of maintaining bindings for those sessions across failover. Operators may also consider the Average Revenue Per User (ARPU) factors to deploy suitable redundancy mode. Warm standby may still be adopted to cover most services while hot standby could be used to upgrade Quality of Experience (QoE) using DNS64 to generate different synthetic responses for limited traffic or destinations. Further considerations are discussed at Section 6.

4.2. Load Balancing

Load balancing is used to accompany redundancy design so that better scalability and resiliency could be achieved. Stateless NAT64s allow asymmetric routing while anycast-based solutions are recommended in [I-D.ietf-softwire-map-deployment]. The deployment of load balancing may make more sense to stateful NAT64s for the sake of single-point failure avoidance. Since the NAT64-CGN and NAT64-FE have distinct facilities, the following lists the considerations for each case.

- o NAT64-CGN equipment doesn't typically implement load-balancing functions onboard. Therefore, the gateways have to resort to DNS64 or internal host's behavior. Once DNS64 is deployed, the load balancing can be performed by synthesizing AAAA response with different IPv6 prefixes. For the applications not requiring DNS resolver, internal hosts could learn multiple IPv6 prefixes through the approaches defined in[RFC7050] and then select one based on a given prefix selection policy.
- o A dedicated Load Balancer could be deployed at front of a NAT64-FE farm. Load Balancer uses proxy mode to redirect the flows to the appropriate NAT64 instance. Stateful NAT64s require a deterministic pattern to arrange the traffic in order to ensure outbound/inbound flows traverse the identical NAT64. Therefore, static scheduling algorithms, for example source-address based policy, is preferred. A dynamic algorithm, for example Round-Robin, may have impacts on applications seeking session continuity, which described in the Table 1.

5. Source Address Transparency

5.1. Traceability

Traceability is required in many cases such as identifying malicious attacks sources and accounting requirements. Operators are asked to record the NAT64 log information for specific periods of time. In our lab testing, the log information from 200,000 subscribers have been collected from a stateful NAT64 gateway for 60 days. Syslog[RFC5424] has been adopted to transmit log message from NAT64 to a log station. Each log message contains transport protocol, source IPv6 address:port, translated IPv4 address: port and timestamp. It takes almost 125 bytes in ASCII format. It has been verified that the rate of traffic flow is around 72 thousand flows per second and the volume of recorded information reaches up to 42.5 terabytes in the raw format. The volume is 29.07 terabytes in a compact format. At scale, operators have to build up dedicated transport links, storage system and servers for the purpose of managing such logging.

There are also several improvements that can be made to mitigate the issue. For example, stateful NAT64 could configure with bulk port allocation method. Once a subscriber creates the first session, a number of ports are pre-allocated. A bulk allocation message is logged indicating this allocation. Subsequent session creations will use one of the pre-allocated port and hence does not require logging. The log volume in this case may be only one thousandth of dynamic port allocation. Some implementations may adopt static port-range allocations [I-D.donley-behave-deterministic-cgn] which eliminates the need for per-subscriber logging. As a side effect, the IPv4 multiplexing efficiency is decreased regarding to those methods. For example, the utilization ratio of public IPv4 address is dropped approximately to 75% when NAT64 gateway is configured with bulk port allocation (The lab testing allocates each subscriber with 400 ports). In addition, port-range based allocation should also consider port randomization described in [RFC6056]. A trade-off among address multiplexing efficiency, logging storage compression and port allocation complexity should be considered. More discussions could be found in [I-D.chen-sunset4-cgn-port-allocation]. The decision can balance usable IPv4 resources against investments in log systems.

5.2. Geo-location

IP addresses are usually used as inputs to geo-location services. The use of address sharing prevents these systems from resolving the location of a host based on IP address alone. Applications that assume such geographic information may not work as intended. The possible solutions listed in [RFC6967] are intended to bridge the gap. However, those solutions can only provide a sub-optimal substitution to solve the problem of host identification, in particular it may not today solve problems with source identification through translation. The following lists current practices to mitigate the issue.

- o Operators who adopt NAT64-FE may leverage the application layer proxies, e.g. X-Forwarded-For (XFF) [I-D.ietf-appsawg-http-forwarded], to convey the IPv6 source address in HTTP headers. Those messages would be passed on to web-servers. The log parsing tools are required to be able to support IPv6 and may lookup Radius servers for the target subscribers based on IPv6 addresses included in XFF HTTP headers. XFF is the de facto standard which has been integrated in most Load Balancers. Therefore, it may be superior to use in a NAT-FE environment. In the downsides, XFF is specific to HTTP. It restricts the usages so that the solution can't be applied to requests made over HTTPs. This makes geo-location problematic for HTTPs based services.

- o The NAT64-CGN equipment may not implement XFF. Geo-location based on shared IPv4 address is rather inaccurate in that case. Operators could subdivide the outside IPv4 address pool so an IPv6 address can be translated depending on their geographical locations. As consequence, location information can be identified from a certain IPv4 address range. [RFC6967] also enumerates several options to reveal the host identifier. Each solution likely has their-own specific usage. For the geo-location systems relying on a Radius database[RFC5580], we have investigated to deliver NAT64 BIBs and Session Table Entries (STEs) to a Radius server[I-D.chen-behave-nat64-radius-extension]. This method could provide geo-location system with an internal IPv6 address to identify each user. It can get along with [RFC5580] to convey original source address through same message bus.

6. Quality of Experience

6.1. Service Reachability

NAT64 is providing a translation capability between IPv6 and IPv4 end-nodes. In order to provide the reachability between two IP address families, NAT64-CGN has to implement appropriate application aware functions, i.e. Application Layer Gateway (ALG), where address translation is not itself sufficient and security mechanisms do not render it infeasible. Most NAT64-CGNs mainly provide FTP-ALG[RFC6384]. NAT64-FEs may have functional richness on Load Balancer, for example HTTP-ALG, HTTPS-ALG, RTSP-ALG and SMTP-ALG have been supported. Those application protocols exchange IP address and port parameters within control session, for example the "Via" field in a HTTP header, "Transport" field in a RTSP SETUP message and "Received: " header in a SMTP message. ALG functions will detect those fields and make IP address translations. It should be noted that ALGs may impact the performance on a NAT64 box to some extent. ISPs as well as content providers might choose to avoid situations where the imposition of an ALG might be required. At the same time, it is also important to remind customers and application developers that IPv6 end-to-end usage does not require ALG imposition and therefore results in a better overall user experience.

The service reachability is also subject to the IPv6 support in the client side. We tested several kinds of applications as shown in the below table to verify the IPv6 supports. The experiences of some applications are still align with [RFC6586]. For example, we have tested P2P file sharing and streaming applications including eMule v0.50a, Thunder v7.9 and PPS TV v3.2.0. It has been found there are some software issues to support IPv6 at this time. The application software would benefit from 464xlat[RFC6877] until the software adds IPv6 support.. A SIP based voice call has been tested in LTE mobile

environment as specified in [IR.92]. The voice call is failed due to the lack of NAT64 traversal when an IPv6 SIP user agent communicates with an IPv4 SIP user agent. In order to address the failure, Interactive Connectivity Establishment (ICE) described in [RFC5245] is recommended to be supported for the SIP IPv6 transition. [RFC6157] describes both signaling and media layer process, which should be followed. In addition, it may be worth to notice that ICE is not only useful for NAT traversal, but also firewall[RFC6092] traversal in native IPv6 deployment.

Different IPsec modes for VPN services have been tested, including IPsec-AH and IPsec-ESP. It has been testified IPsec-AH can't survive since the destination host detects the IP header changes and invalidate the packets. IPsec-ESP failed in our testing because the NAT64 does not translate IPsec ESP (i.e. protocol 50) packets. It has been suggested that IPsec ESP should succeed if the IPsec client supports NAT-Traversal in the IKE[RFC3947] and uses IPsec ESP over UDP[RFC3948].

Table 1: The tested applications

APPs	Results and Found Issues
Webservice	Mostly pass, some failure cases due to IPv4 Literals
Instant Message	Mostly fail, software can't support IPv6
Games	Mostly pass for web-based games; mostly fail for standalone games due to the lack of IPv6 support in software
SIP-VoIP	Fail, due to the lack of NAT64 traversal
IPsec-VPN	Fail, the translated IPsec packets are invalidated
P2P file sharing and streaming	Mostly fail, software can't support IPv6, e.g. eMule, Thunder and PPS TV
FTP	Pass
Email	Pass

6.2. Resource Reservation

Session status normally is managed by a static timer. For example, the value of the "established connection idle-timeout" for TCP sessions must not be less than 2 hours 4 minutes[RFC5382] and 5

minutes for UDP sessions[RFC4787]. In some cases, NAT resource maybe significantly consumed by largely inactive users. The NAT translator and other customers would suffer from service degradation due to port consummation by other subscribers using the same NAT64 device. A flexible NAT session control is desirable to resolve the issues. PCP[RFC6887] could be a candidate to provide such capability. A NAT64-CGN should integrate with a PCP server, to allocate available IPv4 address/port resources. Resources could be assigned to PCP clients through PCP MAP/PEER mode. Such ability can be considered to upgrade user experiences, for example assigning different sizes of port ranges for different subscribers. Those mechanisms are also helpful to minimize terminal battery consumption and reduce the number of keep-alive messages to be sent by mobile terminal devices.

Subscribers can also benefit from network reliability. It has been discussed that hot-standby offers satisfactory experience once outage of primary NAT64 is occurred. Operators may rightly be concerned about the considerable investment required for NAT64 equipment relative to low ARPU income. For example, transport links may cost much, because primary NAT64 and backup are normally located at different locations, separated by a relatively large distance. Additional cost has to be assumed to ensure the connectivity quality. However, that may be necessary to some applications, which are delay-sensitive and seek session continuity, for example on-line games and live-streaming. Operators may be able to get added-values from those services by offering first-class services. It can be pre-configured on the gateway to hot-standby modes depending on subscriber's profile. The rest of other sessions can be covered by cold/warm standby.

7. MTU Considerations

IPv6 requires that every link in the internet have an Maximum Transmission Unit (MTU) of 1280 octets or greater[RFC2460]. However, in case of NAT64 translation deployment, some IPv4 MTU constrained link will be used in some communication path and originating IPv6 nodes may therefore receive an ICMP Packet Too Big (PTB) message, reporting a Next-Hop MTU less than 1280 bytes. The result would be that IPv6 allows packets to contain a fragmentation header, without the packet being fragmented into multiple pieces. A NAT64 would receive IPv6 packets with fragmentation header in which "M" flag equal to 0 and "Fragment Offset" equal to 0. Those packets likely impact other fragments already queued with the same set of {IPv6 Source Address, IPv6 Destination Address, Fragment Identification}. If the NAT64 box is compliant with [RFC5722], there is risk that all the fragments have to be dropped.

[RFC6946] discusses how this situation could be exploited by an attacker to perform fragmentation-based attacks, and also proposes an improved handling of such packets. It required enhancements on NAT64 gateway implementations to isolate packet's processing. NAT64 should follow the recommendation and take steps to prevent the risks of fragmentation.

Another approach that potentially avoids this issue is to configure IPv4 MTU more than 1260 bytes. It would forbid the occurrence of PTB smaller than 1280 bytes. Such an operational consideration is hard to universally apply to the legacy "IPv4 Internet" NAT64-CGN bridged. However, it's a feasible approach in NAT64-FE cases, since a IPv4 network NAT64-FE connected is rather well-organized and operated by a IDC operator or content provider. Therefore, the MTU of IPv4 network in NAT64-FE case are strongly recommended to set to more than 1260 bytes.

8. ULA Usages

Unique Local Addresses (ULAs) are defined in [RFC4193] to be renumbered within a network site for local communications. Operators may use ULAs as NAT64 prefixes to provide site-local IPv6 connectivity. Those ULA prefixes are stripped when the packets going to the IPv4 Internet, therefore ULAs are only valid in the IPv6 site. The use of ULAs could help in identifying the translation traffic. [I-D.ietf-v6ops-ula-usage-recommendations] provides further guidance for the ULAs usages.

We configure ULAs as NAT64 prefixes on a NAT64-CGN. If a host is only assigned with an IPv6 address and connected to NAT64-CGN, when connect to an IPv4 service, it would receive AAAA record generated by the DNS64 with the ULA prefix. A Global Unicast Address (GUA) will be selected as the source address to the ULA destination address. When the host has both IPv4 and IPv6 address, it would initiate both A and AAAA record lookup, then both original A record and DNS64-generated AAAA record would be received. A host, which is compliant with [RFC6724], will never prefer ULA over IPv4. An IPv4 path will be always selected. It may be undesirable because the NAT64-CGN will never be used. Operators may consider to add additional site-specific rows into the default policy table for host address selection in order to steer traffic flows going through NAT64-CGN. However, it involves significant costs to change terminal's behavior. Therefore, operators are not suggested to configure ULAs on a NAT64-CGN.

ULAs can't work when hosts transit the Internet to connect with NAT64. Therefore, ULAs are inapplicable to the case of NAT64-FE.

9. Security Considerations

This document presents the deployment experiences of NAT64 in CGN and FE scenarios. In general, RFC 6146[RFC6146] provides TCP-tracking, address-dependent filtering mechanisms to protect NAT64 from Distributed Denial of Service (DDoS). In NAT64-CGN cases, operators also could adopt unicast Reverse Path Forwarding (uRPF)[RFC3704] and black/white-list to enhance the security by specifying access policies. For example, NAT64-CGN should forbid establish NAT64 BIB for incoming IPv6 packets if uRPF in Strict or Loose mode check does not pass or whose source IPv6 address is associated to black-lists.

The stateful NAT64-FE creates state and maps that connection to an internally-facing IPv4 address and port. An attacker can consume the resources of the NAT64-FE device by sending an excessive number of connection attempts. Without a DDoS limitation mechanism, the NAT64-FE is exposed to attacks. Load Balancer is recommended to enable the capabilities of line rate DDOS defense, such as the employment of SYN PROXY-COOKIE. Security domain division is necessary as well in this case. Therefore, Load Balancers could not only serve for optimization of traffic distribution, but also prevent service from quality deterioration due to security attacks.

The DNS64 process will potentially interfere with the DNSSEC functions[RFC4035], since DNS response is modified and DNSSEC intends to prevent such changes. More detailed discussions can be found in [RFC6147].

10. IANA Considerations

This memo includes no request to IANA.

11. Acknowledgements

The authors would like to thank Jari Arkko, Dan Wing, Remi Despres, Fred Baker, Hui Deng, Iljitsch van Beijnum, Philip Matthews, Randy Bush, Mikael Abrahamsson, Lorenzo Colitti, Sheng Jiang, Nick Heatley, Tim Chown, Gert Doering and Simon Perreault for their helpful comments.

Many thanks to Wesley George, Lee Howard and Satoru Matsushima for their detailed reviews.

The authors especially thank Joel Jaeggli and Ray Hunter for his efforts and contributions on editing which substantially improves the legibility of the document.

Thanks to Cameron Byrne who was an active co-author of some earlier versions of this draft.

12. Additional Author List

The following are extended authors who contributed to the effort:

Qiong Sun
China Telecom
Room 708, No.118, Xizhimennei Street
Beijing 100035
P.R.China
Phone: +86-10-58552936
Email: sunqiong@ctbri.com.cn

QiBo Niu
ZTE
50,RuanJian Road.
YuHua District,
Nan Jing 210012
P.R.China
Email: niu.qibo@zte.com.cn

13. References

13.1. Normative References

- [I-D.ietf-appsawg-http-forwarded]
Petersson, A. and M. Nilsson, "Forwarded HTTP Extension",
draft-ietf-appsawg-http-forwarded-10 (work in progress),
October 2012.
- [RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6
(IPv6) Specification", RFC 2460, December 1998.
- [RFC3704] Baker, F. and P. Savola, "Ingress Filtering for Multihomed
Networks", BCP 84, RFC 3704, March 2004.
- [RFC3947] Kivinen, T., Swander, B., Huttunen, A., and V. Volpe,
"Negotiation of NAT-Traversal in the IKE", RFC 3947,
January 2005.
- [RFC3948] Huttunen, A., Swander, B., Volpe, V., DiBurro, L., and M.
Stenberg, "UDP Encapsulation of IPsec ESP Packets", RFC
3948, January 2005.

- [RFC4035] Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "Protocol Modifications for the DNS Security Extensions", RFC 4035, March 2005.
- [RFC4193] Hinden, R. and B. Haberman, "Unique Local IPv6 Unicast Addresses", RFC 4193, October 2005.
- [RFC4787] Audet, F. and C. Jennings, "Network Address Translation (NAT) Behavioral Requirements for Unicast UDP", BCP 127, RFC 4787, January 2007.
- [RFC5245] Rosenberg, J., "Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator (NAT) Traversal for Offer/Answer Protocols", RFC 5245, April 2010.
- [RFC5382] Guha, S., Biswas, K., Ford, B., Sivakumar, S., and P. Srisuresh, "NAT Behavioral Requirements for TCP", BCP 142, RFC 5382, October 2008.
- [RFC5424] Gerhards, R., "The Syslog Protocol", RFC 5424, March 2009.
- [RFC5580] Tschofenig, H., Adrangi, F., Jones, M., Lior, A., and B. Aboba, "Carrying Location Objects in RADIUS and Diameter", RFC 5580, August 2009.
- [RFC5722] Krishnan, S., "Handling of Overlapping IPv6 Fragments", RFC 5722, December 2009.
- [RFC5798] Nadas, S., "Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6", RFC 5798, March 2010.
- [RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", RFC 5880, June 2010.
- [RFC6052] Bao, C., Huitema, C., Bagnulo, M., Boucadair, M., and X. Li, "IPv6 Addressing of IPv4/IPv6 Translators", RFC 6052, October 2010.
- [RFC6145] Li, X., Bao, C., and F. Baker, "IP/ICMP Translation Algorithm", RFC 6145, April 2011.
- [RFC6146] Bagnulo, M., Matthews, P., and I. van Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", RFC 6146, April 2011.

- [RFC6147] Bagnulo, M., Sullivan, A., Matthews, P., and I. van Beijnum, "DNS64: DNS Extensions for Network Address Translation from IPv6 Clients to IPv4 Servers", RFC 6147, April 2011.
- [RFC6157] Camarillo, G., El Malki, K., and V. Gurbani, "IPv6 Transition in the Session Initiation Protocol (SIP)", RFC 6157, April 2011.
- [RFC6384] van Beijnum, I., "An FTP Application Layer Gateway (ALG) for IPv6-to-IPv4 Translation", RFC 6384, October 2011.
- [RFC6555] Wing, D. and A. Yourtchenko, "Happy Eyeballs: Success with Dual-Stack Hosts", RFC 6555, April 2012.
- [RFC6724] Thaler, D., Draves, R., Matsumoto, A., and T. Chown, "Default Address Selection for Internet Protocol Version 6 (IPv6)", RFC 6724, September 2012.
- [RFC6887] Wing, D., Cheshire, S., Boucadair, M., Penno, R., and P. Selkirk, "Port Control Protocol (PCP)", RFC 6887, April 2013.
- [RFC6946] Gont, F., "Processing of IPv6 "Atomic" Fragments", RFC 6946, May 2013.
- [RFC7050] Savolainen, T., Korhonen, J., and D. Wing, "Discovery of the IPv6 Prefix Used for IPv6 Address Synthesis", RFC 7050, November 2013.

13.2. Informative References

- [Alexa] Alexa, "<http://www.alexa.com/topsites>", April 2013.
- [Cisco-VNI] Cisco, "Cisco Visual Networking Index: Forecast and Methodology, 2012-2017, <http://ciscovni.com/forecast-widget/index.html>", May 2013.
- [I-D.anderson-siit-dc] Anderson, T., "Stateless IP/ICMP Translation in IPv6 Data Centre Environments", draft-anderson-siit-dc-00 (work in progress), November 2012.
- [I-D.chen-behave-nat64-radius-extension] Chen, G. and D. Binet, "Radius Attributes for Stateful NAT64", draft-chen-behave-nat64-radius-extension-00 (work in progress), July 2013.

- [I-D.chen-sunset4-cgn-port-allocation]
Chen, G., Tsou, T., Donley, C., and T. Taylor, "Analysis of NAT64 Port Allocation Method", draft-chen-sunset4-cgn-port-allocation-03 (work in progress), February 2014.
- [I-D.donley-behave-deterministic-cgn]
Donley, C., Grundemann, C., Sarawat, V., Sundaresan, K., and O. Vautrin, "Deterministic Address Mapping to Reduce Logging in Carrier Grade NAT Deployments", draft-donley-behave-deterministic-cgn-07 (work in progress), January 2014.
- [I-D.ietf-software-map-deployment]
Qiong, Q., Chen, M., Chen, G., Tsou, T., and S. Perreault, "Mapping of Address and Port (MAP) - Deployment Considerations", draft-ietf-software-map-deployment-03 (work in progress), October 2013.
- [I-D.ietf-software-map-t]
Li, X., Bao, C., Dec, W., Troan, O., Matsushima, S., and T. Murakami, "Mapping of Address and Port using Translation (MAP-T)", draft-ietf-software-map-t-05 (work in progress), February 2014.
- [I-D.ietf-software-stateless-4v6-motivation]
Boucadair, M., Matsushima, S., Lee, Y., Bonness, O., Borges, I., and G. Chen, "Motivations for Carrier-side Stateless IPv4 over IPv6 Migration Solutions", draft-ietf-software-stateless-4v6-motivation-05 (work in progress), November 2012.
- [I-D.ietf-v6ops-ula-usage-recommendations]
Liu, B. and S. Jiang, "Recommendations of Using Unique Local Addresses", draft-ietf-v6ops-ula-usage-recommendations-02 (work in progress), February 2014.
- [I-D.kaliwoda-sunset4-dual-ipv6-coexist]
Kaliwoda, A. and D. Binet, "Co-existence of both dual-stack and IPv6-only hosts", draft-kaliwoda-sunset4-dual-ipv6-coexist-01 (work in progress), October 2012.
- [I-D.wing-dhc-dns-reconfigure]
Patil, P., Boucadair, M., Wing, D., and T. Reddy, "DHCPv6 Dynamic Reconfiguration", draft-wing-dhc-dns-reconfigure-02 (work in progress), September 2013.

- [IR.92] Global System for Mobile Communications Association (GSMA), , "IMS Profile for Voice and SMS Version 7.0", March 2013.
- [RFC6036] Carpenter, B. and S. Jiang, "Emerging Service Provider Scenarios for IPv6 Deployment", RFC 6036, October 2010.
- [RFC6056] Larsen, M. and F. Gont, "Recommendations for Transport-Protocol Port Randomization", BCP 156, RFC 6056, January 2011.
- [RFC6092] Woodyatt, J., "Recommended Simple Security Capabilities in Customer Premises Equipment (CPE) for Providing Residential IPv6 Internet Service", RFC 6092, January 2011.
- [RFC6144] Baker, F., Li, X., Bao, C., and K. Yin, "Framework for IPv4/IPv6 Translation", RFC 6144, April 2011.
- [RFC6346] Bush, R., "The Address plus Port (A+P) Approach to the IPv4 Address Shortage", RFC 6346, August 2011.
- [RFC6459] Korhonen, J., Soininen, J., Patil, B., Savolainen, T., Bajko, G., and K. Iisakkila, "IPv6 in 3rd Generation Partnership Project (3GPP) Evolved Packet System (EPS)", RFC 6459, January 2012.
- [RFC6586] Arkko, J. and A. Keranen, "Experiences from an IPv6-Only Network", RFC 6586, April 2012.
- [RFC6877] Mawatari, M., Kawashima, M., and C. Byrne, "464XLAT: Combination of Stateful and Stateless Translation", RFC 6877, April 2013.
- [RFC6883] Carpenter, B. and S. Jiang, "IPv6 Guidance for Internet Content Providers and Application Service Providers", RFC 6883, March 2013.
- [RFC6967] Boucadair, M., Touch, J., Levis, P., and R. Penno, "Analysis of Potential Solutions for Revealing a Host Identifier (HOST_ID) in Shared Address Deployments", RFC 6967, June 2013.

Appendix A. Testing Results of Application Behavior

We test several application behaviors in a lab environment to evaluate the impact when a primary NAT64 is out of service. In this testing, participants are asked to connect a IPv6-only WiFi network

using laptops, tablets or mobile phones. NAT64 is deployed as the gateway to connect Internet service. The tested applications are shown in the below table. Cold standby, warm standby and hot standby are taken turn to be tested. The participants may experience service interruption due to the NAT64 handover. Different interruption intervals are tested to gauge application behaviors. The results are illuminated as below.

Table 2: The acceptable delay of applications

APPs	Acceptable Interrupt Recovery	Session Continuity
Web Browse	As maximum as 6s	No
Http streaming	As maximum as 10s(cache)	Yes
Gaming	200ms~400ms	Yes
P2P streaming, file sharing	10~16s	Yes
Instant Message	1 minute	Yes
Mail	30 seconds	No
Downloading	1 minutes	No

Authors' Addresses

Gang Chen
 China Mobile
 Xuanwumenxi Ave. No.32,
 Xuanwu District,
 Beijing 100053
 China

Email: phdgang@gmail.com

Zhen Cao
China Mobile
Xuanwumenxi Ave. No.32,
Xuanwu District,
Beijing 100053
China

Email: caozhen@chinamobile.com, zehn.cao@gmail.com

Chongfeng Xie
China Telecom
Room 708 No.118, Xizhimenneidajie
Beijing 100035
P.R.China

Email: xiechf@ctbri.com.cn

David Binet
France Telecom-Orange
Rennes
35000
France

Email: david.binet@orange.com

IPv6 Operations (V6OPS)
Internet-Draft
Obsoletes: 3316 (if approved)
Intended status: Informational
Expires: April 18, 2013

J. Korhonen, Ed.
Nokia Siemens Networks
J. Arkko, Ed.
Ericsson
T. Savolainen
Nokia
S. Krishnan
Ericsson
October 15, 2012

IPv6 for 3GPP Cellular Hosts
draft-korhonen-v6ops-rfc3316bis-00.txt

Abstract

As the deployment of third and fourth generation cellular networks progresses, a large number of cellular hosts are being connected to the Internet. Standardization organizations are making Internet Protocol version 6 (IPv6) mandatory in their specifications. However, the concept of IPv6 covers many aspects and numerous specifications. In addition, the characteristics of cellular links in terms of bandwidth, cost and delay put special requirements on how IPv6 is used. This document considers IPv6 for cellular hosts that attach to the General Packet Radio Service (GPRS), Universal Mobile Telecommunications System (UMTS), or Evolved Packet System (EPS) networks (Hereafter collectively referred to as 3GPP networks). This document also lists out specific IPv6 functionality that needs to be implemented in addition what is already prescribed in the IPv6 Node Requirements document. It also discusses some issues relating to the use of these components when operating in these networks. This document obsoletes RFC 3316.

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 April 18, 2013.

Copyright Notice

Copyright (c) 2012 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	4
1.1. Scope of this Document	4
1.2. Abbreviations	5
1.3. Cellular Host IPv6 Features	6
2. Basic IP	7
2.1. Internet Protocol Version 6	7
2.2. Neighbor Discovery in 3GPP Networks	7
2.3. IPv6 Stateless Address Autoconfiguration	8
2.4. Stateless Address Autoconfiguration in 3GPP Networks	8
2.5. IP version 6 over PPP in 3GPP Networks	9
2.6. MLD in 3GPP Networks	9
2.7. Privacy Extensions for Address Configuration in IPv6	9
2.8. Dynamic Host Configuration Protocol for IPv6 (DHCPv6)	10
2.9. DHCPv6 Prefix Delegation	10
2.10. Router preferences and more specific routes	10
2.11. Neighbor Discovery and additional host configuration	10
3. IP Security	11
3.1. Extension header considerations	11
4. Mobility	11
5. IANA Considerations	11
6. Acknowledgements	12
7. Security Considerations	12
8. References	14
8.1. Normative references	14
8.2. Informative references	14
Appendix A. Cellular Host IPv6 Addressing in the 3GPP Model	15
Appendix B. Changes to RFC 3316	16
B.1. Version -00	16
Authors' Addresses	17

1. Introduction

Technologies such as GPRS (General Packet Radio Service), UMTS (Universal Mobile Telecommunications System), Evolved Packet System (EPS), CDMA2000 (Code Division Multiple Access 2000) and eHRPD (Enhanced High Rate Packet Data) are making it possible for cellular hosts to have an always-on connection to the Internet. IPv6 [RFC2460] has become essential to such networks as the number of such cellular hosts is increasing rapidly. Standardization organizations working with cellular technologies have recognized this and made IPv6 mandatory in their specifications.

Support for IPv6 and the introduction of UMTS started with 3GPP Release-99 networks and hosts. For detailed description of IPv6 in 3GPP networks including the Evolved Packet System, see [RFC6459].

1.1. Scope of this Document

For the purposes of this document, a cellular interface is considered to be the interface to a cellular access network based on the following standards: 3GPP GPRS and UMTS Release-99, Release-4 to Release-11, and EPS Release-8 to Release-11 as well as future UMTS or EPS releases. A cellular host is considered to be a host with such a cellular interface.

This document complements the IPv6 node requirements [RFC6434] in places where clarifications are needed with discussion on the use of these selected IPv6 specifications when operating over cellular interfaces. Such a specification is necessary in order for the optimal use of IPv6 in a cellular environment. The description is made from a cellular host point of view. Complementary access technologies may be available in the cellular host, but those are not discussed in detail. Important considerations are given in order to eliminate unnecessary user confusion over configuration options, ensure interoperability and to provide an easy reference for those implementing IPv6 in a cellular host. It is necessary to ensure that cellular hosts are good citizens of the Internet.

This document is informational in nature, and it is not intended to replace, update, or contradict any IPv6 standards documents or the IPv6 node requirements [RFC6434].

This document is mainly targeted towards the implementers of cellular hosts that will be used with the cellular networks listed in the scope. The document provides guidance on which IPv6 related specifications are to be implemented in such cellular hosts. Parts of this document may also apply to other cellular link types, but this document does not provide any detailed analysis on other link

types. This document should not be used as a definitive list of IPv6 functionality for cellular links other than those listed above. Future changes in 3GPP networks that impact host implementations may result in updates to this document.

There are different ways to implement cellular hosts:

- o The host can be a "closed" device with optimized applications, with no possibility to add or download applications that can have IP communications. An example of such a host is a very simple form of a mobile phone.
- o The host can be an open device, e.g., a "smart phone" where it is possible to download applications to expand the functionality of the device.
- o The cellular radio modem part can be separated from the host IP stack with an interface. On example of such host is a laptop computer that uses a USB cellular modem for the cellular access.

If a cellular host has additional interfaces on which IP is used, (such as Ethernet, WLAN, Bluetooth, etc.) then there may be additional requirements for the device, beyond what is discussed in this document. Additionally, this document does not make any recommendations on the functionality required on laptop computers having a cellular interface such as an embedded modem or a USB modem stick, other than recommending link specific behavior on the cellular link.

This document discusses IPv6 functionality as of the time when this document has been written. Ongoing work on IPv6 may affect what is required of future hosts.

Transition mechanisms used by cellular hosts are not described in this document and are left for further study. The primary transition mechanism supported by 3GPP is dual-stack [RFC4213]. Dual-stack capable bearers were added to GPRS starting from 3GPP Release-9 and to EPS starting from Release-8 [RFC6459], whereas in earlier releases 3GPP multiple single IP version bearers had to be used to support dual stack.

1.2. Abbreviations

- | | |
|----|---|
| 2G | Second Generation Mobile Telecommunications, such as GSM and GPRS technologies. |
| 3G | Third Generation Mobile Telecommunications, such as UMTS technology. |

4G	Fourth Generation Mobile Telecommunications, such as LTE technology.
3GPP	3rd Generation Partnership Project. Throughout the document, the term 3GPP (3rd Generation Partnership Project) networks refers to architectures standardized by 3GPP, in Second, Third and Fourth Generation releases: 99, 4, and 5, as well as future releases.
APN	Access Point Name. The APN is a logical name referring to a GGSN and/or a PGW, and an external network.
EPC	Evolved Packet Core.
EPS	Evolved Packet System.
ESP	Encapsulating Security Payload
GGSN	Gateway GPRS Support Node (a default router for 3GPP IPv6 cellular hosts in GPRS).
GPRS	General Packet Radio Service.
LTE	Long Term Evolution.
MT	Mobile Terminal, for example, a mobile phone handset.
MTU	Maximum Transmission Unit.
PDN	Packet Data Network.
PDP	Packet Data Protocol.
PGW	Packet Data Network Gateway (the default router for 3GPP IPv6 cellular hosts in EPS).
SGW	Serving Gateway. The user plane equivalent of an SGSN in EPS (and the default router for 3GPP IPv6 cellular hosts when using PMIPv6).
TE	Terminal Equipment, for example, a laptop attached through a 3GPP handset.
UMTS	Universal Mobile Telecommunications System.
WLAN	Wireless Local Area Network.

1.3. Cellular Host IPv6 Features

This specification defines IPv6 features for cellular hosts in three groups.

Basic IP

In this group, basic parts of IPv6 are described.

IP Security

In this group, the IP Security parts are described.

Mobility

In this group, IP layer mobility issues are described.

2. Basic IP

For most parts refer to the IPv6 Node Requirements document [RFC6434].

2.1. Internet Protocol Version 6

The Internet Protocol Version 6 (IPv6) is specified in [RFC2460]. This specification is a mandatory part of IPv6. A cellular host must conform the generic IPv6 Host Requirements [RFC6434], unless specifically pointed out otherwise in this document.

2.2. Neighbor Discovery in 3GPP Networks

A cellular host must support Neighbor Solicitation and Neighbor Advertisement messages. Some further notes on how these are applied in the particular type of an interface can be useful, however:

In GPRS, UMTS and EPS networks, some Neighbor Discovery messages can be unnecessary in certain cases. GPRS, UMTS and EPS links resemble a point-to-point link; hence, the cellular host's only neighbor on the cellular link is the default router that is already known through Router Discovery. The cellular host always solicits for routers when the cellular interface is enabled (as described in [RFC4861], Section 6.3.7).

There are no link layer addresses. Therefore, address resolution and next-hop determination are not needed. If the cellular host still attempts the address resolution e.g., for the default router, it must be understood that the GGSN/PGW may not even answer the address resolution Neighbor Solicitations. And even if it does, the Neighbor Advertisement is unlikely to contain the Target link-layer address option as there are no link-layer addresses.

The cellular host must support Neighbor Unreachability Detection (NUD) as specified in [RFC4861]. Note that the link-layer address considerations above also apply to the Neighbor Unreachability Detection. The NUD triggered Neighbor Advertisement is also unlikely to contain the Target link-layer address option as there are no link-layer addresses.

In GPRS, UMTS and EPS networks, it is very desirable to reduce any additional periodic signaling. Therefore, the cellular host should include a mechanism in upper layer protocols to provide reachability confirmations when two-way IP layer reachability can be confirmed (see [RFC4861], Section 7.3.1). These confirmations would allow the suppression of NUD-related messages in most cases.

Host TCP implementation should provide reachability confirmation in the manner explained in [RFC4861], Section 7.3.1.

The widespread use of UDP in 3GPP networks poses a problem for providing reachability confirmation. As UDP itself is unable to provide such confirmation, applications running on top of UDP should provide the confirmation where possible. In particular, when UDP is used for transporting DNS, the DNS response should be used as a basis for reachability confirmation. Similarly, when UDP is used to transport RTP, the RTCP protocol feedback should be used as a basis for the reachability confirmation. If an RTCP packet is received with a reception report block indicating some packets have gone through, then packets are reaching the peer. If they have reached the peer, they have also reached the neighbor.

When UDP is used for transporting SIP, responses to SIP requests should be used as the confirmation that packets sent to the peer are reaching it. When the cellular host is acting as the server side SIP node, no such confirmation is generally available. However, a host may interpret the receipt of a SIP ACK request as confirmation that the previously sent response to a SIP INVITE request has reached the peer.

2.3. IPv6 Stateless Address Autoconfiguration

IPv6 Stateless Address Autoconfiguration is defined in [RFC4862]. This specification is a mandatory part of IPv6 and also the only mandatory method to configure an IPv6 address in a 3GPP cellular host.

2.4. Stateless Address Autoconfiguration in 3GPP Networks

A cellular host in a 3GPP network must process a Router Advertisement as stated in [RFC4862]. The Router Advertisement contains a maximum of one prefix information option and the advertised prefix cannot ever be used for on-link determination (see [RFC6459], Section 5.2).

Hosts in 3GPP networks can set DupAddrDetectTransmits equal to zero, as each delegated prefix is unique within its scope when advertised using the 3GPP IPv6 Stateless Address Autoconfiguration. In addition, the default router (GGSN/PGW) will not configure any addresses on its interfaces based on prefixes advertised to IPv6 cellular hosts on those interfaces. Thus, the host is not required to perform Duplicate Address Detection on the cellular interface.

Furthermore, the GGSN/PGW will provide the cellular host with an interface identifier that must be used for link-local address configuration. The link-local address configured from this interface

identifier is guaranteed not to collide with the link-local address that the GGSN/PGW uses. Thus, the cellular host is not required to perform Duplicate Address Detection for the link-local address either on the cellular interface.

See Appendix A for more details on 3GPP IPv6 Stateless Address Autoconfiguration.

2.5. IP version 6 over PPP in 3GPP Networks

A cellular host in a 3GPP network that supports PPP, must support the IPv6CP interface identifier option. This option is needed to be able to connect other devices to the Internet using a PPP link between the cellular device (MT) and other devices (TE, e.g., a laptop). The MT performs the PDP Context activation based on a request from the TE. This results in an interface identifier being suggested by the MT to the TE, using the IPv6CP option. To avoid any duplication in link-local addresses between the TE and the GGSN/PGW, the MT must always reject other suggested interface identifiers by the TE. This results in the TE always using the interface identifier suggested by the GGSN for its link-local address.

The rejection of interface identifiers suggested by the TE is only done for creation of link-local addresses, according to 3GPP specifications. The use of privacy addresses [RFC4941] for unique local IPv6 unicast addresses (ULA) [RFC4193] and global addresses is not affected by the above procedure. The above procedure is only concerned with assigning the interface identifier used for forming link-local addresses, and does not preclude TE from using other interface identifiers for addresses with larger scopes (i.e., ULAs and global).

2.6. MLD in 3GPP Networks

Within 3GPP networks, hosts connect to their default routers (GGSN/PGW) via point-to-point links. Moreover, there are exactly two IP devices connected to the point-to-point link, and no attempt is made (at the link-layer) to suppress the forwarding of multicast traffic. Consequently, sending MLD reports for link-local addresses in a 3GPP environment may not always be necessary.

MLD is needed for multicast group knowledge that is not link-local.

2.7. Privacy Extensions for Address Configuration in IPv6

Privacy Extensions for Stateless Address Autoconfiguration [RFC4941] should be supported. RFC 4941, and privacy in general, is important for the Internet. Cellular hosts may use the temporary addresses as

described in RFC 4941. However, the use of the Privacy Extension in an environment where IPv6 addresses are short-lived may not be necessary. At the time this document has been written, there is no experience on how long-lived cellular network address assignments (i.e., attachments to the network) are. The length of the address assignments depends upon many factors such as radio coverage, device status and user preferences. Additionally, the use of temporary address with IPsec may lead to more frequent renegotiation for the Security Associations.

Refer to Section 7 for a discussion of the benefits of privacy extensions in a 3GPP network.

2.8. Dynamic Host Configuration Protocol for IPv6 (DHCPv6)

The Dynamic Host Configuration Protocol for IPv6 (DHCPv6) [RFC3315] may be used. As of 3GPP Release-11 DHCPv6 is neither required nor supported for address autoconfiguration. The IPv6 stateless autoconfiguration still remains the only mandatory address configuration method. However, DHCPv6 may be useful for other configuration needs on a cellular host. e.g. Stateless DHCPv6 [RFC3736] may be used to configure DNS and SIP server addresses, and DHCPv6 prefix delegation [RFC3633] may be used to delegate a prefix to the cellular host for use on its non-cellular links.

2.9. DHCPv6 Prefix Delegation

Starting from Release-10 DHCPv6 Prefix Delegation was added as an optional feature to the 3GPP system architecture [RFC3633]. The prefix delegation model defined for Release-10 requires that the /64 IPv6 prefix assigned for the cellular host on the 3GPP link must aggregate with the shorter delegated IPv6 prefix. The cellular host should implement the Prefix Exclude Option for DHCPv6 Prefix Delegation [RFC6603] (see [RFC6459], Section 5.3 for further discussion).

2.10. Router preferences and more specific routes

The cellular host should implement the Default Router Preferences and More-Specific Routes extension to extension to Router Advertisement messages [RFC4191]. These options may be useful for cellular hosts that also have additional interfaces on which IPv6 is used.

2.11. Neighbor Discovery and additional host configuration

The DNS server configuration is learned from 3GPP link layer signaling. However, the cellular host should also implement the IPv6 Router Advertisement Options for DNS Configuration [RFC6106]. DHCPv6

is still optional for cellular hosts, and learning the DNS server addresses from the link layer signaling can be cumbersome when the MT and the TE are separated using other techniques than PPP interface.

The cellular host should also honor the MTU option in the Router Advertisement (see [RFC4861], Section 4.6.4). 3GPP system architecture uses extensive tunneling in its packet core network below the 3GPP link and this may lead to packet fragmentation issues. Therefore, the GGSN/PGW may propose a MTU to the cellular host that takes the additional tunneling overhead into account.

3. IP Security

IPsec [RFC4301] is a fundamental but not mandatory part of IPv6. Refer IPv6 Node Requirements Section 11 of [RFC6434] for the security requirements that also apply to cellular hosts.

3.1. Extension header considerations

The support for the Routing Header Type 0 (RH0) has been deprecated [RFC5095]. Therefore, the cellular host should as a default setting follow the RH0 processing described in Section 3 of RFC 5095.

IPv6 packet fragmentation has known security concerns. The cellular host must follow the handling of overlapping fragments as described in [RFC5722] and the cellular host must not fragment any neighbor discovery messages as described in [I-D.ietf-6man-nd-extension-headers].

4. Mobility

For the purposes of this document, IP mobility is not relevant. The movement of cellular hosts within 3GPP networks is handled by link layer mechanisms in majority of cases. 3GPP Release-8 introduced the dual-stack Mobile IPv6 (DSMIPv6) for a client based mobility [RFC5555]. Client based IP mobility is optional in 3GPP architecture.

5. IANA Considerations

This document has no IANA actions.

6. Acknowledgements

The authors would like to thank the original authors for their grounding work on this documents: Gerben Kuijpers, John Loughney, Hesham Soliman and Juha Wiljakka.

The original RFC 3316 document was based on the results of a team that included Peter Hedman and Pertti Suomela in addition to the authors. Peter and Pertti have contributed both text and their IPv6 experience to this document.

The authors would like to thank Jim Bound, Brian Carpenter, Steve Deering, Bob Hinden, Keith Moore, Thomas Narten, Erik Nordmark, Michael Thomas, Margaret Wasserman and others at the IPv6 WG mailing list for their comments and input.

We would also like to thank David DeCamp, Karim El Malki, Markus Isomaki, Petter Johnsen, Janne Rinne, Jonne Soininen, Vlad Stirbu and Shabnam Sultana for their comments and input in preparation of this document.

7. Security Considerations

This document does not specify any new protocols or functionality, and as such, it does not introduce any new security vulnerabilities. However, specific profiles of IPv6 functionality are proposed for different situations, and vulnerabilities may open or close depending on which functionality is included and what is not. There are also aspects of the cellular environment that make certain types of vulnerabilities more severe. The following issues are discussed:

- o The suggested limitations (Section 3.1) in the processing of extension headers limits also exposure to Denial-of-Service (DoS) attacks through cellular hosts.
- o IPv6 addressing privacy [RFC4941] may be used in cellular hosts. However, it should be noted that in the 3GPP model, the network would assign new addresses, in most cases, to hosts in roaming situations and typically, also when the cellular hosts activate a PDP context. This means that 3GPP networks will already provide a limited form of addressing privacy, and no global tracking of a single host is possible through its address. On the other hand, since a GGSN's coverage area is expected to be very large when compared to currently deployed default routers (no handovers between GGSNs are possible), a cellular host can keep an address for a long time. Hence, IPv6 addressing privacy can be used for additional privacy during the time the host is on and in the same area. The privacy features can also be used to e.g., make

different transport sessions appear to come from different IP addresses. However, it is not clear that these additional efforts confuse potential observers any further, as they could monitor only the network prefix part.

- o The use of various security services such as IPsec or TLS in the connection of typical applications in cellular hosts is discussed in Section 3 and further pointer for recommendations are given there.
- o The airtime used by cellular hosts is expensive. In some cases, users are billed according to the amount of data they transfer to and from their host. It is crucial for both the network and the users that the airtime is used correctly and no extra charges are applied to users due to misbehaving third parties. The cellular links also have a limited capacity, which means that they may not necessarily be able to accommodate more traffic than what the user selected, such as a multimedia call. Additional traffic might interfere with the service level experienced by the user. While Quality of Service mechanisms mitigate these problems to an extent, it is still apparent that DoS aspects may be highlighted in the cellular environment. It is possible for existing DoS attacks that use for instance packet amplification to be substantially more damaging in this environment. How these attacks can be protected against is still an area of further study. It is also often easy to fill the cellular link and queues on both sides with additional or large packets.
- o Within some service provider networks, it is possible to buy a prepaid cellular subscription without presenting personal identification. Attackers that wish to remain unidentified could leverage this. Note that while the user hasn't been identified, the equipment still is; the operators can follow the identity of the device and block it from further use. The operators must have procedures in place to take notice of third party complaints regarding the use of their customers' devices. It may also be necessary for the operators to have attack detection tools that enable them to efficiently detect attacks launched from the cellular hosts.
- o Cellular devices that have local network interfaces (such as WLAN or Bluetooth) may be used to launch attacks through them, unless the local interfaces are secured in an appropriate manner. Therefore, local network interfaces should have access control to prevent others from using the cellular host as an intermediary.
- o The 3GPP link model mitigates most of the known IPv6 on-link and neighbor cache targeted attacks (see Section 2.2 and Appendix A).
- o Advice for implementations in the face of Neighbor Discovery DoS attacks may be useful in some environments [RFC6583].
- o Section 9 of RFC 6459 discusses further some recent concerns related to cellular hosts security.

8. References

8.1. Normative references

- [I-D.ietf-6man-nd-extension-headers]
Gont, F., "Security Implications of the Use of IPv6 Extension Headers with IPv6 Neighbor Discovery", draft-ietf-6man-nd-extension-headers-00 (work in progress), June 2012.
- [RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.
- [RFC4213] Nordmark, E. and R. Gilligan, "Basic Transition Mechanisms for IPv6 Hosts and Routers", RFC 4213, October 2005.
- [RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, December 2005.
- [RFC4861] Narten, T., Nordmark, E., Simpson, W., and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, September 2007.
- [RFC4862] Thomson, S., Narten, T., and T. Jinmei, "IPv6 Stateless Address Autoconfiguration", RFC 4862, September 2007.
- [RFC4941] Narten, T., Draves, R., and S. Krishnan, "Privacy Extensions for Stateless Address Autoconfiguration in IPv6", RFC 4941, September 2007.
- [RFC5095] Abley, J., Savola, P., and G. Neville-Neil, "Deprecation of Type 0 Routing Headers in IPv6", RFC 5095, December 2007.
- [RFC5722] Krishnan, S., "Handling of Overlapping IPv6 Fragments", RFC 5722, December 2009.
- [RFC6434] Jankiewicz, E., Loughney, J., and T. Narten, "IPv6 Node Requirements", RFC 6434, December 2011.

8.2. Informative references

- [RFC3315] Droms, R., Bound, J., Volz, B., Lemon, T., Perkins, C., and M. Carney, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", RFC 3315, July 2003.
- [RFC3633] Troan, O. and R. Droms, "IPv6 Prefix Options for Dynamic Host Configuration Protocol (DHCP) version 6", RFC 3633,

December 2003.

- [RFC3736] Droms, R., "Stateless Dynamic Host Configuration Protocol (DHCP) Service for IPv6", RFC 3736, April 2004.
- [RFC4191] Draves, R. and D. Thaler, "Default Router Preferences and More-Specific Routes", RFC 4191, November 2005.
- [RFC4193] Hinden, R. and B. Haberman, "Unique Local IPv6 Unicast Addresses", RFC 4193, October 2005.
- [RFC5555] Soliman, H., "Mobile IPv6 Support for Dual Stack Hosts and Routers", RFC 5555, June 2009.
- [RFC6106] Jeong, J., Park, S., Beloeil, L., and S. Madanapalli, "IPv6 Router Advertisement Options for DNS Configuration", RFC 6106, November 2010.
- [RFC6459] Korhonen, J., Soininen, J., Patil, B., Savolainen, T., Bajko, G., and K. Iisakkila, "IPv6 in 3rd Generation Partnership Project (3GPP) Evolved Packet System (EPS)", RFC 6459, January 2012.
- [RFC6583] Gashinsky, I., Jaeggli, J., and W. Kumari, "Operational Neighbor Discovery Problems", RFC 6583, March 2012.
- [RFC6603] Korhonen, J., Savolainen, T., Krishnan, S., and O. Troan, "Prefix Exclude Option for DHCPv6-based Prefix Delegation", RFC 6603, May 2012.

Appendix A. Cellular Host IPv6 Addressing in the 3GPP Model

The appendix aims to very briefly describe the 3GPP IPv6 addressing model for 2G (GPRS), 3G (UMTS) and 4G (EPS) cellular networks from Release-99 onwards. More information for 2G and 3G can be found from 3GPP Technical Specifications 23.060 and T29.061. The equivalent documentation for 4G can be found from 3GPP Technical Specifications 23.401, 23.402 and 29.061.

There are two possibilities to allocate the address for an IPv6 node: stateless and stateful autoconfiguration. The stateful address allocation mechanism needs a DHCP server to allocate the address for the IPv6 node. On the other hand, the stateless autoconfiguration procedure does not need any external entity involved in the address autoconfiguration (apart from the GGSN/PGW). At the time of writing this document, the IPv6 stateless address autoconfiguration mechanism is still the only mandatory and supported address configuration method

for the cellular 3GPP link.

In order to support the standard IPv6 stateless address autoconfiguration mechanism as recommended by the IETF, the GGSN/PGW shall assign a prefix that is unique within its scope to each primary PDP context that uses IPv6 stateless address autoconfiguration. This avoids the necessity to perform Duplicate Address Detection (DAD) at the network level for every address built by the mobile host. The GGSN/PGW always provides an Interface Identifier to the mobile host. The Mobile host uses the interface identifier provided by the GGSN to generate its link-local address. The GGSN/PGW provides the cellular host with the interface identifier, usually in a random manner. It must ensure the uniqueness of such identifier on the link (i.e., no collisions between its own link-local address and the cellular host's).

In addition, the GGSN/PGW will not use any of the prefixes assigned to cellular hosts to generate any of its own addresses. This use of the interface identifier, combined with the fact that each PDP Context or PDN Connection is allocated a unique prefix, will eliminate the need for DAD messages over the air interface, and consequently reduces inefficient use of radio resources. Furthermore, the allocation of a prefix to each PDP context will allow hosts to implement the privacy extensions in RFC 4941 without the need for further DAD messages.

In practice, the GGSN/PGW only needs to route all traffic to the cellular host that fall under the prefix assigned to it. This implies the GGSN/PGW may implement a minimal neighbor discovery protocol subset; since, due the point-to-point link model and the absence of link-layer addressing the address resolution can be entirely statically configured per PDP Context or PDN Connection, and there is no need to defend any other address than the link-local address for very unlikely duplicates.

See Sections 5 of RFC 6459 for further discussion on 3GPP address allocation and link model.

Appendix B. Changes to RFC 3316

B.1. Version -00

- o Removal of all sections that can be directly found from RFC 6434.
- o Clarifications to 3GPP link model and how Neighbor Discovery works on it.

- o Addition of RFC 4191 recommendations.
- o Addition of DHCPv6-based Prefix Delegation recommendations.
- o Addition of RFC 6106 recommendations.
- o Addition of RFC 5555 regarding client based mobility.
- o Addition of Router Advertisement MTU option handling.
- o Addition of Evolved Packet System text.
- o Clarification on the primary 3GPP IPv6 transition mechanism.
- o Addition of RFC 5095 that deprecates the RH0
- o Addition of RFC 5722 and draft-ietf-6man-nd-extension-headers regarding the IPv6 fragmentation handling.
- o Addition of RFC 6583 for Neighbor Discovery denial-of-service attack considerations.
- o Made the PPP IPV6CP support text conditional.

Authors' Addresses

Jouni Korhonen (editor)
Nokia Siemens Networks
Linnoitustie 6
FIN-02600 Espoo
Finland

Email: jouni.nospam@gmail.com

Jari Arkko (editor)
Ericsson
Jorvas 02420
Finland

Email: jari.arkko@piuha.net

Teemu Savolainen
Nokia
Hermiankatu 12 D
FI-33720 Tampere
FINLAND

Email: teemu.savolainen@nokia.com

Suresh Krishnan
Ericsson
8400 Decarie Blvd.
Town of Mount Royal, QC
Canada

Phone: +1 514 345 7900 x42871
Email: suresh.krishnan@ericsson.com

v6ops
Internet-Draft
Intended status: Informational
Expires: January 15, 2014

D. Lopez
Telefonica I+D
Z. Chen
China Telecom
T. Tsou
Huawei Technologies (USA)
C. Zhou
Huawei Technologies
A. Servin
LACNIC
July 14, 2013

IPv6 Operational Guidelines for Datacenters
draft-lopez-v6ops-dc-ipv6-05

Abstract

This document is intended to provide operational guidelines for datacenter operators planning to deploy IPv6 in their infrastructures. It aims to offer a reference framework for evaluating different products and architectures, and therefore it is also addressed to manufacturers and solution providers, so they can use it to gauge their solutions. We believe this will translate in a smoother and faster IPv6 transition for datacenters of these infrastructures.

The document focuses on the DC infrastructure itself, its operation, and the aspects related to DC interconnection through IPv6. It does not consider the particular mechanisms for making Internet services provided by applications hosted in the DC available through IPv6 beyond the specific aspects related to how their deployment on the Data Center (DC) infrastructure.

Apart from facilitating the transition to IPv6, the mechanisms outlined here are intended to make this transition as transparent as possible (if not completely transparent) to applications and services running on the DC infrastructure, as well as to take advantage of IPv6 features to simplify DC operations, internally and across the Internet.

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 15, 2014.

Copyright Notice

Copyright (c) 2013 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	4
2. Architecture and Transition Stages	4
2.1. General Architecture	5
2.2. Experimental Stage. Native IPv4 Infrastructure	7
2.2.1. Off-shore v6 Access	8
2.3. Dual Stack Stage. Internal Adaptation	8
2.3.1. Dual-stack at the Aggregation Layer	10
2.3.2. Dual-stack Extended OS/Hypervisor	12
2.4. IPv6-Only Stage. Pervasive IPv6 Infrastructure	12
3. Other Operational Considerations	13
3.1. Addressing	13
3.2. Management Systems and Applications	14
3.3. Monitoring and Logging	15
3.4. Costs	15
4. Security Considerations	15
4.1. Neighbor Discovery Protocol attacks	16
4.2. Addressing	16
4.3. Edge filtering	17
4.4. Final Security Remarks	17
5. IANA Considerations	17
6. Acknowledgements	17
7. Informative References	17
Authors' Addresses	19

1. Introduction

The need for considering the aspects related to IPv4-to-IPv6 transition for all devices and services connected to the Internet has been widely mentioned elsewhere, and it is not our intention to make an additional call on it. Just let us note that many of those services are already or will soon be located in Data Centers (DC), what makes considering the issues associated to DC infrastructure transition a key aspect both for these infrastructures themselves, and for providing a simpler and clear path to service transition.

All issues discussed here are related to DC infrastructure transition, and are intended to be orthogonal to whatever particular mechanisms for making the services hosted in the DC available through IPv6 beyond the specific aspects related to their deployment on the infrastructure. General mechanisms related to service transition have been discussed in depth elsewhere (see, for example [I-D.ietf-v6ops-icp-guidance] and [I-D.ietf-v6ops-enterprise-incremental-ipv6]) and are considered to be independent to the goal of this discussion. The applicability of these general mechanisms for service transition will, in many cases, depend on the supporting DC's infrastructure characteristics. However, this document intends to keep both problems (service vs. infrastructure transition) as different issues.

Furthermore, the combination of the regularity and controlled management in a DC interconnection fabric with IPv6 universal end-to-end addressing should translate in simpler and faster VM migrations, either intra- or inter-DC, and even inter-provider.

2. Architecture and Transition Stages

This document presents a transition framework structured along transition stages and operational guidance associated with the degree of penetration of IPv6 into the DC communication fabric. It is worth noting we are using these stages as a classification mechanism, and they have not to be associated with any a succession of steps from a v4-only infrastructure to full-fledged v6, but to provide a framework that operators, users, and even manufacturers could use to assess their plans and products.

There is no (explicit or implicit) requirement on starting at the stage describe in first place, nor to follow them in successive order. According to their needs and the available solutions, DC operators can choose to start or remain at a certain stage, and freely move from one to another as they see fit, without contravening this document. In this respect, the classification intends to

support the planning in aspects such as the adaptation of the different transition stages to the evolution of traffic patterns, or risk assessment in what relates to deploying new components and incorporating change control, integration and testing in highly-complex multi-vendor infrastructures.

Three main transition stages can be considered when analyzing IPv6 deployment in the DC infrastructure, all compatible with the availability of services running in the DC through IPv6:

- o Experimental. The DC keeps a native IPv4 infrastructure, with gateway routers (or even application gateways when services require so) performing the adaptation to requests arriving from the IPv6 Internet.
- o Dual stack. Native IPv6 and IPv4 are present in the infrastructure, up to whatever the layer in the interconnection scheme where L3 is applied to packet forwarding.
- o IPv6-Only. The DC has a fully pervasive IPv6 infrastructure, including full IPv6 hypervisors, which perform the appropriate tunneling or NAT if required by internal applications running IPv4.

2.1. General Architecture

The diagram in Figure 1 depicts a generalized interconnection schema in a DC.

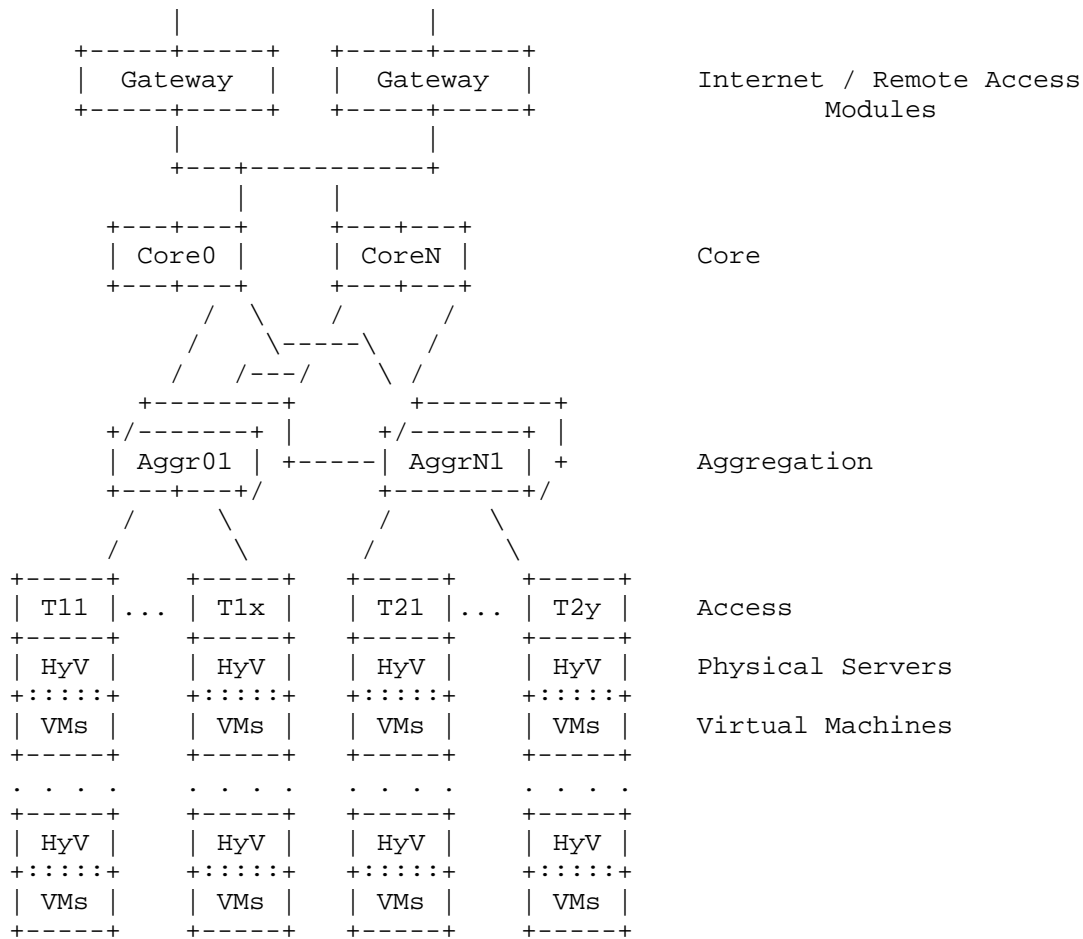


Figure 1: DC Interconnection Schema

- o Hypervisors provide connection services (among others) to virtual machines running on physical servers.
- o Access elements provide connectivity directly to/from physical servers. The access elements are typically placed either top-of-rack (ToR) or end-of-row(EoR).
- o Aggregation elements group several (many) physical racks to achieve local integration and provide as much structure as possible to data paths.
- o Core elements connect all aggregation elements acting as the DC backbone.

- o One or several gateways connecting the DC to the Internet, Branch Offices, Partners, Third-Parties, and/or other DCs. The interconnectivity to other DC may be in the form of VPNs, WAN links, metro links or any other form of interconnection.

In many actual deployments, depending on DC size and design decisions, some of these elements may be combined (core and gateways are provided by the same routers, or hypervisors act as access elements) or virtualized to some extent, but this layered schema is the one that best accommodates the different options to use L2 or L3 at any of the different DC interconnection layers, and will help us in the discussion along the document.

2.2. Experimental Stage. Native IPv4 Infrastructure

This transition stage corresponds to the first step that many datacenters may take (or have taken) in order to make their external services initially accessible from the IPv6 Internet and/or to evaluate the possibilities around it, and corresponds to IPv6 traffic patterns totally originated out of the DC or their tenants, being a small percentage of the total external requests. At this stage, DC network scheme and addressing do not require any important change, if any.

It is important to remark that in no case this can be considered a permanent stage in the transition, or even a long-term solution for incorporating IPv6 into the DC infrastructure. This stage is only recommended for experimentation or early evaluation purposes.

The translation of IPv6 requests into the internal infrastructure addressing format occurs at the outmost level of the DC Internet connection. This can be typically achieved at the DC gateway routers, that support the appropriate address translation mechanisms for those services required to be accessed through native IPv6 requests. The policies for applying adaptation can range from performing it only to a limited set of specified services to providing a general translation service for all public services. More granular mechanisms, based on address ranges or more sophisticated dynamic policies are also possible, as they are applied by a limited set of control elements. These provide an additional level of control to the usage of IPv6 routable addresses in the DC environment, which can be especially significant in the experimentation or early deployment phases this stage is applicable to.

Even at this stage, some implicit advantages of IPv6 application come into play, even if they can only be applied at the ingress elements:

- o Flow labels can be applied to enhance load-balancing, as described in [I-D.ietf-6man-flow-ecmp]. If the incoming IPv6 requests are adequately labeled the gateway systems can use the flow labels as a hint for applying load-balancing mechanisms when translating the requests towards the IPv4 internal network.
- o During VM migration (intra- or even inter-DC), Mobile IP mechanisms can be applied to keep service availability during the transient state.

2.2.1. Off-shore v6 Access

This model is also suitable to be applied in an "off-shore" mode by the service provider connecting the DC infrastructure to the Internet, as described in [I-D.sunq-v6ops-contents-transition].

When this off-shore mode is applied, the original source address will be hidden to the DC infrastructure, and therefore identification techniques based on it, such as geolocation or reputation evaluation, will be hampered. Unless there is a specific trust link between the DC operator and the ISP, and the DC operator is able to access equivalent identification interfaces provided by the ISP as an additional service, the off-shore experimental stage cannot be considered applicable when source address identification is required.

2.3. Dual Stack Stage. Internal Adaptation

This stage requires dual-stack elements in some internal parts of the DC infrastructure. This brings some degree of partition in the infrastructure, either in a horizontal (when data paths or management interfaces are migrated or left in IPv4 while the rest migrate) or a vertical (per tenant or service group), or even both.

Although it may seem an artificial case, situations requiring this stage can arise from different requirements from the user base, or the need for technology changes at different points of the infrastructure, or even the goal of having the possibility of experimenting new solutions in a controlled real-operations environment, at the price of the additional complexity of dealing with a double protocol stack, as noted in [I-D.ietf-v6ops-icp-guidance] and elsewhere.

This transition stage can accommodate different traffic patterns, both internal and external, though it better fits to scenarios of a clear differentiation of different types of traffic (external vs. internal, data vs management...), and/or a more or less even distribution of external requests. A common scenario would include native dual stack servers for certain services combined with single

stack ones for others (web server in dual stack and database servers only supporting v4, for example).

At this stage, the advantages outlined above on load balancing based on flow labels and Mobile IP mechanisms are applicable to any L3-based mechanism (intra- as well as inter-DC). They will translate into enhanced VM mobility, more effective load balancing, and higher service availability. Furthermore, the simpler integration provided by IPv6 to and from the L2 flat space to the structured L3 one can be applied to achieve simpler deployments, as well as alleviating encapsulation and fragmentation issues when traversing between L2 and L3 spaces. With an appropriate prefix management, automatic address assignment, discovery, and renumbering can be applied not only to public service interfaces, but most notably to data and management paths.

Other potential advantages include the application of multicast scopes to limit broadcast floods, and the usage of specific security headers to enhance tenant differentiation.

On the other hand, this stage requires a much more careful planning of addressing (please refer to ([RFC5375]) schemas and access control, according to security levels. While the experimental stage implies relatively few global routable addresses, this one brings the advantages and risks of using different kinds of addresses at each point of the IPv6-aware infrastructure.

2.3.1. Dual-stack at the Aggregation Layer

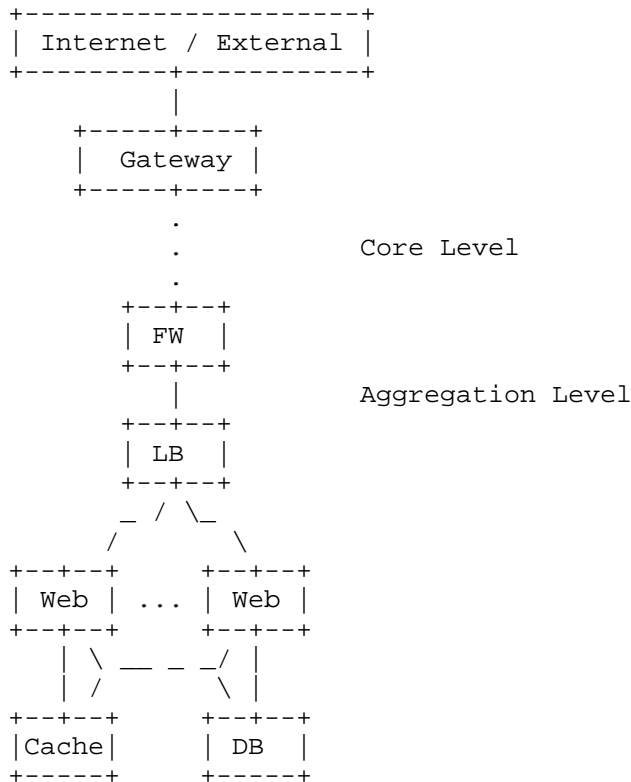


Figure 2: Data Center Application Scheme

An initial approach corresponding to this transition stage relies on taking advantage of specific elements at the aggregation layer described in Figure 1, and make them able to provide dual-stack gatewaying to the IPv4-based servers and data infrastructure.

Typically, firewalls (FW) are deployed as the security edge of the whole service domain and provides safe access control of this service domain from other function domains. In addition, some application optimization based on devices and security devices (e.g. Load Balancers, SSL VPN, IPS and etc.) may be deployed in the aggregation level to alleviate the burden of the server and to guarantee deep security, as shown in Figure 2.

The load balancer (LB) or some other boxes could be upgraded to support the data transmission. There may be two ways to achieve this

at the edge of the DC: Encapsulation and NAT. In the encapsulation case, the LB function carries the IPv6 traffic over IPv4 using an encapsulation (IPv6-in-IPv4). In the NAT case, there are already some technologies to solve this problem. For example, DNS and NAT device could be concatenated for IPv4/IPv6 translation if IPv6 host needs to visit IPv4 servers. However, this may require the concatenation of multiple network devices, which means the NAT tables needs to be synchronized at different devices. As described below, a simplified IPv4/IPv6 translation model can be applied, which could be implemented in the LB device. The mapping information of IPv4 and IPv6 will be generated automatically based on the information of the LB. The host IP address will be translated without port translation.

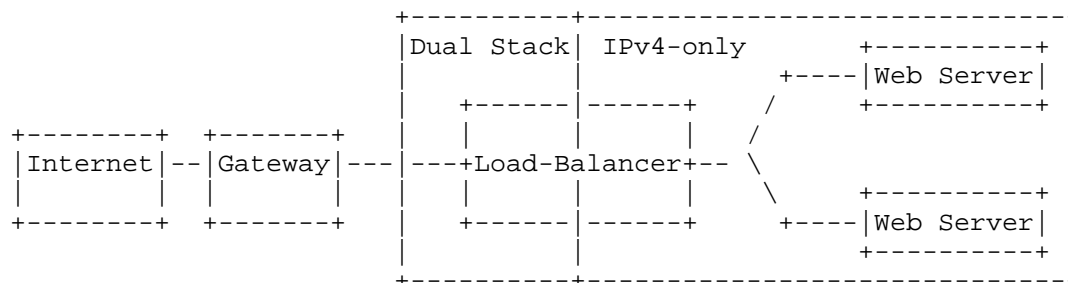


Figure 3: Dual Stack LB mechanism

As shown in Figure 3, the LB can be considered divided into two parts: The dual-stack part facing the external border, and the IPv4-only part which contains the traditional LB functions. The IPv4 DC is allocated an IPv6 prefix which is for the VSIPv6 (Virtual Service IPv6 Address). We suggest that the IPv6 prefix is not the well-known prefix in order to avoid the IPv4 routings of the services in different DCs spread to the IPv6 network. The VSIPv4 (Virtual Service IPv4 Address) is embedded in VSIPv6 using the allocated IPv6 prefix. In this way, the LB has the stateless IP address mapping between VSIPv6 and VSIPv4, and synchronization is not required between LB and DNS64 server.

The dual-stack part of the LB has a private IPv4 address pool. When IPv6 packets arrive, the dual-stack part does the one-on-one SIP (source IP address) mapping (as defined in [I-D.sunq-v6ops-contents-transition]) between IPv4 private address and IPv6 SIP. Because there will be too many UDP/TCP sessions between the DC and Internet, the IP addresses binding tables between IPv6 and IPv4 are not session-based, but SIP-based. Thus, the dual-stack part of LB builds IP binding stateful tables for the host IPv6 address and private IPv4 address of the pool. When the following

IPv6 packets of the host come from Internet to the LB, the dual stack part does the IP address translation for the packets. Thus, the IPv6 packets were translated to IPv4 packets and sent to the IPv4 only part of the LB.

2.3.2. Dual-stack Extended OS/Hypervisor

Another option for deploying a infrastructure at the dual-stack stage would bring dual-stack much closer to the application servers, by requiring hypervisors, VMs and applications in the v6-capable zone of the DC to be able to operate in dual stack. This way, incoming connections would be dealt in a seamless manner, while for outgoing ones an OS-specific replacement for system calls like `gethostbyname()` and `getaddrinfo()` would accept a character string (an IPv4 literal, an IPv6 literal, or a domain name) and would return a connected socket or an error message, having executed a happy eyeballs algorithm ([RFC6555]).

If these hypothetical system call replacements were smart enough, they would allow the transparent interoperation of DCs with different levels of v6 penetration, either horizontal (internal data paths are not migrated, for example) or vertical (per tenant or service group). This approach requires, on the other hand, all the involved DC infrastructure to become dual-stack, as well as some degree of explicit application adaptation.

2.4. IPv6-Only Stage. Pervasive IPv6 Infrastructure

We can consider a DC infrastructure at the final stage when all network layer elements, including hypervisors, are IPv6-aware and apply it by default. Conversely with the experimental stage, access from the IPv4 Internet is achieved, when required, by protocol translation performed at the edge infrastructure elements, or even supplied by the service provider as an additional network service.

There are different drivers that could motivate DC managers to transition to this stage. In principle the scarcity of IPv4 addresses may require to reclaim IPv4 resources from portions of the network infrastructure which no longer need them. Furthermore, the unavailability of IPv4 address would make dual-stack environments not possible anymore and careful assessments will be perfumed to asses where to use the remaining IPv4 resources.

Another important motivation to move DC operations from dual-stack to IPv6-only is to save costs and operation activities that managing a single-stack network could bring in comparison with managing two stacks. Today, besides of learning to manage two different stacks, network and system administrators require to duplicate other tasks

such as IP address management, firewalls configuration, system security hardening and monitoring among others. These activities are not just costly for the DC management, they may also may lead to configuration errors and security holes.

This stage can be also of interest for new deployments willing to apply a fresh start aligned with future IPv6 widespread usage, when a relevant amount of requests are expected to be using IPv6, or to take advantage of any of the potential benefits that an IPv6 support infrastructure can provide. Other, and probably more compelling in many cases, drivers for this stage may be either a lack of enough IPv4 resources (whether private or globally unique) or a need to reclaim IPv4 resources from portions of the network which no longer need them. In these circumstances, a careful evaluation of what still needs to speak IPv4 and what does not will need to happen to ensure judicious use of the remaining IPv4 resources.

The potential advantages mentioned for the previous stages (load balancing based on flow labels, mobility mechanisms for transient states in VM or data migration, controlled multicast, and better mapping of L2 flat space on L3 constructs) can be applied at any layer, even especially tailored for individual services. Obviously, the need for a careful planning of address space is even stronger here, though the centralized protocol translation services should reduce the risk of translation errors causing disruptions or security breaches.

[V6DCS] proposes an approach to a next generation DC deployment, already demonstrated in practice, and claims the advantages of materializing the stage from the beginning, providing some rationale for it based on simplifying the transition process. It relies on stateless NAT64 ([RFC6052], [RFC6145]) to enable access from the IPv4 Internet.

3. Other Operational Considerations

In this section we review some operation considerations related addressing and management issues in V6 DC infrastructure.

3.1. Addressing

There are different considerations related on IPv6 addressing topics in DC. Many of these considerations are already documented in a variety of IETF documents and in general the recommendations and best practices mentioned on them apply in IPv6 DC environments. However we would like to point out some topics that we consider important to mention.

The first question that DC managers often have is the type of IPv6 address to use; that is Provider Aggregated (PA), Provider Independent (PI) or Unique Local IPv6 Addresses (ULAs) [RFC4193]. Related to the use of PA vs. PI, we concur with [I-D.ietf-v6ops-icp-guidance] and [I-D.ietf-v6ops-enterprise-incremental-ipv6] that PI provides independence from the ISP and decreases renumbering issues, it may bring up other considerations as a fee for the allocation, a request process and allocation maintenance to the Regional Internet Registry, etc. In this respect, there is not a specific recommendation to use either PI vs. PA as it would depend also on business and management factors rather than pure technical.

ULAs should be used only in DC infrastructure that does not require access to the public Internet; such devices may be databases servers, application-servers, and management interfaces of web servers and network devices among others. This practice may decrease the renumbering issues when PA addressing is used, as only public faced devices would require an address change. Also we would like to know that although ULAs may provide some security the main motivation for it used should be address management.

Another topic to discuss is the length of prefixes within the DC. In general we recommend the use of subnets of 64 bits for each vlan or network segment used in the DC. Although subnet with prefixes longer than 64 bits may work, it is necessary that the reader understand that this may break stateless autoconfiguration and at least manual configuration must be employed. For details please read [RFC5375].

Address plans should follow the principles of being hierarchical and able to aggregate address space. We recommend at least to have a /48 for each data-center. If the DC provides services that require subassignment of address space we do not offer a single recommendation (i.e. request a /40 prefix from an RIR or ISP and assign /48 prefixes to customers), as this may depend on other no technical factors. Instead we refer the reader to [RFC6177].

For point-to-point links please refer to the recommendations in [RFC6164].

3.2. Management Systems and Applications

Data-centers may use Internet Protocol address management (IPAM) software, provisioning systems and other variety of software to document and operate. It is important that these systems are prepared and possibly modified to support IPv6 in their data models. In general, if IPv6 support for these applications has not been previously done, changes may take sometime as they may be not just

adding more space in input fields but also modifying data models and data migration.

3.3. Monitoring and Logging

Monitoring and logging are critical operations in any network environment and they should be carried at the same level for IPv6 and IPv4. Monitoring and management operations in V6 DC are by no means different than any other IPv6 networks environments. It is important to consider that the collection of information from network devices is orthogonal to the information collected. For example it is possible to collect data from IPv6 MIBs using IPv4 transport. Similarly it is possible to collect IPv6 data generated by Netflow9/IPFIX agents in IPv4 transport. In this way the important issue to address is that agents (i.e. network devices) are able to collect data specific to IPv6.

And as final note on monitoring, although IPv6 MIBs are supported by SNMP versions 1 and 2, we recommend to use SNMP version 3 instead.

3.4. Costs

It is very possible that moving from a single stack data-center infrastructure to any of the IPv6 stages described in this document may incur in capital expenditures. This may include but it is not confined to routers, load-balancers, firewalls and software upgrades among others. However the cost that most concern us is operational. Moving the DC infrastructure operations from a single-stack to a dual-stack may infer in a variety of extra costs such as application development and testing, operational troubleshooting and service deployment. At the same time, this extra cost may be seeing as saving when moving from a dual-stack DC to an IPv6-Only DC.

Depending of the complexity of the DC network, provisioning and other factors we estimate that the extra costs (and later savings) may be around between 15 to 20%.

4. Security Considerations

A thorough collection of operational security aspects for IPv6 network is made in [I-D.ietf-opsec-v6] . Most of them, with the probable exception of those specific to residential users, are applicable in the environment we consider in this document.

4.1. Neighbor Discovery Protocol attacks

The first important issue that V6 DC manager should be aware is the attacks against Neighbor Discovery Protocol [RFC6583]. This attack is similar to ARP attacks [RFC4732] in IPv4 but exacerbated by the fact that the common size of an IPv6 subnet is /64. In principle an attacker would be able to fill the Neighbor Cache of the local router and starve its memory and processing resources by sending multiple ND packets requesting information of non-existing hosts. The result would be the inability of the router to respond to ND requests, to update its Neighbor Cache and even to forward packets. The attack does need to be launched with malicious purposes; it could be just the result of bad stack implementation behavior.

R[RFC6583] mentions some options to mitigate the effects of the attacks against NDP. For example filtering unused space, minimizing subnet size when possible, tuning rate limits in the NDP queue and to rely in router vendor implementations to better handle resources and to prioritize NDP requests.

4.2. Addressing

Other important security considerations in V6 DC are related to addressing. Because of the large address space is commonly thought that IPv6 is not vulnerable to reconnaissance techniques such as scanning. Although that may be true to force brute attacks, [I-D.ietf-opsec-ipv6-host-scanning] shows some techniques that may be employed to speed up and improve results in order to discover IPv6 address in a subnet. The use of virtual machines and SLACC aggravate this problem due the fact that they tend to use automatically-generated MAC address well known patterns.

To mitigate address-scanning attacks it is recommended to avoid using SLAAC and if used stable privacy-enhanced addresses [I-D.ietf-6man-stable-privacy-addresses] should be the method of address generation. Also, for manually assigned addresses try to avoid IID low-byte address (i.e. from 0 to 256), IPv4-based addresses and wordy addresses especially for infrastructure without a fully qualified domain name.

In spite of the use of manually assigned addresses is the preferred method for V6 DC, SLACC and DHCPv6 may be also used for some special reasons. However we recommend paying special attention to RA [RFC6104] and DHCP [I-D.gont-opsec-dhcpv6-shield] hijack attacks. In these kinds of attacks the attacker deploys rogue routers sending RA messages or rogue DHCP servers to inject bogus information and possibly to perform a man in the middle attack. In order to mitigate this problem it is necessary to apply some techniques in access

switches such as RA-Guard [RFC6105] at least.

Another topic that we would like to mention related to addressing is the use of ULAs. As we previously mentioned, although ULAs may be used to hide host from the outside world we do not recommend to rely on them as a security tool but better as a tool to make renumbering easier.

4.3. Edge filtering

In order to avoid being used as a source of amplification attacks is it important to follow the rules of BCP38 on ingress filtering. At the same time it is important to filter-in on the network border all the unicast traffic and routing announcement that should not be routed in the Internet, commonly known as "bogus prefixes".

4.4. Final Security Remarks

Finally, let us just emphasize the need for careful configuration of access control rules at the translation points. This latter one is specially sensitive in infrastructures at the dual-stack stage, as the translation points are potentially distributed, and when protocol translation is offered as an external service, since there can be operational mismatches.

5. IANA Considerations

None.

6. Acknowledgements

We would like to thank Tore Anderson, Wes George, Ray Hunter, Joel Jaeggli, Fred Baker, Lorenzo Colitti, Dan York, Carlos Martinez, Lee Howard, Alejandro Acosta, Alexis Munoz, Nicolas Fiumarelli, Santiago Aggio and Hans Velez for their questions, suggestions, reviews and comments.

7. Informative References

[I-D.gont-opsec-dhcpv6-shield]
Gont, F. and W. Liu, "DHCPv6-Shield: Protecting Against Rogue DHCPv6 Servers", draft-gont-opsec-dhcpv6-shield-01 (work in progress), October 2012.

[I-D.ietf-6man-flow-ecmp]

Carpenter, B. and S. Amante, "Using the IPv6 flow label for equal cost multipath routing and link aggregation in tunnels", draft-ietf-6man-flow-ecmp-05 (work in progress), July 2011.

[I-D.ietf-6man-stable-privacy-addresses]

Gont, F., "A method for Generating Stable Privacy-Enhanced Addresses with IPv6 Stateless Address Autoconfiguration (SLAAC)", draft-ietf-6man-stable-privacy-addresses-10 (work in progress), June 2013.

[I-D.ietf-opsec-ipv6-host-scanning]

Gont, F. and T. Chown, "Network Reconnaissance in IPv6 Networks", draft-ietf-opsec-ipv6-host-scanning-01 (work in progress), April 2013.

[I-D.ietf-opsec-v6]

Chittimaneni, K., Kaeo, M., and E. Vyncke, "Operational Security Considerations for IPv6 Networks", draft-ietf-opsec-v6-02 (work in progress), February 2013.

[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-ipv6-03 (work in progress), July 2013.

[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-05 (work in progress), January 2013.

[I-D.sunq-v6ops-contents-transition]

Sun, Q., Liu, D., Zhao, Q., Liu, Q., Xie, C., Li, X., and J. Qin, "Rapid Transition of IPv4 contents to be IPv6-accessible", draft-sunq-v6ops-contents-transition-03 (work in progress), March 2012.

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

[RFC4732] Handley, M., Rescorla, E., and IAB, "Internet Denial-of-Service Considerations", RFC 4732, December 2006.

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

Considerations", RFC 5375, December 2008.

- [RFC6052] Bao, C., Huitema, C., Bagnulo, M., Boucadair, M., and X. Li, "IPv6 Addressing of IPv4/IPv6 Translators", RFC 6052, October 2010.
- [RFC6104] Chown, T. and S. Venaas, "Rogue IPv6 Router Advertisement Problem Statement", RFC 6104, February 2011.
- [RFC6105] Levy-Abegnoli, E., Van de Velde, G., Popoviciu, C., and J. Mohacsi, "IPv6 Router Advertisement Guard", RFC 6105, February 2011.
- [RFC6145] Li, X., Bao, C., and F. Baker, "IP/ICMP Translation Algorithm", RFC 6145, April 2011.
- [RFC6164] Kohno, M., Nitzan, B., Bush, R., Matsuzaki, Y., Colitti, L., and T. Narten, "Using 127-Bit IPv6 Prefixes on Inter-Router Links", RFC 6164, April 2011.
- [RFC6177] Narten, T., Huston, G., and L. Roberts, "IPv6 Address Assignment to End Sites", BCP 157, RFC 6177, March 2011.
- [RFC6555] Wing, D. and A. Yourtchenko, "Happy Eyeballs: Success with Dual-Stack Hosts", RFC 6555, April 2012.
- [RFC6583] Gashinsky, I., Jaeggli, J., and W. Kumari, "Operational Neighbor Discovery Problems", RFC 6583, March 2012.
- [V6DCS] "The case for IPv6-only data centres", <https://ripe64.ripe.net/presentations/67-20120417-RIPE64-The_Case_for_IPv6_Only_Data_Centres.pdf>.

Authors' Addresses

Diego R. Lopez
Telefonica I+D
Don Ramon de la Cruz, 84
Madrid 28006
Spain

Phone: +34 913 129 041
Email: diego@tid.es

Zhonghua Chen
China Telecom
P.R.China

Phone:
Email: 18918588897@189.cn

Tina Tsou
Huawei Technologies (USA)
2330 Central Expressway
Santa Clara, CA 95050
USA

Phone: +1 408 330 4424
Email: Tina.Tsou.Zouting@huawei.com

Cathy Zhou
Huawei Technologies
Bantian, Longgang District
Shenzhen 518129
P.R. China

Phone:
Email: cathy.zhou@huawei.com

Arturo Servin
LACNIC
Rambla Republica de Mexico 6125
Montevideo 11300
Uruguay

Phone: +598 2604 2222
Email: aservin@lacnic.net

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

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 April 25, 2013.

Copyright Notice

Copyright (c) 2012 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	3
2. Design Choices	3
2.1. Mix IPv4 and IPv6 on the Same Link?	4
2.2. Links with Only Link-Local Addresses?	4
2.3. Link-Local Next-Hop in a Static Route?	6
2.4. Separate or Combined eBGP Sessions?	6
2.5. eBGP Endpoints: Global or Link-Local Addresses?	7
3. General Observations	8
3.1. Use of Link-Local Addresses	8
3.2. Separation of IPv4 and IPv6	9
4. IANA Considerations	9
5. Security Considerations	10
6. Acknowledgements	10
7. History	10
8. Informative References	10
Author's Address	12

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:

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

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:

- 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 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-to-impossible when using LLAs only.
- o 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:

- 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 redistributing 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 an extra difference 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 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, 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

Internet Engineering Task Force
Internet-Draft
Updates: 4291,5156,6303,6724
(if approved)
Intended status: Standards Track
Expires: August 24, 2013

M. Smith
IMOT
February 20, 2013

A Larger Loopback Prefix for IPv6
draft-smith-v6ops-larger-ipv6-loopback-prefix-04

Abstract

During the development and testing of a network application, it can be useful to run multiple instances of the application using the same transport layer protocol port on the same development host, while also having network access to the application instances limited to the local host. Under IPv4, this has commonly been possible by using different loopback addresses within 127/8. It is not possible under IPv6, as the loopback prefix of ::1/128 only provides a single loopback address. This memo proposes a new larger loopback prefix that will provide many IPv6 loopback addresses. The processing rules for this new larger loopback prefix also allow sending or forwarding of packets containing these addresses beyond the originating router under certain circumstances.

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 August 24, 2013.

Copyright Notice

Copyright (c) 2013 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	3
1.1. Requirements Language	4
2. Larger Loopback Prefix Requirements	4
3. Proposed Larger Loopback Prefix	4
4. Address Assignment and Configuration	5
5. Larger Loopback Prefix Processing Rules	6
5.1. Host Rules	6
5.1.1. Packets Originated with Larger Loopback Source and/or Destination Addresses	6
5.1.2. Packets Received Externally With Larger Loopback Source and/or Destination Addresses	7
5.2. Router Rules	7
5.2.1. Packets Originated with Larger Loopback Source and/or Destination Addresses	8
5.2.2. Packets Received Externally With Larger Loopback Source and/or Destination Addresses	8
6. Default Address Selection	8
7. DNS Considerations	9
8. Acknowledgements	9
9. IANA Considerations	9
10. Security Considerations	9
11. Change Log [RFC Editor please remove]	10
12. References	11
12.1. Normative References	11
12.2. Informative References	11
Author's Address	12

1. Introduction

During the development and testing of a network application, it can be useful to run multiple instances of the application on the same development host. It may also be useful or important for network access to these application instances to be limited to only the development host itself.

Networked applications that use fixed and usually well known transport layer protocol ports will typically accept incoming traffic on that port for any address assigned to the host. This will prevent multiple instances of the application running on the same port. This port reuse limitation can be overcome by having each application instance bind to different individual addresses available on the host.

Under IPv4, the 127/8 loopback prefix [RFC1122] provides many addresses that have commonly been able to be used to run multiple instances of an application on the same port, while also limiting access to the local host.

The IPv6 addressing architecture [RFC4291] only specifies a single loopback address (::1/128). Multiple IPv6 loopback addresses are not available to bind application instances to when using the same port on the same host.

The IPv4-Mapped IPv6 Address form of 127/8, ::ffff:127.0.0.0/104 [RFC4291], could be used to provide more host local loopback addresses. However these addresses do not have native IPv6 address properties. For example, they cannot accommodate 64 bit Interface Identifiers. Other current and future IPv6 address forms that contain IPv4 addresses or prefixes, such as IPv4-Embedded IPv6 Addresses [RFC6052], have or are likely to have similar or other drawbacks.

A Unique Local IPv6 Unicast Address (ULA) prefix [RFC4193] could be used to increase the number of addresses available on the local host. However this prefix would need to be generated and configured at least once by a system administrator or operator. Without additional configuration, traffic towards addresses not assigned to the local host would not be prevented from leaving the host, and access may not be limited to the local host. A ULA prefix would not be well known, and would not be easy to remember and type accurately without violating the randomness requirements of the Global ID component of a ULA prefix. Using hostnames in DNS or the local host's name resolution file (e.g., /etc/hosts) to overcome the effort required to remember or type a ULA prefix may not be possible for some types of applications which directly deal with IPv6 addresses.

This memo proposes a new larger IPv6 loopback prefix that provides many more loopback addresses, has properties of native IPv6 addresses, and is easy to remember and type accurately. As with `::1/128`, it is intended to be automatically configured during system initialisation, making it ubiquitous. Unlike `::1/128`, the processing rules for this prefix match those of IPv4's `127/8`. These rules allow sending or forwarding of packets with the new larger loopback prefix addresses beyond the originating router under certain circumstances.

This memo, if published, updates [RFC4291], [RFC5156], [RFC6303] and [RFC6724].

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. Larger Loopback Prefix Requirements

A new larger loopback prefix should attempt to satisfy all of the following requirements. It should:

- o be a well known prefix,
- o be within an existing special purpose prefix, such as `0000::/8` (the parent prefix of the current IPv6 loopback address),
- o be easy for a human to remember [EASY-NUMBERS],
- o be easy for a human to type accurately [DOET],
- o cover the existing IPv6 loopback prefix,
- o support 64 bit Interface Identifiers,
- o provide a large number of /64 subnets.

3. Proposed Larger Loopback Prefix

Ideally, the prefix length of `::1/128` could be shortened, resulting in a new single larger loopback prefix for IPv6, such as `::/48`. However, if the existing loopback prefix length is shortened enough to satisfy all of the larger loopback prefix requirements, it would then cover the IPv4 Mapped IPv6 Address prefix, `::ffff:0.0.0.0/96`, and prevent its use described in [RFC4038].

Giving up the requirement of covering the existing IPv6 loopback prefix, the proposed new larger loopback prefix is:

```
0001:0000:0000:0000:0000:0000:0000:0000/32
```

or concisely,

```
1::/32
```

This prefix satisfies all remaining larger loopback prefix requirements.

Allocating a /32 prefix for the loopback function may seem excessive, as a /48 length prefix would satisfy the larger loopback prefix requirements. However, within the parent 0000::/8 special purpose prefix, there are approximately 16 million /32 prefixes, so a single /32 for the larger loopback prefix is easily afforded. A /32 larger loopback prefix will satisfy all current and likely future uses of the loopback function.

4. Address Assignment and Configuration

Consistent with the IPv6 Addressing Model [RFC4291], each address within the larger loopback prefix is always logically assigned to one of the node's interfaces, although not necessarily the same interface for all addresses. This means that the node acts as though all addresses within the larger loopback prefix have been configured on one or more interfaces. Applications will accept packets destined to any of the larger loopback prefix addresses, unless the application is bound to a specific larger loopback address. Typically the addresses will be logically assigned to one or more virtual "loopback" interfaces, which locally returns or loops outgoing packets back to the same node that originated the packets.

It is also common to configure a well known loopback address on the loopback interface during system initialisation, making a loopback address visible to the system operator or user [DOET]. For IPv4, this address is 127.0.0.1/8; for IPv6, it is ::1/128. For the new larger loopback prefix, the address automatically configured on the loopback interface should be:

```
1::1/64
```

This address will be easy for a human to both remember [EASY-NUMBERS][DOET] and type accurately [DOET].

A /64 prefix length has been chosen over /32 to provide a 64 bit

Interface Identifier for the loopback interface. This is different from the use of the whole loopback prefix length when configuring 127.0.0.1/8 or ::1/128.

Some nodes may support more than one loopback interface. These subsequent loopback interfaces, when initialised, should be assigned a larger loopback /64 prefix locally unique within the node. All addresses within the assigned /64 are logically assigned to the interface. Additionally, the "::1" address for the subnet should be configured on the loopback interface, making it visible to a system operator or user [DOET].

/64 subnet identifier uniqueness could be achieved by using the loopback interface instance number as the subnet identifier, with the first instance numbered 0 to suit the use of 1::1/64 on the first loopback interface. For example, the second loopback interface could be assigned 1:0:0:1::/64, while the forth loopback interface could be assigned 1:0:0:3::/64. Alternatively, the interfaces' ifIndex [RFC1213] could be used to determine these subsequent interfaces' loopback /64 subnet identifier. Other schemes which ensure subnet identifier uniqueness would be acceptable.

It should be possible for an operator to remove these automatically configured loopback addresses. It should also be possible for an operator to configure further loopback addresses from within the assigned /64, or addresses from other parts of the larger loopback prefix, including other /64s assigned to other loopback interfaces. Other addresses within the assigned /64(s) would continue to be logically assigned to the subsequent loopback interface. Configuration of addresses is for operational visibility and convenience [DOET], and does not change the behaviour of non-visible logically assigned addresses.

The larger loopback prefix addresses that are outside of the subsequent loopback interface assigned /64s would continue to be logically assigned to the oldest loopback interface.

5. Larger Loopback Prefix Processing Rules

5.1. Host Rules

5.1.1. Packets Originated with Larger Loopback Source and/or Destination Addresses

Packets originated with larger loopback source and/or destination addresses MUST be returned to the origin host for standard processing by the local IPv6 protocol implementation. They MUST NOT be sent

over any external links attached to the host.

If the implementation supports multiple loopback interfaces, and they have been assigned prefixes and addresses from within the larger loopback prefix, the egress loopback interface SHOULD be the interface assigned the matching destination loopback address. The ingress loopback interface MUST be the interface assigned the matching destination loopback address. This will facilitate loopback interface specific handling of the looped traffic, such as traffic filtering or traffic conditioning, which may be useful during network application development. Note that standard IPv6 longest match packet forwarding will facilitate this multiple loopback interface processing.

All addresses within the larger loopback prefix MUST always be considered assigned to one of the host's interfaces, consistent with IPv6's Addressing Model [RFC4291]. Ingress packets, once they have passed any interface specific policies, MUST be delivered to the appropriate protocol module (e.g., such as TCP, SCTP, UDP or ICMPv6) interested in packets with the destination larger loopback prefix address for further processing.

5.1.2. Packets Received Externally With Larger Loopback Source and/or Destination Addresses

Packets with larger loopback source and/or destination addresses received over any of the external links attached to the host MUST be dropped. ICMPv6 error messages, such as Destination Unreachable messages, MUST NOT be generated for these dropped packets.

Implementation suggestion: For these dropped packets, it may be useful to generate an appropriate system log message, indicating a packet with an invalid source or destination address (a "martian" [RFC1812]) was received over an external interface. By default, these messages should be suppressed. If they are enabled, they should be appropriately rate limited to prevent a system log denial-of-service attack.

5.2. Router Rules

IPv4 loopback packet processing rules for routers, specified in [RFC1812], by default prohibited forwarding of packets with 127/8 destinations, other than those originated locally and returned back to the router itself. A software switch could be provided to disable this prohibition. This special case of allowing forwarding of packets towards 127/8 destinations has been taken advantage of by [RFC4379], for MPLS troubleshooting purposes. An equivalent function for IPv6 is provided by using the IPv4-Mapped IPv6 prefix of ::ffff:

127.0.0.0/104.

The existing ::1/128 packet processing rules for routers are the same as those for IPv6 hosts [RFC4291].

For the new larger loopback prefix, the IPv6 router processing rules are changed to match those of IPv4, to suit future uses similar to the MPLS troubleshooting case.

5.2.1. Packets Originated with Larger Loopback Source and/or Destination Addresses

By default, a router MUST follow the host processing rules, described previously, for packets originated with larger loopback source and/or destination addresses.

A software switch MAY be provided to permit packets with larger loopback source and/or destination addresses to be sent via an external interface. If provided, this software switch MUST default to being switched off.

5.2.2. Packets Received Externally With Larger Loopback Source and/or Destination Addresses

By default, a router MUST follow the host processing rules, described previously, for packets received externally with larger loopback source and/or destination addresses.

A software switch MAY be provided to permit received packets with larger loopback source and/or destination addresses to be forwarded via an external interface. This software switch MUST default to being switched off.

6. Default Address Selection

For the purposes of default address selection [RFC6724], as with ::1/128, addresses within the larger loopback prefix MUST be treated as having link-local scope, and must have a "preferred" configuration status.

Within the address selection default policy table [RFC6724], the larger loopback prefix is to be assigned a precedence value of 60. As the existing ::1/128 loopback address has a precedence value of 50, given a choice, a larger loopback prefix address will be chosen as a destination address over ::1/128.

Within the address selection default policy table [RFC6724], the

larger loopback prefix is to be assigned a label value of 14, for use during source address selection.

These default address selection changes should be enabled at the same time that the larger loopback prefix and corresponding processing rules are enabled on a node.

7. DNS Considerations

The DNS zone for 1::/32, 0.0.0.0.1.0.0.0.IP6.ARPA, SHOULD be served locally. [RFC6303] provides further discussion regarding local serving of DNS zones for non-global IP address spaces.

8. Acknowledgements

Nick Hilliard provided valuable review, comments and advice on this memo.

Review and comments were provided by, in alphabetical order, Bill Atwood, Brian Carpenter, Roland Chan, Chris Chaundy, Owen DeLong, Chris Donovan, Matts Kallioniemi, Erik Kline and Tina Tsou. Thanks to Bill for persisting with advice on grammar errors. Owen DeLong does not agree with what is proposed in this memo, however his review and comments, as with the other reviewers' comments, have helped improve it.

This memo was prepared using the xml2rfc tool.

9. IANA Considerations

IANA is requested to allocate 0001::/32 from within 0000::/8 of the Internet Protocol Version 6 Address Space, for use as a larger loopback prefix for IPv6, as detailed in this memo, and to record it in the [IANA-IPV6REG].

10. Security Considerations

During deployment of a new larger loopback prefix, there will be a transition period where some hosts and routers have implemented the larger loopback processing rules defined in this memo while others haven't. These legacy hosts and routers will forward larger loopback prefix traffic using conventional unicast processing. For traffic towards non-local larger loopback addresses, traffic will most likely leave the legacy originating host via its default route, and may be

forwarded by legacy routers using their default route. This may unintentionally disclose sensitive information.

Packet filters, matching traffic with larger loopback source and/or destination addresses, should be used to prevent unintended forwarding of loopback traffic. They should be deployed at the following locations:

- o on the legacy hosts themselves,
- o on legacy routers interconnecting different networks, such as on a router interconnecting a private network and the Internet,
- o on appropriate security domain boundary legacy routers within the local network, if not all legacy routers within the local network.

Routes for the new larger loopback prefix should not be announced or accepted if received, unless necessary for special cases where packets with larger loopback prefix addresses are allowed to be forwarded.

11. Change Log [RFC Editor please remove]

draft-smith-larger-ipv6-loopback-prefix-00, initial version,
2012-07-24

draft-smith-larger-ipv6-loopback-prefix-01, much less verbose
version, 2012-08-17

draft-smith-larger-ipv6-loopback-prefix-02, clarifications,
2013-01-07

- o clarification that the larger loopback prefix should fall within `::/8`, the parent prefix of `::/128` and `::1/128`
- o Change from `1::/48` to `1::/32`
- o text about logically assigning addresses to interface(s), as per IPv6 addressing model
- o automatic loopback address configuration to multiple loopback interfaces
- o local serving of `0.0.0.1.0.0.0.IP6.ARPA` zone in DNS

draft-smith-larger-ipv6-loopback-prefix-03, clarifications,
2013-02-07

- o default address selection precedence and label values
- o comment about other IPv4 in IPv6 address forms
- o more clarifications
- o grammar corrections

draft-smith-larger-ipv6-loopback-prefix-04, minor fixups, 2013-02-20

- o usability references (DOET and EASY-NUMBERS)
- o minor clarifications
- o grammar corrections

12. References

12.1. Normative References

[IANA-IPV6REG]

Internet Assigned Numbers Authority, "IPv6 Special Purpose Address Registry", 2013, <<http://www.iana.org/assignments/iana-ipv6-special-registry>>.

[RFC1122] Braden, R., "Requirements for Internet Hosts - Communication Layers", STD 3, RFC 1122, October 1989.

[RFC1213] McCloghrie, K. and M. Rose, "Management Information Base for Network Management of TCP/IP-based internets:MIB-II", STD 17, RFC 1213, March 1991.

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

12.2. Informative References

[DOET] Norman, D., "The Design of Everyday Things", 2002, <<http://www.jnd.org/books/the-design-of-everyday-things.html>>.

[EASY-NUMBERS]

Milikowski, M. and J. Elshout, "What makes a number easy to remember?", 1995, <http://http://www.rekencentrale.nl/bestanden/Andere_artikelen_MM/1995_1999/pdf_files/What_makes_a_number_easy.pdf>.

[RFC1812] Baker, F., "Requirements for IP Version 4 Routers",

RFC 1812, June 1995.

- [RFC4038] Shin, M-K., Hong, Y-G., Hagino, J., Savola, P., and E. Castro, "Application Aspects of IPv6 Transition", RFC 4038, March 2005.
- [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.
- [RFC4379] Kompella, K. and G. Swallow, "Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures", RFC 4379, February 2006.
- [RFC5156] Blanchet, M., "Special-Use IPv6 Addresses", RFC 5156, April 2008.
- [RFC6052] Bao, C., Huitema, C., Bagnulo, M., Boucadair, M., and X. Li, "IPv6 Addressing of IPv4/IPv6 Translators", RFC 6052, October 2010.
- [RFC6303] Andrews, M., "Locally Served DNS Zones", BCP 163, RFC 6303, July 2011.
- [RFC6724] Thaler, D., Draves, R., Matsumoto, A., and T. Chown, "Default Address Selection for Internet Protocol Version 6 (IPv6)", RFC 6724, September 2012.

Author's Address

Mark Smith
In My Own Time
PO BOX 521
HEIDELBERG, VIC 3084
AU

Email: markzzzsmith@yahoo.com.au

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: June 6, 2014

J. Jaeggli
Zynga
L. Colitti
W. Kumari
Google
E. Vyncke
Cisco
M. Kaeo
Double Shot Security
T. Taylor, Ed.
Huawei Technologies
December 3, 2013

Why Operators Filter Fragments and What It Implies
draft-taylor-v6ops-fragdrop-02

Abstract

This memo was written to make application developers and network operators aware of the significant possibility that IPv6 packets containing fragmentation extension headers may fail to reach their destination. Some protocol or application assumptions about the ability to use messages larger than a single packet may accordingly not be supportable in all networks or circumstances.

This memo provides observational evidence for the dropping of IPv6 fragments along a significant number of paths, explores the operational impact of fragmentation and the reasons and scenarios where drops occur, and considers the effect of fragment drops on applications where fragmentation is known to occur, particularly including DNS.

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 June 6, 2014.

Copyright Notice

Copyright (c) 2013 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. Observations and Rationale	3
2.1. Possible Causes	3
2.1.1. Stateful inspection	4
2.1.2. Stateless ACLs	4
2.1.3. Performance considerations	4
2.1.4. Other considerations	4
2.1.5. Conclusions	5
2.2. Impact on Applications	5
3. Acknowledgements	5
4. IANA Considerations	5
5. Security Considerations	6
6. Informative References	6
Authors' Addresses	6

1. Introduction

Measurements of whether Internet Service Providers and edge networks deliver IPv6 fragments to their destination reveal that for IPv6 in particular, fragments are being dropped along a substantial number of paths. The filtering of IPv6 datagrams with fragmentation headers is presumed to be a non-issue in the core of the Internet, where fragments are routed just like any other IPv6 datagram. However, fragmentation can create operational issues at the edges of the Internet that may lead to administratively imposed filtering or inadvertent failure to deliver the fragment to the end-system or application.

Section 2 begins with some observations on how often IPv6 fragment loss occurs in practice. We go on to look at the operational reasons for filtering fragments, a key aspect of which is the limitations they expose in the application of security policy, at resource bottlenecks and in forwarding decisions. Section 2.2 then looks at the impact on key applications, particularly DNS.

In the longer run, as network operators gain a better understanding of the risks and non-risks of fragmentation and as middlebox, customer premise equipment (CPE), and host implementations improve, we believe that some incidence of fragment dropping currently required will diminish. Some of the justifications for filtering will persist in the long-term, and application developers and network operators must remain aware of the implications.

This document deliberately refrains from discussing possible responses to the problem posed by the dropping of IPv6 fragments. Such a discussion will quickly turn up a number of possibilities, application-specific or more general; but the amount of time needed to specify and deploy a given resolution will be a major constraint in choosing amongst them. In any event, that discussion is likely to proceed in multiple directions, occur in different areas and is therefore considered beyond the scope of this memo.

2. Observations and Rationale

[Blackhole] is a good public reference for some empirical data on IPv6 fragment filtering. It describes experiments run to determine the incidence and location of ICMP Packet Too Big and fragment filtering. The authors used fragmented DNS packets to determine the latter, setting the servers to an IPv6 minimum of 1280 bytes to avoid any PMTU issues. The tests found for IPv6 that filtering appeared to be occurring on some 10% of the tested paths. The filtering appeared to be located at the edge (enterprise and customer networks) rather than in the core.

2.1. Possible Causes

Why does such filtering happen? One cause is non-conforming implementations in CPE and low-end routers. Some network managers filter fragments on principle, thinking this is an easier way to deter realizable attacks utilizing IPv6 fragments without thinking of other network impacts, similar to the practice of filtering ICMP Packet Too Big. Both implementations and management should improve over time, reducing the problem somewhat.

Some filtering and dropping of fragments is known to be done for hardware, performance, or topological considerations.

2.1.1. Stateful inspection

Stateful inspection devices or destination hosts can readily experience resource exhaustion if they are flooded with fragments that are not followed in a timely manner by the remaining fragments of the original datagram. Holding fragments for reassembly even on end-system firewalls can readily result in an effective denial of service by memory and CPU exhaustion even if techniques, such as virtual re-assembly exist.

2.1.2. Stateless ACLs

Stateless ACLs at layer 4 and up may be difficult to apply to fragments other than the first one in which enough of the upper layer header is present. As [Attacks] demonstrates, inconsistencies in reassembly logic between middleboxes or CPEs and hosts can cause fragments to be wrongfully discarded, or can allow exploits to pass undetected through middleboxes. Stateless load balancing schemes may hash fragmented datagrams from the same flow to different paths because the 5-tuple may be available on only the initial fragment. While rehashing has the possibility of reordering packets in ISP cores it is not disastrous. However, in front of a stateful inspection device, load balancer tier, or anycast service instance, where headers other than the L3 header -- for example, the L4 header, interface index (for traffic already rehashed onto different paths), DS fields -- are considered as part of the hash, rehashing may result in the fragments being delivered to different end-systems

2.1.3. Performance considerations

Leaving aside these incentives towards fragment dropping, other considerations may weigh on the operator's mind. One example cited on the NANOG list was that of a router where fragment processing was done by the control plane processor rather than in the forwarding plane hardware, with a consequent hit on performance.

2.1.4. Other considerations

Another incentive toward dropping of fragments is the disproportionate number of software errors still being encountered in fragment processing. Since this code is exercised less frequently than the rest of the stack, bugs remain longer in the code before they are detected. Some of these software errors can introduce vulnerabilities subject to exploitation. It is common practice [RFC6192] to recommend that control-plane ACLs protecting routers and network devices be configured to drop all fragments.

2.1.5. Conclusions

Operators weigh the risks associated with each of the considerations just enumerated, and come up with the most suitable policy for their circumstances. It is likely that at least some operators will find it desirable to drop fragments in at least some cases.

The IETF and operators can help this effort by identifying specific classes of fragments that do not represent legitimate use cases and hence should always be dropped. Examples of this work are given by [RFC6946] and [I-D.ietf-6man-oversized-header-chain]. The problem of inconsistent implementations may also be mitigated by providing further advice on the more difficult points. However, some cases will remain where legitimate fragments are discarded for legitimate reasons. The potential problems these cases pose for applications is our next topic.

2.2. Impact on Applications

Some applications can live without fragmentation, some cannot. UDP DNS is one application that has the potential to be impacted when fragment dropping occurs. EDNS0 extensions [RFC2671] allow for responses in UDP PDUs that are greater than 512 bytes. Particularly with DNSSEC [RFC4033], responses may be larger than the link MTU and fragmentation would therefore occur at the sending host in order to respond using UDP. The current choices open to the operators of DNS servers in this situation are to defer deployment of DNSSEC, fragment responses, or use TCP if there are cases where the rreset would be expected to exceed the MTU. The use of fallback to TCP will impose a major resource and performance hit and increases vulnerability to denial of service attacks.

Other applications, such as the Network File System, NFS, are also known to fragment large UDP packets for datagrams larger than the MTU. NFS is most often restricted to the internal networks of organizations. In general, managing NFS connectivity should not be impacted by decisions managing fragment drops at network borders or end-systems.

3. Acknowledgements

The authors of this document would like to thank the RIPE Atlas project and NLNetlabs whose conclusions ignited this document.

4. IANA Considerations

This memo includes no request to IANA.

5. Security Considerations

The potential for denial of service attacks, as well as limitations inherent in upper-layer filtering when dealing with non-initial fragments are significant issues under consideration by operators and end-users filtering fragments. This document does not offer alternative solutions to that problem, it does describe the impact of those filtering practices.

6. Informative References

- [Attacks] Atlasis, A., "Attacking IPv6 Implementation Using Fragmentation", March 2012.
- http://media.blackhat.com/bh-eu-12/Atlasis/bh-eu-12-Atlasis-Attacking_IPv6-WP.pdf
- [Blackhole] de Boer, M. and J. Bosma, "Discovering Path MTU black holes on the Internet using RIPE Atlas", July 2012.
- <http://www.nlnetlabs.nl/downloads/publications/pmtu-black-holes-msc-thesis.pdf>
- [I-D.ietf-6man-oversized-header-chain] Gont, F., Manral, V., and R. Bonica, "Implications of Oversized IPv6 Header Chains", draft-ietf-6man-oversized-header-chain-08 (work in progress), October 2013.
- [RFC2671] Vixie, P., "Extension Mechanisms for DNS (EDNS0)", RFC 2671, August 1999.
- [RFC4033] Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "DNS Security Introduction and Requirements", RFC 4033, March 2005.
- [RFC6192] Dugal, D., Pignataro, C., and R. Dunn, "Protecting the Router Control Plane", RFC 6192, March 2011.
- [RFC6946] Gont, F., "Processing of IPv6 "Atomic" Fragments", RFC 6946, May 2013.

Authors' Addresses

Joel Jaeggli
Zynga
630 taylor ct #10
Mountain View, CA 94043
USA

Email: jjaeggli@zynga.com

Lorenzo Colitti
Google

Email: lorenzo@google.com

Warren Kumari
Google
1600 Amphitheatre Parkway
Mountain View, CA 94043
USA

Email: warren@kumari.net

Eric Vyncke
Cisco
De Kleetlaan 6A
Diegem 1831
Belgium

Email: evyncke@cisco.com

Merike Kaeo
Double Shot Security

Email: merike@doubleshotsecurity.com

Tom Taylor (editor)
Huawei Technologies
Ottawa, Ontario
Canada

Email: tom.taylor.stds@gmail.com

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: Jan 2, 2016

T. Yang
L. Li
Q. Ma
China Mobile
July 2, 2015

Scenario of IPv6 Transition Technologies Selection
draft-yang-v6ops-ipv6tran-select-03

Abstract

Many IPv6 transition technologies has been proposed, such as Dual-Stack, 6rd and so on. An CPE may support some of them instead of only one. But the ISPs always support different kinds of transition technologies. So they must control all the CPEs to match the exact transition tech through the CPEs' management system or configuring them before issuing to the customers.

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 Jan 2, 2016

Copyright Notice

Copyright (c) 2013 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. Problem Statement	2
2. Security Considerations	3
Authors' Addresses	3

1. Problem Statement

Nowadays, many IPv6 transitioning technologies has been proposed such as Dual-Stack, DS-Lite, 6rd and so on. Each of them proposes individual requirment to the CPEs. To promote the competitive ability of products,the CPE manufacturers certainly will try to support more technologies as much as possible. Meanwhile, the operators tend to use single or less technologies. Moreover, users can buy and use their own equipments instead of using the one which operator gives them, that will bring the diversity of CPEs.

Assume that an operator uses one or more transitioning strategies in its network. There are two ways to make the CPEs available. The first one is to make a pre-configuration for each CPE in advance. But, when the users modify the configuration or change to their own equipment, the connection will fail. The Second method is to deploy Network Management System (NMS) to configurate all the CPEs. Various CPEs from different manufactories usually need different NMS which means either the operator needs to maintains multiple NMS in their network or operator can only use one manufacturer's product in a subnet. What's worse, when users buy and use their own CPE instead of using the original one, there will be no any solutions to configurate correctly except visiting service.

Some specific messages need to be define between CPE and DHCP Server to communicate the IPv6 transition technologies.

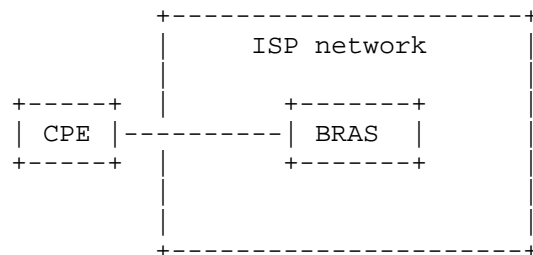


Figure1 IPv6 trans tech configuration Scenario

2. Security Considerations

The security problem is under disscussion.

Authors' Addresses

Tianle Yang
China Mobile
32, Xuanwumenxi Ave.
Xicheng District, Beijing 100053
China

Email: yangtianle@chinamobile.com

Li Lianyuan
China Mobile
32, Xuanwumenxi Ave.
Xicheng District, Beijing 100053
China

Email: lilianyuan@chinamobile.com

Qiongfang Ma
China Mobile
32, Xuanwumenxi Ave.
Xicheng District, Beijing 100053
China

Email: maqiongfang@chinamobile.com