IPv6 Operations
Internet-Draft
Intended status: Informational
Expires: April 21, 2011

F. Baker
Cisco Systems

October 18, 2010

Opening TCP Sessions in Complex Environments
draft-baker-v6ops-session-start-time-01

Abstract

A barrier to the deployment of IPv6 is the amount of time it takes to open a session using common transport APIs. This note addresses issues and requests solutions that may respond to them.

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 21, 2011.

Table of Contents

1. Introduction

One of the issues in IPv6 deployment is the time, from a user's perspective, that it takes to open a standard application, which is to say the time it takes to open a TCP session that the application can use to accomplish its mission.

One thing to understand is that each source/destination pair of addresses (IPv4 and IPv6 addresses, including link-local, organizational scope such as [RFC1918] or ULA [RFC4193], and global addresses) defines a path between those interfaces. The path may or may not actually work (the two addresses may not be in the same domain or the same scope, or routing may not be defined, or forwarding may be filtered), and even if the network workds, the peer may or may not be willing to respond to any given address. Hence, in the worst case, every pair of addresses may need to be tried in the process of finding a pair that enables communication.

In the immortal words of [RFC1958],

> The current exponential growth of the network seems to show that connectivity is its own reward, and is more valuable than any individual application such as mail or the World-Wide Web. This connectivity requires technical cooperation between service providers, and flourishes in the increasingly liberal and competitive commercial telecommunications environment.

An application or API that fails to quickly enable connectivity between any two systems that are authorized to communicate has fundamentally missed the point, and can expect its customers to migrate to solutions that don't miss the point.

Part of the issue has to do with source address choice in multihomed networks, as described in [I-D.troan-multihoming-without-nat66]; if the host selects the wrong source address for a session with a peer, BCP 38 [RFC2827] ingress filtering will prevent its delivery. Any delay in selecting an alternative source address will irritate the user, making IPv6 appear less desirable.

Part of it has to do with the standard response of TCP and SCTP clients to RST and ICMP Unreachable messages; if another address pair exists, any delay in selecting an alternative source address will irritate the user, making IPv6 appear less desirable.

Part of it has to do with the rate of session attempts; if one takes multiple seconds per attempt and, present implementations require as much as 40 seconds to open a basic web page. Again, such delays irritate the user, making IPv6 appear less desirable.

2. Possible Solutions

TCP’s standard reaction to soft errors, which includes its response to an abrupt RST from the peer and its response to ICMP "unreachable messages", doesn’t help. [RFC5461] makes pragmatic suggestions to address the issues. From an operator’s perspective, it is felt that the fundamental suggestion is a good one, and either should be standardized and widely deployed or a better suggestion should be standardized and widely deployed.

The Happy Eyeballs [I-D.wing-v6ops-happy-eyeballs-ipv6] draft addresses the startup question. From an operator’s perspective, it is felt that the fundamental suggestion is a good one, and either should be standardized and widely deployed or a better suggestion should be standardized and widely deployed.

3. IANA Considerations

This memo asks the IANA for no new parameters.

Note to RFC Editor: This section will have served its purpose if it correctly tells IANA that no new assignments or registries are required, or if those assignments or registries are created during the RFC publication process. From the author"s perspective, it may therefore be removed upon publication as an RFC at the RFC Editor"s discretion.

4. Security Considerations

This note doesn’t address security-related issues.

5. Acknowledgements

This note was discussed with Joel Jaeggli, Dan Wing, and Fernando Gont.

6. Change Log

-00 Version: October 6, 2010

-01 Version: update Happy Eyeballs reference.

7. Informative References

[I-D.troan-multihoming-without-nat66]
Troan, O., Miles, D., Matsushima, S., Okimoto, T., and D. Wing, "IPv6 Multihoming without Network Address Translation", draft-troan-multihoming-without-nat66-01 (work in progress), July 2010.

[I-D.wing-v6ops-happy-eyeballs-ipv6]
Wing, D. and A. Yourtchenko, "Happy Eyeballs: Trending Towards Success with Dual-Stack Hosts", draft-wing-v6ops-happy-eyeballs-ipv6-00 (work in progress), October 2010.

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

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

[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.

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

[RFC5461] Gont, F., "TCP's Reaction to Soft Errors", RFC 5461, February 2009.

Author's Address

Fred Baker
Cisco Systems
Santa Barbara, California 93117
USA

Email: fred@cisco.com

IPv6 Operations
Internet-Draft
Intended status: Informational
Expires: May 11, 2011

F. Baker
Cisco Systems

November 7, 2010

Opening TCP Sessions in Complex Environments
draft-baker-v6ops-session-start-time-02

Abstract

A barrier to the deployment of IPv6 is the amount of time it takes to open a session using common transport APIs. This note addresses issues and requests solutions that may respond to them.

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 May 11, 2011.

Table of Contents

1. Introduction

One of the issues in IPv6 deployment is the time, from a user's perspective, that it takes to open a standard application, which is to say the time it takes to open a TCP session that the application can use to accomplish its mission.

One thing to understand is that each source/destination pair of addresses (IPv4 and IPv6 addresses, including link-local, organizational scope such as [RFC1918] or ULA [RFC4193], and global addresses) defines a path between those interfaces. The path may or may not actually work (the two addresses may not be in the same domain or the same scope, or routing may not be defined, or forwarding may be filtered), and even if the network workds, the peer may or may not be willing to respond to any given address. Hence, in the worst case, every pair of addresses may need to be tried in the process of finding a pair that enables communication.

In the immortal words of [RFC1958],

> The current exponential growth of the network seems to show that connectivity is its own reward, and is more valuable than any individual application such as mail or the World-Wide Web. This connectivity requires technical cooperation between service providers, and flourishes in the increasingly liberal and competitive commercial telecommunications environment.

An application or API that fails to quickly enable connectivity between any two systems that are authorized to communicate has fundamentally missed the point, and can expect its customers to migrate to solutions that don't miss the point.

Part of the issue has to do with source address choice in multihomed networks, as described in [I-D.troan-multihoming-without-nat66]; if the host selects the wrong source address for a session with a peer, BCP 38 [RFC2827] ingress filtering will prevent its delivery. Any delay in selecting an alternative source address will irritate the user, making IPv6 appear less desirable.

Part of it has to do with the standard response of TCP and SCTP clients to RST and ICMP Unreachable messages; if another address pair exists, any delay in selecting an alternative source address will irritate the user, making IPv6 appear less desirable.

Part of it has to do with the rate of session attempts; if one takes multiple seconds per attempt and, present implementations require as much as 40 seconds to open a basic web page. Again, such delays irritate the user, making IPv6 appear less desirable.

2. Possible Solutions

TCP’s standard reaction to soft errors, which includes its response to an abrupt RST from the peer and its response to ICMP "unreachable messages", doesn’t help. [RFC5461] makes pragmatic suggestions to address the issues. From an operator’s perspective, it is felt that the fundamental suggestion is a good one, and either should be standardized and widely deployed or a better suggestion should be standardized and widely deployed.

The Happy Eyeballs [I-D.wing-v6ops-happy-eyeballs-ipv6] draft addresses the startup question. From an operator’s perspective, it is felt that the fundamental suggestion is a good one, and either should be standardized and widely deployed or a better suggestion should be standardized and widely deployed.

The Testing Eyeball Happiness [I-D.baker-bmwg-testing-eyeball-happiness] draft outlines a relatively simple test that can be applied to determine whether a given application is likely to meet the operational intent of the Happy Eyeballs draft. It does not test for correct implementation of the algorithm per se; it tests whether the algorithm implemented addresses the operational concern.

3. IANA Considerations

This memo asks the IANA for no new parameters.

Note to RFC Editor: This section will have served its purpose if it correctly tells IANA that no new assignments or registries are required, or if those assignments or registries are created during the RFC publication process. From the author"s perspective, it may therefore be removed upon publication as an RFC at the RFC Editor"s discretion.

4. Security Considerations

This note doesn’t address security-related issues.

5. Acknowledgements

This note was discussed with Joel Jaeggli, Dan Wing, Andrew Yourtchecnko, and Fernando Gont.

6. Change Log

-00 Version: October 6, 2010

-01 Version: update Happy Eyeballs reference.

-02 Version: Add Happy Eyeballs test proposal.

7. Informative References

[I-D.baker-bmwg-testing-eyeball-happiness]
Baker, F., "Testing Eyeball Happiness", draft-baker-bmwg-testing-eyeball-happiness-00 (work in progress), November 2010.

[I-D.troan-multihoming-without-nat66]
Troan, O., Miles, D., Matsushima, S., Okimoto, T., and D. Wing, "IPv6 Multihoming without Network Address Translation", draft-troan-multihoming-without-nat66-01 (work in progress), July 2010.

[I-D.wing-v6ops-happy-eyeballs-ipv6]
Wing, D. and A. Yourtchenko, "Happy Eyeballs: Trending Towards Success with Dual-Stack Hosts", draft-wing-v6ops-happy-eyeballs-ipv6-01 (work in progress), October 2010.

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

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

[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.

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

[RFC5461] Gont, F., "TCP’s Reaction to Soft Errors", RFC 5461, February 2009.

Author’s Address

Fred Baker
Cisco Systems
Santa Barbara, California 93117
USA

Email: fred@cisco.com

V6OPS
Internet-Draft
Intended status: Informational
Expires: February 19, 2011

B. Carpenter
Univ. of Auckland
S. Jiang
Huawei Technologies Co., Ltd
V. Kuarsingh
Rogers Communications
August 18, 2010

Framework for IP Version Transition Scenarios
draft-carpenter-v4v6tran-framework-00

Abstract

This document sets out a framework for the presentation of scenarios and recommendations for a variety of approaches to the transition from IPv4 to IPv6, given the necessity for a long period of co-existence of the two protocols.

Table of Contents

1. Introduction

This document sets out a framework for the presentation of scenarios and recommendations for a variety of approaches to the transition from IPv4 to IPv6, given the necessity for a long period of co-existence of the two protocols. A general "call to arms" for transition is found in [RFC5211], and a recommendation for four principal scenarios is given in [I-D.arkko-ipv6-transition-guidelines]. A report on experience and plans of various Internet Service Providers (ISPs) is given in [I-D.ietf-v6ops-isp-scenarios]. However, it is clear that operators require more detailed technical recommendations than are available so far. A companion document [reference TBD] provides a technical problem statement. Unfortunately, the number of different combinations of existing IPv4 deployment models, customer profiles and requirements, and possible coexistence and transition models, is enormous, so it is quite impracticable to produce either a set of recommendations for each case, or a recommended "one size fits all" model. That is why this document proposes a set of topics or dimensions, as a framework for a reasonable number of recommendation documents.

The reader is assumed to be familiar with IPv6. The IETF's view of core IPv6 requirements is to be found in [RFC4294] (currently being updated as [I-D.ietf-6man-node-req-bis]). However, this does not give a complete view of mechanisms an ISP may need to deploy, since it considers the requirements for an individual node, not for a network or service infrastructure as a whole.

[RFC4029] discussed scenarios for introducing IPv6 into ISP networks, as the problem was viewed some years ago. Its end goal was simply a dual-stack ISP backbone. Today's view is that this is insufficient, as it does not allow for prolonged interworking between IPv6-only and legacy (IPv4-only) hosts. Indeed, the end goal today might be an IPv6-only ISP backbone, with some form of legacy IPv4 support [I-D.arkko-ipv6-transition-guidelines].

Although the basic IPv6 standards are stable, considerable work continues in several IETF working groups, on issues such as multihoming, tunneling, and IP layer interworking between IPv6-only and IPv4-only hosts. However, operators faced with IPv4 address exhaustion in the coming few years need immediate guidance.

2. Document Topics

On the assumption that a series of documents are produced describing and recommending transition scenarios, there are two basic

conditions:

1. The documents will not be primary protocol specifications, because those are the outcome of IETF working groups chartered to work on specific protocol mechanisms.
2. The documents are addressed to service providers who have taken the decision to support IPv6, have acquired basic knowledge and skills, have determined how they will obtain upstream IPv6 connectivity, and are ready to write their operational plan for transition.

The documents should each cover some or all of the following aspects or dimensions:

- For the convenience of readers, each document should briefly describe its network model in the Abstract (or Introduction) for quick reference.
- The documents should explain how certain technology components fit together in a given transition and co-existence scenario.
- They will present major generic network models, and their subsets, which exist (or are firmly planned) today, including network topologies and/or architectures.
- They should specify their scope: the range of technologies that they do or do not apply to (e.g. specific access network technologies, core network technologies and topologies, mobile vs fixed hosts, business vs private customers, etc.).
- They should develop analysis criteria on how to recognize appropriate transition technologies for existing provider networks within their scope. This should include information related to deployed protocols and functions which may assist or hinder various transition technologies from being deployed.
- If multiple transition technologies are needed for provider environments where access networks differ and have various capabilities, the documents should show how these technologies can be deployed simultaneously.
- They should describe how multiple technologies can co-exist, if necessary, during all stages of migration (e.g., moving from IPv4 Only to Dual-Stack to DS-Lite to NAT64).
- They should cover considerations for legacy operation while moving to IPv6 and its transition technologies. Many operators will have large quantities of IPv4-only equipment which cannot feasibly be upgraded until the end of its economic life, or which is under customer control.
- They should cover considerations which apply when retro-fitting various technologies to existing networks. Included in this would be impacts on ancillary protocols, routing platforms/systems, security policies, provisioning systems, network services (i.e. DHCP, DNS etc), law enforcement procedures and more.

- o They should quantify scaling characteristics of deployment modes for each technology model and intersections during co-existence (e.g. if some of the Network is DS-Lite and some is classical Dual Stack; peak load on NAT64; etc.).
- o The documents should include security considerations for their specific transition scenario(s).

A desirable outcome would be a set of Best Current Practice (BCP) or advisory (Informational) documents for a range of generic deployment models and how they fit into a network, including key services such as subscriber authentication, DHCP, and DNS. However, it must not be forgotten that every service provider is different and such documents can never replace specific deployment plans drawn up by each individual service provider.

3. Security Considerations

Service providers will insist on having security for IPv6 services, and for all transition technologies, that is at least as good as for IPv4 services in all respects. Particular attention must be paid to security exposures that are specific to transition and coexistence mechanisms. Thus, all recommendations for transition scenarios must include security aspects.

4. IANA Considerations

This document makes no request of the IANA.

5. Acknowledgements

Useful comments and contributions were made by ... and others.

This document was produced using the xml2rfc tool [RFC2629].

6. Change log

draft-carpenter-v4v6tran-framework-00: original version, 2010-08-18

7. Informative References

[I-D.arkko-ipv6-transition-guidelines]
Arkko, J. and F. Baker, "Guidelines for Using IPv6 Transition Mechanisms",

draft-arkko-ipv6-transition-guidelines-03 (work in progress), July 2010.

[I-D.ietf-6man-node-req-bis]
Jankiewicz, E., Loughney, J., and T. Narten, "IPv6 Node Requirements RFC 4294-bis", draft-ietf-6man-node-req-bis-05 (work in progress), July 2010.

[I-D.ietf-v6ops-isp-scenarios]
Carpenter, B. and S. Jiang, "Emerging Service Provider Scenarios for IPv6 Deployment", draft-ietf-v6ops-isp-scenarios-00 (work in progress), April 2010.

[RFC2629] Rose, M., "Writing I-Ds and RFCs using XML", RFC 2629, June 1999.

[RFC4029] Lind, M., Ksinant, V., Park, S., Baudot, A., and P. Savola, "Scenarios and Analysis for Introducing IPv6 into ISP Networks", RFC 4029, March 2005.

[RFC4294] Loughney, J., "IPv6 Node Requirements", RFC 4294, April 2006.

[RFC5211] Curran, J., "An Internet Transition Plan", RFC 5211, July 2008.

Authors' Addresses

Brian Carpenter
Department of Computer Science
University of Auckland
PB 92019
Auckland, 1142
New Zealand

Email: brian.e.carpenter@gmail.com

Sheng Jiang
Huawei Technologies Co., Ltd
KuiKe Building, No.9 Xinxi Rd.,
Shang-Di Information Industry Base, Hai-Dian District, Beijing
P.R. China

Email: shengjiang@huawei.com

Victor Kuarsingh
Rogers Communications
Canada

Email: Victor.Kuarsingh@rci.rogers.com

Internet Engineering Task Force G. Chen
Internet-Draft H. Deng
Intended status: Informational China Mobile
Expires: April 21, 2011 October 18, 2010

NAT64-CPE Mode Operation for Opening Residential Service
draft-chen-v6ops-nat64-cpe-00

Abstract

The document has proposed an approach of NAT64-CPE mode, which would give residential service opportunities to be accessed by remote subscribers going through IPv6 networks. The document captures the fundamental NAT64 functionalities with special cares to fit into CPE scenarios and don’t need cooperate with DNS64 any more. It will compatible with legacy residential servers and no further updates requirements to DNS.

Table of Contents

1. Introduction

The document is aimed at proposing an approach of NAT64-CPE mode, which would give residential service opportunities to be accessed by remote subscribers going through IPv6 networks. The document captures the fundamental NAT64[NAT64] functionalities with special cares to fit into CPE scenarios. In these scenarios, the NAT64-CPE don't need cooperate with DNS64[DNS64] any more, whereby this mechanism allows an IPv6-only client (i.e. either a host with only IPv6 stack, or a host with both IPv4 and IPv6 stack, but only with IPv6 connectivity or a host running an IPv6 only application) to initiate communications to an IPv4-only residential service server.

Recently,IPv6 transition is fairly prevalent due to the depletion of IPv4 soon enough. However, the large number of installed CPE is IPv4-only based and likely to remain for several years. Considering the existing deployment approaches, majority of ISP assigned private IPv4 address to their customers, including residential servers. The nature of private IPv4 would block the end-to-end bi-directional communications. On the other hand, the goal of Internet services is to offer users ubiquitous experiences. User will be certainly supposed to able to enjoy such conveniences regardless of where we are. Therefore, ISP would take advantage of the accessibilities of residential services to provide plenty of services. Fortunately, IPv6 will get ISP end-to-end benefits. During IPv6 migration period, NAT64-CPE mode could overcome the obstacles to achieve final goals.

The document is structured as follows. Section 2 describes appropriate scenario the NAT64-CPE mode fit to. Section 3 enumerates various functional parts for NAT64-CPE operation. Section 4 focus on the benefits the NAT64-CPE could bring. Section 5 is further securities consideration.

2. NAT64-CPE Mode Scenario Overviews

Figure 1 illustrates a possible network scenario where an IPv6-only client attached to a dual-stack network, but the destination server is running on a private site where there is NAT64-CPE numbered with public IPv6 addresses and private IPv4 addresses. DNS is located in dual stack Internet for naming-resolving.

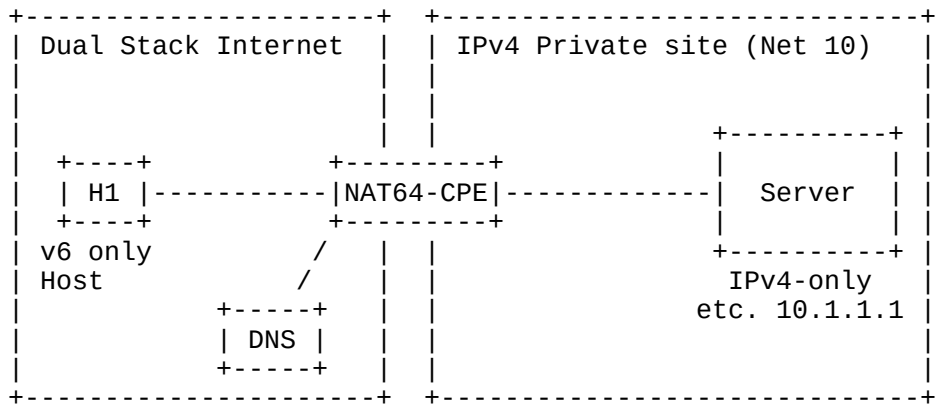


Figure 1: NAT64-CPE Network Scenario

This scenario appears in ISP network quite popular. As the instances, visitors go through distant network to take care of family affairs, like monitoring house security via residential camera, manipulating household appliances remotely prior to comeback home.

3. NAT64-CPE Mode Operation

The whole process of NAT64-CPE operation involves CPE, DNS and addressing mechanism. This section illustrates different parts of functionalities.

3.1. CPE Functionalities Description

Two kinds of functions the NAT64-CPE would take on. First, it will perform the functionalities that normal CPE does except NAT44 forwarding, like assigning private IPv4 address to their attached residential servers. Additionally, CPE will allocate private IPv4 address to the servers depending on the server MAC address. Therefore, the server could always get constant private IPv4 address.

Second, CPE should carry NAT64 capable mode without integrating DNS64. According to normative handing, NAT64-CPE translates in-coming IPv6 destination address by striping NAT64 IPv6-prefix and maintains a IPv4 pool for translating IPv6 sources address. Therein, the NAT64 IPv6 prefix will be NSP specified in IPv6 Addressing of IPv4/IPv6 Translators [IPv6 Addressing of IPv4/IPv6 Translators]. And, ISP will reserve distinct NSP for each CPE.

The prerequisite here is that NAT64-CPE should maintain address

mapping between inner IP address and outer IP address. PCP [PCP] could handle such problems. But that goes beyond the scope of this draft. Also, NAT64-CPE would install ALG, but it is optional.

3.2. DNS Configuration Consideration

Each residential services should be represented by FQDN format so as to users could easily remember and understand. The corresponding naming resource record should be stored as AAAA. The record's IPv6 address is synthesized by NAT64 prefix and private IPv4 address. The IPv6 format is compliant with assembling IPv6 address in DNS64.

The deployed DNS just follow regular DNS handling. There is no demands for performing DNS64 process.

3.3. NAT64-CPE Mode Operation Example

Figure 2 demonstrates the NAT64-CPE Mode operation flow, in which IPv6 host initiate service interaction with residential server remotely. The detailed actions that different entities performed was described afterwards.

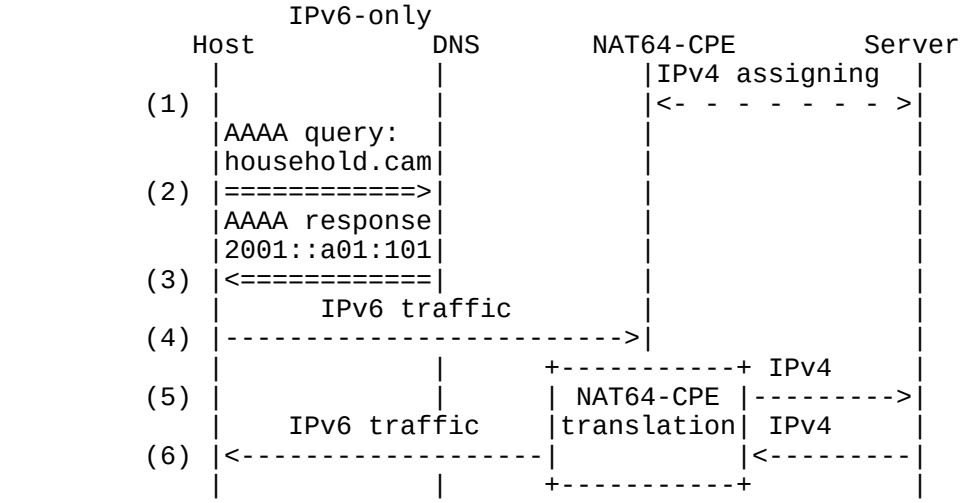


Figure 2: NAT64-CPE Mode Operation Example

(1) NAT64-CPE should be configured with NAT64 prefix, which is allocated by ISP. In that case, the NAT64 prefix is 2001::/64. NAT64-CPE assign private IPv4 address to the servers depending on the server MAC address. And, NAT64-CPE already has maintained the

mapping between inner IP address and outer IP address.

(2)IPv6-only host initiates AAAA query for resolving service name, for example, that is household.cam.

(3)DNS response AAAA record to the previous query. The IPv6 address of this service is synthesized by NAT64 prefix and assigned private IPv4 address. That is 2001::a01:101

(4)IPv6-only host send IPv6 traffic targeting to 2001::a01:101. The IPv6 traffic is routed to CPE.

(5)NAT64-CPE detects incoming IPv6 packets and algorithmically translated to IPv4 addresses by using the algorithm defined in [I-D.ietf-behave-address-format]. The translated IPv4 traffic is headed to IPv4-only residential server and perform somehow process.

(6)The residential IPv4 server responses these requests by IPv4 traffic, which will be sent to CPE. CPE performs reversed algorithm to translate IPv4 to IPv6 based on the maintained mapping information. And then, CPE generate IPv6 traffic and transmit to IPv6-only Host.

4. NAT64-CPE Approach Discussion

Considering above description, NAT64-CPE has following specific features.

- NAT64-CPE is capable of making residential server to be accessed, by means of which users could visit the IPv4-only server remotely.
- NAT64-CPE is a solely NAT64 deployed solution in CPE environment. It will compatible with legacy residential servers and no further updates requirements to DNS. Therefore, it’s liable to be deployed.

5. Security Considerations

Essentially, there are strong demands to have thorough security mechanism to prevent privacy invasion in CPE scenario. The detailed considerations need to be further identified.

6. IANA Considerations

This memo includes no request to IANA.

7. Informative References

[DNS64] Bagnulo, M., "DNS64: DNS extensions for Network Address Translation from IPv6 Clients to IPv4 Servers", IETF Internet-draft draft-ietf-behave-dns64-10.txt, July 2010.

[IPv6 Addressing of IPv4/IPv6 Translators] Bao, C., "IPv6 Addressing of IPv4/IPv6 Translators", draft-ietf-behave-address-format-10.txt (work in progress), August 2010.

[NAT64] Bagnulo, M., "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", draft-ietf-behave-v6v4-xlate-stateful-12.txt (work in progress), July 2010.

[PCP] Wing, D., "Pinhole Control Protocol (PCP)", draft-wing-softwire-port-control-protocol-02.txt (work in progress), July 2010.

Authors' Addresses

Gang Chen
China Mobile
53A,Xibianmennei Ave.,
Xuanwu District,
Beijing 100053
China

Email: phdgang@gmail.com

Hui Deng
China Mobile
53A,Xibianmennei Ave.
Beijing 100053
P.R.China

Phone: +86-13910750201
Email: denghui02@gmail.com

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: April 3, 2012

G. Chen
China Mobile

Oct 2011

NAT64 Operational Considerations
draft-chen-v6ops-nat64-cpe-03

Abstract

The document has summarized NAT64 usages on different modes, in which NAT64 may serve for a large-scale network or would give enterprise or residential service opportunities to be accessed by IPv6 remote subscribers. The document has described different operations for each usage and proposed operational considerations for each particular NAT64-mode.

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 3, 2012.

Copyright Notice

Table of Contents

1. Introduction

With fast developments of global Internet, the demands for IP address are rapidly increasing at present.This year, IANA announced that the global free pool of IPv4 depleted on 3 February. IPv6 is the only real option on the table. Operators have to accelerate the process of deploying IPv6 networks in order to address IP address strains.IPv6 deployment normally involves a step-wise approach where parts of the network should properly updated gradually. As IPv6 deployment progresses it may be simpler for operators and ICP/ISP to employ NAT64[RFC6146] functionalities at edge of IPv4 and IPv6 networks, since a significant part of network will still stay in IPv4 for long time. Especially, NAT64 could facilitate large ICP/ISP IPv6 transition process by eliminating upgrations of tremendous legacy IPv4 servers. Therefore, it’s quite popular to deploy NAT64 at the front of IDC to shift the entire service to be IPv6-enable.

Depending on different usage, NAT64 could be deployed on different places. The document has summarized NAT64 usages on different modes. Considering the existing deployment approaches, the memo has proposed different operational consideration for each particular NAT64-mode.

2. NAT64-CGN Deployment

2.1. Deployment in IDC

NAT has widely used in data center environments whenever IDC have to make your IPv4-only content available to IPv6 clients.

Figure 1 illustrates the usage where an IPv6-only host would like to initiate communications with IDC in IPv4 domain through NAT64. The NAT64 would accept IPv6 incoming session and distribute them to multiple IPv4 servers.

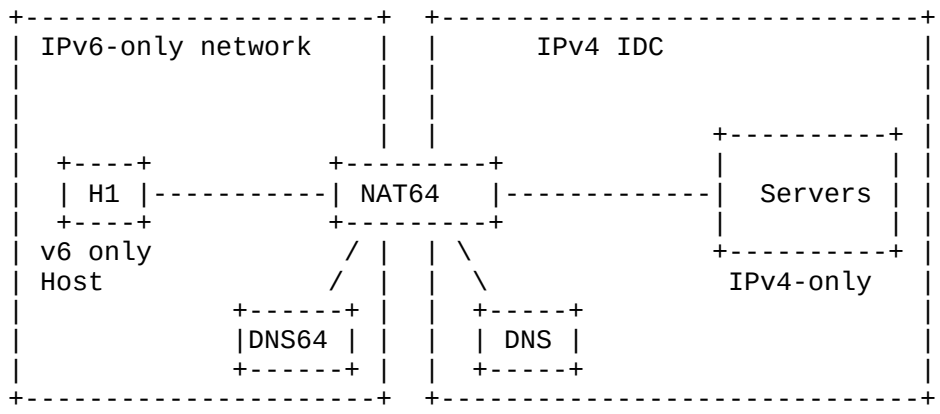


Figure 1: NAT64-CGN Mode Usage

NAT64 device in IDC may also take responsibilities of load balancer, which can accept incoming TCP/UDP sessions on a single virtual IPv6 interface or multiple IPv6 interfaces. Afterwards, it distributes them according to a specific algorithm it uses to multiple IPv4 servers. Ideally you could have a mix of IPv4 and IPv6 servers sitting behind the virtual IPv6 address.

Therein, NAT64 has to pick a new source IPv4 address and associated port number from local IPv4 address pool. DNS64 is a logical function that synthesizes DNS resource records(e.g., AAAA records containing IPv6 addresses) from DNS resource records actually contained in the DNS (e.g., A records containing IPv4 addresses).

2.2. Connecting with IPv4 Internet

NAT64 may also be used to connecting IPv6 users with IPv4 Internet. In this cases, NAT64 could colocated with BNG or Core Router to map legacy IPv4 servers into a NAT64 prefix and performs 6-to-4 address.

Therein, NAT64 would perform protocol translation mechanism and address translation mechanism. Protocol translation from an IPv4 packet header to an IPv6 packet header and vice versa is performed according to the IP/ICMP Translation Algorithm [RFC6145]. Address translation maps IPv6 transport addresses to IPv4 transport addresses and vice versa.

Following illustrates normal process for this usage.

- Step1: IPv6-only host performs an AAAA DNS query to DNS64 for the IPv6 address of the Pv4-only sever.
- Step2: DNS64 could not find the IPv6 address of the IPv4-only sever. So it tries to get the IPv4 address of the Pv4-only sever by sending A DNS query to DNS4.
- Step3: DNS4 return the A record to the DNS64.
- Step4: DNS64 map the IPv4 address to IPv6 address and send a synthetic AAAA record which is translated from A record to IPv6-only host.
- Step5: IPv6-only host send the IPv6 packet to the NAT64. NAT64 translates the IPv6 packet to IPv4 packet and send it to IPv4-only server.

2.3. NAT64-CGN Mode Requirements

According to above description for NAT64-CGN, the NAT64-CGN requirements are listed as following.

NAT64-CGN-R1: Each NAT64 device MUST have at least one unicast IPv6 prefix assigned to it, denoted Pref64::/n.

NAT64-CGN-R2:A NAT64 MUST have one or more unicast IPv4 addresses assigned to it.

NAT64-CGN-R3:Irrespective of the transport protocol used, the NAT64 MUST silently discard all incoming IPv6 packets containing a source address that contains the Pref64::/n.

NAT64-CGN-R4:The NAT64 MUST only process incoming IPv6 packets that contain a destination address that contains Pref64::/n. Likewise, the NAT64 MUST only process incoming IPv4 packets that contain a destination address that belongs to the IPv4 pool assigned to the NAT64.

NAT64-CGN-R5:NAT64 MUST support the algorithm for generating IPv6 representations of IPv4 addresses defined in RFC6052 as Address Translation Algorithms.

NAT64-CGN-R6:For incoming packets carrying TCP or UDP fragments with a non-zero checksum, NAT64 MAY elect to queue the fragments as they arrive and translate all fragments at the same time.

NAT64-CGN-R7: For incoming IPv4 packets carrying UDP packets with a zero checksum, if the NAT64 has enough resources, the NAT64 MUST

reassemble the packets and MUST calculate the checksum. If the NAT64 does not have enough resources, then it MUST silently discard the packets.

NAT64-CGN-R8: The NAT64 MAY require that the UDP, TCP, or ICMP header be completely contained within the fragment that contains fragment offset equal to zero.

NAT64-CGN-R9:The NAT64 MUST limit the amount of resources devoted to the storage of fragmented packets in order to protect from DoS attacks.

NAT64-CGN-R10:The NAT64 MUST make fragmentation process when MTU of incoming IPv4 traffic exceed maximum MTU on IPv6 side.

NAT64-CGN-R11:The NAT64 MAY let hosts and applications know IPv6 prefix used by the NAT64 and DNS64 so as to hosts have knowledge whether synthetic IPv6 address is targeted.

NAT64-CGN-R12:The NAT64 MAY decouple with DNS64 in order to establish communication with IPv4-only servers.

NAT64-CGN-R13:The NAT64 MAY take load-balancing functionalities incorporating with DNS64.

3. NAT64-CE Mode

NAT64-CE mode represents usages where there NAT64 is closed to customer edges, like enterprise network edge or residential network edge.

3.1. NAT64 at Enterprise Network Edge

Some enterprise would like to offers their employees with IPv6 access. However, the service may still stay in IPv4 domain. NAT64 useges in enterprise network could help shift all enterprise service to be IPv6 enable.

Figure 2 illustrates a network usage where an IPv6-only client attached to a dual-stack network, but the destination server is running on a private site where there is NAT64-CE numbered with public IPv6 addresses and private IPv4 addresses.

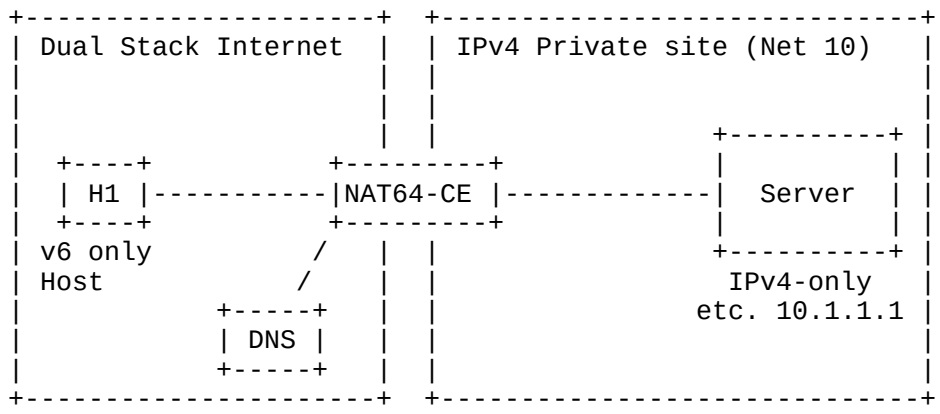


Figure 2: NAT64-CPE Mode Usage

3.2. NAT64 at Residential Network Edge

Residential servers are usually going beyond the operator's management. They may not able to IPv6-enable due to limitations of application supporting. In this case, ISP is still assigning private IPv4 address to servers. However, The nature of private IPv4 would block the end-to-end bi-directional communications. On the other hand, IPv6 will bring end-to-end benefits to operators. NAT64-CPE mode could let IPv6 users to access such IPv6-disable services in residential areas.

This scenario may appears in ISP network for several cases. As the instances, visitors go through distant network to take care of family affairs, like monitoring house security via residential camera, manipulating household appliances remotely prior to comeback home.

4. Security Considerations

Essentially, there are strong demands to have thorough security mechanism to prevent privacy invasion in NAT64-CPE scenario. The detailed considerations need to be further identified.

5. IANA Considerations

This memo includes no request to IANA.

6. Normative References

[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.

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

Author's Address

Gang Chen
China Mobile
53A,Xibianmennei Ave.,
Xuanwu District,
Beijing 100053
China

Email: chengang@chinamobile.com

Internet Draft
Category: Proposed Standard
Updates: 4291, 5952
Expires: April 6, 2011

L. Donnerhacke
Editor (DENOG)
Richard Hartmann
Editor (DENOG)

October 6, 2010

Naming IPv6 address parts
draft-denog-v6ops-addresspartnaming-02.txt

Status of this Memo

This Internet-Draft will expire on April 6, 2011.

Copyright Notice

Abstract

In the daily communication between technicians, engineers and other people who need to deal with computer networks, it is often necessary to refer to particular parts of IP addresses. In the world of IPv4, the term "octet" is well established, however as the use of IPv6 is spreading, it becomes apparent that there is no such commonly accepted term for IPv6 addresses.

Discussing and explaining technical matters become difficult when different people use different terms for the same thing. Therefore, this document discusses several naming proposal for those 16bit pieces of IPv6 addreses.

Table of Contents

1. Introduction

Verbal and written communication requires a common set of terms, easily understood by every potential party. While deploying IPv6, when refering to segments of IPv6 addresses, confusion regularly arises due to the usage of different and sometimes conflicting nomenclature for the same pieces of information.

[IPV6Addr] is the normative reference to IPv6 addressing and avoids to coin a special term for the subject of this document itself:

> The preferred form is x:x:x:x:x:x:x:x, where the ’x’s are one to four hexadecimal digits of the eight 16-bit pieces of the address.

[IPv6Rep] is the normative reference to IPv6 address text representation and introduces the term "16-bit field" or short "field".

2. Rationale

While we readily agree that the naming of IPv6 address parts is not the most pressing concern the Internet is facing today, a common nomenclature is important for efficient communication.

In IPv6 deployments the delimiting colons are regularly used to facilitate the separation of labels discerning not only administrative boundaries but also network segments and distinct infrastructure components. Consequently the values between the colons are frequently refered to especially in communication regarding coordinative matters.

Time spent explaining what one is referring to is wasted and conflicting names can lead to misunderstanding while the usage of a common term helps facilitating quick understanding.

To solve this problem, the specification of a precise and recognizable term is advised.

A typical ambiguity occurs in [IPv6Rep] which uses the term "field" or "16-bit field" for the term in question. This case is interesting because there was a short IETF WG discussion which term should be used.

> If an IPv6 address field in a certificate was incorrectly verified by converting it to text ...

Since parts of the internet community only accept authoritative advice substantiated by a published document, also known as the ’citation needed’ approach, it is helpful to have a definite source.

3. Naming Considerations

Any term that can be confused with other technical terms due to phonetic similarities can lead to misconfiguration causing reachability and security risks to the involved parties. Even with English being the preferred language in the IT world today, a good name should describe the technical matter precisely while being easy to remember, spell and pronounce in as many languages as possible.

4. Naming Proposals

We are presenting a broadest selection of mostly serious proposals which needs to be narrowed in the future by straw polls and finally select one using normal IETF consensus.

4.1. Chazwazza

"Chazwazza" was proposed as a Simpsons reference, see [greg]. While this is certainly a unique term in the networking world, it is not particularly meaningful nor easy to pronounce.

4.2. Chunk

A chunk is commonly understood to be a specific amount of data. The term is not unique to IPv6, however easy to remember and pronounce.

4.3. Column

The colons in an IPv6 address’ text representation make it similar to a table. Besides that, the meaning of the word "column" has very little to do with the actual technical meaning of a 16bit piece of an IPv6 address, though.

4.4. Colonade, Colonnade

Based on the colon as seperator the word sounds English (using a single ’n’ to make it an artificial word) and is easy to spell and pronounce. Alternatively, "colonnade" could be used, overloading the existing, yet unrelated word with a new meaning.

4.5. Doctet

Derived from "double octet", thus accurately describes the technical matter, as an octet is a standard term for a sequence of 8 bits.

4.6. Field

A "field" describes a form of a data structure in many programming languages. The term stresses the fact that a field is one of multiple fractions of a bigger subject, just like countryside is divided into fields, or like IPv6 addresses into 16bit long pieces. A drawback of that similarity is the lack of uniqueness to IPv6, though.

4.7. Hexadectet

"Hexadectet" is directly derived from IPv4's "octet", thus technically correct and probably convenient to get used to. On the other hand, it is much harder to pronounce.

4.8. Hit

Short for "hex-bit", short and convenient to pronounce, however usually associated with a completely different meaning.

4.9. Orone

Initially started as a typo in [greg], "orone" is a short, unique word without a specific meaning yet.

4.10. Part

The word "part" has been used throughout this document to describe the subject until there is a better term for this. It is very unspecific and can be used in countless ways, not only to describe 16bit long parts of an IPv6 address.

4.11. Provider number, customer number, network number

These terms provide semantic descriptions of the different parts of an IPv6 address. However, it is not within the scope of this document to find terms describing semantic, but rather syntactic elements.

Furthermore, naming the 16bit pieces of IPv6 addresses in a semantic way would introduce new problems, like limited applicability, e.g. it would not work for multicast addresses.

4.12. Quad nibble, qibble, quibble

A nibble is a 4bit entity, hence 16 bits are a quad nibble. This is a rather bulky word, however, so "quibble" is a convenient abbreviation. Also, it is a unique term, thus eliminating any chances of misinterpretation.

4.13. Segment

"Segment" is another obvious choice, however it is also quite unspecific and used in different contexts, e.g. "network segments".

4.14. Tuple

A tuple is a sequence of typically heterogenous elements considered as a new entity by itself. It is also a short, descriptive word that is not yet associated with anything networking related. Usually a tuple exceeds grouping by creating a new semantic level.

4.15. Word

A "word" usually refers to a fixed group of bits that are processed at a time, and especially on legacy x86 systems is a synonym for 16 bits. It has a different and much more unspecific meaning to less technically skilled people, which might be problematic.

5. Security Considerations

This memo does not directly discuss security issues, however the lack of a common, well established term could theoretically lead to misinterpretation, possible leading to insecure configuration of computer systems.

6. IANA Considerations

No assignments by the IANA are required. However it is considered desirable that the IANA adopts the term in future documents.

7. References

7.1. Normative References

[IPV6Addr] Deering, S. and R. Hinden, "IP Version 6 Addressing Architecture", RFC 4291, February 2006

[IPv6Rep] Kawamura, S. and Kawashima, M., "A Recommendation for IPv6 Address Text Representation", RFC 5952, August 2010

[Q.6] ITU-T, "Advantages of international automatic working", Fascicle VI.1 of the Blue Book, 1988

7.2. Informal References

[greg] http://etherealmind.com/network-dictionary-chazwazza/, Sept 5, 2010

8. Acknowledgements

Thanks go to Greg Ferro who initiated the discussion by proposing the term "chazwazza".[greg]

Thanks all the people who read to this point and are willing to provide valuable input instead of simply shaking their heads and moving on.

The inital version of this document was created following the spirit of [Q.6].

Authors' Addresses

Lutz Donnerhacke
Leutragraben 1
07743 Jena
Germany
Tel: 1.6.5.3.7.5.1.4.6.3.9.4.e164.arpa.
EMail: lutz@thur.de

Richard Hartmann
Munich
Germany
Email: richih.mailinglist@gmail.com
http://richardhartmann.de

Michael Horn
Po Box 540153
10042 Berlin
Germany
http://nibbler.tel/

Kay Rechthien
Netsign GmbH
Lindenallee 27
14050 Berlin
Germany
EMail: kre@netsign.eu

Leon Weber
Ahornstrasse 5d
01458 Ottendorf-Okrilla
Germany
EMail: leon@whitejack.org

Supporter's Addresses

Ronny Boesger
Lahnsteiner Strasse 7
07629 Hermsdorf
eMail: rb@isppro.de

Thorsten Dahm
Josefstrasse 21
66265 Heusweiler
Germany
EMail: t.dahm@resolution.de

Joerg Dorchain
Harspergerflur 23
66740 Saarlouis
Germany
EMail: joerg@dorchain.net

Sascha Lenz
s-lz.net
Zum Oberbaeumle 49
97318 Kitzingen
Germany
E-Mail: sascha.lenz@s-lz.net

Jens Link
Freelance Consultant
Foelderichstr. 40
13595 Berlin
Germany
EMail: jl@jenslink.net

Jan Walzer
Kopernikusstrasse 2
68519 Viernheim
Germany
EMail: jan.w@lzer.net

Sebastian Wiesinger
Germany
EMail: sebastian@karotte.org

Appendix A. Change History

00 - inital version

01 - Jens Link moved from Author to Supporter
- Leon Weber moved from Supporter to Author
- numerous typographic fixes
- added "field" from [IPv6Rep] as proposal and as reason
- added "part" for completeness
- dismissed "hextet / hexatet / sixlet"
- created sub sections for each proposal
- added update notification of RFC4291 and RFC5952
- added a "Security considerations" section

02 - Fixing nits
- Propose a selection mechanism

Internet Draft
Category: Proposed Standard
Updates: 4291, 5952
Expires: October 5, 2011

L. Donnerhacke
Editor
Richard Hartmann
Editor

April 7, 2011

Naming IPv6 address parts
draft-denog-v6ops-addresspartnaming-04

Status of this Memo

Copyright Notice

Abstract

In the daily communication between technicians, engineers and other people who need to deal with computer networks, it is often necessary to refer to particular parts of IP addresses. In the world of IPv4, the term "octet" is well established, however as the use of IPv6 is spreading, it becomes apparent that there is no such commonly accepted term for IPv6 addresses.

Discussing and explaining technical matters become difficult when different people use different terms for the same thing. Therefore, this document discusses several naming proposal for those 16bit pieces of IPv6 addreses.

Table of Contents

1. Introduction

Verbal and written communication requires a common set of terms, easily understood by every potential party. While deploying IPv6, when refering to segments of IPv6 addresses, confusion regularly arises due to the usage of different and sometimes conflicting nomenclature for the same pieces of information.

[IPV6Addr] is the normative reference to IPv6 addressing and avoids to coin a special term for the subject of this document itself:

> The preferred form is x:x:x:x:x:x:x:x, where the ’x’s are one to four hexadecimal digits of the eight 16-bit pieces of the address.

[IPv6Rep] is the normative reference to IPv6 address text representation and introduces the term "16-bit field" or short "field".

2. Rationale

While we readily agree that the naming of IPv6 address parts is not the most pressing concern the Internet is facing today, a common nomenclature is important for efficient communication.

In IPv6 deployments the delimiting colons are regularly used to facilitate the separation of labels discerning not only administrative boundaries but also network segments and distinct infrastructure components. Consequently the values between the colons are frequently refered to especially in communication regarding coordinative matters.

Time spent explaining what one is referring to is wasted and conflicting names can lead to misunderstanding while the usage of a common term helps facilitating quick understanding.

To solve this problem, the specification of a precise and recognizable term is advised.

A typical ambiguity occurs in [IPv6Rep] which uses the term "field" or "16-bit field" for the term in question. This case is interesting because there was a short IETF WG discussion which term should be used.

> If an IPv6 address field in a certificate was incorrectly verified by converting it to text ...

Since parts of the internet community only accept authoritative advice substantiated by a published document, also known as the ’citation needed’ approach, it is helpful to have a definite source.

3. Naming Considerations

Any term that can be confused with other technical terms due to phonetic similarities can lead to misconfiguration causing reachability and security risks to the involved parties. Even with English being the preferred language in the IT world today, a good name should describe the technical matter precisely while being easy to remember, spell and pronounce in as many languages as possible.

4. Naming Proposals

4.1. hextet

"hexadectet" is directly derived from IPv4's "octet", thus technically correct and convenient to get used to. Because it is harder to pronounce, the short form "hextet" is used.

"hextet" MUST be used in all technical documents and specifications refering to IPv6 address parts.

4.2. quibble

A nibble is a 4bit entity, hence 16 bits are a quad nibble. This is a rather bulky word, however, so "quibble" is a convenient abbreviation. It is a unique term in networking but has an existing meaning in ordinary English.

"quibble" MAY be used for informal communication.

5. Security Considerations

This memo does not directly discuss security issues, however the lack of a common, well established term could theoretically lead to misinterpretation, possible leading to insecure configuration of computer systems.

6. IANA Considerations

No assignments by the IANA are required. However it is considered desirable that the IANA adopts the term in future documents.

7. References

7.1. Normative References

[IPV6Addr] Deering, S. and R. Hinden, "IP Version 6 Addressing Architecture", RFC 4291, February 2006

[IPv6Rep] Kawamura, S. and Kawashima, M., "A Recommendation for IPv6 Address Text Representation", RFC 5952, August 2010

[Q.6] ITU-T, "Advantages of international automatic working", Fascicle VI.1 of the Blue Book, 1988

[RFC2119] S. Bradner, Key words for use in RFCs to Indicate Requirement Levels

7.2. Informal References

[greg] http://etherealmind.com/network-dictionary-chazwazza/, Sept 5, 2010

8. Acknowledgements

Thanks go to Greg Ferro who initiated the discussion by proposing the term "chazwazza".[greg]

Many thanks to all those people which contribute to our work and participate in the straw poll about all the other propoals, which are described in former versions of this memo: Chazwazza, Chunk, Column, Colonade, Colonnade, Doctet, Field, Hexadectet, Hit, Orone, Part, Provider number, customer number, network number, Quad nibble, qibble, Segment, Tuple, and Word.

The inital version of this document was created following the spirit of [Q.6].

Authors' Addresses

Lutz Donnerhacke
Leutragraben 1
07743 Jena
Germany
Tel: 1.6.5.3.7.5.1.4.6.3.9.4.e164.arpa.
EMail: lutz@thur.de

Richard Hartmann
Munich
Germany
Email: richih.mailinglist@gmail.com
http://richardhartmann.de

Michael Horn
Po Box 540153
10042 Berlin
Germany
http://nibbler.tel/

Kay Rechthien
Netsign GmbH
Lindenallee 27
14050 Berlin
Germany
EMail: kre@netsign.eu

Leon Weber
Ahornstrasse 5d
01458 Ottendorf-Okrilla
Germany
EMail: leon@whitejack.org

Supporter's Addresses

Ronny Boesger
Lahnsteiner Strasse 7
07629 Hermsdorf
eMail: rb@isppro.de

Thorsten Dahm
Josefstrasse 21
66265 Heusweiler
Germany
EMail: t.dahm@resolution.de

Joerg Dorchain
Harspergerflur 23
66740 Saarlouis
Germany
EMail: joerg@dorchain.net

Sascha Lenz
s-lz.net
Zum Oberbaeumle 49
97318 Kitzingen
Germany
E-Mail: sascha.lenz@s-lz.net

Jens Link
Freelance Consultant
Foelderichstr. 40
13595 Berlin
Germany
EMail: jl@jenslink.net

Jan Walzer
Kopernikusstrasse 2
68519 Viernheim
Germany
EMail: jan.w@lzer.net

Sebastian Wiesinger
Germany
EMail: sebastian@karotte.org

Appendix A. Change History

00 - inital version

01 - Jens Link moved from Author to Supporter
- Leon Weber moved from Supporter to Author
- numerous typographic fixes
- added "field" from [IPv6Rep] as proposal and as reason
- added "part" for completeness
- dismissed "hextet / hexatet / sixlet"
- created sub sections for each proposal
- added update notification of RFC4291 and RFC5952
- added a "Security considerations" section

02 - Fixing nits
- Propose a selection mechanism

03 - Added Hextet
- Removed references to DENOG
- Select two proposals based of the straw poll and the WG

04 - Upgraded hextet to MUST
- Corrected formalia & typos
- lower-cased hextet and quibble as that is consensus for octet
- Formatting
- Fixed nits introduced by -03

v6ops
Internet-Draft

D. Sturek
Pacific Gas & Electric
T. Herbst
Silver Spring Networks

Intended status: Informational

October 15, 2010

CPE Considerations in IPv6 Deployments
draft-herbst-v6ops-cpeenhancements-00

Status of this Memo

This Internet-Draft will expire on April 17, 2011.

Copyright Notice

This Internet-Draft will expire on April 17, 2011.

Abstract

Smart metering deployments in residential settings introduce the prospects of ad-hoc deployment of internetworked IPv6 customer premise equipment (CPE). WiFi access points, cable boxes and other home devices with internet access could all be internetworked with smart metering devices by customers with no data networking expertise resulting in a complex multi-segment network with differing prefixes, routing support and service discovery needs.

Table of Contents

1. Introduction

The availability of energy usage information within the Home Area Network, enabled through smart meter deployment, adds a popular interconnection target for electricity customers, service providers and third party suppliers. These opportunities for energy usage management are all assuming a home owner with no data networking expertise can link together a collection of standalone networks and enable a consistent set of services and device addressing modes. This draft starts a discussion on needed standards work to make the deployment of these services a reality in an IPv6 environment.

2. Description

In a regulated utility environment, utilities must deploy energy savings programs accessible to all customers. Broadband internet access cannot be assumed since around 40% of customers don’t have broadband. The smart meter is then architected as a standalone border gateway with a unique prefix supplied by the smart meter.

To fully enable deployment of energy savings applications onto a variety of devices, ad hoc internetworking of smart meters with HAN devices and existing networks employing diverse data links such as IEEE 802.15.4, IEEE P1901,and WiFi must be supported. The set of issues to be addressed include:

- Assignment of /64 prefixes from Globally Unique Address (GUA) and Unique Local Address (ULA) [RFC4193] prefixes available to the residential network
- Introduction of ULAs for a residence
- ULA Delegation and Reassignment when network segments with differing ULAs are combined
- Enablement of intra-network routing when CPEs within the residence are interconnected
- Extensions to multicast DNS to extend local name resolution across a multi-link residential network

2.1 Assignment of /64 prefixes from GUA and ULA prefixes

The residential network that includes multiple links will need a mechanism for assigning /64 prefixes for each link from one or more shorter prefixes assigned to the network. For example, DOCSIS 3.0 uses DHCPv6-PD [RFC3633] to delegate a prefix to the residential gateway. /64 prefixes from this delegated prefix must be assigned to every link within the residential network.

Similarly, the residential network may have a ULA prefix for local traffic if the residential network does not have any GUA prefixes (see section 2.2). /64 prefixes from the ULA must be assigned to the links in the residential network.

2.2 Unique Local Addresses (ULAs)

IPv6 offers three types of addressing prefixes: GUA, ULA and link-local. ULA prefixes are useful in the residential network scenario for local communication when no GUA prefixes are availalbe; e.g., when the external link to the ISP is unavailable and no delegated prefixes are available.

The first requirement is that gateways in the residential network create a ULA for use within the network rooted at the gateway and no other ULA prefix is available.

The second requirement is that when multiple networks are created, then interconnected in a home, multiple ULAs may be present. When these networked are interconnected (by a homeowner without networking skills), the ULAs for these network segements should be harmonized without user interaction into a single set of ULAs and notification made to hosts holding references to the previous ULAs.

2.3 Intra-network routing with multiple internet connected CPEs

As network segments are interconnected, and CPE devices become border gateways for new bordering network segments, a routing protocol like RIPng needs to be supported. As noted for ULA delegation, the CPE needs to automatically detect the need in support for inter-segment routing and provide support automatically.

2.4 Extensions to multicast DNS for sitewide name resolution

For service discovery, two alternatives exist: user agent based

discovery and directory agent based discovery described as follows:

- User agent: Devices hold service discovery information themselves and respond to discovery requests based on matching criteria in the request. DNS Service Discovery [DNS-SD] resolved over Multicast DNS [mDNS] is an example of this type of solution.
- Directory agent: Devices register service discovery information with a central repository. A well known example of this type of solution includes uPnP [uPnP] which uses the Simple Service Discovery Protocol (SSDP) [SSDP]. Note that uPnP supports both user agent and directory agent service discovery methods.

mDNS only provides link-local name resolution. Use of mDNS in the residential network requires extensions so that mDNS can use site-local multicast that spans multiple hops using IP forwarding for sitewide local name resolution.

3. Future Work

The following work items are proposed:

- Create extensions to DHCPv6-PD to delegate prefixes across multiple links
- Define precoedures for gateways to generate a ULA if required
- Create procedures for HAN devices to join the ULA and procedures to combine network segments with different ULAs into a single ULA.
- Define mechanisms for automated provisioning and operation of routing across multiple links in a residential networl
- Create extensions to multicast DNS to support sitewide local name resolution across multiple links

4. Conclusions

To realize deployment requirements of self installed, ad hoc networking where different segments can be installed and provisioned at different times and where various link technologies may be used, additional features are needed in CPE.

5. Security Considerations

This requirements document introduces no security considerations.

6. IANA Considerations

This requirements document introduces no IANA considerations.

7. Acknowledgments

8. References

8.1. Normative References

8.2. Informative References

[mDNS] Cheshire, S. and Krochmal, M., "Multicast DNS", draft-cheshire-dnsext-multicastdns-11 (work in progress), March 2010.

[DNS-SD] Cheshire, S. and Krochmal, M., "DNS-Based Service Discovery", draft-cheshire-dnsext-dns-sd-06.txt (work in progress), March 2010.

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

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

[uPnP] uPnP Forum, "uPnP Device Architecture v1.1", 15 October, 2008

[SSDP] Goland, Y., Cai, T., Gu, Y., Albright, S., "Simple Service Discovery Protocol/1.0 Operating without an Arbiter", October 1999 (expired April 2000)

Authors' Addresses

Tom Herbst
Silver Spring Networks
Redwood City, CA
USA

Phone: +1 650-542-4782
Email: therbst@silverspringnet.com

Don Sturek
Pacific Gas & Electric
77 Beale Street
San Francisco, CA
USA

Phone: +1-619-504-3615
Email: d.sturek@att.net

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: April 23, 2011

C. Huang, Ed.
X. Li
L. Hu
China Telecom
October 20, 2010

Use Case For IPv6 Transition For a Large-Scale Broadband network
draft-huang-v6ops-v4v6tran-bb-usecase-01

Abstract

This document describes a use case for the migration from IPv4 to IPv6 for one of the typical broadband networks. The contant is orgnized by various scenarios we can foresee during the migration.

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 23, 2011.

Copyright Notice

Table of Contents

1. Introduction

The situations of IPv6 transition for developing countries are different from the developed countries. The developed countries,which is the originate place of Internet, hold the majority part of IPv4 address space. However, according to the considerably high growth speed of the potential subscribers in the developing countries due to the large base number of users and booming economies, e.g. China, Brazil and India, the small IPv4 address space do put these developing countries under great pressure.

Generally speaking, developing countries have a large number of subscribers. Some of them with more than dozen millions of broadband subscribers, increase at a rate of 20+ percent of subscribers annually.

Developing countries are facing unprecedented pressure in business aspect, with the Global IPv4 address exhaustion. Therefore developing countries' network and services will migrate to IPv6 eventually. However, during the transition procedure, developing coutries should seriously concerned about the transition of existing services in order to support the v4v6 coexistent environment, along with the researches and developments of new services and applications. Developing countries broadband networks are so large with various types of service that the transition to IPv6 is doomed to be complicated and difficult.

One of the typical network structure is shown in Figure 1. As this figure shows, the network comprises three segments: backbone network which usually includes IP backbone and MPLS backbone, metro network (MAN) which basically includes Core Router(CR),Aggregation Router(AR)and Broadband Remote Access Server(BRAS), and access network covers from BRAS to users' Customer Premises Equipment(CPE).

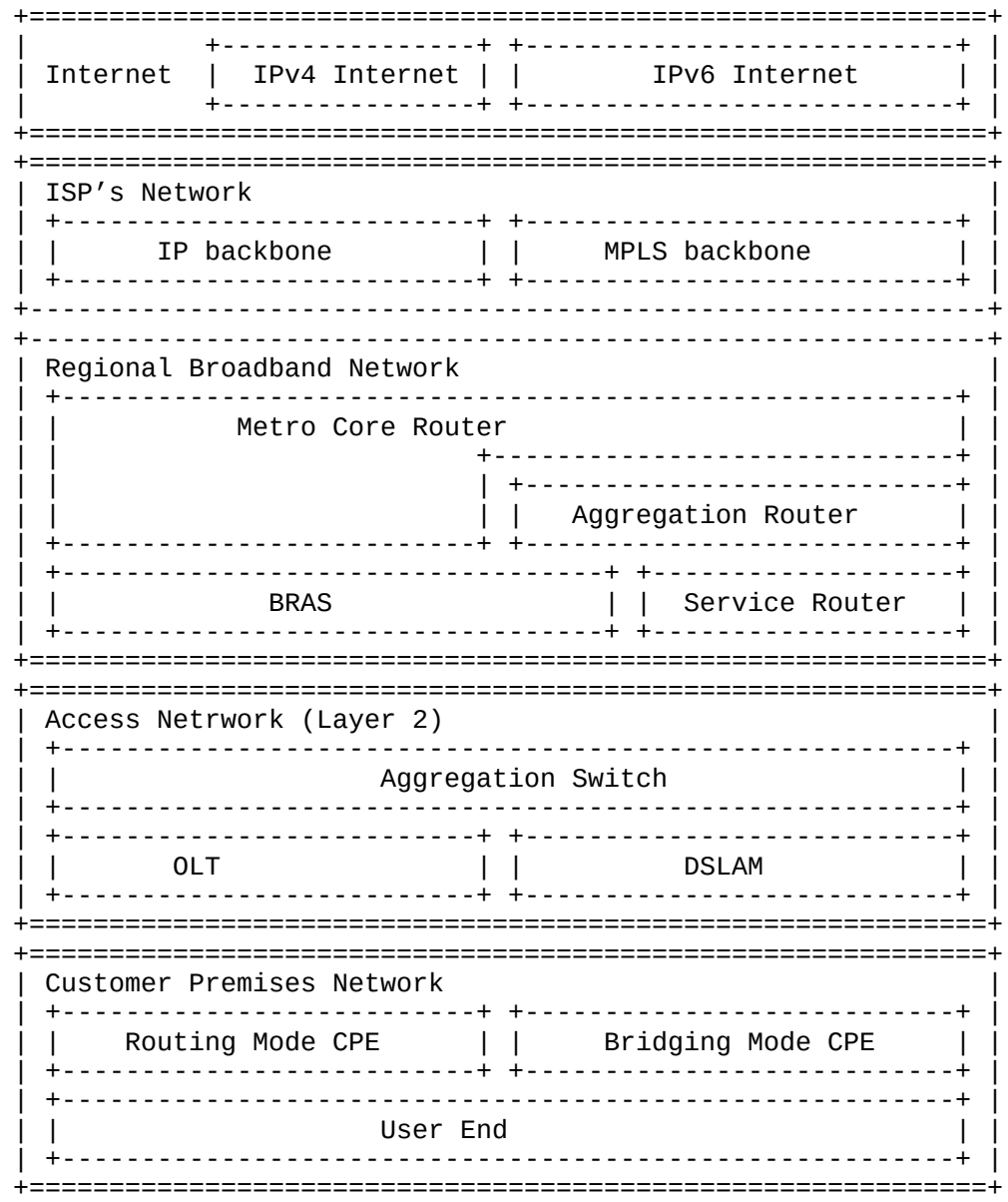


Figure 1: Typical Large-scale Broadband Network Architecture

The backbone network provides long distance transmission for the broadband traffic, which includes an IP backbone and a MPLS one. The former one provides Internet service for home users and SME (Small and Medium Enterprise) users while the latter one provides VPN and

leased line services for the enterprise customers.

The metro network is mainly composed of Core Routers, Aggregation Routers, and BRASs. CR, as the egress router of the metro network, is connected to both IP backbone and MPLS backbone. Most BRASs are connected directly to Core Routers, but a small portion of the BRASs in some large metro networks are connected to Core Routers via Aggregation Routers.

The access network provides broadband access service for users. It mainly includes layer 2 devices (e.g., DSLAM, Aggregation Switch).Broadband access service is provided over ADSL, LAN, PON and so on.In the access network, the IPv6 capability is limited due to some security considerations (IP-Based ACLs or policies). The best part of UE access the network using the PPPOE dial up method. So, hosts can acquire IPv6 address through PPPoE to avoid the problem.

With respect to the terminals, the Windows(TM) OS dominates in the market, the Windows Vista and Windows 7 is the minority while the Windows XP is majority. The WindowsXP cannot support IPv6 PPPOE.

The situation of terminals in some developing countries, e.g. China is somewhat different from the situation of terminals in Europe. CPE can operates either in routed or in bridged mode. The major part of existing Routed mode CPE cannot support IPv6 PPPOE dial up. The bridged mode CPE allows the users to dial up from PCs. CPE Home Gateways in some cases are purchased by the users themselves, which are unmanageable by the service provider.

During the migration to IPv6, the transition strategies and technologies selection becomes one of the most significant issues due to the complexity of the network and the services, the large number of scenarios and the multiple methods of terminals and user access. However, there is no universally applicable solution or guideline for the migration of network and services to native IPv6 in the industry yet. Each operator is looking for the appropriate transition strategy for itself.

To investigate the impact of various transition technologies on network and services and select correct migration procedures. A lot of experiments were conducted on its existing networks, covering the backbone network, metro network, terminals, service platforms, and the provisioning systems.

In this document, several possible migration scenarios applicable to the typical network are introduced. Related solutions for these scenarios are also introduced.

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. Backbone Network Migration

As stated in Section 1, there are usually two types of backbones: IP backbone and an MPLS backbone.The migration solutions could be enabling 6PE in MPLS backbone, dual stack capability in IP backbone, and building up a new IPv6-only backbone network.

2.1. Solution 1: 6PE in the MPLS Network

MPLS backbone provides VPN service for the large scale enterprise customers.When the whole network deploys MPLS, 6PE [rfc4798] can be used to provide IPv6 transmission.The IPv6 routing information is marked with MPLS labels through IBGP and is distributed into IPv4/ MPLS backbone network.The communication of IPv6 is achieved by the LSP among PEs.Using 6PE and implementing IPv4 and IPv6 protocol stack at the PE device connecting to IPv6 network, the original IPv4/ MPLS network in the backbone network could be adopted to provide access capability for the distributed IPv6 only user.

Technology saying, there is no problem with 6PE. However, the effect of large scale deployment of 6PE(over thousands nodes)should be evaluated in the future. In addition, since the IPv6 packet is encapsulated in IPv4 tunnel, it is not easy for trouble shooting.

2.2. Solution 2: Dual-stack IP Backbone

The device enables IPv4/IPv6 protocol stack at the same time.IPv4 and IPv6 routing are both in the routers which forward IPv4/IPv6 packet separately based on the IPv4/IPv6 routing tables.

At present, functionally, the new equipments has better support for DS while the left old devices do not.

2.3. Solution 3: IPv6-Only Backbone

Newly establish a native IPv6 network according to the scale of current backbone network. The device enables IPv6 protocol stack only. There is IPv6 routing only in the router which does not carry IPv4 traffic.

From technology aspect, the IPv6 is running well in various existing

experimental network. But from the business aspect, according to many large-scale ISP which have no commercial experiences and examples to reference, establishing a pure IPv6 network has the risk impacting existing services. Besides, the IPv6 information resource is extremely limited in recent time. It means IPv4 traffic dominates the backbone traffic and consequently cause waste for the pure IPv6 network.

2.4. Conclusion

In conclusion, for the migration of backbone network, the most reliable method is enable IPv6 in all the new-added devices, since "support IPv6" in the manual script is far from well running in current network.

3. Regional IP Network migration

The metro network here covers the network from BRAS to metro egress router. Generally speaking, there are 4 mainstream programs.

The Overview of the solutions in the Regional Broadband Network is reduced to the following three types:

Upgrading the existing BRASs and CRs for existing users, and adding new BRASs for new users;

Upgrading the existing BRASs and CRs for all users;

Building a completely new Regional Broadband Network for new users;

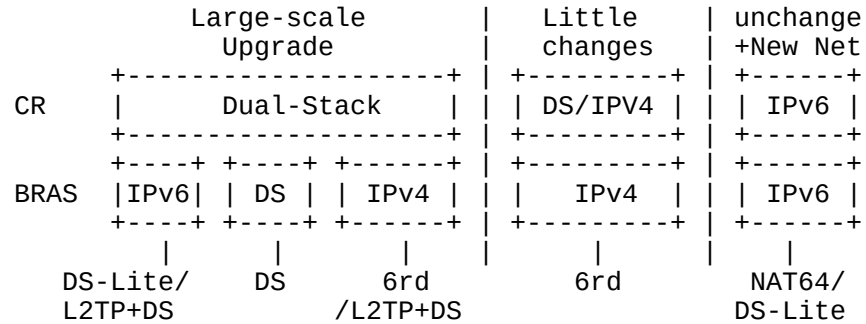


Figure 2: Overview of Solutions in Regional Broadband Network

In each transition solution of regional broadband network, it can connect to one of the following backbone described in section 3.1:

- Connect to the Dual-stack IP backbone;
- Connect to the IPv6-only backbone for IPv6 traffic and to the existing IP backbone for IPv4 traffic if it has;;
- Connect to the MPLS backbone for IPv6 traffic and to the existing IP backbone for IPv4 traffic if it has.

Some less possible transition solutions haven't been listed above:

- Upgrade the existing regional broadband network to IPv6-only; It will lead to a huge influence to existing network and services. ;
- Create a new regional broadband network with native IPv6 CRs and Dual-Stack BRASs; It has very low possibilities because if we create a new regional broadband network to provide dual-stack service with new dual-stack BRAS, the simplest solution will be let the new CRs to be dual-stack too. If the new CRs are IPv6-only, they need other transition technologies working together which seem to be more complicated. ;
- Create a new regional broadband network with Dual-stack CRs and native IPv6 BRASs; It also has very low possibilities and the reason is as same as the above one.

In the following sections, the technical solutions based on the scenarios in Figure 2 are discussed. Although there may be many technical options in each scenario, the discussion will focus on one of them.

The possible solutions referred to Figure 2 that we will discuss:

- Solution 1: Dual-Stack and L2TP ;
- Solution 2: Dual-Stack over IPv6 - DS-lite;
- Solution 3: Dual-Stack over IPv4 - 6rd
- Solution 4: IPv6 and protocol translation

In this document, we consider that the CPE is basically purchased by customers themselves. The access method of subscribers in each technical solution will be also discussed in this section. In the PPPoE dial-up cases, most users dial-up from PC, but there is some deployed a Home Gateway (e.g. WLAN AP) by themselves and set up an automatically dial-up from it. Until now, most terminals, including PCs and CPEs, will still be IPv4-only. Even if most PC operating system (OS) declared that they already supported IPv6, there is still

a problem on supporting PPPoE with IPv6. Not only the most widely used OS, Windows(TM) XP, doesn't support PPPoE with IPv6, but also nearly all CPEs in the market does not support PPPoE in IPv6 environment. These problems will be a significant bottleneck of the development of IPv6 broadband.

3.1. Solution1:Dual-Stack and L2TP

In this solution, both the CRs and BRASs will be transition to Dual-stack by upgrading or replacing the existing devices. However, there are so many different BRASs with diverse IPv6 capability in a large-scale broadband network. So there is a possibility that some BRASs cannot upgrade to Dual-stack and PPPoE with IPv6.

In the Figure 3 below, there are three scenarios in this solution.

- Scenario 1: A Dual-stack, IPv4 or IPv6 terminal accessing to a Dual-stack BRAS;
- Scenario 2: A Dual-stack or IPv6 terminal accessing to a legacy IPv4-only BRAS
- Scenario 3: An IPv4 terminal accessing to a legacy IPv4-only BRAS

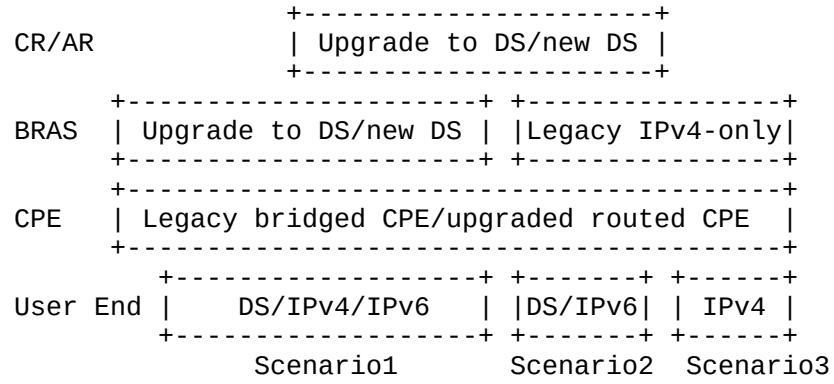


Figure 3: Dual-stack Transition Solution

The Scenario 1 is very simple. But the routing CPE at the edge of customer premises network need to be upgraded to support IPv6 and PPPoE with IPv6. And the PC operation system (OS) also need to support PPPoE with IPv6.

The Scenario 3 is as the same as the access method currently.

The Scenario 2 is a little bit complicated. The BRAS which the subscriber is connecting to is not support IPv6 and PPPoE with IPv6. So, one possible solution could be terminating the point-to-point protocol (PPP) [RFC1661] link at a remote Dual-stack BRAS. A tunnel technology like Layer 2 Tunnel Protocol (L2TP) [RFC2661] [RFC2661] can be used in this scenario. Other technologies could be an alternative. But considering the device capabilities and the maturity of the technology, the following discussion will focus on the solution that Dual-stack network with L2TP to provide Dual-stack services. [SeeFigure 4]

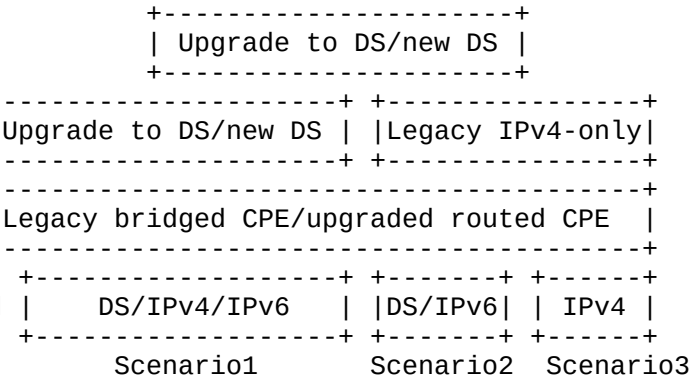


Figure 4: The L2TP Solution in partly Dual-Stack network

Although tunnel technologies can solve this problem, it is considered as a temporary solution. The legacy IPv4-only BRASs will be replaced eventually.

For the Dual-stack service, IPv4 address is still need to allocate to terminal. After the IPv4 addresses exhaustion, Dual-stack BRASs could allocate private IPv4 addresses for broadband subscribers, and a NAT44 Large Scale NAT (LSN) [I D.kuarsingh lsn deployment] device will be deployed to provide IPv4 NAT services for subscribers who are using private IPv4 addresses.

The operating system (OS) of the new subscriber is recommended to support PPPoE with IPv6 [TR 187]. Third-party dial-up software could be provided if the OS is not support PPPoE with IPv6.

The routing mode CPE of new subscriber is required to support PPPoE with IPv6. Otherwise, they are required to turn off the auto-dialup function, and initial the PPPoE dial-up session from the host that supports PPPoE with IPv6.

The legacy subscribers are recommended to upgrade their OSs and CPEs, but not required. They can still access by IPv4-only. Third-party dial-up software could also be provided to support PPPoE with IPv6.

Applicable scenarios: This solution could be suitable for the initial stage or the intermediate stage of the IPv6 transition when the IPv4 traffic is still very large in the network. And the broadband network is going to provide Dual-stack services with incremental deployment. It is also suitable when the number of subscribers is increasing very fast, and there is a large amount of CPEs and OSs that do not support PPPoE with IPv6.

In conclusion, for dual stack aspect, The newly-added equipments have few problem with the dual stack while the old ones do not. In these new devices, the aggregation routers have much more problems than the core routers. The switches has poor support capacities of IPv6. The BRASs have many problems, like no well-formed access protocols, no specificated address allocation policies, too many uncertain factors which may influence the network operation and maintenance. Moreover, most metro network is formed through continually devices expansion. So, there are mounts of equipments which are too old to be upgraded. Moreover, with IPv6 information resource increasing, how to provide IPv6 internet visiting for the existing IPv4 subscribers should be put into consideration.

3.2. Solution2: Dual-Stack over IPv6 - DS-lite

In this solution, the CRs in the regional broadband network are Dual-stack or IPv6-only and the BRASs are IPv6-only. This network is providing a Dual-stack service or an IPv6-only service for subscribers. This section will discuss the dual-stack subscriber accessing this network. [See Figure 5]

A Dual-stack, IPv4 or IPv6 terminal accessing to an IPv6-only BRAS.

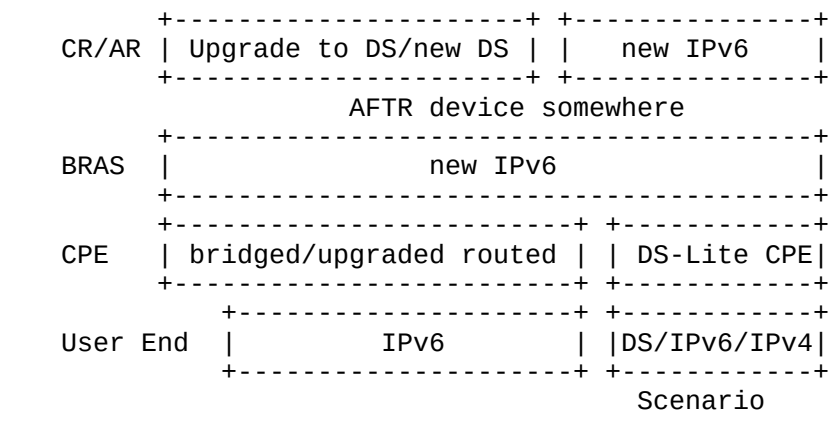


Figure 5: The DS-Lite Solution in IPv6 Infrastructure

The Scenario for IPv6 subscriber is very simple. But the routing CPE at the edge of customer premises network need to be upgraded to support IPv6 and PPPoE with IPv6. And the PC operation system (OS) also needs to support PPPoE with IPv6. This scenario will exist when the IPv6 traffic is already dominant in the network. The little IPv4 traffic will be translated by a NAT64 device located at the edge of IPv6 Ocean.

The Scenario for Dual-stack subscriber is a little bit complicated. It provides Dual-stack service over an IPv6-only infrastructure. The technologies like DS-Lite [I D.ietf softwire dual stack lite] can be deployed in this scenario. This section will discuss focus on this technology.

DS-Lite is a tunnel technology with a point-to-multipoint IPv4-in-IPv6 tunnel between B4 element and AFTR. According to the definition in [I D.ietf softwire dual stack lite], the B4 element is a function implemented on a dual-stack capable node, either a directly connected device or a CPE, which creates a tunnel to an AFTR.

Any locally unique IPv4 address could be configured on the IPv4-in-IPv6 tunnel to represent the B4 element. IANA has defined a well-known range, 192.0.0.0/29.

DS-Lite technology is designed for an IPv6 infrastructure with a layer 3 (L3) access network. However, [I-D.zhou-softwire-ds-lite-p2p] describes the Point-to-Point access method scenario. For a layer 2 (L2) access network with PPPoE access method, each CPE has a unique PPP link. The link information can be used to identify the CPE and any IPv4 or IPv6 address does not need

to be allocated to the CPE. However, the CPE can allocate an internal IPv4 address to a host. It simply puts the packets to the point-to-point link and forward to the BRAS. When BRAS receives the packet, it maps the point-to-point identifier to the IPv6 Flow Label [RFC3697] and send to the AFTR for NAT.

The operating system (OS) of the new subscriber is recommended to support PPPoE with IPv6. Third-party dial-up software could be provided if the OS is not support PPPoE with IPv6. Subscribers need to replace the existing CPEs for DS-Lite services.

No matter the L3 or L2 access network DS-Lite is deploying on, the DS-Lite Address Family Translation Router (AFTR) needs to be deployed somewhere in the regional broadband network.

In all, currently, DS-lite, which offers IPv4 internet visiting ability to dual stack subscribers through pure IPv6 accessing network, does not be well integrated into devices. At the same time, technically, it can not assign IPv4 DNS to the subscribers in the first version. Moreover, how to assign addresses to users, considering PPPOE accessing methods, should also be thought about carefully.

3.2.1. The Location of AFTR

There are mainly three locations can be deployed an AFTR:

- Deploying a centralized AFTR connecting to the IPv6 CR; Figure 6
- Deploying a centralized AFTR card at the Dual-stack CR;[Figure 7]
- Deploying distributed AFTRs connecting to IPv6 BRASs;[Figure 8]

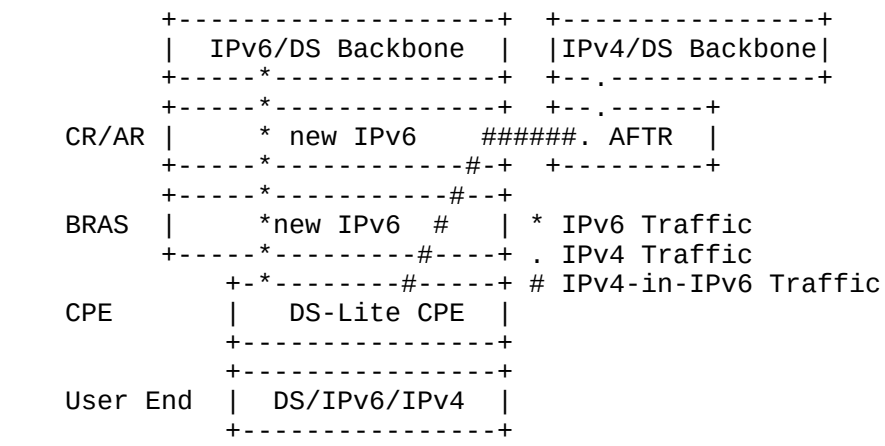


Figure 6: Centralized AFTR connecting to the IPv6 CR

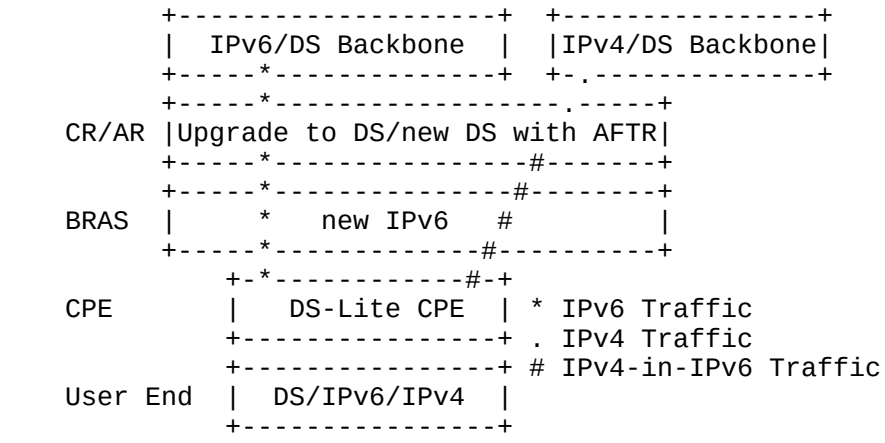


Figure 7: Centralized AFTR card at the Dual-stack CR

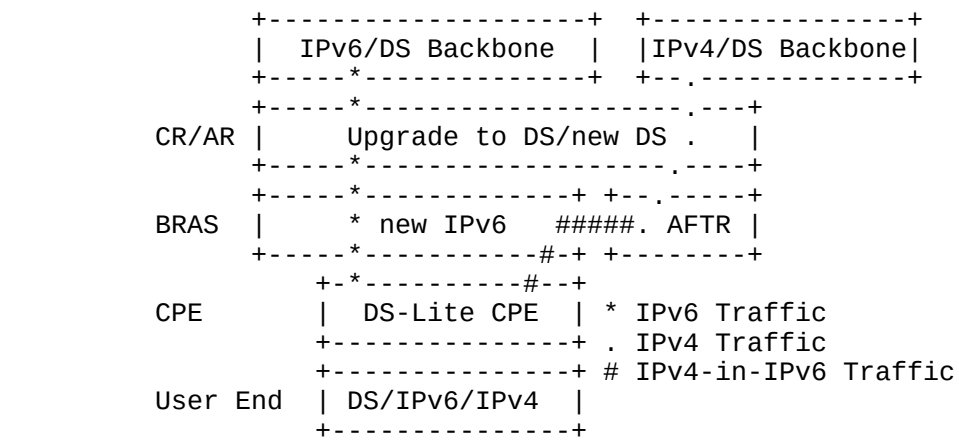


Figure 8: Distributed AFTRs connecting to IPv6 BRASs

3.3. Solution3: Dual-Stack over IPv4 - 6rd

In this solution, the CRs in the regional broadband network are IPv4-only, Dual-stack or IPv6-only and the BRASs are IPv4-only. It provides a Dual-stack service or an IPv6-only service with a completely/partly IPv4 infrastructure for subscribers.

The discussion will focus on scenario in Figure 9. " An IPv6 or Dual-stack terminal accessing to an IPv4-only BRAS.

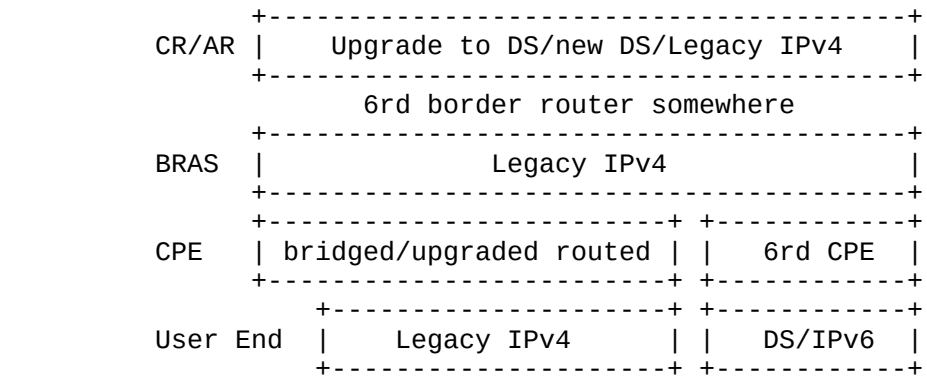


Figure 9: The 6rd Solution in an IPv4 infrastructure

A possible technical solution for this scenario is IPv6 Rapid

Deployment (6rd) [RFC5969]. There are two components in this solution. 6rd CPEs support Ipv4 on their customer premise side and support 6rd on the provider side. 6rd gateway (a.k.a 6rd border router or 6rd relay) is operated at the border between IPv4 infrastructure and the IPv6 Internet. The 6rd mechanism operates statelessly, which ensures simplicity and scalability. The IPv4 address in the IPv4 infrastructure could be a private address, 6rd mechanism can support the private IPv4 address.

In all, currently, 6RD, which offer Ipv6 internet visiting ability to dual stack subscribers through pure Ipv4 accessing network, does not be well integrated into devices. At the same time, technically, Moreover, how to assign addresses to users, considering PPPOE accessing methods, should also be thought about carefully.

3.3.1. The Location of 6rd Gateway

There are mainly two locations can be deployed a 6rd gateway:

- Deploying a centralized 6rd gateway at the edge of IPv4 regional broadband network;
- Deploying distributed 6rd gateways connecting to IPv4 BRASs;

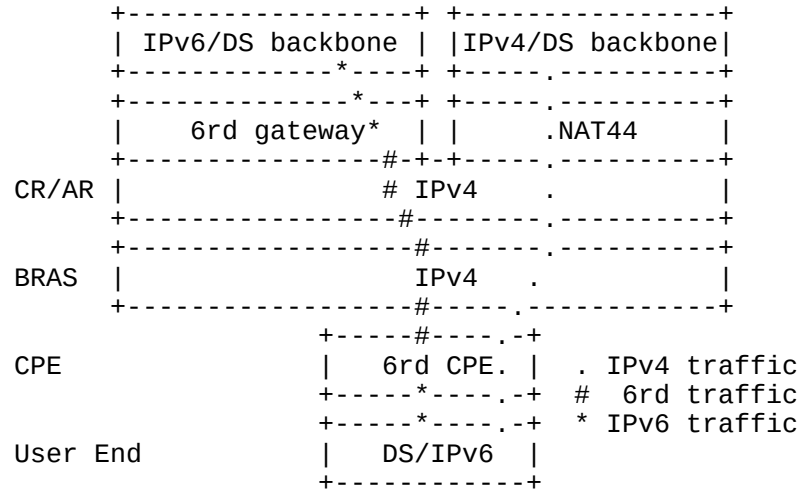


Figure 10: 6rd gateway at the edge of IPv4 network

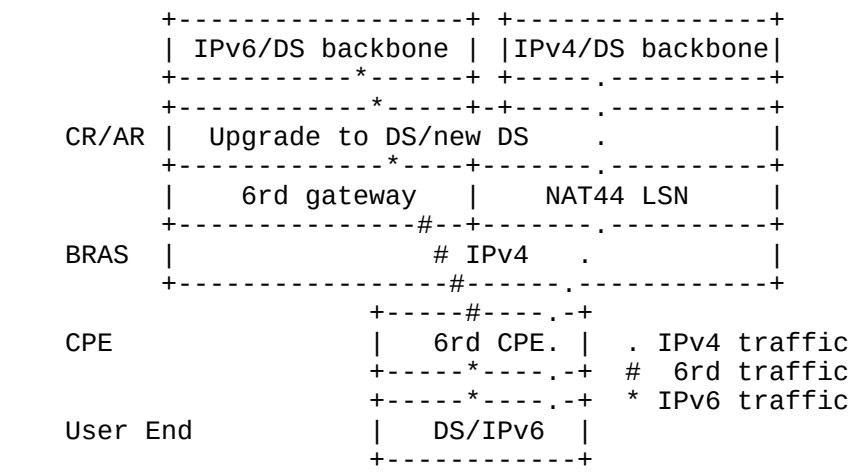


Figure 11: Distributed 6rd gateways connecting to IPv4 BRASs

3.4. Solution4: IPv6 and protocol translation

This solution is for the IPv6-only subscribers that are accessing to the new built IPv6-only regional broadband network. Basically, only IPv6 address is allocated to the subscribers. And for the requirement of IPv4 services, it is needed to deploy a NAT64 (stateful/IVI) [I D.ietf behave v6v4 xlate stateful] [I D.xli behave ivi] device to solve the intercommunication problem between IPv6 and IPv4 for the IPv6-only subscribers.

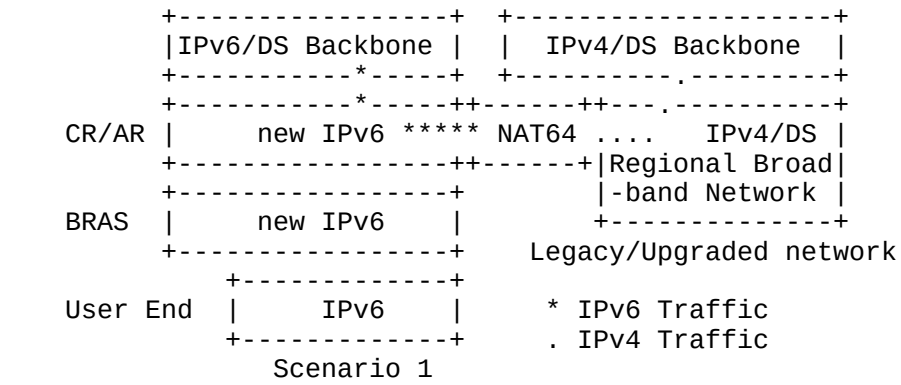


Figure 12: The NAT64 Solution in an IPv6 infrastructure - Scenario 1

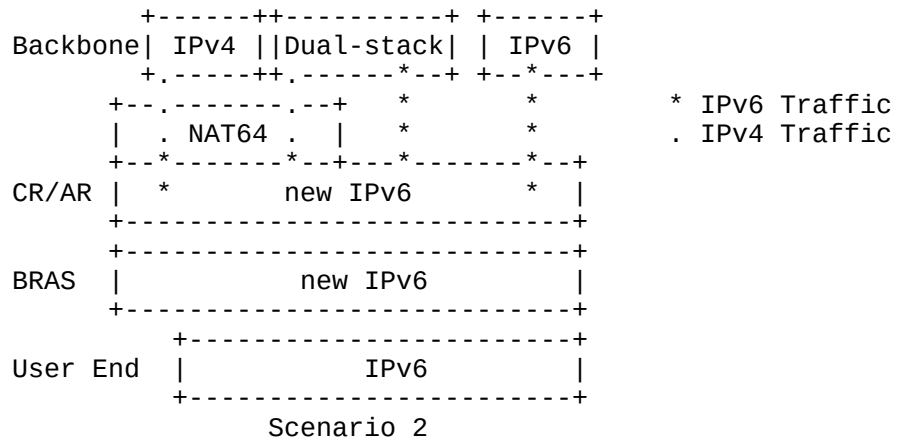


Figure 13: The NAT64 Solution in an IPv6 infrastructure - Scenario 2

The operating system (OS) is required to support PPPoE with IPv6. Third-party dial-up software could be provided if the OS of the new hosts is unable to support PPPoE with IPv6.

The routing mode CPE that is purchased by subscribers is also required to support PPPoE with IPv6 as well, or to turn off the auto-dialup function, and initial the PPPoE with IPv6 dial-up session from the host.

In conclusion, protocol translation, just like NAT64, IVI,etc, is widely agreed to be used in mobile network. But due to the vast types of application in the internet accessed through fixed network, protocol translation technologies can not deal with all the applications. As far as we know, conclude from the experimental data, only mail and http service protocol can be well translated.

4. Terminal migration

From the terminal aspect, there are two types of CPE: Routed CPE and switched CPE. The switched CPE, which can transparently switch the Ipv6 traffic from user PC to the network access device, does not block the Ipv6 PPPOE dial up .

But the routed one, which always automatically initiates a PPPOE , cumbers the normal Ipv6 accessing since almost most part of the routed CPE do not support Ipv6 PPPOE dial up. For the ISP, who can provide CPE to the subscribers and manage them, can offer Ipv6 accessing service by upgrading the CPE. But for the ISPs in this

case, who have large mount of users and have to allow the users to purchase the CPE by themselves, it is nearly impossible for them to replace all the CPE for the cost sake.

From OS aspect, there are some problem with windows XP operation system, which dominate the over 80% market, to support the Ipv6 PPPOE dial up.

5. ICP migration

Internet Content Provider (ICP) migration to IPv6 is the most important step to break the deadlock in the IPv6 industry chain. The ICPs can migrate by themselves, or can be assisted by others. For the transition of ICP itself, some ICPs have modified their proprietary web service.

This solution requires a protocol translation technology, like NAT64 [ID_behave-v6v4-xlate-stateful] device deployed at the edge of IDC(Internet Data Center), but not requires to deploy a DNS64. The solution could include the following steps:

- Configure AAAA records with IPv6 addresses for the ICP's sites on its Authoritative DNS Server. These IPv6 addresses are constituted by combining the ICP's IPv4 public address and an specific prefix according to the Prefix+v4 address style described in SIIT [RFC2765] or in IVI [ID_xli-behave-ivi].

- When user initials an http request to the website, the host will query the corresponding IP address from DNS cache server which is usually located in Metro Network.

- The DNS cache server replies with the ICP's IPv6 address in an AAAA record which is from the ICP's Authoritative DNS Server. In some circumstances, these records could be configured in a centralized DNS server of the IDC.

- The user will access the IPv4 services using the IPv6 address.The BRAS will route it to the NAT64 device located at the IDC edge, and remove the prefix, then translating the IPv6 client address by installing mappings in the normal NAT manner.

6. Challenges Faced In Migrating To IPv6

There is a long list of challenges during the transition from IPv4 to IPv6:

- The remaining stock of IPv4 addresses cannot support the development of existing services. Future service's development is not the only thing which is concerned about, but the transition of existing services to IPv6 is also considered.

- Due to the long transition period, it will be highly possible that one thing taken into consideration while the other be neglected when the different parts of end-to-end network are migrated. A balanced strategy is needed to guide the transition.

- Lack of IPv6 Internet resources. ICP seldom deploy IPv6 and almost no Content Provider/Service Provider (CP/SP) considers IPv6 when developing proprietary services due to the large volume of recoding. The applications of ICPs are of various types and so complex that they cannot all support IPv6 in a short time. In addition, many business websites are always linking to each other, creating a complex topology which will lead to many problems when one website migrates to IPv6 only. Another reason ICP migration lacks motivation is that the CP/SP does not realize how urgent it is to migrate to IPv6.

- From the perspective of terminals, some specific terminals (e.g.,set top boxes) do not support IPv6 even if the main operating systems can do so. Some operation systems for mobile terminals officially claimed they don't support IPv6.

- No accumulated experience with IPv6 transition. Large scale network and large number of subscribers are the two key problems.IPv6 transition should be seriously considered. With large scale network and various service platforms, the transition involves multiple levels and broad scope, so the cost of modification will be huge and the return on investment will not be so evident. The selection of transition technology and network modification solution is not clear for the transition roadmap of the whole network.

7. IANA Considerations

This memo includes no request to IANA.

8. Security Considerations

The IETF is specifying security considerations for the tools that it is providing for IPv6 migration.However, it is possible that additional considerations arise due to the interaction of these tools, and the fact that the network is in a transitional state.

Security considerations should be incentive concerned about because of the potential loss, which is caused by the IPv6 security issues, e.g., dual-stack routing security, network expandability, device reliability, network anti-attack, user tracing, government supervision and etc.

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.

[min_ref] authSurName, authInitials., "Minimal Reference", 2006.

9.2. Informative References

[I-D.baker-behave-ivi]
Li, X., Bao, C., Baker, F., and K. Yin, "IVI Update to SIIT and NAT-PT", draft-baker-behave-ivi-01 (work in progress), September 2008.

[I-D.durand-softwire-dual-stack-lite]
Durand, A., Droms, R., Haberman, B., and J. Woodyatt, "Dual-stack lite broadband deployments post IPv4 exhaustion", draft-durand-softwire-dual-stack-lite-01 (work in progress), November 2008.

[I-D.ietf-behave-v6v4-xlate-stateful]
Bagnulo, M., Matthews, P., and I. Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", draft-ietf-behave-v6v4-xlate-stateful-12 (work in progress), July 2010.

[I-D.ietf-softwire-gateway-init-ds-lite]
Brockners, F., Gundavelli, S., Speicher, S., and D. Ward, "Gateway Initiated Dual-Stack Lite Deployment", draft-ietf-softwire-gateway-init-ds-lite-01 (work in progress), October 2010.

[I-D.zhou-softwire-ds-lite-p2p]
Zhou, C., ZOU), T., Lee, Y., and G. Yang, "Deployment DS-lite in Point-to-Point Access Network", draft-zhou-softwire-ds-lite-p2p-02 (work in progress), July 2010.

[RFC2516] Mamakos, L., Lidl, K., Evarts, J., Carrel, D., Simone, D., and R. Wheeler, "A Method for Transmitting PPP Over Ethernet (PPPoE)", RFC 2516, February 1999.

[RFC2629] Rose, M., "Writing I-Ds and RFCs using XML", RFC 2629, June 1999.

[RFC2765] Nordmark, E., "Stateless IP/ICMP Translation Algorithm (SIIT)", RFC 2765, February 2000.

[RFC3552] Rescorla, E. and B. Korver, "Guidelines for Writing RFC Text on Security Considerations", BCP 72, RFC 3552, July 2003.

[RFC3931] Lau, J., Townsley, M., and I. Goyret, "Layer Two Tunneling Protocol - Version 3 (L2TPv3)", RFC 3931, March 2005.

[RFC4798] De Clercq, J., Ooms, D., Prevost, S., and F. Le Faucheur, "Connecting IPv6 Islands over IPv4 MPLS Using IPv6 Provider Edge Routers (6PE)", RFC 4798, February 2007.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.

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

Authors' Addresses

CanCan Huang (editor)
China Telecom
109, Zhongshan Ave. West,
Guangzhou, Tianhe District 510630
P.R. China

Phone:
Email: huangcc@gsta.com

XiaoYang Li
China Telecom
109, Zhongshan Ave. West,
Guangzhou, Tianhe District 510630
P.R. China

Phone:
Email: hz_lxy@gsta.com

LeMing Hu
China Telecom
109, Zhongshan Ave. West,
Guangzhou, Tianhe District 510630
P.R. China

Phone:
Email: hulm@gsta.com

Individual Submission
Internet Draft
Intended status: Informational
Expires: April 2011

E. Jankiewicz (Ed.)
SRI International, Inc.

October 25, 2010

An Annotated Bibliography for IPv4-IPv6 Transition and Coexistence
draft-jankiewicz-v6ops-v4v6biblio-03.txt

Abstract

The Internet is in the early stages of what may be a protracted period of coexistence of IPv4 and IPv6. Network operators are challenged with the task of activating IPv6 without negative impact on operating IPv4 networks and their customers. This draft is an informational "annotated bibliography" compiled to help in the analysis and development of basic guidelines and recommendations for network operators. The goal of this document is to survey the current state of RFCs, Internet-Drafts and external reference materials that define the use cases, problem statements, protocols, transition mechanisms and coexistence tools that will be of interest to a network operator planning to turn on IPv6.

Status of this Memo

Table of Contents

1. Introduction

Since the IPv6 protocol was defined in 1995 as RFC 1883 (replaced in 1998 by RFC 2460) the Internet has been in a long transition from IPv4 to IPv6. In reality, we are still in the early stages of what is likely to be a protracted period of coexistence, where IPv6 penetration in hosts (both servers and clients) will gradually ramp up as networks make IPv6 available through their infrastructures.

Network operators face a daunting task to design and implement plans to activate IPv6 without negative impact on large (in some cases very large) operating IPv4 networks with many live customers. Some basic guidelines and recommendations for network operators are being developed (http://tools.ietf.org/html/draft-lee-v4v6tran-problem) and this draft is an informational companion to that effort. The goal of this document is to survey the current state of RFCs, active (and expired but still relevant) Internet-Drafts and external reference materials that define the use cases, problem statements, protocols, transition mechanisms and coexistence tools that will be of interest to a network operator planning to turn on IPv6.

This is a dynamic and evolving marketplace of ideas. At best, this draft is a blurry snapshot of the landscape near to the time of its publication. The editor intends this compendium to be merely the starting point for an active database or wiki available for community contribution including feedback on the real-world experience of network operators as they turn on IPv6. Note that the links to RFCs and drafts are based on the IETF Tools view of the repository at http://tools.ietf.org/html/. The links for active drafts are not for a specific revision but should link to the last or latest version.

The following sections comprise an annotated bibliography of the currently available documentation to knowledge of the editor. It is provided as informational guidance only, and any network operator contemplating an IPv6 implementation will of course exercise due

diligence in researching all the issues, standards and recommendations and analyze applicability to the particular network operation.

Note that as the body of this text includes full reference information for the bibliography entries these are not included in the normal Reference section.

[Editor's note to be removed before publication:

While this draft is circulating, the editor is interested in any and all pointers to additional useful references. Contributions of capsule summaries and applicability for any of the listed entries would also be appreciated and will be graciously acknowledged. If I have missed anyone who already chipped in, this will be cheerfully rectified upon your reminder via a private e-mail.]

1.1. The Three Laws of IPv4/IPv6 Coexistence Mechanisms

The Editor of this draft thought it might be helpful to briefly explore the motivations driving the current profusion of coexistence mechanisms. In the not so distant past little or no discussion of this topic was going on in the IETF, as many felt the case was closed. A discussion in the Intarea meeting at IETF 71 in Dublin and a presentation at the plenary at that meeting led to a reawakening of interest in coexistence and transition tools. This discussion continued at a special meeting in Montreal in October 2008, and has occupied substantial time on the mailing lists and meetings of several Working Groups since then. The Internet Area, IPv6 Operation (v6ops), Softwires and Behave WGs have generated many contributions, and an ad-hoc discussion mailing list has been established at https://www.ietf.org/mailman/listinfo/v4tov6transition.

Early in the life of IPv6, the assumption was made that IPv6 deployment, based on dual-stack implementations, would be ubiquitous long before the IPv4 address pool would run out. For special cases, tunneling through dissimilar networks or use of an external translation box such as NAT-PT would allow interim operation of legacy equipment. At present, this has not yet come to pass. The impending exhaustion of IPv4 address space renders dual-stack impossible in some deployments and issues have resulted in NAT-PT being deprecated to Historic status.

Nature (and your average Internet-Draft author) abhors a vacuum. With the demise of NAT-PT and the increasing urgency to get moving on IPv6 transition, we are now in a period of "Let 1000 Flowers Bloom" where many ideas are being advanced, and a lot of IETF brainpower is

being spent debating the relative merits and evilness of various approaches. The spectrum of opinion on coexistence mechanisms has two extremes:

IPv4 is so Over: Concentrate on deploying native IPv6 and managing it effectively, rather than spinning more complex webs of IPv4 accommodation. Deploying anything that delays IPv6 and enables more IPv4 usage at this point is irresponsible.

Where’s the Business Case: Real customers need IPv4, there is no IPv6 content, no demand for IPv6. Scale up NAT to keep IPv4 viable, provide some sort of artificial IPv6 access, if and when customers ask. No plans for native IPv6 in the foreseeable future.

A reasonable position recognizes the valid motivation on both sides. An ISP may not be able to dictate updates to customer computers and routers, and must provide access to all legacy customers, not just eager IPv6 adopters, so an interim mechanism that minimizes their inconvenience is needed. One size will never fit all, so some solutions may be a good fit for one ISP, and not for others. While evaluating all the alternative documented here, the principle to keep in mind is that the IETF should provide good engineering opinions on all these alternatives, to permit things that will help, and prevent things that will cause problems.

This can be summed up in the "Three Laws of IPv4/IPv6 Coexistence Mechanisms":

1. First, do no harm.

2. Keep it simple.

3. Keep moving towards more native IPv6.

"No harm" in this case means that a good solution will not unduly interfere with good experience for the legacy IPv4 customer, nor will it impede the eager IPv6 adopter. The solution must not cause problems for peer or backbone networks or for the Internet community at large.

"Simple" means to solve particular problems with specific solutions focused to the point of need rather than attempting broad and complex methods that impinge on all traffic. However, do not simplify any more than necessary to avoid harm.

The compulsion to move towards native IPv6 follows from the first two laws. Over time, even minimal harm and complexity that even a good

mechanism presents can and should be reduced over time by continuing to enable, promote and encourage transition to native IPv6. Design and deploy your interim solution(s) with a clear migration path that will eventually render them redundant. Set a date after which you will not deploy any new equipment that does not support IPv6. Set a date to sunset IPv4 access, giving legacy customers plenty of time (and incentive) to upgrade their old equipment.

In summary, it seems that the Robustness Principle (Postel's Law) would apply, as it does in many situations:

> "Be conservative in what you do, be liberal in what you accept from others." [RFC 793]

Following the Robustness Principle and the Three Laws should allow an operator complete freedom to manage their own network and to chose and operate any coexistence mechanism as long as they need to for supporting their customers, except where those choices cause harm to someone else. Of course, there is no universal definition of "harm" so reasonable people can disagree, e.g. if a mechanism in use on the access side causes additional delay, content providers may see that as "harming" their users' experience. That's why Working Group mailing lists and IETF meetings are just so much fun.

Oh, and by the way, the Fourth Law should be "Don't reinvent the wheel" so please explore the RFCs, drafts and other citations to see if someone has already proposed something similar to your idea. Your contributions are needed, but time and energy is better spent exploring novel approaches and building on what has already been proposed.

2. IPv6 and related Protocol Specifications

"IPv6 Node Requirements" J. Loughney, Ed. April 2006
http://tools.ietf.org/html/rfc4294

"IPv6 Node Requirements RFC 4294-bis" E. Jankiewicz, J. Loughney, T. Narten
http://tools.ietf.org/html/draft-ietf-6man-node-req-bis

RFC 4294 and its update draft are included by reference. These provide a comprehensive overview of the IPv6 baseline specifications and the reader is directed to them to avoid a redundant listing here.

3. Problem Statements and Use Cases

"Problem Statements of IPv6 Transition of ISP" Y. Lee, Ed.
http://tools.ietf.org/html/draft-lee-v4v6tran-problem

This draft is being developed by an ad-hoc group interested in providing guidance to network operators on the IPv6 transition. It will include high level use cases (as contributed by IETF participants with network operator experience) and a problem statement documenting what additional work IETF could do to provide sufficient tools and guidance for the network operators

"Mobile Networks Considerations for IPv6 Deployment" R. Koodli
http://tools.ietf.org/html/draft-ietf-v6ops-v6-in-mobile-networks

Mobile Internet access from smartphones and other mobile devices is accelerating the exhaustion of IPv4 addresses. IPv6 is widely seen as crucial for the continued operation and growth of the Internet, and in particular, it is critical in mobile networks. This document discusses the issues that arise when deploying IPv6 in mobile networks. Hence, this document can be a useful reference for service providers and network designers.

"Routing Loop Attack using IPv6 Automatic Tunnels: Problem Statement and Proposed Mitigations", G. Nakibly and F. Templin
http://tools.ietf.org/html/draft-ietf-v6ops-tunnel-loops

This document is concerned with security vulnerabilities in IPv6-in-IPv4 automatic tunnels. These vulnerabilities allow an attacker to take advantage of inconsistencies between the IPv4 routing state and the IPv6 routing state. The attack forms a routing loop which can be abused as a vehicle for traffic amplification to facilitate DoS attacks. If automatic tunnels are used in a deployment the warnings and mitigations in this draft should be considered.

"Use Case for IPv6 Transition for a Large-Scale Broadband Network" CC. Huang (Ed.), XY. Li and LM. Hu
http://tools.ietf.org/html/draft-huang-v6ops-v4v6tran-bb-usecase

"IPv6 Transition Cable Access Network Use Cases" Y. Lee and V. Kuarsingh
http://tools.ietf.org/html/draft-lee-v4v6tran-usecase-cable

"IPv6 Transition Use Case for a Large Mobile Network" C. Zhou (Ed.) and T. Taylor
http://tools.ietf.org/html/draft-zhou-v6ops-mobile-use-case-00

Each of these use case drafts is focused on a particular deployment model for a specific market segment. While each may be based on a singular operator's experience or planning, the intention is to develop the set of use cases drafts to be of interest to any network operator in the segment.

"Considerations for Stateless Translation (IVI/dIVI) in Large SP Network" Q. Sun et al.
http://tools.ietf.org/html/draft-sunq-v6ops-ivi-sp

"dIVI" is a prefix-specific and stateless address mapping method based on IVI which can directly translate IPv4 packet to IPv6 packet. This document describes the challenges and requirements for large Service Provider to deploy IPv6 in an operational network and specifically considerations for dIVI deployment.

4. Transition and Coexistence Scenarios and Architectures

RFC 5211 "An Internet Transition Plan." J. Curran, July 2008
http://tools.ietf.org/html/rfc5211

While the abstract for this RFC humbly describes it as just "one possible plan" for the IPv6 transition, it provides very good context and a common language to use when talking about transition plans, and can be seen as a call to action. It describes three phases of the transition, and proposes a timeline based on predictions of the imminent exhaustion of the IPv4 address space. The phases are:

1. Preparation, where IPv4 predominates while service providers trial and experiment with IPv6, and end-users prepare to provide Internet-facing IPv6 services in the future. The timeline in the RFC described this phase as in progress, and optimally this phase would have ended already.

2. Transition, where both IPv4 and IPv6 services are offered and used, with production level support for IPv6, although this may be via transition mechanisms rather than native IPv6. The RFC targeted this phase to end in 2011.

3. Post-Transition, where native IPv6 services should be offered while IPv4 services may still be supported.

"Guidelines for Using Transition Mechanisms During IPv6 Deployment" J. Arkko and F. Baker
http://tools.ietf.org/html/draft-arkko-ipv6-transition-guidelines

IPv6 deployment requires some effort, resources, and expertise. The availability of many different deployment models is one reason why expertise is required. This draft discusses the IPv6 deployment models and migration tools, and recommends ones that have been found to work well in operational networks in many common situations.

"IPv6 Transition Guide For A Large ISP Providing Broadband Access", G. Yang (Ed.), L. Hu and J. Lin
http://tools.ietf.org/html/draft-yang-v6ops-v4v6tran-bb-transition-guide

This draft is a product of the current v4tov6transition effort and it examines IPv6 migration solutions for each part of the Large-scale broadband infrastructure with a layer 2 access network. The analysis is based on the requirements for providing existing broadband services in v4v6-coexisting or IPv6-only situations. The draft describes the suitable scenarios for each solution.

"IPv6 Transition Guide for a Large Mobile Operator" T. Tsou (Ed.) and T. Taylor
http://tools.ietf.org/html/draft-tsou-v6ops-mobile-transition-guide

Similarly, this draft examines IPv6 migration solutions for a large mobile network.

RFC 6036 "Emerging Service Provider Scenarios for IPv6 Deployment", B. Carpenter, S. Jiang
http://www.rfc-editor.org/rfc/rfc6036.txt

This document describes practices and plans that are emerging among Internet Service Providers for the deployment of IPv6 services. They are based on practical experience so far, as well as current plans and requirements, reported in a survey of a number of ISPs carried out in early 2010. The document identifies a number of technology gaps, but does not make recommendations.

"Framework for IP Version Transition Scenarios", B. Carpenter, S. Jiang and V. Kuarasingh
http://tools.ietf.org/html/draft-carpenter-v4v6tran-framework

This document sets out a framework for the presentation of scenarios and recommendations for a variety of approaches to the transition from IPv4 to IPv6, given the necessity for a long period of co-existence of the two protocols.

5. Transition/Coexistence Tools

As network operators and end-users independently proceed with transition to IPv6 while others continue to use IPv4, a potentially long period of coexistence will ensue. Variations on terminology have been used since the specification of IPv6; transition implies a process whereby the star of IPv6 rises and the star of IPv4 sets; coexistence implies that both will operate together. Due to thoroughly discussed limits to the growth of an Internet using only IPv4, IPv6 is a necessary technology for the future of the Internet. However, nothing compels the elimination of IPv4; no protocol police will forbid its use in the foreseeable future. IPv4 may disappear due to irrelevance when IPv6 is so pervasive to make it redundant, but network operators should be prepared to operate IPv4 and IPv6 in a mixed deployment for some time. However, the techniques and mechanisms supported by a network operator can be expected to evolve and change over time as a rational goal would be to gradually shift coexistence costs (real operational expense as well as convenience) from "early adopters" of IPv6 to the shrinking pool of IPv4 maintainers.

Various techniques are required for coexistence, roughly divided into three categories:

1. Address Mapping: Many situations will require the use of address mapping to maintain scalability in the face of dwindling IPv4 global address space and to support translation and tunneling approaches.

2. Tunneling: A method for the encapsulation and transport of one protocol over or through the infrastructure that favors the other, e.g. IPv6 traffic via an IPv4 infrastructure

3. Protocol Translation: A mechanism for rewriting packets from one protocol to the other so they can be delivered as native (non-encapsulated) packets typically due to incompatible end nodes, e.g. an IPv6 client to an IPv4 server.

These categories are not mutually exclusive, as some scenarios and solutions incorporate aspects of multiple approaches.

RFC 4213 "Basic Transition Mechanisms for IPv6 Hosts and Routers" E. Nordmark and R. Gilligan October 2005
http://tools.ietf.org/html/rfc4213

5.1. Address Mapping

The introduction of address family translation presents challenges similar to those experienced with Network Address Translation (NAT) as it has evolved in the IPv4 Internet. The depletion of IPv4 global address space conspires with the continuing need for routable IPv4 address in some coexistence approaches to further press proliferation and scale of NAT. While alternatives exist, some network operators will continue to see the various flavors of NAT as a necessary evil, so it remains important to understand the impact on network operations, on the end-user and on applications.

Dual-Stack Lite (DS-lite) is one of the alternatives to providing dual-stack support to end-users in the face of limited global IPv4 address space.

RFC 2663 "IP Network Address Translator (NAT) Terminology and Considerations" P. Srisuresh and M. Holdrege August 1999 http://tools.ietf.org/html/rfc2663

This document attempts to describe the operation of NAT devices and the associated considerations in general, and to define the terminology used to identify various flavors of NAT.

5.1.1. Address Translation in Network Operations

"Common Requirements for IP Address Sharing Schemes" I. Yamagati et al. http://tools.ietf.org/html/draft-ietf-behave-lsn-requirements

This document defines common requirements of multiple types of Large Scale Network Address Translation (NAT) that handles Unicast UDP, TCP and ICMP.

"Issues with IP Address Sharing" M. Ford (Ed.) et al. http://tools.ietf.org/html/draft-ietf-intarea-shared-addressing-issues

The completion of IPv4 address allocations from IANA and the RIRs is causing service providers around the world to question how they will continue providing IPv4 connectivity service to their subscribers when there are no longer sufficient IPv4 addresses to allocate them one per subscriber. Several possible solutions to this problem are now emerging based around the idea of shared IPv4 addressing. These solutions give rise to a number of issues and this memo identifies those common to all such address sharing approaches.

"An Incremental Carrier-Grade NAT (CGN) for IPv6 Transition", Sheng Jiang, Dayong Guo, Brian Carpenter
http://tools.ietf.org/html/draft-ietf-v6ops-incremental-cgn

Carrier-Grade NAT (CGN) devices with integrated transition mechanisms can reduce the operational change required during the IPv4 to IPv6 migration or coexistence period. This document proposes an incremental CGN approach for IPv6 transition. It can provide IPv6 access services for IPv6-enabled hosts and IPv4 access services for IPv4 hosts while leaving much of a legacy IPv4 ISP network unchanged. It is suitable for the initial stage of IPv4 to IPv6 migration. Unlike NAT444 based CGN alone, Incremental CGN also supports and encourages transition towards dual-stack or IPv6-only ISP networks. A smooth transition to IPv6 deployment is also described in this document.

"Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers" Bagnulo, Matthews, van Beijnum
http://tools.ietf.org/html/draft-ietf-behave-v6v4-xlate-stateful

This document describes stateful NAT64 translation, which allows IPv6-only clients to contact IPv4 servers using unicast UDP, TCP, or ICMP. The public IPv4 address can be shared among several IPv6-only clients. When the stateful NAT64 is used in conjunction with DNS64 no changes are usually required in the IPv6 client or the IPv4 server.

"NAT64-CPE Mode Operation for Opening Residential Service" G. Chen and H. Deng
http://tools.ietf.org/html/draft-chen-v6ops-nat64-cpe

The authors of this draft describe the application of fundamental NAT64 functionality in CPE deployment scenarios. The approach is intended to eliminate the need for CPE to cooperate with DNS64, and to be compatible with legacy residential servers without changes to DNS requirements.

"Flexible IPv6 Migration Scenarios in the Context of IPv4 Address Shortage" M. Boucadair (Ed.) et al, October 20, 2009 (expired)
http://tools.ietf.org/html/draft-boucadair-behave-ipv6-portrange-04

This memo presents a solution to solve IPv4 address shortage and ease IPv4-IPv6 interconnection. The document presents a set of incremental steps for the deployment of IPv6 as a means to solve IPv4 address exhaustion. Stateless IPv4/IPv6 address mapping functions are introduced and IPv4-IPv6 interconnection scenarios presented.

This memo advocates for a more proactive approach for the deployment of IPv6 into operational networks. This memo specifies the IPv6 variant of the A+P. Both encapsulation and translation scheme are covered. Moreover, two modes are elaborated: the binding mode (compatible mode with DS-lite) and the stateless mode.

"A Note on NAT64 Interaction with Mobile IPv6" W. Haddad and C. Perkins
http://tools.ietf.org/html/draft-haddad-mext-nat64-mobility-harmful

This memo discusses potential NAT64 technology repercussions for mobile nodes using Mobile IPv6. An ambiguity is identified related to the use of DNS during bootstrapping, which is likely to inhibit proper signaling between mobile node and home agent.

"NAT64 for Dual Stack Mobile IPv6" B. Sarikaya and F. Xia
http://tools.ietf.org/html/draft-sarikaya-behave-mext-nat64-dsmip

This memo specifies how IPv6 only mobile nodes (MN) receiving host-based mobility management using Dual Stack Mobile IPv6 (DSMIPv6) can communicate with IPv4 only servers. The protocol is based on home agents maintaining a table similar to NAT64 and linking it to the binding cache. This technique avoids the problems encountered when NAT64 is used for mobile nodes in Dual Stack Mobile IPv6. How IPv6 only mobile nodes can receive multicast data from IPv4 only content providers is also explained.

"NAT64 for Proxy Mobile IPv6" B. Sarikaya and F. Xia
http://tools.ietf.org/html/draft-sarikaya-behave-netext-nat64-pmip

Similarly, this memo specifies how IPv6 only mobile nodes (MN) receiving network-based mobility management using Proxy Mobile IPv6 (PMIPv6) can communicate with IPv4 only servers.

5.1.2. Application and End-User Considerations With NAT

"Problem Statement for Referrals" B. Carpenter, S. Jiang and B. Zhou
http://tools.ietf.org/html/draft-carpenter-referral-ps

The purpose of a referral is to enable a given entity in a multiparty Internet application to pass information to another party. It enables a communication initiator to be aware of relevant information of its destination entity before launching the communication. This memo discusses the problems involved in referral scenarios.

"Referrals Across an IPv6/IPv4 Translator" D. Wing, October 19, 2009
http://tools.ietf.org/html/draft-wing-behave-nat64-referrals-01

While this draft is expired, this issue remains a topic of conversation, including a Bar-BoF at IETF 78. Referrals across disparate address domains may be needed for provision of services such as SIP during transition.

"Legacy NAT Traversal for IPv6: Simple Address Mapping for Premises Legacy Equipment (SAMPLE)"
http://tools.ietf.org/html/draft-carpenter-softwire-sample

IPv6 deployment is delayed by the existence of millions of subscriber network address translators (NATs) that cannot be upgraded to support IPv6. This document specifies a mechanism for traversal of such NATs. It is based on an address mapping and on a mechanism whereby suitably upgraded hosts behind a NAT may obtain IPv6 connectivity via a stateless server, known as a SAMPLE server, operated by their Internet Service Provider. SAMPLE is an alternative to the Teredo protocol.

"Some Considerations on the Load-Balancer for NAT64" D. Zhang et al.
http://tools.ietf.org/html/draft-wang-behave-nat64-load-balancer

This draft investigates issues with deploying load-balancers with NAT64 devices.

"An FTP ALG for IPv6-to-IPv4 Translation" I. van Beijnum
http://tools.ietf.org/html/draft-ietf-behave-ftp64

The File Transfer Protocol (FTP) has a very long history, and despite the fact that today, other options exist to perform file transfers, FTP is still in common use. As such, it is important that in the situation where some client computers are IPv6-only while many servers are still IPv4-only and IPv6-to-IPv4 translators are used to bridge that gap, FTP is made to work through these translators as best it can. This document specifies a middlebox that enables legacy usage of FTP with translation.

"Assessing the Impact of NAT444 on Network Applications" C. Donley et al. http://tools.ietf.org/html/draft-donley-nat444-impacts

NAT444 is an IPv4 extension technology being considered by Service Providers to continue offering IPv4 service to customers while transitioning to IPv6. This technology adds an extra Large-Scale NAT ("LSN") in the Service Provider network, thereby resulting in two NATs. CableLabs, Time Warner Cable, and Rogers Communications independently tested the impacts of NAT444 on many popular Internet services using a variety of test scenarios, network topologies, and vendor equipment. This document identifies areas where adding a

second layer of NAT disrupts the communication channel for common Internet applications.

5.1.3. Dual-Stack Lite (DS-lite)

"Understanding Dual-Stack Lite" Jeff Doyle, Network World October 22, 2009 http://www.networkworld.com/community/node/46600

This article provides a good introduction to DSlite, at the time of its publication. Please see the following drafts for details and more current work.

"Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion" A. Durand et al.
http://tools.ietf.org/html/draft-ietf-softwire-dual-stack-lite

This document revisits the dual-stack model and introduces the dual-stack lite technology aimed at better aligning the costs and benefits of deploying IPv6 in service provider networks. Dual-stack lite enables a broadband service provider to share IPv4 addresses among customers by combining two well-known technologies: IP in IP (IPv4-in-IPv6) and Network Address Translation (NAT).

"Dual-stack Lite Mobility Solutions" B. Sarikaya and F. Xia October 11, 2009 (expired)
http://tools.ietf.org/html/draft-sarikaya-softwire-dslitemobility-01

Two solutions are presented to show how to use Dual-Stack Lite transition technique in mobile networks: one for Proxy Mobile IPv6 and the other for Dual-Stack Mobile IPv6. Proxy Mobile IPv6 allows IPv4 nodes to receive mobility services using an IPv4 home address. In case of client based mobility using DSMIPv6, mobile node is a dual-stack node and it can receive an IPv4 home address from the home agent which is co-located with DS-lite carrier-grade NAT.

"Scalable Operation of Address Translators with Per-Interface Bindings" J. Arkko and L. Eggert February 9, 2009 (expired)
http://tools.ietf.org/html/draft-arkko-dual-stack-extra-lite-00

This document explains how to employ address translation in networks that serve a large number of individual customers without requiring a correspondingly large amount of private IPv4 address space.

"Gateway Initiated Dual-Stack Lite Deployment" F. Brockners et al.
http://tools.ietf.org/html/draft-ietf-softwire-gateway-init-ds-lite

Gateway-Initiated Dual-Stack lite (GI-DS-lite) is a modified approach to the original Dual-Stack lite (DS-lite) applicable to certain tunnel-based access architectures. GI-DS-lite extends existing access tunnels beyond the access gateway to an IPv4-IPv4 NAT using softwires with an embedded context identifier, that uniquely identifies the end-system the tunneled packets belong to. The access gateway determines which portion of the traffic requires NAT using local policies and sends/receives this portion to/from this softwire tunnel.

"Deployment DS-lite in Point-to-Point Access Network" Y. Lee (Ed.) et al. http://tools.ietf.org/html/draft-zhou-softwire-ds-lite-p2p

Gateway-Initiated Dual-Stack lite (GI-DS-lite) is a proposal to logically extend existing access tunnels beyond the access gateway to DS-Lite Address Family Transition Router element (AFTR) using softwires with an embedded context identifier. This memo describes a deployment model using GI-DS-lite in Point-to-Point access network.

"Deploying Dual-Stack Lite in IPv6 Network" M. Boucadair (Ed.) et al. http://tools.ietf.org/html/draft-boucadair-dslite-interco-v4v6

Dual-Stack lite requires that the AFTR must have IPv4 connectivity. This forbids a service provider who wants to deploy AFTR in an IPv6-only network. This memo proposes an extension to implement a stateless IPv4-in-IPv6 encapsulation in the AFTR so that AFTR can be deployed in an IPv6-only network.

"IPv6 RA Option for DS-lite AFTR Element" Y. Lee, M. Boucadair and X. Xu http://tools.ietf.org/html/draft-lee-6man-ra-dslite

This document specifies a new optional extension to IPv6 Router Advertisement messages to allow IPv6 routers to advertise DS-Lite AFTR addresses to IPv6 hosts (i.e., a default IPv6 route for DS-Lite traffic). The provisioning of the AFTR address is crucial to access IPv4 connectivity services in a DS-Lite context. Means to ensure reliable delivery of this information to connecting hosts is a must.

Furthermore, this RA option can be used as a means to distribute DS-Lite serviced customers among a set of deployed AFTRs without requiring a central knowledge of the underlying topology and deployed AFTRs.

5.2. Tunneling Mechanisms

RFC 2473 "Generic Packet Tunneling in IPv6 Specification." A. Conta and S. Deering, December 1998
http://tools.ietf.org/html/rfc2473

This document defines the model and generic mechanisms for IPv6 encapsulation of Internet packets, such as IPv6 and IPv4. The model and mechanisms can be applied to other protocol packets as well, such as AppleTalk, IPX, CLNP, or others.

RFC 2529 "Transmission of IPv6 over IPv4 Domains without Explicit Tunnels" B. Carpenter and C. Jung March 1999.
http://tools.ietf.org/html/rfc2529

This memo specifies the frame format for transmission of IPv6 packets and the method of forming IPv6 link-local addresses over IPv4 domains. The motivation for this method is to allow isolated IPv6 hosts, located on a physical link which has no directly connected IPv6 router, to become fully functional IPv6 hosts by using an IPv4 domain that supports IPv4 multicast as their virtual local link.

RFC 3056 "Connection of IPv6 Domains via IPv4 Clouds" B. Carpenter and K. Moore February 2001
http://tools.ietf.org/html/rfc3056

This memo specifies an optional interim mechanism for IPv6 sites to communicate with each other over the IPv4 network without explicit tunnel setup, and for them to communicate with native IPv6 domains via relay routers.

RFC 3053 "IPv6 Tunnel Broker" A. Durand, I. Guardini and D. Lento January 2001
http://tools.ietf.org/html/rfc3053

The IPv6 global Internet as of today uses a lot of tunnels over the existing IPv4 infrastructure. Those tunnels are difficult to configure and maintain in a large scale environment, and the process is too complex for the isolated end user. The motivation for the development of the tunnel broker model is to help early IPv6 adopters to hook up to an existing IPv6 network with stable, permanent IPv6 addresses and DNS names.

5.2.1. Teredo

RFC 4380 "Teredo: Tunneling IPv6 over UDP" C. Huitema February 2006
http://tools.ietf.org/html/rfc4380

This RFC defined a service that enables nodes located behind one or more IPv4 Network Address Translations (NATs) to obtain IPv6 connectivity by tunneling packets over UDP; we call this the Teredo service. Running the service requires the help of "Teredo servers" and "Teredo relays". The Teredo servers are stateless, and only have to manage a small fraction of the traffic between Teredo clients; the Teredo relays act as IPv6 routers between the Teredo service and the "native" IPv6 Internet. The relays can also provide interoperability with hosts using other transition mechanisms such as "6to4". Teredo client capability has been included in Windows operating systems since Windows XP and public servers are available.

RFC 5991 "Teredo Security Extensions" D. Thaler, S. Krishnan and J. Hoagland September 2010
http://tools.ietf.org/html/rfc5991

The Teredo protocol defines a set of flags that are embedded in every Teredo IPv6 address. This document specifies a set of security updates that modify the use of this flags field, but are backward compatible.

"Teredo Extensions", D. Thaler
http://tools.ietf.org/html/draft-thaler-v6ops-teredo-extensions

This document specifies a set of extensions to the Teredo protocol. These extensions provide additional capabilities to Teredo, including support for more types of Network Address Translations (NATs), and support for more efficient communication.

5.2.2. IPv6 Rapid Deployment (6rd)and Extensions

IPv6 Rapid Deployment (6rd) is an approach that allows a service provider to quickly roll out an IPv6 service offering. Free, a large French ISP, successfully deployed a 6rd offering in 5 weeks. It is also being used in a current IPv6 trial offered by Comcast in the USA.

"How 6rd Eases the Transition to IPv6" Mike Capuano on Cisco SP360 blog, August 5, 2010
http://blogs.cisco.com/sp/how_6rd_eases_the_transition_to_ipv6/

This article provides a quick overview of 6rd. The fundamental protocol specification and initial implementation experience can be found in RFC 5969 and 5569.

RFC 5969 "IPv6 Rapid Deployment on IPv4 Infrastructures (6rd)- Protocol Specification" W. Townsley and O. Troan August 2010 http://tools.ietf.org/html/rfc5969

RFC 5569 "IPv6 Rapid Deployment on IPv4 Infrastructures (6rd)" R. Despres January 2010 http://tools.ietf.org/html/rfc5569

"IPv6 Across NAT44 CPEs (6a44)" R. Despres, B. Carpenter and S. Jiang http://tools.ietf.org/html/draft-despres-softwire-6a44

IPv6 Across NAT44 CPEs (6a44) 6a44 is based on an address mapping and on a mechanism whereby suitably upgraded hosts behind a NAT may obtain IPv6 connectivity via a stateless 6a44 server function operated by their Internet Service Provider. With it, traffic between two 6a44 hosts in a single site remains within the site. Except for IANA numbers that remain to be assigned, the specification is intended to be complete enough for running codes to be independently written and interwork.

[Note that this draft converges and supersedes work started in two separate drafts, which are no longer relevant: http://tools.ietf.org/html/draft-despres-softwire-6rdplus-00 http://tools.ietf.org/html/draft-carpenter-softwire-sample-00]

"UDP Encapsulation of 6rd" Y. Lee and P. Kapoor http://tools.ietf.org/html/draft-lee-softwire-6rd-udp-02

This memo specifies the UDP encapsulation to IPv6 Rapid Deployment (6rd) protocol which enables hosts behind unmodified Home Gateway device to access 6rd service. One variation (Server Model) avoids host modification by offloading the implementation to a small server (relay) on the home LAN.

"Gateway Initiated 6rd" T. Tsou et al. http://tools.ietf.org/html/draft-tsou-softwire-gwinit-6rd

This document proposes an alternative to the deployment model defined in RFC 5969 for 6rd. This model extends existing access tunnels beyond an operator-owned gateway collocated with the operator’s IPv4 network edge to the Border Router. This modification makes it unnecessary to provide IPv4 routes to IPv6 UEs. The gateway serves as an aggregation point for IPv4 routing.

5.2.3. Tunnel Support Protocol (TSP)

RFC 5572 "IPv6 Tunnel Broker with the Tunnel Setup Protocol (TSP)" M. Blanchet and F. Parent, February 2010
http://tools.ietf.org/html/rfc5572

TSP is an Experimental RFC defining a method for a tunnel client to negotiate tunnel characteristics with a tunnel broker. It enables tunnels in various deployment architectures including NAT traversal and mobility, and for user authentication it utilizes:

RFC 4422 "Simple Authentication and Security Layer (SASL)" A. Melikov ad K. Zeilenga(Eds.) June 2006
http://tools.ietf.org/html/rfc4422

5.2.4. Residual IPv4 Deployment over IPv6-only Infrastructure

Further down the transition road, operators may desire to retire IPv4 routing support and move their backbone networks to IPv6-only. There may be residual IPv4 legacy customers (clients and servers) still requiring the delivery of IPv4 packets. While the previously proposed Dual-Stack Transition Mechanism (DSTM) approach attempted to satisfy this use case, it was complex and stateful. A stateless approach to IPv4 residual deployment (4rd) is defined in section 3.2 of the Stateless Address Mapping (SAM) draft. At the time of this publication, several network operators in Japan are planning implementation to support residual IPv4 customers.

"Stateless Address Mapping (SAM) - a Simplified Mesh-Softwire Model" Despres, R. July 12, 2010
http://tools.ietf.org/html/draft-despres-softwire-sam

"IPv4 Residual Deployment across IPv6-Service networks (4rd): A NAT-less Solution" R. Despres
http://tools.ietf.org/html/draft-despres-softwire-4rd

5.2.5. Address Plus Port (AplusP)

"The A+P Approach to the IPv4 Address Shortage" R. Bush (Ed.) October 27, 2009 (expired, but authors indicate a new draft is coming)
http://tools.ietf.org/html/draft-ymbk-aplusp

This draft discusses the possibility of address sharing by treating some of the port number bits as part of an extended IPv4 address (Address plus Port, or A+P). Instead of assigning a single IPv4

address to a customer device, we propose to extended the address by "stealing" bits from the port number in the TCP/UDP header, leaving the applications a reduced range of ports. This means assigning the same IPv4 address to multiple clients (e.g., CPE, mobile phones), each with its assigned port-range. In the face of IPv4 address exhaustion, the need for addresses is stronger than the need to be able to address thousands of applications on a single host. If address translation is needed, the end-user should be in control of the translation process - not some smart boxes in the core.

"Aplusp Lite - A light weight aplusp approach" Z. Xiaoyu http://tools.ietf.org/html/draft-xiaoyu-aplusp-lite

This document proposes a solution aimed at providing IPv4 continuity in IPv6 environment. The proposed solution is expected to alleviate the public IPv4 depletion problem while maximize the benefits from IPv6 deployment, and meet the desired service availability and reliability with affordable cost.

5.2.6. IRON-RANGER and ISATAP Solutions

A body of RFCs and drafts in progress provide an alternative approach to IPv4/IPv6 coexistence. This approach utilizes tunneling techniques to create "overlay" networks. While currently considered "Experimental" it may be of interest to network operators as an alternative network architecture.

RFC 5214 "Intra-Site Automatic Tunnel Addressing Protocol (ISATAP)" F. Templin et al. March 2008 http://tools.ietf.org/html/rfc5214

RFC 5579 "Transmission of IPv4 Packets over Intra-Site Automatic Tunnel Addressing Protocol (ISATAP) Interfaces" F. Templin (Ed.) February 2010 http://tools.ietf.org/html/rfc5579

RFC 5320 "The Subnetwork Encapsulation and Adaptation Layer (SEAL)" F. Templin (Ed.) February 2010 http://tools.ietf.org/html/rfc5320

Fred Templin originally published SEAL as an Experimental RFC, and is currently updating with the intention to publish as Standards Track: http://tools.ietf.org/html/draft-templin-intarea-seal

RFC 5558 "Virtual Enterprise Traversal (VET)" F. Templin (Ed.) February 2010 http://tools.ietf.org/html/rfc5558

Fred Templin originally published VET as an Informational RFC, and is currently updating with the intention to publish as Standards Track: http://tools.ietf.org/html/draft-templin-intarea-vet

RFC 5720 "Routing and Addressing in Networks with Global Enterprise Recursion (RANGER)" F. Templin (Ed.) February 2010 http://tools.ietf.org/html/rfc5720

"The Internet Routing Overlay Network (IRON)" F. Templin (Ed.) http://tools.ietf.org/html/draft-templin-iron

5.2.7. Softwires Hub and Spoke with L2TP

RFC 5571 "Softwire Hub and Spoke Deployment Framework with Layer Two Tunneling Protocol Version 2 (L2TPv2)" B. Storer et al. June 2009 http://tools.ietf.org/html/rfc5571

This document describes the framework of the Softwire "Hub and Spoke" solution with the Layer Two Tunneling Protocol version 2 (L2TPv2). The implementation details specified in this document should be followed to achieve interoperability among different vendor implementations.

5.3. Translation

From the earliest specification of IPv6 IETF contributors have recognized that translation would be a necessary tool for transition and coexistence, as IPv6 was designed as an incompatible replacement rather than an extension of IPv4. The original approach to stateless translation defined in RFC 2765 and its implementation as NA(P)T-PT as described in RFC 2766 had a number of issues that resulting in the approach being deprecated by RFC 4966. Recently the Behave WG has taken on the work of defining a set of scenarios covering the use cases for translation, prioritizing the work and defining new solutions that overcome the deficiencies of the historic approach.

5.3.1. Historic Approach

RFC 2765 "Stateless IP/ICMP Translation (SIIT)." E. Nordmark, February 2000 http://tools.ietf.org/html/rfc2765

This document specifies a transition mechanism algorithm in addition to the mechanisms already specified in RFC 1933 (note that this reference was subsequently obsoleted by RFC 2893 which in turn was obsoleted by RFC 4213). The algorithm translates between IPv4 and IPv6 packet headers (including ICMP headers) in separate translator "boxes" in the network without requiring any per-connection state in those "boxes". This new algorithm can be used as part of a solution that allows IPv6 hosts, which do not have a permanently assigned IPv4 addresses, to communicate with IPv4-only hosts. The document neither

specifies address assignment nor routing to and from the IPv6 hosts when they communicate with the IPv4-only hosts.

SIIT has been applied in several translation implementations, including the historic NAT-PT specified in RFC 2766 and deprecated by RFC 4966. SIIT is currently being revised in "IP/ICMP Translation Algorithm" X. Li, C. Bao and F. Baker
http://tools.ietf.org/html/draft-ietf-behave-v6v4-xlate

RFC 2766 "Network Address Translation - Protocol Translation (NAT-PT)." G. Tsirtsis and P. Srisresh, February 2000
http://tools.ietf.org/html/rfc2766

This solution attempted to provide transparent routing to end-nodes in an IPv6 realm trying to communicate with end-nodes in an IPv4 realm and vice versa. This combined Network Address Translation and Protocol Translation. While it did mandate dual-stack support or special purpose routing requirements (such as requiring tunneling support) on end nodes, it did introduce issues that were considered harmful enough to lead to its deprecation in July 2007 by RFC 4966 "Reasons to Move the Network Address Translator - Protocol Translator (NAT-PT) to Historic Status" http://tools.ietf.org/html/rfc4966.

RFC 2767 "Dual-Stack Hosts Using ’Bump in the Stack’ Technique (BIS)" K. Tsuchiay, H. Higuchi and Y. Atarashi February 2000

RFC 3338 "Dual-Stack Hosts Using ’Bump in the API’ (BIA)" S. Lee, et al. October 2002
http://tools.ietf.org/html/rfc3338

These two RFCs are proposed for obsolescence by a draft that combines both:

"Dual-Stack Hosts Using ’Bump in the Host’(BIH)" B. Huang, H. Deng and T. Savolainen
http://tools.ietf.org/html/draft-ietf-behave-v4v6-bih

5.3.2. Current Translation Approaches

A renewed effort to define new translation mechanisms started with discussions in the Internet Area (intarea) meeting and the Technical Plenary at IETF 71 in Dublin, and continued at a special meeting in Montreal in October 2008. This led to a commitment by contributors in the Behave WG to take on the work. A set of scenarios were defined along with a framework for the translation solutions.

"IPv4 Run-Out and IPv4-IPv6 Co-Existence Scenarios" J. Arkko and M. Townsley
http://tools.ietf.org/html/draft-arkko-townsley-coexistence

When IPv6 was designed, it was expected that the transition from IPv4 to IPv6 would occur more smoothly and expeditiously than experience has revealed. The growth of the IPv4 Internet and predicted depletion of the free pool of IPv4 address blocks on a foreseeable horizon has highlighted an urgent need to revisit IPv6 deployment models. This document provides an overview of deployment scenarios with the goal of helping to understand what types of additional tools the industry needs to assist in IPv4 and IPv6 co-existence and transition.

This document was originally created as input to the Montreal co-existence interim meeting in October 2008, which led to the rechartering of the Behave and Softwire working groups to take on new IPv4 and IPv6 coexistence work. This document is published as a historical record of the thinking at the time.

"A Framework for IPv4/IPv6 Translation" F. Baker et al.
http://tools.ietf.org/html/draft-ietf-behave-v6v4-framework

This draft (Framework) is the place to start to understand the historic context for translation, the definition and rationale for the set of translation scenarios and canonical definitions for some of the terminology that arises when talking about translation and coexistence in general.

The 4 deployment modes for these scenarios are:

1. Connecting between the IPv4 Internet and the IPv6 Internet

2. Connecting an IPv6 network to the IPv4 Internet

3. Connecting an IPv4 network to the IPv6 Internet

4. Connecting between an IPv4 network and an IPv6 network

As solutions may differ with respect to the initiating end of the conversation, 8 scenarios are defined in the Framework draft, as recapped in the following sections along with specifications that fit each scenario.

Some general specifications that are cited in the various solution specifications (or may be in subsequent revisions) are:

"IPv6 Addressing of IPv4/IPv6 Translators" C. Bao et al. August 16, 2010 http://tools.ietf.org/html/draft-ietf-behave-address-format-10

"DNS64: DNS extensions for Network Address Translation from IPv6 Clients to IPv4 Servers" M. Bagnulo et al. July 5, 2010 http://tools.ietf.org/html/draft-ietf-behave-dns64-10

"Analysis of 64 Translation" R. Penno, T. Saxena and D. Wing http://tools.ietf.org/html/draft-penno-behave-64-analysis

Due to specific problems, NAT-PT was deprecated by the IETF as a mechanism to perform IPv6-IPv4 translation. Since then, new effort has been undertaken within IETF to standardize alternative mechanisms to perform IPv6-IPv4 translation. This document evaluates how the new translation mechanisms avoid the problems that caused the IETF to deprecate NAT-PT.

5.3.2.1. An IPv6 network to the IPv4 Internet

The Framework defines Scenario 1 for an early adopter (end user or network operator) which establishes an IPv6 network and needs to maintain access to the global IPv4 Internet, preferably without assigning IPv4 addresses to the nodes of the IPv6 network. Either the Stateful or Stateless solutions proposed may satisfy this deployment scenario.

"Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers" M. Bagnulo, P. Matthews and I. van Beijnum http://tools.ietf.org/html/draft-ietf-behave-v6v4-xlate-stateful

"IP/ICMP Translation Algorithm" X. Li, C. Bao and F. Baker http://tools.ietf.org/html/draft-ietf-behave-v6v4-xlate

5.3.2.2. The IPv4 Internet to an IPv6 network

The Framework defines Scenario 2 for a node on the IPv4 Internet initiating a transmission to a node on an IPv6 network. The original approach to this deployment was the NAT-PT implementation of SIIT (as defined in RFC 2766) which has been deprecated (by RFC 4966). The Stateless Translation solution for Scenario 1 also would work for this case as it does support IPv4-initiated communication with a subset of IPv6 addresses.

5.3.2.3. The IPv6 Internet to an IPv4 network

The Framework defines Scenario 3 where a legacy IPv4 network has a requirement to provide services to users in the IPv6 Internet.

Stateful Translation with static AAAA records in DNS to represent the IPv4-only hosts will work.

"Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers" M. Bagnulo, P. Matthews and I. van Beijnum http://tools.ietf.org/html/draft-ietf-behave-v6v4-xlate-stateful

"DNS64: DNS extensions for Network Address Translation from IPv6 Clients to IPv4 Servers" M. Bagnulo et al. http://tools.ietf.org/html/draft-ietf-behave-dns64

Alternatively, host-based translation (BIH) or tightly-coupled translators may be considered.

5.3.2.4. An IPv4 network to the IPv6 Internet

Scenario 4 is not easy to solve but fortunately will not arise until significant IPv6 uptake. In-network translation is not viable, and other techniques should be considered including host-based translation (BIH) or tightly-coupled translators that adapt legacy hosts or networks to the IPv6 Internet.

5.3.2.5. An IPv6 network to an IPv4 network

Scenario 5 describes a configuration where both the IPv6 network and IPv4 network are within the administrative control of the same organization. It appears amenable to the same solutions proposed for Scenario 1.

5.3.2.6. An IPv4 network to an IPv6 network

Scenario 6 is the mirror image of Scenario 5, with communication initiated from the IPv4 side. It appears amenable to the same solution proposed for Scenario 2.

5.3.2.7. The IPv6 Internet to the IPv4 Internet

The Framework indicates that Scenario 7, the interconnection of the IPv4 Internet with the IPv6 Internet may appear to be an ideal case for an in-network translator (such as the deprecated NAT-PT), but there is no viable way to map the immense IPv6 address space onto IPv4. This situation would not entail until significant IPv6 adoption, and has not been a priority for solution.

5.3.2.8. The IPv4 Internet to the IPv6 Internet

Scenario 8 presents a challenge similar to Scenario 7.

5.4. Connectivity Checking and Delay Avoidance

One important issue that arises in a coexistence environment is negative impact on the initiation of peer-to-peer connections, such as VoIP, video, etc. The initiator doesn't know a priori whether the peer is using the same address family incurring a possible delay as the first attempt may fail. There is also ambiguity, as the IPv6 path may be temporarily broken.

"IPv6 Connectivity Check and Redirection by HTTP Servers" E. Vyncke
http://tools.ietf.org/html/draft-vyncke-http-server-64aware

Rather than forcing the client to decide whether IPv4 or IPv6 is more convenient to reach a web server; this document proposes to let the web server check whether there is IPv6 connectivity to the client; then the web server can do a HTTP redirect to the force the client to use IPv6.

This is done easily by a script within the server HTML pages and does not require any change in the client applications or configuration. The client still can control whether he/she wants to enabled IPv6.

"Happy Eyeballs: Trending Towards Success (IPv6 and SCTP)", D. Wing, A. Yourtchenko, P. Natarajan.
http://tools.ietf.org/html/draft-wing-http-new-tech

This draft makes several recommendations to ensure user satisfaction and a smooth transition from HTTP's pervasive IPv4 to IPv6 and from TCP to SCTP. While the target audience is app developers and content providers, network operators should be aware of techniques needed to maintain peaceful coexistence without negative impact on end-user perception of service level.

"Migrating SIP to IPv6 Media Without Connectivity Checks" D. Wing, A. Yourtchenko
http://tools.ietf.org/html/draft-wing-dispatch-v6-migration

During the migration from IPv4 to IPv6, it is anticipated that an IPv6 path might be broken for a variety of reasons, causing endpoints to not receive RTP data. Connectivity checks would detect and avoid the user noticing such a problem, but there is industry reluctance to implement connectivity checks.

This document describes a mechanism allowing dual-stack SIP endpoints to attempt communications over IPv6 and fall back to IPv4 if the IPv6 path is not working. The mechanism does not require connectivity checks.

6. Prefix and Address Assignment and Distribution

RFC 4291 "IP Version 6 Addressing Architecture." R. Hinden, S. Deering. February 2006.
http://tools.ietf.org/html/rfc4291

RFC 5952 "A Recommendation for IPv6 Text Representation" S. Kawamura and M. Kawashima, August 2010
http://tools.ietf.org/html/rfc5952

RFC 4291 defines the addressing architecture of the IP Version 6 (IPv6) protocol. The document includes the IPv6 addressing model, text representations of IPv6 addresses, definition of IPv6 unicast addresses, anycast addresses, and multicast addresses, and an IPv6 node's required addresses. RFC 5952 updates RFC 4291 with a recommended method for rendering IPv6 addresses in a standard form for user interfaces, logging and reporting.

"IPv6 Addressing of IPv4/IPv6 Translators" C. Bao et al. (Status: Standards Track, in RFC Editor will update RFC 4291)
http://tools.ietf.org/html/draft-ietf-behave-address-format

This document discusses the algorithmic translation of an IPv6 address to a corresponding IPv4 address, and vice versa, using only statically configured information. It defines a well-known prefix for use in algorithmic translations, while allowing organizations to also use network-specific prefixes when appropriate. Algorithmic translation is used in IPv4/IPv6 translators, as well as other types of proxies and gateways (e.g., for DNS) used in IPv4/IPv6 scenarios.

RFC 3177 "IAB/IESG Recommendations on IPv6 Address Allocations to Sites." IAB, IESG. September 2001.
http://tools.ietf.org/html/rfc3177

RFC 3177 provides recommendations to the addressing registries (APNIC, ARIN and RIPE-NCC) on policies for assigning IPv6 address blocks to end sites. In particular, it recommends the assignment of /48 in the general case, /64 when it is known that one and only one subnet is needed and /128 when it is absolutely known that one and only one device is connecting.

"IPv6 Address Assignment to End Sites", T. Narten, G. Huston, R. Roberts, 12-Jul-10
http://tools.ietf.org/html/draft-ietf-v6ops-3177bis-end-sites

The proposed update to RFC 3177 revises the recommendation to leave the exact choice to the operational community. The role of the IETF

is limited to providing guidance on IPv6 architectural and operational considerations. This document reviews the architectural and operational considerations of end site assignments as well as the motivations behind the original 3177 recommendations. Moreover, the document clarifies that a one-size-fits-all recommendation of /48 is not nuanced enough for the broad range of end sites and is no longer recommended as a single default.

RFC 4192 "Procedures for Renumbering an IPv6 Network without a Flag Day" F. Baker, E. Lear and R. Droms
http://www.ietf.org/rfc/rfc4192.txt

RFC 5942 "IPv6 Subnet Model: The Relationship between Links and Subnet Prefixes." H. Singh, W. Beebee, E. Nordmark. July 2010.
http://tools.ietf.org/html/rfc5942

IPv6 specifies a model of a subnet that is different than the IPv4 subnet model. The subtlety of the differences has resulted in incorrect implementations that do not interoperate. This document spells out the most important difference: that an IPv6 address isn't automatically associated with an IPv6 on-link prefix. This document also updates (partially due to security concerns caused by incorrect implementations) a part of the definition of "on-link" from RFC 4861.

RFC 4862 "IPv6 Stateless Address Autoconfiguration." S. Thomson, T. Narten, T. Jinmei. September 2007.
http://tools.ietf.org/html/rfc4862

RFC 4941 "Privacy Extensions for Stateless Address Autoconfiguration in IPv6." T. Narten, R. Draves, S. Krishnan. September 2007.
http://tools.ietf.org/html/rfc4941

The IPv6 addressing architecture presumes that the remaining 64 bits are an endpoint interface identifier. This could be the MAC Address (EUI-64 Address) in an appropriate encoding, or it could be what is called a "privacy address", which is a random number. You will find the most common approach to that, for hosts, in this RFC.

RFC 3315 "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)." R. Droms (Ed.), J. Bound, B. Volz, T. Lemon, C. Perkins, M. Carney. July 2003. http://tools.ietf.org/html/rfc3315

"Analysis of Solution Proposals for hosts to learn NAT64 Prefixes" J. Korhonen (Ed.) and T. Savolainen
http://tools.ietf.org/html/draft-korhonen-behave-nat64-learn-analysis

Hosts and applications may benefit from the knowledge if an IPv6 address is synthesized, which would mean a NAT64 is used to reach the IPv4 network or Internet. This document analyses number of proposed solutions for communicating if the synthesis is taking place, used address format, and the IPv6 prefix used by the NAT64 and DNS64. This enables both NAT64 avoidance and intentional utilization by allowing local IPv6 address synthesis.

7. How-to, Whitepapers and FAQs

"IPv6 Rollout: Where do we start?" O. Crepin-Leblond
http://www.slideshare.net/ocl999/suggestion-for-an-ipv6-roll-out

"Everything Sysadmin" T. Limoncelli
http://everythingsysadmin.com/2009/01/google-enables-ipv6-for-most-s.html
http://everythingsysadmin.com/2010/08/methods-of-converting-to-ipv6.html

"IPv6 Deployment in Internet Exchange Points (IXPs)", Roque Gagliano
http://tools.ietf.org/html/draft-ietf-v6ops-v6inixp

This draft suggests that in an Internet Exchange Point one might use an address that helps in debugging routing exchanges. One could also look at what other folks do, embedding identifying marks in addresses. For example, Facebook includes "face:b00c" in the IID portion of their address.

8. Experiments, Trials and Prototypes

6bone (concluded)
http://go6.net/ipv6-6bone/

Hurricane Electric (ongoing)
http://www.he.net/

T-Mobile USA (ongoing)
http://groups.google.com/group/tmoipv6beta

Comcast (ongoing)
http://www.comcast6.net/

Internode ADSL (Ongoing)
http://ipv6.internode.on.net/access/adsl/

Verizon FiOS (small scale test - concluded)
http://newscenter.verizon.com/press-releases/verizon/2010/verizon-begins-testing-ipv6.html

"Considerations for Stateless Translation (IVI/dIVI) in Large SP Network" Q. Sun et al.
http://tools.ietf.org/html/draft-sunq-v6ops-ivi-sp

In addition to the deployment use case this draft describes, the draft documents an experimental use of the translation in a research network.

Measurements of IPv6 Path MTU Discovery Behavior
http://www.ripe.net/ripe/meetings/ripe-60/presentations/Stasiewicz-Measurements_of_IPv6_Path_MTU_Discovery_Behaviour.pdf

9. Implementation Reports

"A Basic Guideline for Listing ISPs that Run IPv6" S. Kawamura
http://tools.ietf.org/html/draft-kawamura-ipv6-isp-listings

This draft attempts to gather information about currently known sites that rate ISP readiness for IPv6 and to look at their evaluation methods. This document also summarizes basic guidelines that these listings may consider when checking an ISPs IPv6 readiness. As the draft says, there are many opinions about what it means to be ready for IPv6, and it would be helpful to evaluate ISPs based on some common criteria.

IPv6 Rapid Deployment
http://tools.ietf.org/html/rfc5569

Google has hosted a meeting of IPv6 Implementers in 2009 and 2010, several presentations covered experimental or live transition experience.
https://sites.google.com/site/ipv6implementors/2009/agenda
https://sites.google.com/site/ipv6implementors/2010/agenda

10. Books on IPv6

Blanchet, Marc. "Migrating to IPv6: a Practical Guide to Implementing IPv6 in Mobile and Fixed Networks." Chichester, England: J. Wiley & Sons, 2006. Print.

Hagen, Silvia. "IPv6 Essentials - Second Edition" Sebastapol, CA: O’Reilly Media, Inc, 2006. Print.

Loshin, Peter. "IPv6, Second Edition: Theory, Protocol and Practice" Morgan Kaufmann Publishing, 2003

Popoviciu, Ciprian, Eric Levy-Abengoli and Patrick Grossetete "Deploying IPv6 Networks" Indianapolis, IN: Cisco Press, 2006. Print.

Siil, Karl A. "IPv6 Mandates: Choosing a Transition Strategy, Preparing Transition Plans, and Executing the Migration of a Network to IPv6." Indianapolis, IN: Wiley, 2008. Print.

11. Miscellaneous

See the Dancing Turtle, but only if you have native IPv6!
http://www.kame.net/

A little more detail than a Dancing Turtle, on your IPv6 readiness can be obtained by using this site put up by Jason Fesler:
http://test-ipv6.com/

There is an extension for Firefox (and perhaps other browsers) that displays the IP address of web pages you visit, clearly indicating when you are connected via IPv4 or IPv6. In Firefox, click on Tools..Add-ons..Extensions and search for ShowIP.

Eric Vyncke is collecting some statistics on IPv6 penetration.
http://www.vyncke.org/ipv6status/

A reasonable estimation of how fast the sky is falling.
http://www.potaroo.net/tools/ipv4/

A graphical representation of IPv4 depletion.
http://www.ipv4depletion.com/old.html

"IPv6 Adoption Remains Slow, Survey Says" W. Jackson, GCN Sept. 5, 2101
http://gcn.com/articles/2010/09/14/adoption-of-ipv6-is-slow.aspx
http://www.nro.net/documents/GlobalIPv6SurveySummaryv2.pdf

Some troubling, yet interesting news about what operators and end-user organizations are thinking about IPv6 adoption at this time.

A study of some of the brokenness around Path MTU Discovery
http://www.ripe.net/ripe/meetings/ripe-60/presentations/Stasiewicz-Measurements_of_IPv6_Path_MTU_Discovery_Behaviour.pdf

Cluenet hosts a mailing list with IPv6 operator participation. Various transition-related topics are brought up there from time to time.http://lists.cluenet.de/mailman/listinfo/ipv6-ops

"IPv6 for Dummies, Part 1: It's Time!"
http://www.xtranormal.com/watch/7201125/

"IPv6 for Dummies, Part 2: Comparing IPv4 and IPv6"
http://www.xtranormal.com/watch/7210035/

12. Security Considerations

This draft does not introduce any security considerations.

13. IANA Considerations

This draft does not require any action from IANA.

[Note to RFC Editor: this section may be removed.]

14. Conclusions

This draft is merely the starting point for a network operator planning an IPv6 rollout. The intention of the editor was to document the great work that is already available that can help in the process and to perhaps save a few hours of redundant effort for someone to find this information. Of course, this will be out of date before it is published as active research continues in coexistence and transition tools. The editor hopes it is at least a useful "You Are Here" map to help navigate the thrill rides available in the IPv6 theme park.

This compendium could serve as an initial set of data to populate an active database or wiki. This would allow continuing community contribution including feedback on the real-world experience of network operators as they turn on IPv6.

15. References

15.1. Normative References

None.

15.2. Informative References

Complete reference information is included in the body of the draft.

16. Acknowledgments

This bibliography is a recapitulation of the contributions of the authors of the cited RFCs, drafts, websites and other publications and many folks on the v6ops and v4v6transition mailing lists, the editor has freely borrowed abstract and summary text from the cited works and e-mail postings. In addition, the editor wishes to acknowledge significant contributions and suggestions from Fred Baker, Brian Carpenter, Remi Despres, Suresh Krishnan, Tina Tsou, Yiu Lee, Marc Blanchet, Med Boucadair, Fred Templin, Andrew Yourtchenko and many contributors on the v4v6trans mailing list. All credit is due to those contributors while the editor takes responsibility for any errors, omissions or mischaracterization of the work in the process of abstracting and summarizing it here.

The IPv4-IPv6 Transition mailing list archive can be found at: https://www.ietf.org/mailman/listinfo/v4tov6transition and the readers are also directed to the mailing list archives of the various IETF Working Groups mentioned for the history of the cited drafts and RFCs.

This document was prepared using 2-Word-v2.0.template.dot.

Author’s Address

Edward J. Jankiewicz
SRI International, Inc.
333 Ravenswood Ave
Menlo Park, CA USA

Phone: 732-389-1003 or 650-859-2000
Email: edward.jankiewicz@sri.com

Individual Submission
Internet-Draft
Intended status: Informational
Expires: April 24, 2011

J. Korhonen, Ed.
J. Soininen
Nokia Siemens Networks
B. Patil
T. Savolainen
G. Bajko
K. Iisakkila
Nokia

October 21, 2010

IPv6 in 3GPP Evolved Packet System
draft-korhonen-v6ops-3gpp-eps-04

Abstract

The increased use of data services, growth of subscribers in 3GPP based mobile networks, and the impending exhaustion of available IPv4 addresses from the registries is driving the need to specify the transition to IPv6 solutions in 3GPP network architectures. This document describes the support for IPv6 in 3GPP network architectures and a solution to transition to IPv6 using a dual-stack approach.

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 24, 2011.

Copyright Notice

Table of Contents

1. Introduction

IPv6 has been specified in the 3rd Generation Partnership Project (3GPP) standards since the early architectures developed for R99 General Packet Radio Service (GPRS). However, the support for IPv6 in commercially deployed networks is nearly non-existent. There are many factors that can be attributed to the lack of IPv6 deployment in 3GPP networks. The most relevant one is essentially the same as the reason for IPv6 not being deployed by other networks as well, i.e. the lack of business and commercial incentives for deployment. 3GPP network architectures have also evolved since 1999 (since R99). The most recent version of the 3GPP architecture, the Evolved Packet System (EPS), which is commonly referred as SAE, LTE or Release-8, is a packet centric architecture. The number of subscribers and devices that are using the 3GPP networks for Internet connectivity and data services has also increased significantly. With the subscriber growth numbers projected to increase even further and the IPv4 addresses depletion problem looming in the near term, 3GPP operators and vendors have started the process of identifying the scenarios and solutions needed to transition to IPv6.

This document describes the establishment of IP connectivity in 3GPP network architectures, specifically in the context of IP bearers for 3GPP GPRS and for 3GPP EPS. It provides an overview of how IPv6 is supported as per the current set of 3GPP specifications. A solution to transitioning to IPv6 based on a dual-stack technology is described as well as some of the issues and concerns with respect to deployment and shortage of private IPv4 addresses within a single network domain.

The IETF has specified a set of tools and mechanisms that can be utilized for transitioning to IPv6. In addition to the dual-stack technology, the two alternative categories for the transition are encapsulation and translation. Most of the mechanisms available in the toolbox can be categorized as belonging to either one of these. The IETF continues to specify additional solutions for enabling the transition based on the deployment scenarios and operator/ISP requirements. The 3GPP scenarios for transition, described in [3GPP.23.975], can be addressed using transition mechanisms that are already available in the toolbox. The objective of transition to IPv6 in 3GPP networks is to ensure that:

1. Legacy devices and hosts which have an IPv4 only stack will continue to be provided with IP connectivity to the Internet and services,

2. Devices which are dual-stack can access the Internet either via IPv6 or IPv4. The choice of using IPv6 or IPv4 depends on the

capability of:

A. the application on the host,

B. the support for IPv4 and IPv6 bearers by the network and/or,

C. the capability of the server(s) and other end points.

3GPP networks are capable of providing a host with IPv4 and IPv6 connectivity today, albeit in many cases with upgrades to network elements such as the SGSN and GGSN.

2. 3GPP Terminology and Concepts

2.1. Terminology

Access Point Name

Access Point Name (APN) is a fully qualified domain name and resolves to a specific gateway in an operators network. The APNs are piggybacked on the administration of the DNS namespace.

Packet Data Protocol Context

A Packet Data Protocol (PDP) Context is the equivalent of a virtual connection between the host and a gateway.

General Packet Radio Service

General Packet Radio Service (GPRS) is a packet oriented mobile data service available to users of the 2G and 3G cellular communication systems Global System for Mobile communications (GSM), and specified by 3GPP.

Packet Data Network

Packet Data Network (PDN) is a packet based network that either belongs to the operator or is an external network such as Internet and corporate intranet. The user eventually accesses services in one or more PDNs. The operator’s packet domain network are separated from packet data networks either by GGSNs or PDN Gateways (PDN-GW).

Gateway GPRS Support Node

Gateway GPRS Support Node (GGSN) is a gateway function in GPRS, which provides connectivity to Internet or other PDNs. The host

attaches to a GGSN identified by an APN assigned to it by an operator. The GGSN also serves as the topological anchor for addresses/prefixes assigned to the mobile host.

Packet Data Network Gateway

Packet Data Network Gateway (PDN-GW) is a gateway function in Evolved Packet System (EPS), which provides connectivity to Internet or other PDNs. The host attaches to a PDN-GW identified by an APN assigned to it by an operator. The PDN-GW also serves as the topological anchor for addresses/prefixes assigned to the mobile host.

Serving Gateway

Serving Gateway (SGW) is a gateway function in EPS, which terminates the interface towards E-UTRAN. The SGW is the Mobility Anchor point for layer-2 mobility (inter-eNodeB handovers). For each User Equipment connected with the EPS, at any given point of time, there is only one SGW. The SGW is essentially the user plane part of the GPRS' SGSN forwarding packets between a PDN-GW.

Serving Gateway Support Node

Serving Gateway Support Node (SGSN) is a network element that is located between the radio access network (RAN) and the gateway (GGSN). A per mobile host point to point (p2p) tunnel between the GGSN and SGSN transports the packets between the mobile host and the gateway.

GPRS tunnelling protocol

GPRS Tunnelling Protocol (GTP) [3GPP.29.060] is a tunnelling protocol defined by 3GPP. It is a network based mobility protocol and similar to Proxy Mobile IPv6 (PMIPv6) [RFC5213]. However, GTP also provides functionality beyond mobility such as inband signaling related to Quality of Service (QoS) and charging among others.

Evolved Packet System

Evolved Packet System (EPS) is an evolution of the 3G GPRS system characterized by higher-data-rate, lower-latency, packet-optimized system that supports multiple Radio Access Technologies (RAT). The EPS comprises the Evolved Packet Core (EPC) together with the evolved radio access network (E-UTRA and E-UTRAN).

Mobility Management Entity

Mobility Management Entity (MME) is a network element that is responsible for control plane functionalities, including authentication, authorization, bearer management, layer-2 mobility, etc. The MME is essentially the control plane part of the GPRS' SGSN and not located on the user plane data path, i.e. user plane traffic bypasses the MME.

UMTS Terrestrial Radio Access Network

UMTS Terrestrial Radio Access Network (UTRAN) is communications network, commonly referred to as 3G, and consists of NodeBs (3G base station) and Radio Network Controllers (RNC) which make up the UMTS radio access network. The UTRAN allows connectivity between the mobile host/device and the core network.

Evolved UTRAN

Evolved UTRAN (E-UTRAN) is communications network, sometimes referred to as 4G, and consists of eNodeBs (4G base station) which make up the E-UTRAN radio access network. The E-UTRAN allows connectivity between the mobile host/device and the core network.

GSM EDGE Radio Access Network

GSM EDGE Radio Access Network (GERAN) is communications network, commonly referred to as 2G or 2.5G, and consists of base stations and Base Station Controllers (BSC) which make up the GSM EDGE radio access network. The GERAN allows connectivity between the mobile host/device and the core network.

UE, MS, MN and Mobile

The terms UE (User Equipment), MS (Mobile Station), MN (Mobile Node) and, mobile refer to the devices which are hosts with ability to obtain Internet connectivity via a 3GPP network. The terms UE, MS, MN and devices are used interchangeably within this document.

2.2. The concept of APN

The Access Point Name (APN) essentially refers to a gateway in the 3GPP network. The 'complete' APN is expressed in a form of a Fully Qualified Domain Name (FQDN) and also piggybacked on the administration of the DNS namespace, thus effectively allowing the discovery of gateways using the DNS. Mobile hosts/devices can choose to attach to a specific gateway in the packet core. The gateway

provides connectivity to the Packet Data Network (PDN) such as the Internet. An operator may also include gateways which do not provide Internet connectivity, rather a connectivity to closed network providing a set of operator's own services. A mobile host/device can be attached to one or more gateways simultaneously. The gateway in a 3GPP network is the GGSN or PDN-GW. Figure 1 below illustrates the APN-based network connectivity concept.

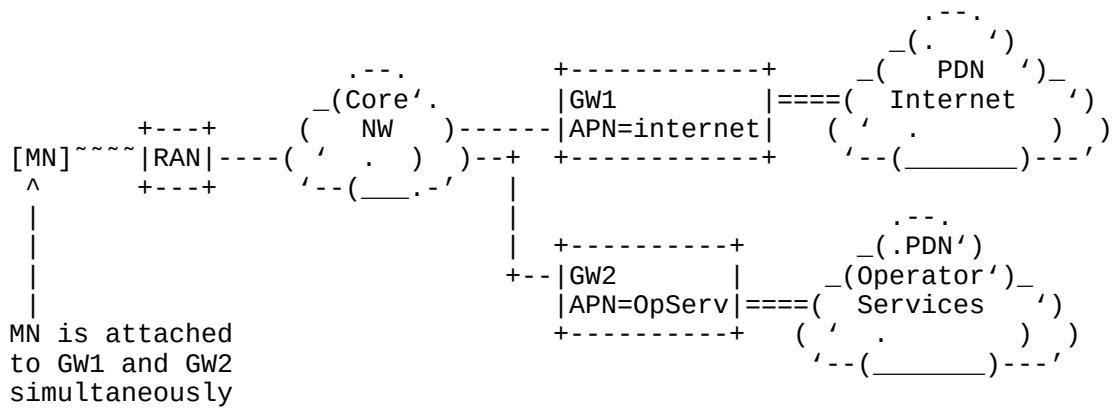


Figure 1: Mobile host/device attached to multiple APNs simultaneously

3. IP over 3GPP GPRS

3.1. Introduction to 3GPP GPRS

A simplified 2G/3G GPRS architecture is illustrated in Figure 2. This architecture basically covers the GPRS core network since R99 to Release-7, and radio access technologies such as GSM (2G), EDGE (2G), WCDMA (3G) and HSPA (3G). The architecture shares obvious similarities with the Evolved Packet System (EPS) as will be seen in Section 4. Based on Gn/Gp interfaces, the GPRS core network functionality is logically implemented on two network nodes, the SGSN and the GGSN.

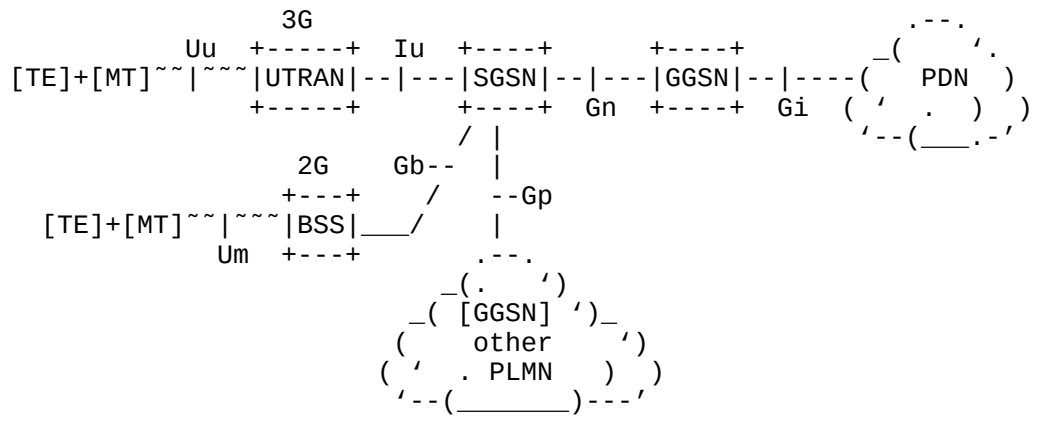


Figure 2: Overview of the 2G/3G GPRS Logical Architecture

Gn/Gp: These interfaces provide a network based mobility service for a mobile host and are used between a SGSN and a GGSN. The Gn interface is used when GGSN and SGSN are located inside one operator (i.e. PLMN). The Gp-interface is used if the GGSN and the SGSN are located in different operator domains (i.e. 'other' PLMN). GTP protocol is defined for the Gn/Gp interfaces (both GTP-C for the control plane and GTP-U for the user plane).

Gb: Is the Base Station System (BSS) to SGSN interface, which is used to carry information concerning packet data transmission and layer-2 mobility management. The Gb-interface is based on either on Frame Relay or IP.

Iu: Is the Radio Network System (RNS) to SGSN interface, which is used to carry information concerning packet data transmission and layer-2 mobility management. The user plane part of the Iu-interface (actually the Iu-PS) is based on GTP-U. The control plane part of the Iu-interface is based on Radio Access Network Application Protocol (RANAP).

Gi: It is the interface between the GGSN and a PDN. The PDN may be an operator external public or private packet data network or an intra-operator packet data network.

Uu/Um: Are either 2G or 3G radio interfaces between a mobile terminal and a respective radio access network.

The SGSN is responsible for the delivery of data packets from and to the mobile hosts within its geographical service area when a direct tunnel option is not used. If the direct tunnel is used, then the

user plane goes directly between the RNS and the GGSN. The control plane traffic always goes through the SGSN. For each mobile host connected with the GPRS, at any given point of time, there is only one SGSN.

3.2. PDP Context

A PDP context is an association between a mobile host represented by one IPv4 address and/or one /64 IPv6 prefix and a PDN represented by an APN. Each PDN can be accessed via a gateway (typically a GGSN or PDN-GW). On the device/mobile host a PDP context is equivalent to a virtual interface/connection. A host may hence be attached to one or more gateways via separate virtual interfaces/connections, i.e. PDP contexts. Each primary PDP context has its own IPv4 address and/or one /64 IPv6 prefix assigned to it by the PDN and anchored in the corresponding gateway. Applications on the host use the appropriate PDP context (virtual interface) for connectivity to a specific PDN. Figure 3 represents a high level view of what a PDP context implies in 3GPP networks.

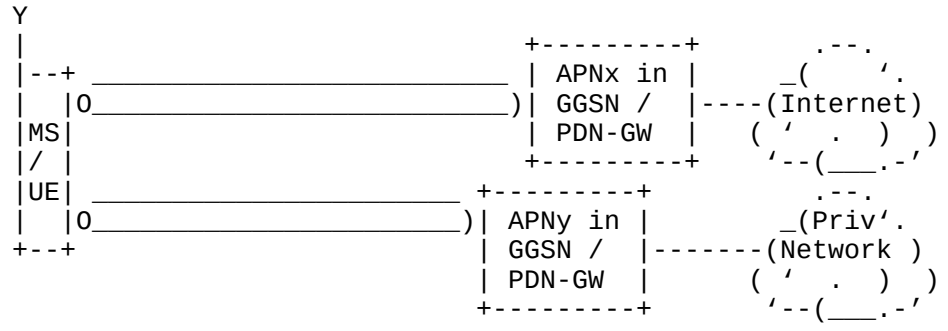


Figure 3: PDP contexts between the MS/UE and gateway

In the above figure there are two PDP contexts at the MS/UE (UE=User Equipment in 3GPP parlance). The PDP context that is connected to APNx provided Internet connectivity and the other PDP context provides connectivity to a private IP network (as an example this network may include operator specific services such as MMS (Multi media service). An application on the host such as a web browser would use the PDP context that provides Internet connectivity for accessing services on the Internet. An application such as MMS would use APNy in the figure above because the service is provided through the private network.

4. IP over 3GPP EPS

4.1. Introduction to 3GPP EPS

In its most basic form, the EPS architecture consists of only two nodes on the user plane, a base station and a core network Gateway (GW). The basic EPS architecture is illustrated in Figure 4. The Mobility Management Entity (MME) node performs control-plane functionality and is separated from the node(s) that performs bearer-plane functionality (GW), with a well-defined open interface between them (S11). The optional interface S5 can be used to split the Gateway (GW) into two separate nodes, the Serving Gateway (SGW) and the PDN-GW. This allows independent scaling and growth of traffic throughput and control signal processing. The functional split of gateways also allows operators to choose optimized topological locations of nodes within the network in order to optimize the network in different aspects.

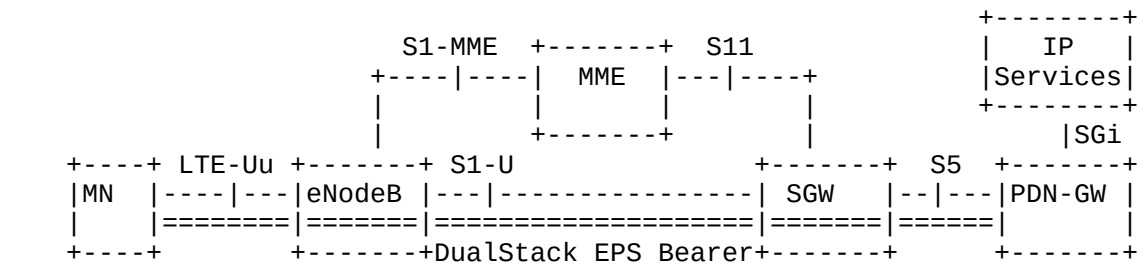


Figure 4: EPS Architecture for 3GPP Access

S5: It provides user plane tunnelling and tunnel management between SGW and PDN-GW, using GTP or PMIPv6 as the network based mobility management protocol.

S1-U: Provides user plane tunnelling and inter eNodeB path switching during handover between eNodeB and SGW, using the GTP-U protocol (GTP user plane).

S1-MME: Reference point for the control plane protocol between eNodeB and MME.

SGi: It is the interface between the PDN-GW and the packet data network. Packet data network may be an operator external public or private packet data network or an intra operator packet data network.

The eNodeB is a base station entity that supports the Long Term Evolution (LTE) air interface and includes functions for radio

resource control, user plane ciphering, and other lower layer functions. MME is responsible for control plane functionalities, including authentication, authorization, bearer management, layer-2 mobility, etc.

The SGW is the Mobility Anchor point for layer-2 mobility. For each MN connected with the EPS, at any given point of time, there is only one SGW.

4.2. PDN Connection

A PDN connection is an association between a mobile host represented by one IPv4 address and/or one /64 IPv6 prefix, and a PDN represented by an APN. Each PDN can be accessed via a gateway (a PDN-GW). PDN is responsible for the IP address/prefix allocation to the mobile host. On the device/mobile host a PDN connection is equivalent to a virtual interface/connection. A host may hence be attached to one or more gateways via separate virtual interfaces/connections, i.e. PDN connection. Each PDP connection has its own IP address/prefix assigned to it by the PDN and anchored in the corresponding gateway. Applications on the host use the appropriate PDN connection (virtual interface) for connectivity. The PDN connection is the EPC equivalent of the GPRS PDP context.

4.3. EPS bearer model

The logical concept of a bearer has been defined to be an aggregate of one or more IP flows related to one or more services. An EPS bearer exists between the Mobile Node (MN i.e. a mobile host) and the PDN-GW and is used to provide the same level of packet forwarding treatment to the aggregated IP flows constituting the bearer. Services with IP flows requiring a different packet forwarding treatment would therefore require more than one EPS bearer. The mobile host performs the binding of the uplink IP flows to the bearer while the PDN-GW performs this function for the downlink packets.

In order to provide low latency for always on connectivity, a default bearer will be provided at the time of startup and an IPv4 address and/or IPv6 prefix gets assigned to the mobile host (this is different from GPRS, where mobile hosts are not automatically assigned with an IP address or prefix). This default bearer will be allowed to carry all traffic which is not associated with a dedicated bearer. Dedicated bearers are used to carry traffic for IP flows that have been identified to require a specific packet forwarding treatment. They may be established at the time of startup; for example, in the case of services that require always-on connectivity and better QoS than that provided by the default bearer. The default bearer and the dedicated bearer(s) associated to it share the same IP

address(es)/prefix.

An EPS bearer is referred to as a GBR bearer if dedicated network resources related to a Guaranteed Bit Rate (GBR) value that is associated with the EPS bearer are permanently allocated (e.g. by an admission control function in the eNodeB) at bearer establishment/ modification. Otherwise, an EPS bearer is referred to as a non-GBR bearer. The default bearer is always non-GBR, with the resources for the IP flows not guaranteed at eNodeB, and with no admission control. However, the dedicated bearer can be either GBR or non-GBR. A GBR bearer has a Guaranteed Bit Rate (GBR) and Maximum Bit Rate (MBR) while more than one non-GBR bearer belonging to the same UE shares an Aggregate Maximum Bit Rate (AMBR). Non-GBR bearers can suffer packet loss under congestion while GBR bearers are immune to such losses.

5. Address Management

5.1. IPv4 Address Configuration

Mobile host’s IPv4 address configuration is essentially always conducted during PDP context/EPS bearer setup procedures (on layer-2). DHCPv4-based [RFC2131] address configuration is supported by the 3GPP specifications, but is not used in wide scale. The mobile host must always support layer-2 based address configuration, since DHCPv4 is optional for both mobile hosts and networks.

5.2. IPv6 Address Configuration

IPv6 Stateless Address Autoconfiguration (SLAAC) is the only supported address configuration mechanisms [RFC4862]. Stateful DHCPv6-based address configuration is not supported by 3GPP specifications [RFC3315]. On the other hand, Stateless DHCPv6-service to obtain other configuration information is supported [RFC3736]. This implies that the M-bit must always be set to zero and the O-bit may be set to one in the Router Advertisement (RA) sent to the UE.

3GPP network allocates each default bearer a unique /64 prefix, and uses layer-2 signaling to suggest user equipment an Interface Identifier that is guaranteed not to conflict with gateway’s Interface Identifier. The UE may configure link local address using this Interface Identifier, but is allowed to use also other Interface Identifiers and as many globally scoped addresses as it needs. There is no restriction, for example, of using Privacy Extension for SLAAC [RFC4941] or other similar types of mechanisms.

In the 3GPP link model the /64 prefix assigned to the UE is always

off-link (i.e. the L-bit in the Prefix Information Option (PIO) in the RA must be set to zero). If the advertised prefix is used for SLAAC then the A-bit in the PIO must be set to one. The details of the 3GPP link-model and address configuration is described in Section 11.2.1.3.2a of [3GPP.29.061].

The current 3GPP architecture limits number of prefixes in each bearer to a single /64 prefix. Therefore, multi-homing within a single bearer is not possible. Renumbering without closing layer-2 connection is also not possible. The lifetime of /64 prefix is bound to lifetime of layer-2 connection even if the advertised prefix lifetime would be longer than the layer-2 connection lifetime.

5.3. Prefix Delegation

IPv6 prefix delegation is a part of Release-10 and is not covered by any earlier release. However, the /64 prefix allocated for each default bearer (and to the user equipment) may be shared to local area network by user equipment implementing Neighbor Discovery proxy (ND proxy) [RFC4389] functionality.

Release-10 prefix delegation uses the DHCPv6-based prefix delegation [RFC3633]. The model defined for Release-10 requires aggregatable prefixes, which means the /64 prefix allocated for the default bearer (and to the user equipment) must be part of the shorter delegated prefix. DHCPv6 prefix delegation has an explicit limitation described in Section 12.1 of [RFC3633] that a prefix delegated to a requesting router cannot be used by the delegating router (i.e., the PDN-GW in this case). This implies the shorter 'delegated prefix' cannot be given to the requesting router (i.e. the user equipment) as such but has to be delivered by the delegating router (i.e. the PDN-GW) in such a way the /64 prefix allocated to the default bearer is not part of the 'delegated prefix'. IETF is working on a solution for DHCPv6-based prefix delegation to exclude a specific prefix from the 'delegated prefix' [I-D.ietf-dhc-pd-exclude], which could actually be used to solve the above problem.

6. 3GPP Dual-Stack Approach to IPv6

6.1. 3GPP Networks Prior to Release-8

3GPP standards prior to Release-8 provide IPv6 access for cellular devices with PDP contexts of type IPv6 [3GPP.23.060]. For dual-stack access, a PDP context of type IPv6 is established in parallel to the PDP context of type IPv4, as shown in Figure 5 and Figure 6. For IPv4-only service, connections are created over the PDP context of type IPv4 and for IPv6-only service connections are created over the

PDP context of type IPv6. The two PDP contexts of different type may use the same APN (and the gateway), however, this aspect is not explicitly defined in standards. Therefore, cellular device and gateway implementations from different vendors may have varying support for this functionality.

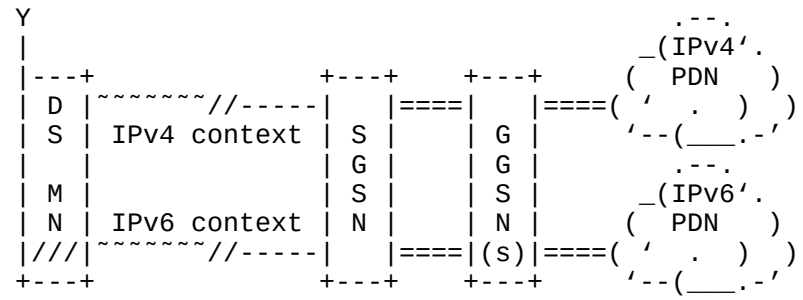


Figure 5: A dual-stack mobile host connecting to both IPv4 and IPv6 Internet using parallel IPv4-only and IPv6-only PDP contexts

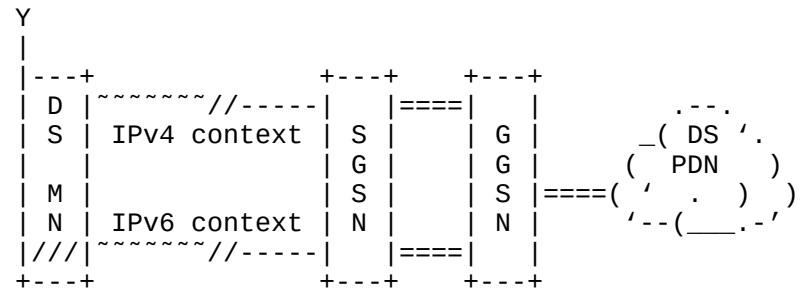


Figure 6: A dual-stack mobile host connecting to dual-stack Internet using parallel IPv4-only and IPv6-only PDP contexts

The approach of having parallel IPv4 and IPv6 type of PDP contexts open is not optimal, because two PDP contexts require double the signaling and consume more network resources than a single PDP context. However, these costs and complexities are lesser than what other transition solutions would incur. In the figure above the IPv4 and IPv6 PDP contexts are attached to the same GGSN. While this is possible, the DS MS may be attached to different GGSNs in the scenario where one GGSN supports IPv4 PDN connectivity while another GGSN provides IPv6 PDN connectivity.

6.2. 3GPP Release-8 and -9 Networks

Since 3GPP Release-8, the powerful concept of a dual-stack type of PDN connection and EPS bearer have been introduced [3GPP.23.401]. This enables parallel use of both IPv4 and IPv6 on a single bearer (IPv4v6), as illustrated in Figure 7, and makes dual stack simpler than in earlier 3GPP releases. As of Release-9, GPRS network nodes also support dual-stack type (IPv4v6) PDP contexts.

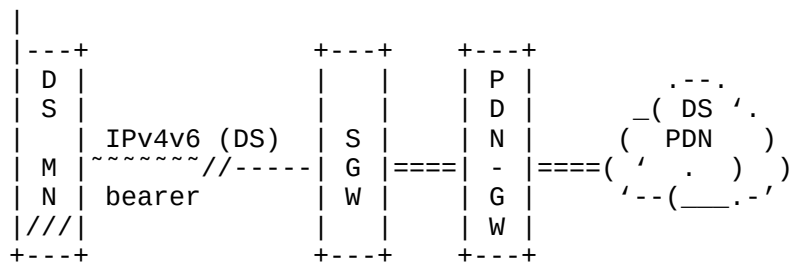


Figure 7: A dual-stack mobile host connecting to dual-stack Internet using a single IPv4v6 type PDN connection

The following is a description of the various PDP contexts/PDN bearer types that are specified by 3GPP:

1. For 2G/3G access to GPRS core (SGSN/GGSN) pre-Release-9 there are two IP PDP Types, IPv4 and IPv6. Two PDP contexts are needed to get dual stack connectivity.

2. For 2G/3G access to GPRS core (SGSN/GGSN) from Release-9 there are three IP PDP Types, IPv4, IPv6 and IPv4v6. Minimum one PDP context is needed to get dual stack connectivity.

3. For 2G/3G access to EPC core (PDN-GW via S4 Release-8 SGSN) from Release-8 there are three IP PDP Types, IPv4, IPv6 and IPv4v6 which gets mapped to PDN Connection type. Minimum one PDP Context is needed to get dual stack connectivity.

4. For LTE (E-UTRAN) access to EPC core from Release-8 there are three IP PDN Types, IPv4, IPv6 and IPv4v6. Minimum one PDN Connection is needed to get dual stack connectivity.

6.3. PDN Connection Establishment Process

The PDN connection establishment process is specified in detail in 3GPP specifications. Figure 8 illustrates the high level process and signaling involved in the establishment of a PDN connection.

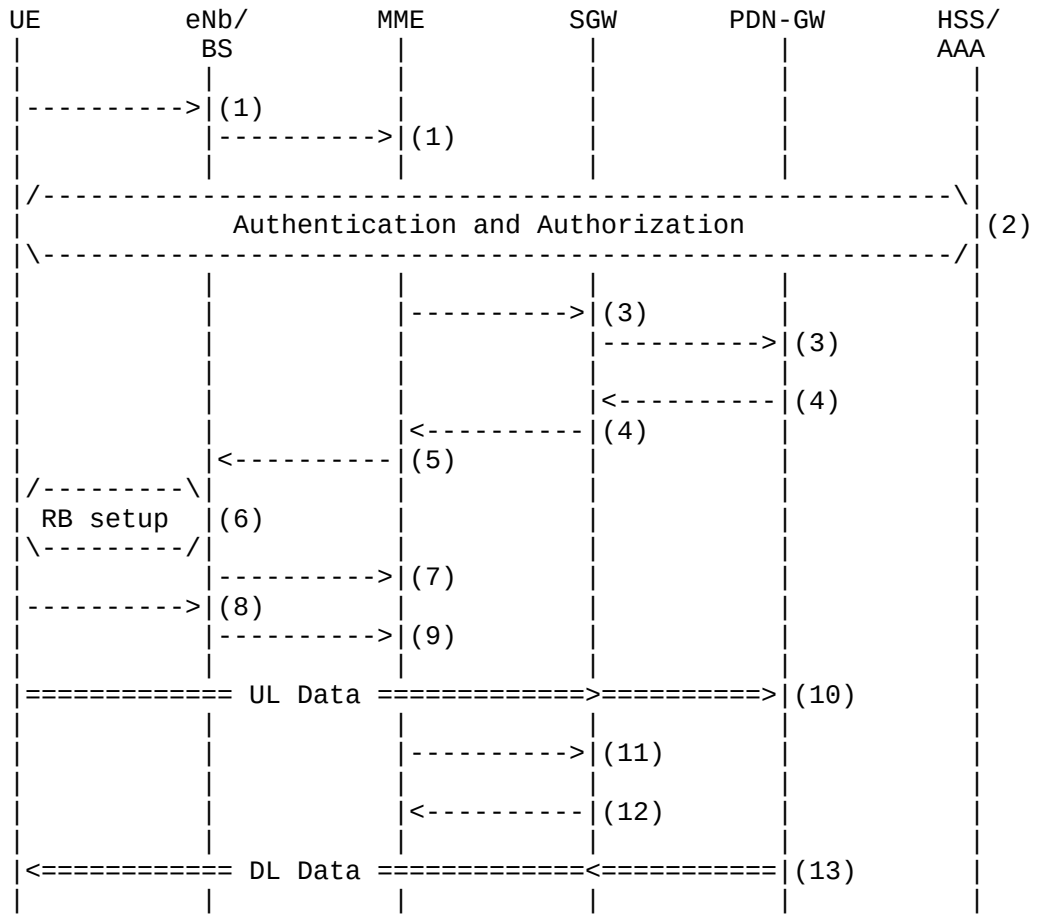


Figure 8: Simplified PDN connection setup procedure in Release-8

1. The UE (i.e the MS) requires a data connection and hence decides to establish a PDN connection with a PDN-GW. The UE sends an "Attach Request" (layer-2) to the BS. The BS forwards this attach request to the MME.

2. Authentication of the UE with the AAA server/HSS follows. If the UE is authorized for establishing a data connection, the following steps continue

3. The MME sends a "Create Session Request" message to the Serving-GW. The SGW forwards the create session request to the PDN-GW. The SGW knows the address of the PDN-GW to forward the create session request to as a result of this information having been obtained by the MME during the authentication/authorization

phase.

The UE IPv4 address and/or IPv6 prefix get assigned during this step. If a subscribed IPv4 address and/or IPv6 prefix is statically allocated for the UE for this APN, then the MME already passes the address information to the SGW and eventually to the PDN-GW in the "Create Session Request" message. Otherwise, the PDN-GW manages the address assignment to the UE (there is another variation to this where IPv4 address allocation is delayed until the UE initiates a DHCPv4 exchange but this is not discussed here).

4. The PDN-GW creates a PDN connection for the UE and sends "Create Session Response" message to the SGW from which the session request message was received from. The SGW forwards the response to the corresponding MME which originated the request.

5. The MME sends the "Attach Accept/Initial Context Setup request" message to the eNodeB/BS.

6. The radio bearer between the UE and the eNb is reconfigured based on the parameters received from the MME

7. The eNb sends "Initial Context Response" message to the MME.

8. The UE sends a "Direct Transfer" message to the eNodeB which includes the Attach complete signal.

9. The eNodeB forwards the Attach complete message to the MME.

10. The UE can now start sending uplink packets to the PDN GW.

11. The MME sends a "Modify Bearer Request" message to the SGW.

12. The SGW responds with a "Modify Bearer Response" message. At this time the downlink connection is also ready

13. The UE can now start receiving downlink packets

The type of PDN connection established between the UE and the PDN-GW can be any of the types described in the previous section. The DS PDN connection, i.e the one which supports both IPv4 and IPv6 packets is the default one that will be established if no specific PDN connection type is specified by the UE in Release-8 networks.

6.4. Mobility of 3GPP IPv4v6 Type of Bearers

3GPP discussed at length various approaches to support mobility between Release-8 and pre-Release-8 networks for the new dual-stack type of bearers.

The chosen approach for mobility is as follows, in short: if a mobile is known to be at risk for doing handovers between Release-8 and pre-Release-8 networks, only single stack bearers are used. Essentially meaning:

1. If a network knows a mobile may do handovers between Release-8 and pre-Release-8 networks (segment), network will only provide single stack bearers, even if the mobile host requests dual-stack bearers. This can happen e.g. if an operator is using pre-Release-8 SGSNs in some parts of the network. The single stack bearers of Release-8 are easy to map one-to-one to pre-Release-8 bearers.

2. If a network knows a mobile will not be able to do handover to pre-Release-8 network (segment), it will provide mobile with dual-stack bearers on request. This can happen e.g. if an operator has upgraded their SGSNs to support dual-stack bearers, or if an operator is running LTE-only network.

The operators should upgrade their, and also if possible roaming partners', networks to Release-8 level in order to support new dual-stack type of bearers. A Release-8 mobile device always requests for a dual-stack bearer, but accepts what is assigned by the network.

7. Dual-Stack Approach to IPv6 Transition in 3GPP Networks

3GPP networks can natively transport IPv4 and IPv6 packets between the mobile station/UE and the gateway (GGSN or PDN-GW) as a result of

establishing either a dual-stack PDP context or

parallel IPv4 and IPv6 PDP contexts.

Current deployments of 3GPP networks primarily support IPv4 only. These networks can be upgraded to also support IPv6 PDP contexts. By doing so devices and applications that are IPv6 capable can start utilizing the IPv6 connectivity. This will also ensure that legacy devices and applications continue to work with no impact. As newer devices start using IPv6 connectivity, the number of IPv4 addresses in use is expected to slowly decrease, providing operators with a smooth transition to IPv6 With a dual-stack approach, there is always

the potential to fallback to IPv4. A device which may be roaming in a network wherein IPv6 is not supported by the visited network would fall back to using IPv4 PDP contexts and hence the end user does not see an interruption to the services.

As the networks evolve to support Release-8 EPS architecture and the dual-stack PDP contexts, newer devices will be able to leverage such capability and have a single bearer which supports both IPv4 and IPv6. Since IPv4 and IPv6 packets are carried as payload within GTP between the MS and the gateway (GGSN/PDN-GW) the transport network capability in terms of whether it supports IPv4 or IPv6 on the interfaces between the eNodeB and SGW or, SGW and PDN-GW is immaterial.

The dual-stack approach enables a systematic migration path to IPv6. From an operational standpoint operators are concerned about ensuring that there is no disruption to the connectivity that subscribers rely on. This can be achieved by upgrading the network to support IPv6 while continuing to maintain IPv4 legacy. Dual-stack capability in the network and devices for the foreseeable future at least is a pragmatic solution.

8. Deployment issues

8.1. Overlapping IPv4 Addresses

Given the shortage of globally routable public IPv4 addresses, operators tend to assign private IPv4 addresses [RFC1918] to hosts when they establish an IPv4 only PDP context or an IPv4v6 type PDN context. About 16 million hosts can be assigned a private IPv4 address that is unique within a domain. However, in case of many operators the number of subscribers is greater than 16 million. The issue can be dealt with by assigning overlapping RFC 1918 IPv4 addresses to hosts. As a result the IPv4 address assigned to a host within the context of a single operator realm would no longer be unique. This has the obvious and know issues of NATed IP connection in the Internet. Direct host to host connectivity becomes complicated, unless the hosts are within the same private address range pool and/or anchored to the same gateway, referrals using IP addresses will have issues and so forth. However, these are generic issues and not only a concern of the EPS. In general this is not seen as a major issue in the EPS for the following reasons:

1. Very large network deployments are partitioned, for example, based on a geographical areas. This partitioning allows overlapping IPv4 addresses ranges to be assigned to hosts that are in different areas. Each area has its own pool of gateways

that are dedicated for a certain overlapping IPv4 address range (referred here later as a zone). Standard NAT44 functionality enables the communication between hosts that are assigned the same IPv4 address but belong to different zones, yet are part of the same operator domain.

2. A mobile host/device attaches to a gateway as part of the attach process. The number of hosts that a gateway supports is in the order of 1 to 10 million. Hence all the hosts assigned to a single gateway can be assigned private IPv4 addresses. Operators with large subscriber bases have multiple gateways and hence the same [RFC1918] IPv4 address space can be reused across gateways. The IPv4 address assigned to a host is unique within the scope of a single gateway.

3. The IPv4 address assigned to a host could also be made irrelevant from a routing perspective at least by the use of protocol solutions such as GI-DSLite [I-D.ietf-softwire-gateway-init-ds-lite]. This requires a Large Scale NAT (LSN) entity that is detached from the gateway (GGSN or PDN-GW). Multiple gateways in an operator domain would attach to a LSN in such an approach and the hosts across these gateways can be assigned overlapping IPv4 addresses.

4. New services requiring direct connectivity between hosts should be build on IPv6. Possible existing IPv4-only services and applications requiring direct connectivity can be ported to IPv6.

8.2. IPv6 for transport

The various reference points of the 3GPP architecture such as S1-U, S5 and S8 are based on either GTP or PMIPv6. The underlying transport for these reference points can be IPv4 or IPv6. GTP has been able to operate over IPv6 transport (optionally) since R99 and PMIPv6 has supported IPv6 transport starting from its introduction in Release-8. The user plane traffic between the mobile host and the gateway can use either IPv4 or IPv6. These packets are essentially treated as payload by GTP/PMIPv6 and transported accordingly with no real attention paid to the information (at least from a routing perspective) contained in the IPv4 or IPv6 headers. The transport links between the eNodeB and the SGW, and the link between the SGW and PDN-GW can be migrated to IPv6 without any direct implications to the architecture.

Currently, the inter-operator (for 3GPP technology) roaming networks are all IPv4 only (see Inter-PLMN Backbone Guidelines [GSMA.IR.34]). Eventually these roaming networks will also get migrated to IPv6, if there is a business reason for that. The migration period can be

prolonged considerably because the 3GPP protocols always tunnel user plane traffic in the core network and as described earlier the transport network IP version is not in any way tied to user plane IP version. Furthermore, the design of the inter-operator roaming networks is such that the user plane and transport network IP addressing is completely separated from each other. The inter-operator roaming network itself is also completely separated from the Internet. Only those core network nodes that must be connected to the inter-operator roaming networks are actually visible there, and be able to send and receive (tunneled) traffic within the inter-operator roaming networks. Obviously, in order the roaming to work properly, the operators have to agree on supported protocol versions so that the visited network does not, for example, unnecessarily drop user plane IPv6 traffic.

8.3. Operational Aspects of Running Dual-Stack Networks

Operating dual-stack networks does imply cost and complexity to a certain extent. However these factors are mitigated by the assurance that legacy devices and services are unaffected and there is always a fallback to IPv4 in case of issues with the IPv6 deployment or network elements. The model also enables operators to develop operational experience and expertise in an incremental manner.

Running dual-stack networks requires the management of multiple IP address spaces. Tracking of hosts needs to be expanded since it can be identified by either an IPv4 address or IPv6 prefix. Network elements will also need to be dual-stack capable in order to support the dual-stack deployment model.

Deployment and migration cases described in Section 6.1 for providing dual-stack like capability may mean doubled resource usage in operator’s network. Also handovers between networks with different capabilities in terms of networks being dual-stack like service capable or not, may turn out hard to comprehend for users and for application/services to cope with. These facts may add other than just technical concerns for operators when planning to roll out dual-stack service offerings.

9. IANA Considerations

This document has no requests to IANA.

10. Security Considerations

This document does not introduce any security related concerns.

11. Summary and Conclusion

The 3GPP network architecture and specifications enable the establishment of IPv4 and IPv6 connections through the use of appropriate PDP context types. The current generation of deployed networks can support dual-stack connectivity if the packet core network elements such as the SGSN and GGSN have the capability. With Release-8, 3GPP has specified a more optimal PDP context type which enables the transport of IPv4 and IPv6 packets within a single PDP context between the mobile station and the gateway.

The authors believe that transitioning to IPv6 in 3GPP networks can be achieved without disruption to legacy devices, networks and services only by taking a dual-stack approach to deployment. As devices and applications are upgraded to support IPv6 they can start leveraging the IPv6 connectivity provided by the networks while maintaining the fallback to IPv4 capability. Enabling IPv6 connectivity in the 3GPP networks by itself will provide some degree of relief to the IPv4 address space as many of the applications and services can start to work over IPv6 right away. However without comprehensive testing of different applications and solutions that exist today and are widely used, for their ability to operate over IPv6 PDN connections, an IPv6 only access would cause disruptions. Hence we recommend adopting the dual-stack approach to IPv6 transition in 3GPP networks.

12. Acknowledgements

The authors thank Shabnam Sultana, Sri Gundavelli, Hui Deng, and Zhenqiang Li for their reviews and comments on this document.

13. Informative References

[3GPP.23.060]
3GPP, "General Packet Radio Service (GPRS); Service description; Stage 2", 3GPP TS 23.060 8.8.0, March 2010.

[3GPP.23.401]
3GPP, "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access", 3GPP TS 23.401 10.1.0, September 2010.

[3GPP.23.975]
3GPP, "IPv6 Migration Guidelines", 3GPP TR 23.975 1.1.1, June 2010.

[3GPP.29.060]
3GPP, "General Packet Radio Service (GPRS); GPRS Tunnelling Protocol (GTP) across the Gn and Gp interface", 3GPP TS 29.060 8.11.0, April 2010.

[3GPP.29.061]
3GPP, "Interworking between the Public Land Mobile Network (PLMN) supporting packet based services and Packet Data Networks (PDN)", 3GPP TS 29.061 8.5.0, April 2010.

[GSMA.IR.34]
GSMA, "Inter-PLMN Backbone Guidelines", GSMA PRD IR.34.4.9, March 2010.

[I-D.ietf-dhc-pd-exclude]
Korhonen, J., Savolainen, T., Krishnan, S., and O. Troan, "Prefix Exclude Option for DHCPv6-based Prefix Delegation", draft-ietf-dhc-pd-exclude-00 (work in progress), October 2010.

[I-D.ietf-softwire-gateway-init-ds-lite]
Brockners, F., Gundavelli, S., Speicher, S., and D. Ward, "Gateway Initiated Dual-Stack Lite Deployment", draft-ietf-softwire-gateway-init-ds-lite-01 (work in progress), October 2010.

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

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

[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.

[RFC4389] Thaler, D., Talwar, M., and C. Patel, "Neighbor Discovery Proxies (ND Proxy)", RFC 4389, April 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.

[RFC5213] Gundavelli, S., Leung, K., Devarapalli, V., Chowdhury, K., and B. Patil, "Proxy Mobile IPv6", RFC 5213, August 2008.

Authors' Addresses

Jouni Korhonen (editor)
Nokia Siemens Networks
Linnoitustie 6
FI-02600 Espoo
FINLAND

Email: jouni.nospam@gmail.com

Jonne Soininen
Nokia Siemens Networks
Linnoitustie 6
FI-02600 Espoo
FINLAND

Email: jonne.soininen@nsn.com

Basavaraj Patil
Nokia
6021 Connection drive
Irving, TX 75019
USA

Email: basavaraj.patil@nokia.com

Teemu Savolainen
Nokia
Hermiankatu 12 D
FI-33720 Tampere
FINLAND

Email: teemu.savolainen@nokia.com

Gabor Bajko
Nokia
323 Fairchild drive 6
Mountain view, CA 94043
USA

Email: gabor.bajko@nokia.com

Kaisu Iisakkila
Nokia
Itamerenkatu 11-13
FI-00180 Helsinki
FINLAND

Email: kaisu.iisakkila@nokia.com

Individual Submission
Internet-Draft
Intended status: Informational
Expires: August 14, 2011

J. Korhonen, Ed.
Nokia Siemens Networks
J. Soininen
Renesas Mobile
B. Patil
T. Savolainen
G. Bajko
Nokia
K. Iisakkila
Renesas Mobile

February 10, 2011

IPv6 in 3GPP Evolved Packet System
draft-korhonen-v6ops-3gpp-eps-06

Abstract

Internet connectivity and use of data services in 3GPP based mobile networks has increased rapidly as a result of smart phones, broadband service via HSPA and HSPA+ networks, competitive service offerings by operators and a large number of applications. Operators who have deployed networks based on 3GPP architectures are facing IPv4 address shortages. With the impending exhaustion of available IPv4 addresses from the registries there is an increased emphasis for operators to migrate to IPv6. This document describes the support for IPv6 in 3GPP network architectures.

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 14, 2011.

Copyright Notice

Table of Contents

1. Introduction

IPv6 has been specified in the 3rd Generation Partnership Project (3GPP) standards since the early architectures developed for R99 General Packet Radio Service (GPRS). However, the support for IPv6 in commercially deployed networks by the end of 2010 is nearly non-existent. There are many factors that can be attributed to the lack of IPv6 deployment in 3GPP networks. The most relevant one is essentially the same as the reason for IPv6 not being deployed by other networks as well, i.e. the lack of business and commercial incentives for deployment. 3GPP network architectures have also evolved since 1999 (since R99). The most recent version of the 3GPP architecture, the Evolved Packet System (EPS), which is commonly referred to as SAE, LTE or Release-8, is a packet centric architecture. The number of subscribers and devices that are using the 3GPP networks for Internet connectivity and data services has also increased significantly. With the subscriber growth numbers projected to increase even further and the IPv4 addresses depletion problem looming in the near term, 3GPP operators and vendors have started the process of identifying the scenarios and solutions needed to transition to IPv6.

This document describes the establishment of IP connectivity in 3GPP network architectures, specifically in the context of IP bearers for 3GPP GPRS and for 3GPP EPS. It provides an overview of how IPv6 is supported as per the current set of 3GPP specifications. Some of the issues and concerns with respect to deployment and shortage of private IPv4 addresses within a single network domain are also discussed.

The IETF has specified a set of tools and mechanisms that can be utilized for transitioning to IPv6. In addition to operating dual-stack networks during the transition from IPv4 to IPv6 phase, the two alternative categories for the transition are encapsulation and translation. Most of the mechanisms available in the toolbox can be categorized into either translation or encapsulation approaches. The IETF continues to specify additional solutions for enabling the transition based on the deployment scenarios and operator/ISP requirements. There is no single approach for transition to IPv6 that can meet the needs for all deployments and models. The 3GPP scenarios for transition, described in [3GPP.23.975], can be addressed using transition mechanisms that are already available in the toolbox. The objective of transition to IPv6 in 3GPP networks is to ensure that:

1. Legacy devices and hosts which have an IPv4 only stack will continue to be provided with IP connectivity to the Internet and services,

2. Devices which are dual-stack can access the Internet either via IPv6 or IPv4. The choice of using IPv6 or IPv4 depends on the capability of:

   A. the application on the host,

   B. the support for IPv4 and IPv6 bearers by the network and/or,

   C. the capability of the server(s) and other end points.

3GPP networks are capable of providing a host with IPv4 and IPv6 connectivity today, albeit in many cases with upgrades to network elements such as the SGSN and GGSN.

2. 3GPP Terminology and Concepts

2.1. Terminology

Access Point Name

Access Point Name (APN) is a fully qualified domain name and resolves to a specific gateway in an operators network. The APNs are piggybacked on the administration of the DNS namespace.

Packet Data Protocol Context

A Packet Data Protocol (PDP) Context is the equivalent of a virtual connection between the host and a gateway.

General Packet Radio Service

General Packet Radio Service (GPRS) is a packet oriented mobile data service available to users of the 2G and 3G cellular communication systems Global System for Mobile communications (GSM), and specified by 3GPP.

Packet Data Network

Packet Data Network (PDN) is a packet based network that either belongs to the operator or is an external network such as Internet and corporate intranet. The user eventually accesses services in one or more PDNs. The operator’s packet domain network are separated from packet data networks either by GGSNs or PDN Gateways (PDN-GW).

Gateway GPRS Support Node

Gateway GPRS Support Node (GGSN) is a gateway function in GPRS, which provides connectivity to Internet or other PDNs. The host attaches to a GGSN identified by an APN assigned to it by an operator. The GGSN also serves as the topological anchor for addresses/prefixes assigned to the mobile host.

Packet Data Network Gateway

Packet Data Network Gateway (PDN-GW) is a gateway function in Evolved Packet System (EPS), which provides connectivity to Internet or other PDNs. The host attaches to a PDN-GW identified by an APN assigned to it by an operator. The PDN-GW also serves as the topological anchor for addresses/prefixes assigned to the mobile host.

Serving Gateway

Serving Gateway (SGW) is a gateway function in EPS, which terminates the interface towards E-UTRAN. The SGW is the Mobility Anchor point for layer-2 mobility (inter-eNodeB handovers). For each User Equipment connected with the EPS, at any given point of time, there is only one SGW. The SGW is essentially the user plane part of the GPRS' SGSN forwarding packets between a PDN-GW.

Serving Gateway Support Node

Serving Gateway Support Node (SGSN) is a network element that is located between the radio access network (RAN) and the gateway (GGSN). A per mobile host point to point (p2p) tunnel between the GGSN and SGSN transports the packets between the mobile host and the gateway.

GPRS tunnelling protocol

GPRS Tunnelling Protocol (GTP) [3GPP.29.060] [3GPP.29.274] is a tunnelling protocol defined by 3GPP. It is a network based mobility protocol and similar to Proxy Mobile IPv6 (PMIPv6) [RFC5213]. However, GTP also provides functionality beyond mobility such as inband signaling related to Quality of Service (QoS) and charging among others.

Evolved Packet System

Evolved Packet System (EPS) is an evolution of the 3GPP GPRS system characterized by higher-data-rate, lower-latency, packet-optimized system that supports multiple Radio Access Technologies

(RAT). The EPS comprises the Evolved Packet Core (EPC) together with the evolved radio access network (E-UTRA and E-UTRAN).

Mobility Management Entity

Mobility Management Entity (MME) is a network element that is responsible for control plane functionalities, including authentication, authorization, bearer management, layer-2 mobility, etc. The MME is essentially the control plane part of the GPRS' SGSN and not located on the user plane data path, i.e. user plane traffic bypasses the MME.

UMTS Terrestrial Radio Access Network

UMTS Terrestrial Radio Access Network (UTRAN) is communications network, commonly referred to as 3G, and consists of NodeBs (3G base station) and Radio Network Controllers (RNC) which make up the UMTS radio access network. The UTRAN allows connectivity between the mobile host/device and the core network. UTRAN comprises of WCDMA, HSPA and HSPA+ radio technologies.

Wideband Code Division Multiple Access

The Wideband Code Division Multiple Access (WCDMA) is the radio interface used in UMTS networks.

High Speed Packet Access

The High Speed Packet Access (HSPA) and the Evolved High Speed Packet Access (HSPA+) are enhanced versions of the WCDMA and UTRAN, thus providing more data throughput and lower latencies.

Evolved UTRAN

Evolved UTRAN (E-UTRAN) is communications network, sometimes referred to as 4G, and consists of eNodeBs (4G base station) which make up the E-UTRAN radio access network. The E-UTRAN allows connectivity between the mobile host/device and the core network.

eNodeB

The eNodeB is a base station entity that supports the Long Term Evolution (LTE) air interface.

GSM EDGE Radio Access Network

GSM EDGE Radio Access Network (GERAN) is communications network, commonly referred to as 2G or 2.5G, and consists of base stations

and Base Station Controllers (BSC) which make up the GSM EDGE radio access network. The GERAN allows connectivity between the mobile host/device and the core network.

UE, MS, MN and Mobile

The terms UE (User Equipment), MS (Mobile Station), MN (Mobile Node) and, mobile refer to the devices which are hosts with ability to obtain Internet connectivity via a 3GPP network. The terms UE, MS, MN and devices are used interchangeably within this document.

PCC

The Policy and Charging Control (PCC) framework is used for QoS policy and charging control. It is optional for 3GPP EPS but needed if dynamic policy and charging control by means of PCC rules based on user and services are desired.

HLR

The Home Location Register (HLR) is a pre-Release-5 database (the reality regarding releases is different, though) for a given subscriber. It is the entity containing the subscription-related information to support the network entities actually handling calls/sessions.

HSS

The Home Subscriber Server (HSS) is a database for a given subscriber and got introduced in 3GPP Release-5. It is the entity containing the subscription-related information to support the network entities actually handling calls/sessions.

2.2. The concept of APN

The Access Point Name (APN) essentially refers to a gateway in the 3GPP network. The 'complete' APN is expressed in a form of a Fully Qualified Domain Name (FQDN) and also piggybacked on the administration of the DNS namespace, thus effectively allowing the discovery of gateways using the DNS. Mobile hosts/devices can choose to attach to a specific gateway in the packet core. The gateway provides connectivity to the Packet Data Network (PDN) such as the Internet. An operator may also include gateways which do not provide Internet connectivity, rather a connectivity to closed network providing a set of operator's own services. A mobile host/device can be attached to one or more gateways simultaneously. The gateway in a 3GPP network is the GGSN or PDN-GW. Figure 1 below illustrates the

APN-based network connectivity concept.

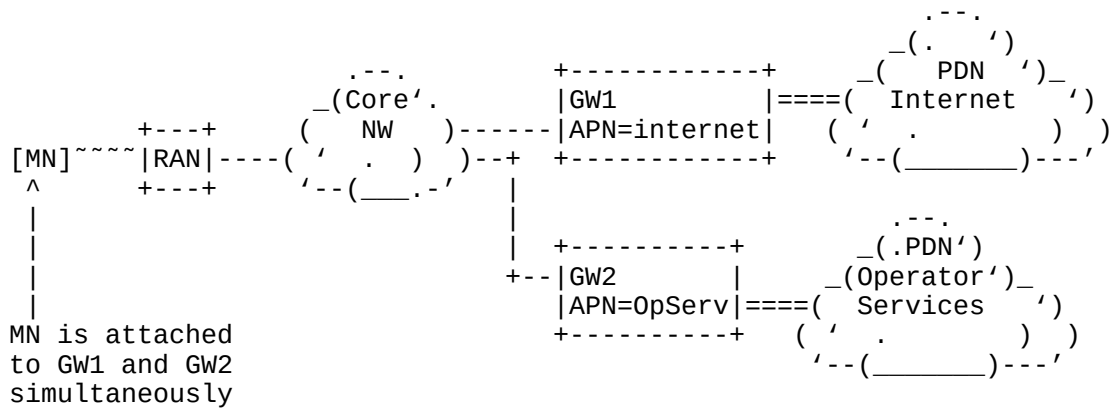


Figure 1: Mobile host/device attached to multiple APNs simultaneously

3. IP over 3GPP GPRS

3.1. Introduction to 3GPP GPRS

A simplified 2G/3G GPRS architecture is illustrated in Figure 2. This architecture basically covers the GPRS core network since R99 to Release-7, and radio access technologies such as GSM (2G), EDGE (2G, ofter referred as 2.5G), WCDMA (3G) and HSPA(+) (3G, often referred as 3.5G). The architecture shares obvious similarities with the Evolved Packet System (EPS) as will be seen in Section 4. Based on Gn/Gp interfaces, the GPRS core network functionality is logically implemented on two network nodes, the SGSN and the GGSN.

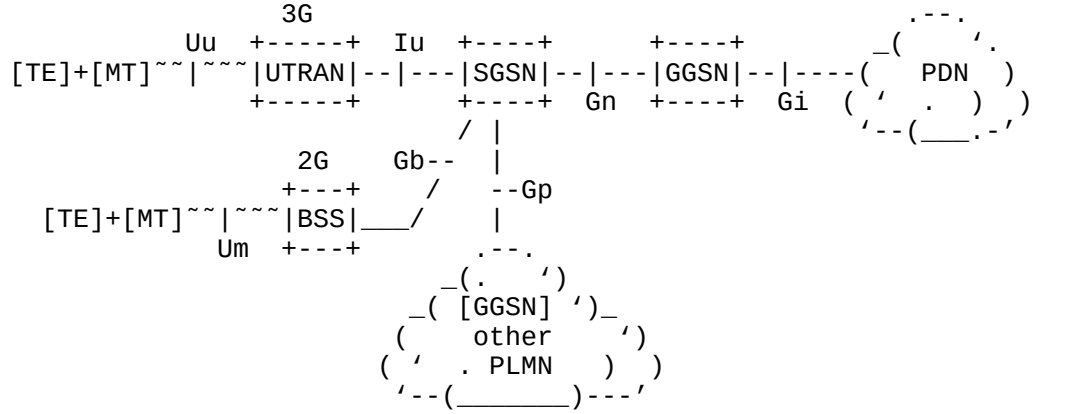


Figure 2: Overview of the 2G/3G GPRS Logical Architecture

Gn/Gp: These interfaces provide a network based mobility service for a mobile host and are used between a SGSN and a GGSN. The Gn interface is used when GGSN and SGSN are located inside one operator (i.e. PLMN). The Gp-interface is used if the GGSN and the SGSN are located in different operator domains (i.e. 'other' PLMN). GTP protocol is defined for the Gn/Gp interfaces (both GTP-C for the control plane and GTP-U for the user plane).

Gb: Is the Base Station System (BSS) to SGSN interface, which is used to carry information concerning packet data transmission and layer-2 mobility management. The Gb-interface is based on either on Frame Relay or IP.

Iu: Is the Radio Network System (RNS) to SGSN interface, which is used to carry information concerning packet data transmission and layer-2 mobility management. The user plane part of the Iu-interface (actually the Iu-PS) is based on GTP-U. The control plane part of the Iu-interface is based on Radio Access Network Application Protocol (RANAP).

Gi: It is the interface between the GGSN and a PDN. The PDN may be an operator external public or private packet data network or an intra-operator packet data network.

Uu/Um: Are either 2G or 3G radio interfaces between a mobile terminal and a respective radio access network.

The SGSN is responsible for the delivery of data packets from and to the mobile hosts within its geographical service area when a direct tunnel option is not used. If the direct tunnel is used, then the user plane goes directly between the RNS and the GGSN. The control plane traffic always goes through the SGSN. For each mobile host connected with the GPRS, at any given point of time, there is only one SGSN.

3.2. PDP Context

A PDP context is an association between a mobile host represented by one IPv4 address and/or one /64 IPv6 prefix and a PDN represented by an APN. Each PDN can be accessed via a gateway (typically a GGSN or PDN-GW). On the device/mobile host a PDP context is equivalent to a network interface. A host may hence be attached to one or more gateways via separate connections, i.e. PDP contexts. Each primary PDP context has its own IPv4 address and/or one /64 IPv6 prefix assigned to it by the PDN and anchored in the corresponding gateway.

Applications on the host use the appropriate network interface (PDP context) for connectivity to a specific PDN. Figure 3 represents a high level view of what a PDP context implies in 3GPP networks.

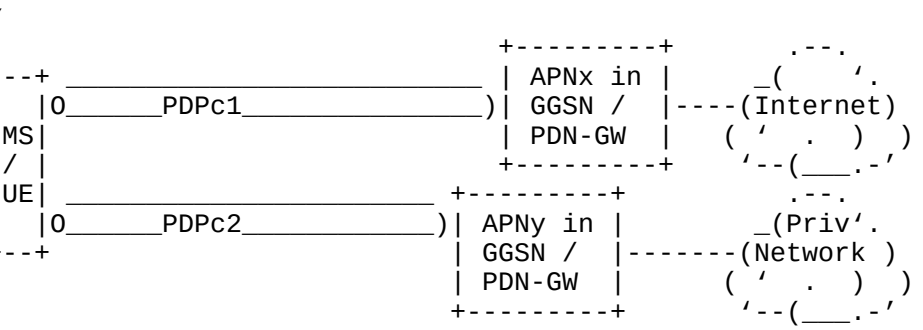


Figure 3: PDP contexts between the MS/UE and gateway

In the above figure there are two PDP contexts at the MS/UE (UE=User Equipment in 3GPP parlance). The 'PDPc1' PDP context that is connected to APNx provided Internet connectivity and the 'PDPc2' PDP context provides connectivity to a private IP network via APNy (as an example this network may include operator specific services such as MMS (Multi media service). An application on the host such as a web browser would use the PDP context that provides Internet connectivity for accessing services on the Internet. An application such as MMS would use APNy in the figure above because the service is provided through the private network.

4. IP over 3GPP EPS

4.1. Introduction to 3GPP EPS

In its most basic form, the EPS architecture consists of only two nodes on the user plane, a base station and a core network Gateway (GW). The basic EPS architecture is illustrated in Figure 4. The Mobility Management Entity (MME) node performs control-plane functionality and is separated from the node(s) that performs bearer-plane functionality (GW), with a well-defined open interface between them (S11). The optional interface S5 can be used to split the Gateway (GW) into two separate nodes, the Serving Gateway (SGW) and the PDN-GW. This allows independent scaling and growth of traffic throughput and control signal processing. The functional split of gateways also allows for operators to choose optimized topological locations of nodes within the network and enables various deployment models including the sharing of radio networks between different operators.

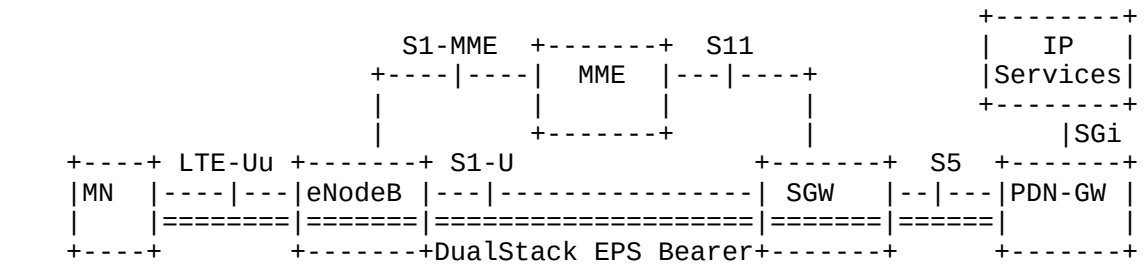


Figure 4: EPS Architecture for 3GPP Access

S5: It provides user plane tunnelling and tunnel management between SGW and PDN-GW, using GTP or PMIPv6 as the network based mobility management protocol.

S1-U: Provides user plane tunnelling and inter eNodeB path switching during handover between eNodeB and SGW, using the GTP-U protocol (GTP user plane).

S1-MME: Reference point for the control plane protocol between eNodeB and MME.

SGi: It is the interface between the PDN-GW and the packet data network. Packet data network may be an operator external public or private packet data network or an intra operator packet data network.

The eNodeB is a base station entity that supports the Long Term Evolution (LTE) air interface and includes functions for radio resource control, user plane ciphering, and other lower layer functions. MME is responsible for control plane functionalities, including authentication, authorization, bearer management, layer-2 mobility, etc.

The SGW is the Mobility Anchor point for layer-2 mobility. For each MN connected with the EPS, at any given point of time, there is only one SGW.

4.2. PDN Connection

A PDN connection is an association between a mobile host represented by one IPv4 address and/or one /64 IPv6 prefix, and a PDN represented by an APN. The PDN connection is the EPC equivalent of the GPRS PDP context. Each PDN can be accessed via a gateway (a PDN-GW). PDN is responsible for the IP address/prefix allocation to the mobile host. On the device/mobile host a PDN connection is equivalent to a network interface. A host may hence be attached to one or more gateways via

separate connections, i.e. PDN connections. Each PDN connection has its own IP address/prefix assigned to it by the PDN and anchored in the corresponding gateway. Applications on the host use the appropriate network interface (PDN connection) for connectivity.

4.3. EPS bearer model

The logical concept of a bearer has been defined to be an aggregate of one or more IP flows related to one or more services. An EPS bearer exists between the Mobile Node (MN i.e. a mobile host) and the PDN-GW and is used to provide the same level of packet forwarding treatment to the aggregated IP flows constituting the bearer. Services with IP flows requiring a different packet forwarding treatment would therefore require more than one EPS bearer. The mobile host performs the binding of the uplink IP flows to the bearer while the PDN-GW performs this function for the downlink packets.

In order to provide low latency for always on connectivity, a default bearer will be provided at the time of startup and an IPv4 address and/or IPv6 prefix gets assigned to the mobile host (this is different from GPRS, where mobile hosts are not automatically assigned with an IP address or prefix). This default bearer will be allowed to carry all traffic which is not associated with a dedicated bearer. Dedicated bearers are used to carry traffic for IP flows that have been identified to require a specific packet forwarding treatment. They may be established at the time of startup; for example, in the case of services that require always-on connectivity and better QoS than that provided by the default bearer. The default bearer and the dedicated bearer(s) associated to it share the same IP address(es)/prefix.

An EPS bearer is referred to as a GBR bearer if dedicated network resources related to a Guaranteed Bit Rate (GBR) value that is associated with the EPS bearer are permanently allocated (e.g. by an admission control function in the eNodeB) at bearer establishment/ modification. Otherwise, an EPS bearer is referred to as a non-GBR bearer. The default bearer is always non-GBR, with the resources for the IP flows not guaranteed at eNodeB, and with no admission control. However, the dedicated bearer can be either GBR or non-GBR. A GBR bearer has a Guaranteed Bit Rate (GBR) and Maximum Bit Rate (MBR) while more than one non-GBR bearer belonging to the same UE shares an Aggregate Maximum Bit Rate (AMBR). Non-GBR bearers can suffer packet loss under congestion while GBR bearers are immune to such losses.

5. Address Management

5.1. IPv4 Address Configuration

Mobile host's IPv4 address configuration is always performed during PDP context/EPS bearer setup procedures (on layer-2). DHCPv4-based [RFC2131] address configuration is supported by the 3GPP specifications, but is not used in wide scale. The mobile host must always support layer-2 based address configuration, since DHCPv4 is optional for both mobile hosts and networks.

5.2. IPv6 Address Configuration

IPv6 Stateless Address Autoconfiguration (SLAAC) as specified in [RFC4862] is the only supported address configuration mechanism. Stateful DHCPv6-based address configuration is not supported by 3GPP specifications [RFC3315]. On the other hand, Stateless DHCPv6-service to obtain other configuration information is supported [RFC3736]. This implies that the M-bit must always be set to zero and the O-bit may be set to one in the Router Advertisement (RA) sent to the UE.

3GPP network allocates each default bearer a unique /64 prefix, and uses layer-2 signaling to suggest user equipment an Interface Identifier that is guaranteed not to conflict with gateway's Interface Identifier. The UE may configure link local address using this Interface Identifier, but is allowed to use also other Interface Identifiers and as many globally scoped addresses as it needs. There is no restriction, for example, of using Privacy Extension for SLAAC [RFC4941] or other similar types of mechanisms.

In the 3GPP link model the /64 prefix assigned to the UE is always off-link (i.e. the L-bit in the Prefix Information Option (PIO) in the RA must be set to zero). If the advertised prefix is used for SLAAC then the A-bit in the PIO must be set to one. The details of the 3GPP link-model and address configuration is described in Section 11.2.1.3.2a of [3GPP.29.061]. More specifically, the GGSN/PDN-GW guarantees that the /64 prefix is unique for the mobile host. Therefore, there is no need to perform any Duplicate Address Detection (DAD) on addresses the mobile host creates (i.e., the 'DupAddrDetectTransmits' variable in the mobile host should be zero). The GGSN/PDN-GW is not allowed to generate any globally unique IPv6 addresses for itself using the /64 prefix assigned to the mobile host in the RA.

The current 3GPP architecture limits number of prefixes in each bearer to a single /64 prefix. If the mobile host finds more than one prefix in the RA, it only considers the first one and silently discard the others [3GPP.29.061]. Therefore, multi-homing within a single bearer is not possible. Renumbering without closing layer-2

connection is also not possible. The lifetime of /64 prefix is bound to lifetime of layer-2 connection even if the advertised prefix lifetime would be longer than the layer-2 connection lifetime.

5.3. Prefix Delegation

IPv6 prefix delegation is a part of Release-10 and is not covered by any earlier release. However, the /64 prefix allocated for each default bearer (and to the user equipment) may be shared to local area network by user equipment implementing Neighbor Discovery proxy (ND proxy) [RFC4389] functionality.

Release-10 prefix delegation uses the DHCPv6-based prefix delegation [RFC3633]. The model defined for Release-10 requires aggregatable prefixes, which means the /64 prefix allocated for the default bearer (and to the user equipment) must be part of the shorter delegated prefix. DHCPv6 prefix delegation has an explicit limitation described in Section 12.1 of [RFC3633] that a prefix delegated to a requesting router cannot be used by the delegating router (i.e., the PDN-GW in this case). This implies the shorter 'delegated prefix' cannot be given to the requesting router (i.e. the user equipment) as such but has to be delivered by the delegating router (i.e. the PDN-GW) in such a way the /64 prefix allocated to the default bearer is not part of the 'delegated prefix'. IETF is working on a solution for DHCPv6-based prefix delegation to exclude a specific prefix from the 'delegated prefix' [I-D.ietf-dhc-pd-exclude].

6. 3GPP Dual-Stack Approach to IPv6

6.1. 3GPP Networks Prior to Release-8

3GPP standards prior to Release-8 provide IPv6 access for cellular devices with PDP contexts of type IPv6 [3GPP.23.060]. For dual-stack access, a PDP context of type IPv6 is established in parallel to the PDP context of type IPv4, as shown in Figure 5 and Figure 6. For IPv4-only service, connections are created over the PDP context of type IPv4 and for IPv6-only service connections are created over the PDP context of type IPv6. The two PDP contexts of different type may use the same APN (and the gateway), however, this aspect is not explicitly defined in standards. Therefore, cellular device and gateway implementations from different vendors may have varying support for this functionality.

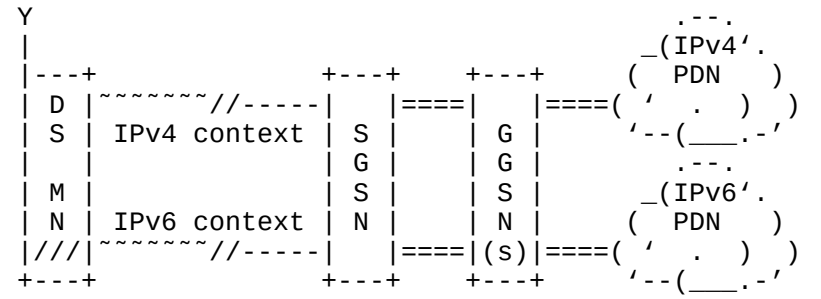


Figure 5: A dual-stack mobile host connecting to both IPv4 and IPv6 Internet using parallel IPv4-only and IPv6-only PDP contexts

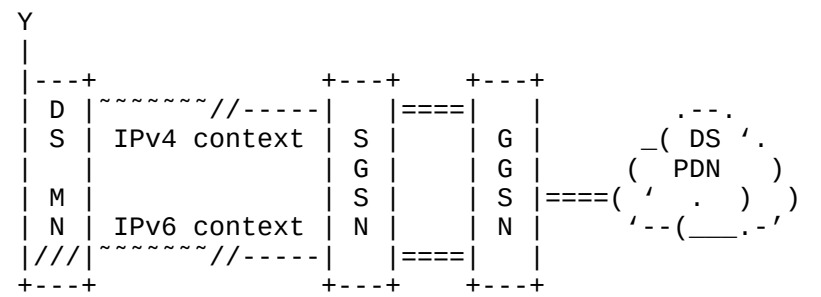


Figure 6: A dual-stack mobile host connecting to dual-stack Internet using parallel IPv4-only and IPv6-only PDP contexts

The approach of having parallel IPv4 and IPv6 type of PDP contexts open is not optimal, because two PDP contexts require double the signaling and consume more network resources than a single PDP context. In the figure above the IPv4 and IPv6 PDP contexts are attached to the same GGSN. While this is possible, the DS MS may be attached to different GGSNs in the scenario where one GGSN supports IPv4 PDN connectivity while another GGSN provides IPv6 PDN connectivity.

6.2. 3GPP Release-8 and -9 Networks

Since 3GPP Release-8, the powerful concept of a dual-stack type of PDN connection and EPS bearer have been introduced [3GPP.23.401]. This enables parallel use of both IPv4 and IPv6 on a single bearer (IPv4v6), as illustrated in Figure 7, and makes dual stack simpler than in earlier 3GPP releases. As of Release-9, GPRS network nodes also support dual-stack type (IPv4v6) PDP contexts.

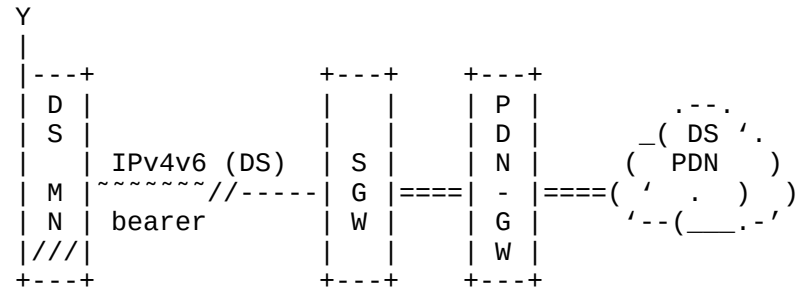


Figure 7: A dual-stack mobile host connecting to dual-stack Internet using a single IPv4v6 type PDN connection

The following is a description of the various PDP contexts/PDN bearer types that are specified by 3GPP:

1. For 2G/3G access to GPRS core (SGSN/GGSN) pre-Release-9 there are two IP PDP Types, IPv4 and IPv6. Two PDP contexts are needed to get dual stack connectivity.

2. For 2G/3G access to GPRS core (SGSN/GGSN) from Release-9 there are three IP PDP Types, IPv4, IPv6 and IPv4v6. Minimum one PDP context is needed to get dual stack connectivity.

3. For 2G/3G access to EPC core (PDN-GW via S4 Release-8 SGSN) from Release-8 there are three IP PDP Types, IPv4, IPv6 and IPv4v6 which gets mapped to PDN Connection type. Minimum one PDP Context is needed to get dual stack connectivity.

4. For LTE (E-UTRAN) access to EPC core from Release-8 there are three IP PDN Types, IPv4, IPv6 and IPv4v6. Minimum one PDN Connection is needed to get dual stack connectivity.

6.3. PDN Connection Establishment Process

The PDN connection establishment process is specified in detail in 3GPP specifications. Figure 8 illustrates the high level process and signaling involved in the establishment of a PDN connection.

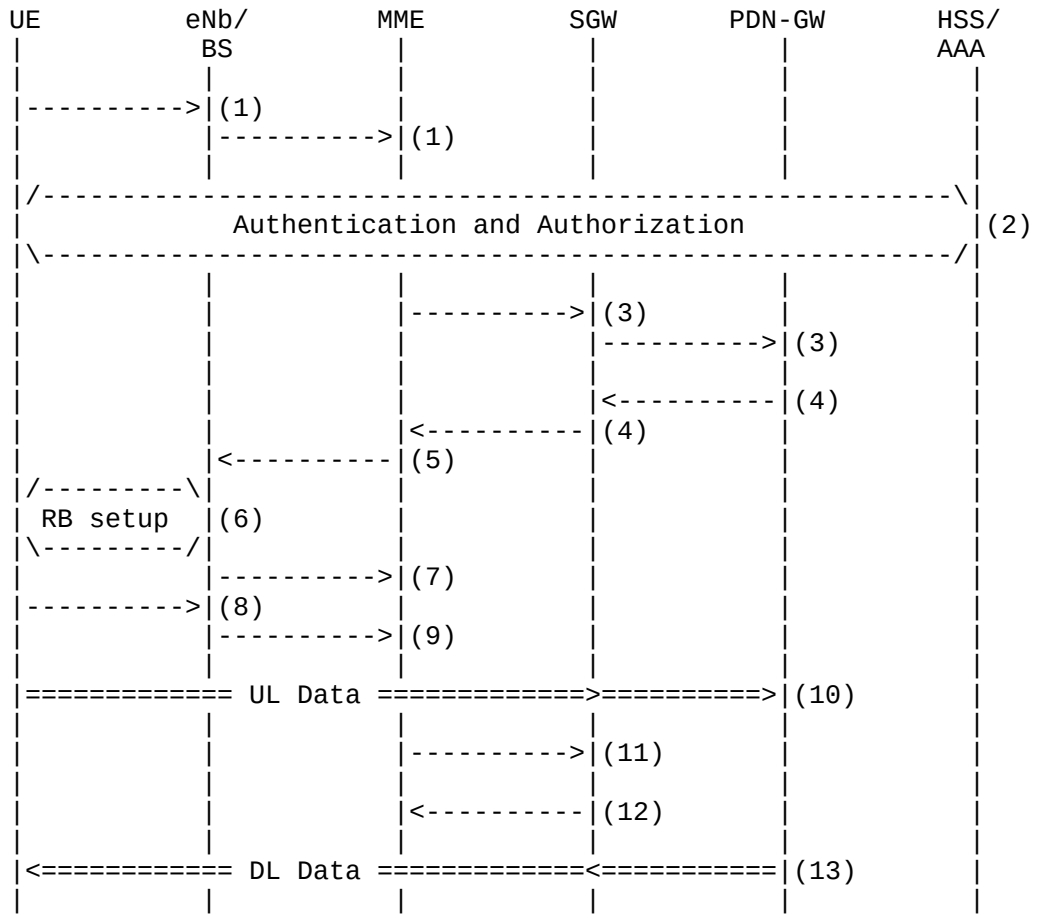


Figure 8: Simplified PDN connection setup procedure in Release-8

1. The UE (i.e the MS) requires a data connection and hence decides to establish a PDN connection with a PDN-GW. The UE sends an "Attach Request" (layer-2) to the BS. The BS forwards this attach request to the MME.

2. Authentication of the UE with the AAA server/HSS follows. If the UE is authorized for establishing a data connection, the following steps continue

3. The MME sends a "Create Session Request" message to the Serving-GW. The SGW forwards the create session request to the PDN-GW. The SGW knows the address of the PDN-GW to forward the create session request to as a result of this information having been obtained by the MME during the authentication/authorization

phase.

The UE IPv4 address and/or IPv6 prefix get assigned during this step. If a subscribed IPv4 address and/or IPv6 prefix is statically allocated for the UE for this APN, then the MME already passes the address information to the SGW and eventually to the PDN-GW in the "Create Session Request" message. Otherwise, the PDN-GW manages the address assignment to the UE (there is another variation to this where IPv4 address allocation is delayed until the UE initiates a DHCPv4 exchange but this is not discussed here).

4. The PDN-GW creates a PDN connection for the UE and sends "Create Session Response" message to the SGW from which the session request message was received from. The SGW forwards the response to the corresponding MME which originated the request.

5. The MME sends the "Attach Accept/Initial Context Setup request" message to the eNodeB/BS.

6. The radio bearer between the UE and the eNb is reconfigured based on the parameters received from the MME

7. The eNb sends "Initial Context Response" message to the MME.

8. The UE sends a "Direct Transfer" message to the eNodeB which includes the Attach complete signal.

9. The eNodeB forwards the Attach complete message to the MME.

10. The UE can now start sending uplink packets to the PDN GW.

11. The MME sends a "Modify Bearer Request" message to the SGW.

12. The SGW responds with a "Modify Bearer Response" message. At this time the downlink connection is also ready

13. The UE can now start receiving downlink packets

The type of PDN connection established between the UE and the PDN-GW can be any of the types described in the previous section. The DS PDN connection, i.e the one which supports both IPv4 and IPv6 packets is the default one that will be established if no specific PDN connection type is specified by the UE in Release-8 networks.

6.4. Mobility of 3GPP IPv4v6 Type of Bearers

3GPP discussed at length various approaches to support mobility between Release-8 and pre-Release-8 networks for the new dual-stack type of bearers.

The chosen approach for mobility is as follows, in short: if a mobile is known to be at risk for doing handovers between Release-8 and pre-Release-8 networks, only single stack bearers are used. Essentially meaning:

1. If a network knows a mobile may do handovers between Release-8 and pre-Release-8 networks (segment), network will only provide single stack bearers, even if the mobile host requests dual-stack bearers. This can happen e.g. if an operator is using pre-Release-8 SGSNs in some parts of the network. The single stack bearers of Release-8 are easy to map one-to-one to pre-Release-8 bearers.

2. If a network knows a mobile will not be able to do handover to pre-Release-8 network (segment), it will provide mobile with dual-stack bearers on request. This can happen e.g. if an operator has upgraded their SGSNs to support dual-stack bearers, or if an operator is running LTE-only network.

When a network operator and their roaming partners have upgraded their networks to Release-8, it is possible to use the new IPv4v6 dual-stack type of bearers. A Release-8 mobile device always requests for a dual-stack bearer, but accepts what is assigned by the network.

7. Dual-Stack Approach to IPv6 Transition in 3GPP Networks

3GPP networks can natively transport IPv4 and IPv6 packets between the mobile station/UE and the gateway (GGSN or PDN-GW) as a result of establishing either a dual-stack PDP context or parallel IPv4 and IPv6 PDP contexts.

Current deployments of 3GPP networks primarily support IPv4 only. These networks can be upgraded to also support IPv6 PDP contexts. By doing so devices and applications that are IPv6 capable can start utilizing the IPv6 connectivity. This will also ensure that legacy devices and applications continue to work with no impact. As newer devices start using IPv6 connectivity, the demand for actively used IPv4 connections is expected to slowly decrease, helping operators with a transition to IPv6. With a dual-stack approach, there is always the potential to fallback to IPv4. A device which may be

roaming in a network wherein IPv6 is not supported by the visited network could fall back to using IPv4 PDP contexts and hence the end user would at least get some connectivity. Unfortunately, dual-stack approach as such does not lower the number of used IPv4 addresses. Every dual-stack bearer still needs to given an IPv4 address, private or public. This is a major concern with dual-stack bearers concerning IPv6 transition. However, if the majority of active IP communication has moved over to IPv6, then in case of NAT44 [RFC1918] IPv4 connections the number of active IPv4 connections can still be expected to gradually decrease and thus giving some level of relief regarding NAT44 function scalability.

As the networks evolve to support Release-8 EPS architecture and the dual-stack PDP contexts, newer devices will be able to leverage such capability and have a single bearer which supports both IPv4 and IPv6. Since IPv4 and IPv6 packets are carried as payload within GTP between the MS and the gateway (GGSN/PDN-GW) the transport network capability in terms of whether it supports IPv4 or IPv6 on the interfaces between the eNodeB and SGW or, SGW and PDN-GW is immaterial.

8. Deployment issues

8.1. Overlapping IPv4 Addresses

Given the shortage of globally routable public IPv4 addresses, operators tend to assign private IPv4 addresses [RFC1918] to hosts when they establish an IPv4 only PDP context or an IPv4v6 type PDN context. About 16 million hosts can be assigned a private IPv4 address that is unique within a domain. However, in case of many operators the number of subscribers is greater than 16 million. The issue can be dealt with by assigning overlapping RFC 1918 IPv4 addresses to hosts. As a result the IPv4 address assigned to a host within the context of a single operator realm would no longer be unique. This has the obvious and know issues of NATed IP connection in the Internet. Direct host to host connectivity becomes complicated, unless the hosts are within the same private address range pool and/or anchored to the same gateway, referrals using IP addresses will have issues and so forth. These are generic issues and not only a concern of the EPS. However, 3GPP as such does not have any mandatory language concerning NAT44 functionality in EPC. Obvious deployment choices apply also to EPC:

1. Very large network deployments are partitioned, for example, based on a geographical areas. This partitioning allows overlapping IPv4 addresses ranges to be assigned to hosts that are in different areas. Each area has its own pool of gateways

that are dedicated for a certain overlapping IPv4 address range (referred here later as a zone). Standard NAT44 functionality enables the communication between hosts that are assigned the same IPv4 address but belong to different zones, yet are part of the same operator domain.

2. A mobile host/device attaches to a gateway as part of the attach process. The number of hosts that a gateway supports is in the order of 1 to 10 million. Hence all the hosts assigned to a single gateway can be assigned private IPv4 addresses. Operators with large subscriber bases have multiple gateways and hence the same [RFC1918] IPv4 address space can be reused across gateways. The IPv4 address assigned to a host is unique within the scope of a single gateway.

3. New services requiring direct connectivity between hosts should be build on IPv6. Possible existing IPv4-only services and applications requiring direct connectivity can be ported to IPv6.

8.2. IPv6 for transport

The various reference points of the 3GPP architecture such as S1-U, S5 and S8 are based on either GTP or PMIPv6. The underlying transport for these reference points can be IPv4 or IPv6. GTP has been able to operate over IPv6 transport (optionally) since R99 and PMIPv6 has supported IPv6 transport starting from its introduction in Release-8. The user plane traffic between the mobile host and the gateway can use either IPv4 or IPv6. These packets are essentially treated as payload by GTP/PMIPv6 and transported accordingly with no real attention paid to the information (at least from a routing perspective) contained in the IPv4 or IPv6 headers. The transport links between the eNodeB and the SGW, and the link between the SGW and PDN-GW can be migrated to IPv6 without any direct implications to the architecture.

Currently, the inter-operator (for 3GPP technology) roaming networks are all IPv4 only (see Inter-PLMN Backbone Guidelines [GSMA.IR.34]). Eventually these roaming networks will also get migrated to IPv6, if there is a business reason for that. The migration period can be prolonged considerably because the 3GPP protocols always tunnel user plane traffic in the core network and as described earlier the transport network IP version is not in any way tied to user plane IP version. Furthermore, the design of the inter-operator roaming networks is such that the user plane and transport network IP addressing is completely separated from each other. The inter-operator roaming network itself is also completely separated from the Internet. Only those core network nodes that must be connected to the inter-operator roaming networks are actually visible there, and

be able to send and receive (tunneled) traffic within the inter-operator roaming networks. Obviously, in order the roaming to work properly, the operators have to agree on supported protocol versions so that the visited network does not, for example, unnecessarily drop user plane IPv6 traffic.

8.3. Operational Aspects of Running Dual-Stack Networks

Operating dual-stack networks does imply cost and complexity to a certain extent. However these factors are mitigated by the assurance that legacy devices and services are unaffected and there is always a fallback to IPv4 in case of issues with the IPv6 deployment or network elements. The model also enables operators to develop operational experience and expertise in an incremental manner.

Running dual-stack networks requires the management of multiple IP address spaces. Tracking of hosts needs to be expanded since it can be identified by either an IPv4 address or IPv6 prefix. Network elements will also need to be dual-stack capable in order to support the dual-stack deployment model.

Deployment and migration cases described in Section 6.1 for providing dual-stack like capability may mean doubled resource usage in operator’s network. This is a major concern against providing dual-stack like connectivity using techniques discussed in Section 6.1. Also handovers between networks with different capabilities in terms of networks being dual-stack like service capable or not, may turn out hard to comprehend for users and for application/services to cope with. These facts may add other than just technical concerns for operators when planning to roll out dual-stack service offerings.

8.4. Operational Aspects of Running a Network with IPv6 Only Bearers

It is possible to allocate IPv6 only type bearers to mobile hosts in 3GPP networks. IPv6 only bearer type has been part of the 3GPP specification since the beginning. In 3GPP Release-8 (and later) it was defined that a dual-stack mobile host (or when the radio equipment has no knowledge of the host IP stack capabilities) must first attempt to establish a dual-stack bearer and then possibly fall back to single IP version bearer. A Release-8 (or later) mobile host with IPv6 only stack can directly attempt to establish an IPv6 only bearer. The IPv6 only behavior is up to a subscription provisioning or a PDN-GW configuration, and the fallback scenarios do not necessarily cause additional signaling.

Although the bullets below introduce IPv6 to IPv4 address translation and specifically discuss NAT64 technology [I-D.ietf-behave-v6v4-framework], the current 3GPP Release-8

architecture does not describe the use of address translation or NAT64. It is up to a specific deployment whether address translation is part of the network or not. Some operational aspects to consider for running a network with IPv6 only bearers:

- The mobile hosts must have an IPv6 capable stack and a radio interface capable of establishing an IPv6 PDP context or PDN connection.

- The GGSN/PDN-GW must be IPv6 capable in order to support IPv6 bearers. Furthermore, the SGSN/MME must allow the creation of PDP Type or PDN Type of IPv6.

- Many of the common applications are IP version agnostic and hence would work using an IPv6 bearer. However, applications that are IPv4 specific would not work.

- Inter-operator roaming is another aspect which causes issues, at least during the ramp up phase of the IPv6 deployment. If the visited network to which outbound roamers attach to does not support PDP/PDN Type IPv6, then there needs to be a fallback option. The fallback option in this specific case is mostly up to the mobile host to implement. Several cases are discussed in the following sections.

- If and when a mobile host using IPv6 only bearer needs to access to IPv4 Internet/network, a translation of some type from IPv6 to IPv4 has to be deployed in the network. NAT64 (and DNS64) is one solution that can be used for this purpose and works for a certain set of protocols (read TCP and UDP, and when applications actually use DNS for resolving name to IP addresses).

8.5. Restricting Outbound IPv6 Roaming

Roaming was briefly touched upon in Sections 8.2 and 8.4. While there is interest in offering roaming service for IPv6 enabled mobile hosts and subscriptions, not all visited networks are prepared for IPv6 outbound roamers. There are basically two issues. First, the visited network (S4-)SGSN does not support the IPv6 PDP Context or IPv4v6 PDP Context types. These should mostly concern pre-Release-8 networks but there is no definitive rule as the deployed feature sets vary depending on implementations and licenses. Second, the visited network might not be commercially ready for IPv6 outbound roamers, while everything might work technically at the user plane level. This would lead to "revenue leakage" especially from the visited operator point of view (note that the use of visited network GGSN/ PDN-GW does not really exist in real deployments today). Therefore, it might be in the interest of operators to prohibit roaming

selectively within specific visited networks.

Unfortunately, it is not mandatory to implement/deploy 3GPP standards based solution to selectively prohibit IPv6 roaming without also prohibiting other packet services (such as IPv4 roaming). However, there are few possibilities how this can be done in real deployments. The examples given below are either optional and/or vendor specific features to the 3GPP EPC:

- Using Policy and Charging Control (PCC) [3GPP.23.203] functionality and its rules to fail, for example, the bearer authorization when a desired criteria is met. In this case that would be PDN/PDP Type IPv6/IPv4v6 and a specific visited network. The rules can be provisioned either in the home network or locally in the visited network.

- Some Home Location Register (HLR) and Home Subscriber Server (HSS) subscriber databases allow prohibiting roaming in a specific (visited) network for a specified PDN/PDP Type.

The obvious problems are that these solutions are not mandatory, are not unified across networks, and therefore also lack well-specified fall back mechanism from the mobile host point of view.

8.6. Inter-rat Handovers and IP Versions

It is obvious that when operators start to incrementally deploy EPS (and E-UTRAN) along with the existing UTRAN/GERAN, handovers between different radio technologies (inter-rat handovers) become inevitable. In case of inter-rat handovers 3GPP supports the following IP addressing scenarios:

- E-UTRAN IPv4v6 bearer has to map one to one to UTRAN/GERAN IPv4v6 bearer.

- E-UTRAN IPv6 bearer has to map one to one to UTRAN/GERAN IPv6 bearer.

- E-UTRAN IPv4 bearer has to map one to one to UTRAN/GERAN IPv4 bearer.

Other types of configurations are considered network planning mistakes. What the above rules essentially imply is that the network migration has to be planned and subscriptions provisioned based on the lowest common nominator, if inter-rat handovers are desired. For example, if some part of the UTRAN network cannot serve anything but IPv4 bearers, then the E-UTRAN is also forced to provide only IPv4 bearers. Various combinations of subscriber provisioning regarding

IP versions are discussed further in Section 8.7.

8.7. Provisioning of IPv6 Subscribers and Various Combinations During Initial Network Attachment

Subscribers' provisioned PDP/PDN Types have multiple configurations. The supported PDP/PDN Type is provisioned per each APN for every subscriber. The following PDN Types are possible in the HSS for a Release-8 subscription [3GPP.23.401]:

- IPv4v6 PDN Type (note that IPv4v6 PDP Type does not exist in HLR).
- IPv6 only PDN Type
- IPv4 only PDN Type.
- IPv4_or_IPv6 PDN Type (note that IPv4_or_IPv6 PDP Type does not exist in HLR).

A Release-8 dual-stack mobile host must always attempt to establish a PDP/PDN Type IPv4v6 bearer. The same also applies when the modem part of the mobile host does not have exact knowledge whether the host operating system IP stack is a dual-stack capable or not. A mobile host that is IPv6 only capable must attempt to establish a PDP/PDN Type IPv6 bearer. Last, a mobile host that is IPv4 only capable must attempt to establish a PDN/PDP Type IPv4 bearer.

In a case the PDP/PDN Type requested by a mobile host does not match what has been provisioned for the subscriber in the HSS (or HLR), the mobile host possibly falls back to a different PDP/PDN Type. The network (i.e. the MME or the SGSN) is able to inform the mobile host during the network attachment signaling why it did not get the requested PDP/PDN Type. These response/cause codes are documented in [3GPP.24.008][3GPP.24.301]. Possible fall back cases include (as documented in [3GPP.23.401]):

- Requested & provisioned PDP/PDN Types match -> requested.
- Requested IPv4v6 & provisioned IPv6 -> IPv6 and a mobile host receives indication that IPv6-only bearer is allowed.
- Requested IPv4v6 & provisioned IPv4 -> IPv4 and the mobile host receives indication that IPv4-only bearer is allowed.
- Requested IPv4v6 & provisioned IPv4_or_IPv6 -> IPv4 or IPv6 is selected by the MME based on an unspecified criteria. The mobile host may then attempt to establish, based on the mobile host implementation, a parallel bearer of a different PDP/PDN Type.

- o Other combinations cause the bearer establishment to fail.

In addition to PDP/PDN Types provisioned in the HSS, it is also possible for a PDN-GW (and a MME) to affect the final selected PDP/PDN Type:

- o Requested IPv4v6 & configured IPv4 or IPv6 in the PDN-GW -> IPv4 or IPv6. If the MME operator had included the "Dual Address Bearer Flag" into the bearer establishment signaling, then the mobile host receives an indication that IPv6-only or IPv4-only bearer is allowed.

- o Requested IPv4v6 & configured IPv4 or IPv6 in the PDN-GW -> IPv4 or IPv6. If the MME operator had not included the "Dual Address Bearer Flag" into the bearer establishment signaling, then the mobile host may attempt to establish, based on the mobile host implementation, a parallel bearer of different PDP/PDN Type.

If for some reason a SGSN does not understand the requested PDP Type, then the PDP Type is handled as IPv4. If for some reason a MME does not understand the requested PDN Type, then the PDN Type is handled as IPv6.

9. IANA Considerations

This document has no requests to IANA.

10. Security Considerations

This document does not introduce any security related concerns.

11. Summary and Conclusion

The 3GPP network architecture and specifications enable the establishment of IPv4 and IPv6 connections through the use of appropriate PDP context types. The current generation of deployed networks can support dual-stack connectivity if the packet core network elements such as the SGSN and GGSN have the capability. With Release-8, 3GPP has specified a more optimal PDP context type which enables the transport of IPv4 and IPv6 packets within a single PDP context between the mobile station and the gateway.

As devices and applications are upgraded to support IPv6 they can start leveraging the IPv6 connectivity provided by the networks while maintaining the fall back to IPv4 capability. Enabling IPv6

connectivity in the 3GPP networks by itself will provide some degree of relief to the IPv4 address space as many of the applications and services can start to work over IPv6. However without comprehensive testing of different applications and solutions that exist today and are widely used, for their ability to operate over IPv6 PDN connections, an IPv6 only access would cause disruptions.

12. Acknowledgements

The authors thank Shabnam Sultana, Sri Gundavelli, Hui Deng, and Zhenqiang Li, Mikael Abrahamsson, James Woodyatt and Cameron Byrne for their reviews and comments on this document.

13. Informative References

[3GPP.23.060]
3GPP, "General Packet Radio Service (GPRS); Service description; Stage 2", 3GPP TS 23.060 8.8.0, March 2010.

[3GPP.23.203]
3GPP, "Policy and charging control architecture (PCC)", 3GPP TS 23.203 8.11.0, September 2010.

[3GPP.23.401]
3GPP, "General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access", 3GPP TS 23.401 10.2.1, January 2011.

[3GPP.23.975]
3GPP, "IPv6 Migration Guidelines", 3GPP TR 23.975 1.1.1, June 2010.

[3GPP.24.008]
3GPP, "Mobile radio interface Layer 3 specification", 3GPP TS 24.008 8.12.0, December 2010.

[3GPP.24.301]
3GPP, "Non-Access-Stratum (NAS) protocol for Evolved Packet System (EPS)", 3GPP TS 24.301 8.8.0, December 2010.

[3GPP.29.060]
3GPP, "General Packet Radio Service (GPRS); GPRS Tunnelling Protocol (GTP) across the Gn and Gp interface", 3GPP TS 29.274 8.8.0, April 2010.

[3GPP.29.061]

3GPP, "Interworking between the Public Land Mobile Network (PLMN) supporting packet based services and Packet Data Networks (PDN)", 3GPP TS 29.061 8.5.0, April 2010.

[3GPP.29.274]
3GPP, "3GPP Evolved Packet System (EPS); Evolved General Packet Radio Service (GPRS) Tunnelling Protocol for Control plane (GTPv2-C)", 3GPP TS 29.060 8.11.0, December 2010.

[GSMA.IR.34]
GSMA, "Inter-PLMN Backbone Guidelines", GSMA PRD IR.34.4.9, March 2010.

[I-D.ietf-behave-v6v4-framework]
Baker, F., Li, X., Bao, C., and K. Yin, "Framework for IPv4/IPv6 Translation", draft-ietf-behave-v6v4-framework-10 (work in progress), August 2010.

[I-D.ietf-dhc-pd-exclude]
Korhonen, J., Savolainen, T., Krishnan, S., and O. Troan, "Prefix Exclude Option for DHCPv6-based Prefix Delegation", draft-ietf-dhc-pd-exclude-01 (work in progress), January 2011.

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

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

[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.

[RFC4389] Thaler, D., Talwar, M., and C. Patel, "Neighbor Discovery Proxies (ND Proxy)", RFC 4389, April 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.

[RFC5213] Gundavelli, S., Leung, K., Devarapalli, V., Chowdhury, K., and B. Patil, "Proxy Mobile IPv6", RFC 5213, August 2008.

Authors' Addresses

Jouni Korhonen (editor)
Nokia Siemens Networks
Linnoitustie 6
FI-02600 Espoo
FINLAND

Email: jouni.nospam@gmail.com

Jonne Soininen
Renesas Mobile

Email: jonne.soininen@renesasmobile.com

Basavaraj Patil
Nokia
6021 Connection drive
Irving, TX 75039
USA

Email: basavaraj.patil@nokia.com

Teemu Savolainen
Nokia
Hermiankatu 12 D
FI-33720 Tampere
FINLAND

Email: teemu.savolainen@nokia.com

Gabor Bajko
Nokia
323 Fairchild drive 6
Mountain view, CA 94043
USA

Email: gabor.bajko@nokia.com

Kaisu Iisakkila
Renesas Mobile

Email: kaisu.iisakkila@renesasmobile.com

V6ops WG
Internet-Draft
Intended status: Informational
Expires: March 26, 2011

V. Kuarsingh, Ed.
Rogers Communications
Y. Lee
Comcast
O. Vautrin
Juniper Networks
September 22, 2010

6to4 Provider Managed Tunnels
draft-kuarsingh-v6ops-6to4-provider-managed-tunnel-00

Abstract

This document provides an overview of a framework describing the management of 6to4 [RFC3056] tunnels within a provider network. The 6to4 provider managed tunnel, or 6to4-PMT, combines the current behavior of 6to4 [RFC3056] utilizing the IPv4 anycast based connectivity defined in [RFC3068] along with IPv6 Prefix Translation. The framework is intended to allow Service Providers an option to translate 6to4 based addresses to provider based addresses utilizing provider assigned prefixes. The framework offers IPv6 connectivity to 6to4 compatible endpoints [RFC3056] with the advantage of a stable provider assigned prefix. The 6to4-PMT operation is not intended to replace Native IPv6 connectivity nor 6RD, but rather provide a connectivity before such options can be deployed by an operator.

Table of Contents

1. Introduction

6to4 [RFC3056] tunneling is widely deployed in modern host OSs and off the shelf gateways sold throughout the retail and OEM channels. 6to4 allows for tunneled IPv6 connectivity through IPv4 clouds, but due to the anycast nature of the ingress and egress flows, flows paths are difficult to determine and often change based on network conditions. The return path is uncontrolled by the local provider and can contribute to poor performance for IPv6, and can also act as a breakage point (i.e. mis-behaving relay/system). For this reason, despite being widely available today, many providers choose not to directly support a 6to4 environment.

Providers which are actively deploying IPv6 networks and operate legacy IPv4 access environments may want to utilize the existing 6to4 behavior in deployed hardware and software and offer a more controlled access to the IPv6 Internet for 6to4 capable endpoints. 6to4-PMT offers a provider the opportunity to utilize IPv6 Prefix Translation to provide a more deterministic path to and from the Internet for 6to4 based traffic.

6to4-PMT translates the prefix portion of the address from the 6to4 address to a provider assigned prefix which is used to represent the source. This translation will then provide a stable forward and return path for the 6to4 traffic by allowing the existing IPv6 routing and policy environment to control the traffic. 6to4-PMT is intended to be used in a stateless manner to maintain many of the elements inherent in normal 6to4 operation.

2. Motivation

Providers endeavor to deploy IPv6 as soon as possible, so as to ensure uninterrupted connectivity to all Internet applications and content through the transition process. The IPv6 preparations within these organizations are often faced with both financial challenges and timing issues related to deploying IPv6 to the network edge and related transition technologies. Many of the new technologies addressing IPv4 to IPv6 transition will require the replacement of the customer CPE to support technologies like 6RD [RFC5569].

Provider initiated replacement of this equipment will take time due to the nature of such mass equipment refresh programs. Additionally, many providers also do not supply CPE related equipment and general lack of awareness in the consumer space may delay the upgrade of many in-home gateway and operating environments. Providers may still be motivated to provide a form of IPv6 connectivity to customers to mitigate potential issues related to IPv6-only deployments elsewhere

on the Internet. After IPv4 run out, IPv6 content may grow rapidly and in some cases, IPv4 may not be a connection option for some web based content providers or the remote host (IPv6 Only).

6to4-PMT allows a provider to help mitigate such challenges by leveraging a protocol which is already found on many CPE home gateways, while maintaining operator control of access to the IPv6 Internet. It is intended for use when better options, such as 6RD or native IPv6, are not yet viable. The 6to4-PMT operation can also be used immediately with existing OS and gateway functionality (in the wild) without the initial costly replacement of consumer equipment. The default 6to4 operation on most consumer grade OSs and gateways will allow for IPv6 connectivity over the IPv4 access network. Once native IPv6 is available to the endpoint, the 6to4-PMT operation is not longer needed. Next step options can include 6RD or Native IPv6. 6to4-PMT offers an opportunity to fill the gap between now and when the provider can feasibly replace the CPE equipment.

3. 6to4 Provider Managed Tunnels

3.1. 6to4 Provider Managed Tunnel Model

The 6to4 managed tunnel model behaves like a standard 6to4 service between the customer IPv6 host or gateway and the 6ot4-PMT Relay (within the provider domain). The 6to4-PMT Relay shares properties with 6RD [RFC5569] by decapsulating and forwarding embedded IPv6 flows, within an IPv4 packet, to the IPv6 Internet. The model provides an additional function which translates the source 6to4 prefix to a provider assigned prefix which is not found in 6RD [RFC5569] or normal 6to4 operation.

The 6to4-PMT Relay is intended to provide a stateless mapping of the 6to4 prefix to a provider supplied prefix by mapping the embedded IPv4 address in the 6to4 prefix to the provider prefix.

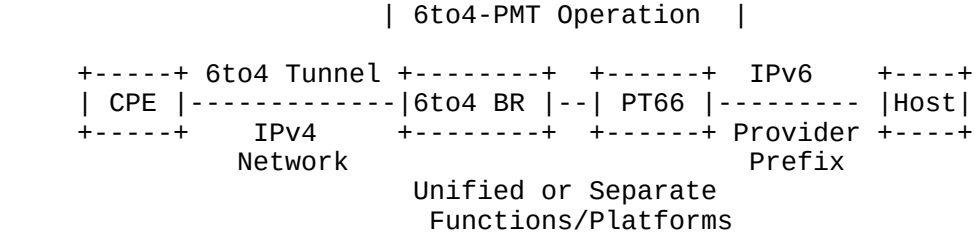


Figure 1: 6to4-PMT Functional Model

3.2. Traffic Flow

Traffic in the 6to4-PMT model is intended to be controlled by the operators IPv6 peering operations. Egress traffic is managed through outgoing routing policy, and incoming traffic is influenced by the operator assigned prefix advertisements.

The routing model is as predictable as native IPv6 traffic and legacy IPv4 based traffic. Figure 1 provides a view of the routing topology needed to support this relay environment. The diagram references PrefixA as 2002::/16 and PrefixB as the example 2001:db8::/32.

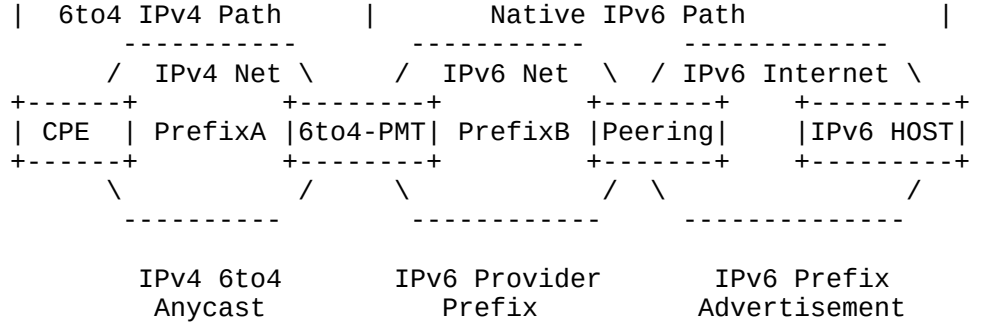


Figure 2: 6to4-PMT Flow Model

Traffic normally between two 6to4 enabled devices would use the IPv4 path for communication according to RFC3056

3.3. Prefix Translation

The IPv6 Prefix Translation is a key part of the system as a whole. The 6to4-PMT framework is a combination of two concepts: 6to4 [RFC3056] and IPv6 Prefix Translation. IPv6 Prefix Translation has some similarities to concepts discussed in [draft-mrw-behave-nat66]. The only change in this particular case is that the provider would build specific rules on the translator to map the 6to4 prefix to an appropriate provider assigned prefix.

The provider can use any prefix mapping strategy they so choose, but the simpler the better. Simple direct bit mapping can be used such as in Figure 2, or more advanced forms of translation can used [reference to I-D here] to achieve higher address compression.

Figure 2 shows a 6to4 Prefix with a Subnet-ID of "0000" mapped to a provider globally unique prefix (2001:db8::/32). With this simple form of translation, there is support for only one Subnet-ID per provider assigned prefix. In characterization of deployed OSs and

gateways, a subnet-id of "0000" is the most common default case.

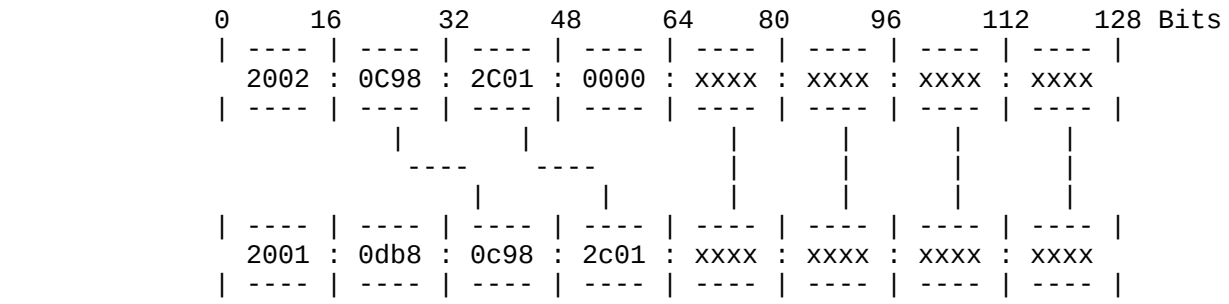


Figure 3: 6to4-PMT Prefix Mapping

Additional prefix compression techniques can be used such as those described in [draft-treamblay-pt66-compression]. These techniques would allow for a more flexible implementation potentially supporting more Subnet-IDs per provider prefix

3.4. Translation State

It is preferred that the overall system use deterministic prefix translation mappings such that stateless operation can be implemented. This allows the provider to place N number of relays within the network without the need to route traffic specifically to support 6to4-PMT.

If stateful operation is used, then the provider would need to force all return traffic to pass back through the same relay as the exit packets. The assumption here is that the provider would operate the 6to4 anycast environment on the IPv4 side of the relay. To control the ingress traffic, the provider would need to craft IPv6 routes (more specific then provider assigned prefix) which match the IPv4 ranges for a given relay.

Maintaining these additional routes would subject the provider’s IPv6 network to maintaining a number of fragmented IPv4 network router (level of fragmentation dependant on network), these routes do not need to be advertised to peer networks, nor should hey.

4. Deployment Issues and Requirements

4.1. Customer Opt-out

A provider enabling this function should provide a method to allow customers to opt-out of such a service should the customer choose to, or wish to maintain normal 6to4 operation.

Since the 6to4-PMT system is targeted at customers who are relatively unaware of IPv6 and IPv4, and normally run network equipment with a default configuration, an opt-out strategy is preferred. This method provides the 6to4-PMT operation for non-IPv6 savvy customers whose equipment may turn on 6to4 automatically.

Customers who are aware of IPv6 operation can request an opt-out, or more appropriately use an automated mechanism to opt-out of the 6to4-PMT operation. One automated opt-out strategy can include the use of Subnet-Id triggers (well known IDs which are determined by policy in the relay to not be translated). Other policy based strategies can be employed by the provider to enable opt-out.

Capable customers can also disable 6to4 entirely and use other tunneling mechanisms if they are so capable. This is not considered the normal case, and most endpoints with auto-6to4 operation will be subject to 6to4-PMT operation operation. 6to4-PMT is targeted as an option for deterministic IPv6 connectivity for average consumers

4.2. ISP Shared Space Interaction

6to4-PMT operation can also be used to mitigate a known problem with 6to4 when ISP Shared Space [draft-weil-opsawg-provider-address-spaces] is used. ISP Shared Space would cause many deployed OSs and network equipment to potentially auto-enable 6to4 operation should non-RFC1918 addressing be used on the CPE IPv4 address assignments.

Such hosts, in normal cases, would send 6to4 traffic to the IPv6 Internet via the IPv4 anycast relay, which would in fact provide broken IPv6 connectivity since the return path is based on an address that is not routed or assigned to the source Network. The use of 6to4-PMT would help reverse these effects by translating the 6to4 prefix to a provided assigned prefix, masking this automatic and undesired behavior. It is conceivable that 6to4-PMT can also used to help provide 6to4 operation with the use of ISP Shared Space.

4.3. End to End Transparency

6to4-PMT mode operation removes the traditional end to end transparency of 6to4. Remote hosts would connect to a translated IPv6 address versus the original 6to4 based prefix. This can be seen as a disadvantage to the 6to4-PMT system. This lack of transparency should also be contrasted with the normal operating state of 6to4 which provides uncontrolled and often high latency prone connectivity

4.4. Routing Requirements

The provider would need to advertise the anycast IP range within the IPv4 routing environment (service customers of interest) to attract the 6to4 upstream traffic. To control this environment and make sure all northbound traffic lands on a provider BR, the operator may filter the anycast range form being advertised form customer endpoints.

The provider would not be able to control route advertisements inside the customer domain, but this use case is out of scope. It is likely in this case the end network/customer understands IPv6 operation and is maintaining their own environment.

The provider would also likely want to advertise the 2002::/16 range within their own network to help bridge within their own network (Native IPv6 to 6to4-IPv6 based endpoint)

5. IANA Considerations

No IANA considerations are defined at this time.

6. Security Considerations

6to4-PMT operation would be subject to the same security concerns as normal 6to4 operation and with the operation of tunnels. Considerations may also include operation modes related to Prefix Translation. Additional considerations may be found after real deployment data is gathered or further analysis is made.

7. Acknowledgements

Thanks to the following people for their textual contributions and/or guidance on 6to4 deployment considerations: Dan Wing, Scott Beuker, JF Tremblay, John Brzozowski and Chris Donley

8. References

8.1. Normative References

[RFC3056] Carpenter, B. and K. Moore, "Connection of IPv6 Domains via IPv4 Clouds", RFC 3056, February 2001.

[RFC3068] Huitema, C., "An Anycast Prefix for 6to4 Relay Routers", RFC 3068, June 2001.

8.2. Informative References

[I-D.mrw-behave-nat66]
Wasserman, M. and F. Baker, "IPv6-to-IPv6 Network Address Translation (NAT66)", draft-mrw-behave-nat66-02 (work in progress), March 2009.

[I-D.weil-opsawg-provider-address-space]
Weil, J., Kuarsingh, V., and C. Donley, "IANA Reserved IPv4 Prefix for IPv6 Transition", draft-weil-opsawg-provider-address-space-01 (work in progress), August 2010.

[RFC5569] Despres, R., "IPv6 Rapid Deployment on IPv4 Infrastructures (6rd)", RFC 5569, January 2010.

Authors' Addresses

Victor Kuarsingh (editor)
Rogers Communications
8200 Dixie Road
Brampton, Ontario L6T 0C1
Canada

Email: victor.kuarsingh@rci.rogers.com
URI: http://www.rogers.com

Yiu L. Lee
Comcast
One Comcast Center
Philadelphia, PA 19103
U.S.A.

Email: yiu_lee@cable.comcast.com
URI: http://www.comcast.com

Olivier Vautrin
Juniper Networks
1194 N Mathilda Avenue
Sunnyvale, CA 94089
U.S.A.

Email: olivier@juniper.net
URI: http://www.juniper.net

v6ops
Internet-Draft
Intended status: Informational
Expires: January 11, 2013

V. Kuarsingh, Ed.
Rogers Communications
Y. Lee
Comcast
O. Vautrin
Juniper Networks
July 10, 2012

6to4 Provider Managed Tunnels
draft-kuarsingh-v6ops-6to4-provider-managed-tunnel-07

Abstract

6to4 Provider Managed Tunnels (6to4-PMT) provide a framework which can help manage 6to4 tunnels operating in an anycast configuration. The 6to4-PMT framework is intended to serve as an option for operators to help improve the experience of 6to4 operation when conditions of the network may provide sub-optimal performance or break normal 6to4 operation. 6to4-PMT provides a stable provider prefix and forwarding environment by utilizing existing 6to4 relays with an added function of IPv6 Prefix Translation. This operation may be particularly important in NAT444 infrastructures where a customer endpoint may be assigned a non-RFC1918 address thus breaking the return path for anycast based 6to4 operation. 6to4-PMT has successfully been used in a production network, has been implemented as open source code, and implemented by a major routing vendor.

Table of Contents

1. Introduction

6to4 [RFC3056] tunnelling along with the anycast operation described in [RFC3068] is widely deployed in modern Operating Systems and off the shelf gateways sold throughout the retail and OEM channels. Anycast [RFC3068] based 6to4 allows for tunnelled IPv6 connectivity through IPv4 clouds without explicit configuration of a relay address. Since the overall system utilizes anycast forwarding in both directions, flow paths are difficult to determine, tend to follow separate paths in either direction, and often change based on network conditions. The return path is normally uncontrolled by the local operator and can contribute to poor performance for IPv6, and can also act as a breakage point. Many of the challenges with 6to4 are described in [RFC6343]. A specific critical use case for problematic anycast 6to4 operation is related to conditions where the consumer endpoints are downstream from a northbound CGN [RFC6264] function when assigned non-RFC1918 IPv4 addresses, which are not routed on interdomain links.

Operators which are actively deploying IPv6 networks and operate legacy IPv4 access environments may want to utilize the existing 6to4 behaviour in customer site resident hardware and software as an interim option to reach the IPv6 Internet in advance of being able to offer full native IPv6. Operators may also need to address the brokenness related to 6to4 operation originating from behind a provider NAT function. 6to4-PMT offers an operator the opportunity to utilize IPv6 Prefix Translation to enable deterministic traffic flow and an unbroken path to and from the Internet for IPv6 based traffic sourced originally from these 6to4 customer endpoints.

6to4-PMT translates the prefix portion of the IPv6 address from the 6to4 generated prefix to a provider assigned prefix which is used to represent the source. This translation will then provide a stable forward and return path for the 6to4 traffic by allowing the existing IPv6 routing and policy environment to control the traffic. 6to4-PMT is primarily intended to be used in a stateless manner to maintain many of the elements inherent in normal 6to4 operation. Alternatively, 6to4-PMT can be used in a stateful translation mode should the operator choose this option.

2. Motivation

Many operators endeavour to deploy IPv6 as soon as possible so as to ensure uninterrupted connectivity to all Internet applications and content through the IPv4 to IPv6 transition process. The IPv6 preparations within these organizations are often faced with both financial challenges and timing issues related to deploying IPv6 to

the network edge and related transition technologies. Many of the new technologies available for IPv4 to IPv6 transition will require the replacement of the customer CPE to support technologies like 6RD [RFC5969], Dual-Stack Lite [RFC6333] and Native Dual Stack.

Operators face a number of challenges related to home equipment replacement. Operator initiated replacement of this equipment will take time due to the nature of mass equipment refresh programs or may require the consumer to replace their own gear. Replacing consumer owned and operated equipment, compounded by the fact that there is also a general unawareness of what IPv6 is, also adds to the challenges faced by operators. It is also important to note that 6to4 is found in much of the equipment found in networks today which do not as of yet, or will not, support 6RD and/or Native IPv6.

Operators may still be motivated to provide a form of IPv6 connectivity to customers and would want to mitigate potential issues related to IPv6-only deployments elsewhere on the Internet. Operators also need to mitigate issues related to the fact that 6to4 operation often is on by default and may be subject to erroneous behaviour. The undesired behaviour may be related to the use of non-RFC1918 addresses on CPE equipment which operate behind large operator NATs, or other conditions as described in a general advisory as laid out in [RFC6343].

6to4-PMT allows an operator to help mitigate such challenges by leveraging the existing 6to4 deployment base, while maintaining operator control of access to the IPv6 Internet. It is intended for use when better options, such as 6RD or Native IPv6, are not yet viable. One of key objectives of 6to4-PMT is to also help reverse the negative impacts of 6to4 in CGN environments. The 6to4-PMT operation can also be used immediately with the default parameters which are often enough to allow it to operate in a 6to4-PMT environment. Once native IPv6 is available to the endpoint, the 6to4-PMT operation is no longer needed and will cease to be used based on correct address selection behaviours in end hosts [RFC3484].

6to4-PMT thus helps operators remove the impact of 6to4 in CGN environments, deals with the fact that 6to4 is often on by default, allows access to IPv6-only endpoints from IPv4-only addressed equipment and provides relief from many challenges related to mis-configurations in other networks which control return flows via foreign relays. Due to the simple nature of 6to4-PMT, it can also be implemented in a cost effective and simple manner allowing operators to concentrate their energy on deploying Native IPv6.

3. 6to4 Provider Managed Tunnels

3.1. 6to4 Provider Managed Tunnel Model

The 6to4 managed tunnel model behaves like a standard 6to4 service between the customer IPv6 host or gateway and the 6to4-PMT Relay (within the provider domain). The 6to4-PMT Relay shares properties with 6RD [RFC5969] by decapsulating and forwarding encapsulated IPv6 flows within an IPv4 packet, to the IPv6 Internet. The model provides an additional function which translates the source 6to4 prefix to a provider assigned prefix which is not found in 6RD [RFC5969] or traditional 6to4 operation.

The 6to4-PMT Relay is intended to provide a stateless (or stateful) mapping of the 6to4 prefix to a provider supplied prefix.

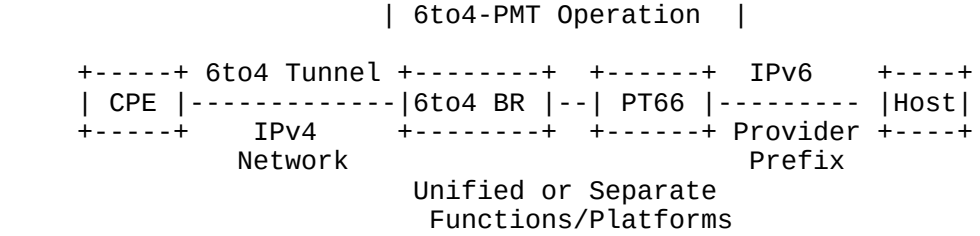


Figure 1: 6to4-PMT Functional Model

This mode of operation is seen as beneficial when compared to broken 6to4 paths and/or environments where 6to4 operation may be functional but highly degraded.

3.2. Traffic Flow

Traffic in the 6to4-PMT model is intended to be controlled by the operator’s IPv6 peering operations. Egress traffic is managed through outgoing routing policy, and incoming traffic is influenced by the operator assigned prefix advertisements using normal interdormain routing functions.

The routing model is as predictable as native IPv6 traffic and legacy IPv4 based traffic. Figure 2 provides a view of the routing topology needed to support this relay environment. The diagram references PrefixA as 2002::/16 and PrefixB as the example 2001:db8::/32.

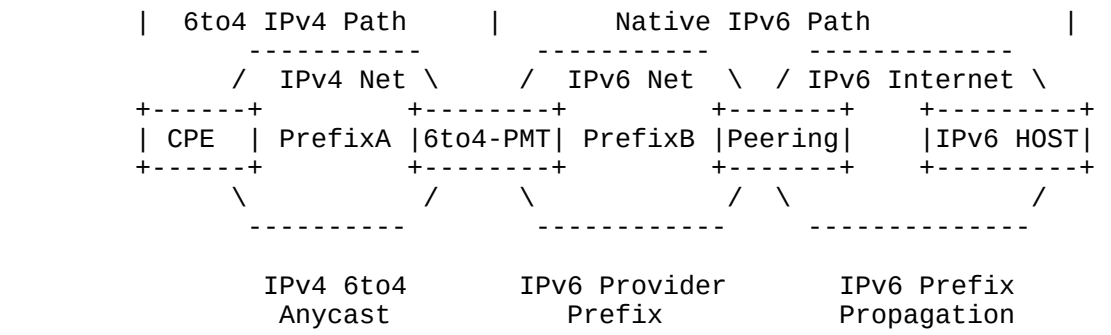


Figure 2: 6to4-PMT Flow Model

Traffic between two 6to4 enabled devices would use the IPv4 path for communication according to RFC3056 unless the local host still prefers traffic via a relay. 6to4-PMT is intended to be deployed in conjunction with the 6to4 relay function in an attempt to help simplify it's deployment. The model can also provide the ability for an operator to forward both 6to4-PMT (translated) and normal 6to4 flows (untranslated) simultaneously based on configured policy.

3.3. Prefix Translation

IPv6 Prefix Translation is a key part of the system as a whole. The 6to4-PMT framework is a combination of two concepts: 6to4 [RFC3056] and IPv6 Prefix Translation. IPv6 Prefix Translation, as used in 6to4-PMT, has some similarities to concepts discussed in [RFC6296]. 6to4-PMT would provide prefix translation based on specific rules configured on the translator which maps the 6to4 2002::/16 prefix to an appropriate provider assigned prefix. In most cases, a ::/32 prefix would work best in 6to4-PMT which matches common RIR prefix assignments to operators.

The provider can use any prefix mapping strategy they so choose, but the simpler the better. Simple direct bit mapping can be used, or more advanced forms of translation should the operator want to achieve higher address compression. More advanced forms of translation may require the use of stateful translation.

Figure 3 shows a 6to4 Prefix with a Subnet-ID of "0000" mapped to a provider assigned globally unique prefix (2001:db8::/32). With this simple form of translation, there is support for only one Subnet-ID per provider assigned prefix. In characterization of deployed OSs and gateways, a Subnet-ID of "0000" is the most common default case followed by Subnet-ID "0001". Use of Subnet-ID can be referenced in [RFC4291]. It should be noted that in normal 6to4 operation the endpoint (network) has access to 65,536 (16-bits) Subnet IDs. In the

6to4-PMT case as described above using the mapping in Figure 3, all but the one Subnet-ID used for 6to4-PMT would still operate under normal 6to4 operation.

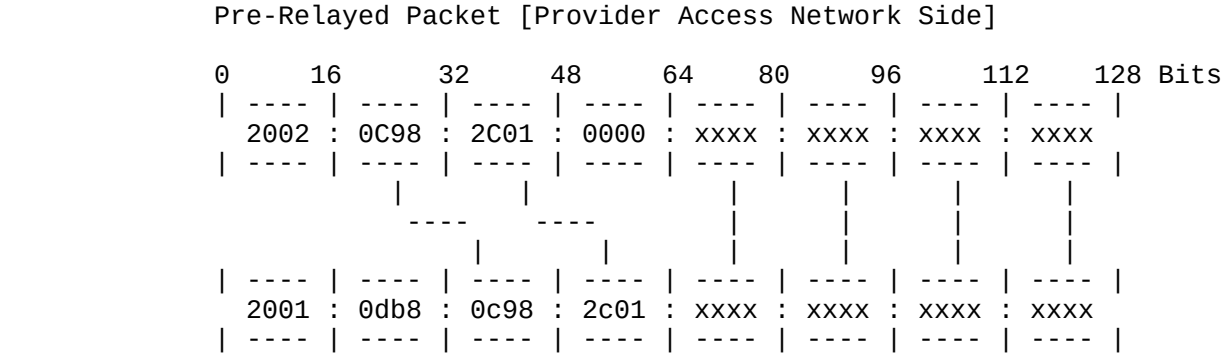


Figure 3: 6to4-PMT Prefix Mapping

3.4. Translation State

It is preferred that the overall system use deterministic prefix translation mappings such that stateless operation can be implemented. This allows the provider to place N number of relays within the network without the need to manage translation state. Deterministic translation also allows a customer to use inward services using the translated (provider prefix) address.

If stateful operation is chosen, the operator would need to validate state and routing requirements particular to that type of deployment. The full body of considerations for this type of deployment are not within this scope of this document.

4. Deployment Considerations and Requirements

4.1. Customer Opt-out

A provider enabling this function should provide a method to allow customers to opt-out of such a service should the customer choose to maintain normal 6to4 operation irrespective of degraded performance. In cases where the customer is behind a CGN device, the customer would not be advised to opt-out and can also be assisted to turn off 6to4.

Since the 6to4-PMT system is targeted at customers who are relatively

unaware of IPv6 and IPv4, and normally run network equipment with a default configuration, an opt-out strategy is recommended. This method provides 6to4-PMT operation for non-IPv6 savvy customers whose equipment may turn on 6to4 automatically and allows savvy customers to easily configure their way around the 6to4-PMT function.

Capable customers can also disable anycast based 6to4 entirely and use traditional 6to4 or other tunnelling mechanisms if they are so inclined. This is not considered the normal case, and most endpoints with auto-6to4 functions will be subject to 6to4-PMT operation since most users are unaware of it’s existence. 6to4-PMT is targeted as an option for stable IPv6 connectivity for average consumers.

4.2. Shared CGN Space Considerations

6to4-PMT operation can also be used to mitigate a known problem with 6to4 when shared address space [RFC6598] or Global Unicast Addresses (GUA) are used behind a CGN and not routed on the Internet. Non-RFC1918, yet un-routed (on interdomain links) address space would cause many deployed OSs and network equipment to potentially auto-enable 6to4 operation even without a valid return path (such as behind a CGN function). The Operators’ desire to use non-RFC1918 addresses, such as shared address space [RFC6598], is considered highly likely based on real world deployments.

Such hosts, in normal cases, would send 6to4 traffic to the IPv6 Internet via the anycast relay, which would in fact provide broken IPv6 connectivity since the return path flow is built using an IPv4 address that is not routed or assigned to the source Network. The use of 6to4-PMT would help reverse these effects by translating the 6to4 prefix to a provider assigned prefix, masking this automatic and undesired behaviour.

4.3. End to End Transparency

6to4-PMT mode operation removes the traditional end to end transparency of 6to4. Remote hosts would connect to a 6to4-PMT serviced host using a translated IPv6 address versus the original 6to4 address based on the 2002::/16 well-known prefix. This can be seen as a disadvantage of the 6to4-PMT system. This lack of transparency should also be contrasted with the normal operating state of 6to4 which provides uncontrolled and often high latency prone connectivity. The lack of transparency is however a better form of operation when extreme poor performance, broken IPv6 connectivity, or no IPv6 connectivity is considered as the alternative.

4.4. Path MTU Discovery Considerations

The MTU will be subject to a reduced value due to standard 6to4 tunnelling operation. Under normal 6to4 operation, the 6to4 service agent would send an ICMP Packet Too Big Message as part of Path MTU Discovery as described in [RFC4443] and [RFC1981] respectively. In 6to4-PMT operation, the PMT Service agent should be aware of the reduced 6to4 MTU and send ICMP messages using the translated address accordingly.

It is also possible to pre-constrain the MTU at the upstream router from the 6to4-PMT service agents which would then have the upstream router send the appropriate ICMP Packet Too Big Messages.

4.5. Checksum Management

Checksum management for 6to4-PMT can be implemented in one of two ways. The first deployment model is based on the stateless 6to4-PMT operational mode. In this case, checksum modifications are made using the method described in [RFC3022] section 4.2. The checksum is modified to match the parameters of the translated address of the source 6to4-PMT host. In the second deployment model where stateful 6to4-PMT translation is used, the vendor can implement checksum neutral mappings as defined in [RFC6296].

4.6. Application Layer Gateways

Vendors can choose to deploy ALGs on their platforms that perform 6to4-PMT if they so choose. No ALGs were deployed as part of the open source and vendor product deployments of 6to4-PMT. In the vendor deployment case, the same rules were used as with their NPTv6 [RFC6296] base code.

4.7. Routing Requirements

The provider would need to advertise the well-known IP address range used for normal anycast 6to4 [RFC3068] operation within the local IPv4 routing environment. This advertisement would attract the 6to4 upstream traffic to a local relay. To control this environment and make sure all northbound traffic lands on a provider controlled relay, the operator may filter the anycast range from being advertised from customer endpoints toward the local network (upstream propagation).

The provider would not be able to control route advertisements inside the customer domain, but that use case is not in scope for this document. It is likely in that case the end network/customer understands 6to4 and is maintaining their own relay environment and

therefore would not be subject to the operators 6to4 and/or PMT operation.

The provider would also likely want to advertise the 2002::/16 range within their own network to help bridge traditional 6to4 traffic within their own network (Native IPv6 to 6to4-PMT based endpoint). It would also be advised that the local 6to4-PMT operator not leak the well-known 6to4 anycast IPv4 prefix to neighbouring Autonomous Systems to prevent PMT operation for neighbouring networks. Policy configuration on the local 6to4-PMT relay can also be used to disallow PMT operation should the local provider service downstream customer networks.

4.8. Relay Deployments

The 6to4-PMT function can be deployed onto existing 6to4 relays (if desired) to help minimize network complexity and cost. 6to4-PMT has already been developed on Linux based platforms which are package add-ons to the traditional 6to4 code. The only additional considerations beyond normal 6to4 relay operation would include the need to route specific IPv6 provider prefix ranges used for 6to4-PMT operation towards peers and transit providers.

5. IANA Considerations

No IANA considerations are defined at this time.

6. Security Considerations

6to4-PMT operation would be subject to the same security concerns as normal 6to4 operation. 6to4-PMT is also not plainly perceptible by external hosts and local entities appear as Native IPv6 hosts to the external hosts.

7. Acknowledgements

Thanks to the following people for their textual contributions and/or guidance on 6to4 deployment considerations: Dan Wing, Wes George, Scott Beuker, JF Tremblay, John Brzozowski, Chris Metz and Chris Donley

Additional thanks to the following for assisting with the coding and testing of 6to4-PMT: Marc Blanchet, John Cianfarani, Tom Jefferd, Nik Lavorato, Robert Hutcheon and Ida Leung

8. References

8.1. Normative References

[RFC3056] Carpenter, B. and K. Moore, "Connection of IPv6 Domains via IPv4 Clouds", RFC 3056, February 2001.

[RFC3068] Huitema, C., "An Anycast Prefix for 6to4 Relay Routers", RFC 3068, June 2001.

8.2. Informative References

[RFC1981] McCann, J., Deering, S., and J. Mogul, "Path MTU Discovery for IP version 6", RFC 1981, August 1996.

[RFC3022] Srisuresh, P. and K. Egevang, "Traditional IP Network Address Translator (Traditional NAT)", RFC 3022, January 2001.

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

[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.

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

[RFC6264] Jiang, S., Guo, D., and B. Carpenter, "An Incremental Carrier-Grade NAT (CGN) for IPv6 Transition", RFC 6264, June 2011.

[RFC6296] Wasserman, M. and F. Baker, "IPv6-to-IPv6 Network Prefix Translation", RFC 6296, June 2011.

[RFC6333] Durand, A., Droms, R., Woodyatt, J., and Y. Lee, "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", RFC 6333, August 2011.

[RFC6343] Carpenter, B., "Advisory Guidelines for 6to4 Deployment", RFC 6343, August 2011.

[RFC6598] Weil, J., Kuarsingh, V., Donley, C., Liljenstolpe, C., and

M. Azinger, "IANA-Reserved IPv4 Prefix for Shared Address Space", BCP 153, RFC 6598, April 2012.

Authors' Addresses

Victor Kuarsingh (editor)
Rogers Communications
8200 Dixie Road
Brampton, Ontario L6T 0C1
Canada

Email: victor.kuarsingh@gmail.com
URI: http://www.rogers.com

Yiu L. Lee
Comcast
One Comcast Center
Philadelphia, PA 19103
U.S.A.

Email: yiu_lee@cable.comcast.com
URI: http://www.comcast.com

Olivier Vautrin
Juniper Networks
1194 N Mathilda Avenue
Sunnyvale, CA 94089
U.S.A.

Email: olivier@juniper.net
URI: http://www.juniper.net

Operations Area
Internet-Draft
Intended status: Informational
Expires: April 15, 2011

Y. Lee
Comcast
V. Kuarsingh
Rogers Communications
October 12, 2010

IPv6 Transition Cable Access Network Use Cases
draft-lee-v6ops-tran-cable-usecase-00

Abstract

This memo describes some use cases to transition to IPv6 in cable access network. This memo discusses enabling dual-stack to users over various types of network infrastructures. It also describes impacts to network, operation, CPE, and applications.

Table of Contents

1. Introduction

The Cable access network primarily uses DOCSIS technology defined by CableLabs to deliver IP services to users. DOCSIS provides an abstraction to deliver IP packets over coxial cable. DOCSIS is a shared media technology and use Ethernet for Layer-2, it doesn't use PPP or ATM for encapsulation.

A Cable Modem which is a DOCSIS enabled modem is the device to transmit the user's Ethernet frames over DOCSIS to the Cable Modem Termination System (CMTS) in the cable operator's network. DOCSIS has gone through few generations. The most current version is DOCSIS 3.0. By specifications, DOCSIS 2.0 and DOCSIS 3.0 both support IPv6 for cable modem management and user's traffic. However, DOCSIS 1.x specification and some older DOCSIS 2.0's implementations do not. Cable operators will take time to retire all the legacy cable modems and replace them to the newer version of cable modems. So there will be a transition period to upgrade all the equipments to support IPv6.

The complexity of upgrading the regional and core network to dual-stack is relatively low compared to upgrading the access network to support IPv6 for thousands of CMTSes and millions of cable modems and CPEs. So this memo focuses on use cases to enable IPv6 in the cable access network. The transition methodology is to provide dual-stack to the users regardless the underneath technology inside a cable operator. When IPv6 services become majority and IPv4 services gradually diminish, the operator may consider to provide only IPv6 to users and provide IPv4-IPv6 translation in the network when users access IPv4 services. This memo describes use cases to provide dual-stack to users because we have more experience.

We divide the use cases into two primary categories. The first category describes dual-stack deployment to the users using the existing access network. The access network could be IPv4-only or dual-stack. The second category describes dual-stack deployment to the users using IPv6-only access network. The goal of these use cases is providing service continuity during the transition.

2. Offer Dual-Stack on Top of Existing Access Network

We discuss three use cases that offer dual-stack to users. The first use case describes the scenario where the access network is IPv4-only and operators utilize tunneling technologies to give dual-stack access to users. The second use case describes the standard native dual-stack deployment model. The third use cases describes native dua-stack where the IPv4 connection may be provided using shared public IPv4 addresses (NAT444).

2.1. IPv4-only Access Network

According to [I-D.arkko-ipv6-transition-guidelines], native dual-stack is the simplest model for transition. However, this requires the entire network to be dual-stack. Moreover, the provisioning system and other support systems must be upgraded to support IPv6. Most operators will need to upgrade the network in phases along with the provisioning system(s) and supporting systems. During the transition period, there will be IPv4-only islands. In order to offer dual-stack access to users over IPv4 islands, operators may consider the use of tunneling technologies such as 6rd and MPLS.

There are incentives to offer IPv6 to users before completing the upgrade. For example: early IPv6 adopters can start experiencing IPv6 services and have connectivity to IPv6-only content should it be available. Operational groups can also being to familiarize themselves with IPv6 and being troubleshooting IPv6. Application developers and content providers can start providing services over IPv6. In the end, this may help to speedup the overall IPv6 adoption.

2.1.1. 6rd

2.1.1.1. Deployment Requirements

6rd [RFC5969] is a technology that provides IPv6 connectivity over the existing IPv4 access network. The idea is simple, it leverages the 6to4 model [RFC3056] and uses the provider’s specific prefix instead of the IANA assigned well-known prefix. This will give the operator’s control of both ingress and egress flows. This technology has been proven to be successful in real operator deployments [RFC5569].

6rd is comprised of two elements: 6rd-CE and 6rd-BR. 6rd-CE initiates an IPv6-in-IP tunnel to the 6rd-BR. 6rd-BR terminates the tunnel and forward the IPv6 packets to the IPv6 Internet. Similar to 6to4, 6rd uses the IPv4 address provisioned to the user to construct the IPv6 address. Since the IPv4 address is stored in the IPv6 prefix, the address translation is stateless.

6rd works when a user was provisioned with a public IPv4 address. It also works with [RFC1918] address when it is combined with a provider NAT44 function in the network. In this use case, we discuss only the public IPv4 address model.

2.1.1.2. Network Impact

This describes the egress connection from the 6rd-CE to the IPv6 Internet. After the IPv6 packet was encapsulated in an IPv4 packet by 6rd-CE, the network will forward the packet similar to any other IPv4 packet. 6rd model is transparent to the IPv4 network. The packet will eventually arrive in the closest 6rd-BR for decapsulation, then it will be forwarded to IPv6 destination. The "closest" 6rd-BR is defined by the IP address used in combination with network routing conditions.

This describes the ingress connection from the IPv6 Internet to 6rd-CE. IPv6 packet with 6rd prefix in the destination address will be forwarded normally and arrive to the closest 6rd-BR. The 6rd-BR extracts the IPv4 information from the IPv6 address and encapsulates the IPv6 packet in an IPv4 packet. Then, it will forward the encapsulated packet to the IPv4 network.

The 6rd prefix is advertised by the 6rd-BR or by an upstream router on it’s behalf. The operator will advertise this prefix within their network and towards the Internet and other neighboring peers. The operator also needs to assign an anycast address to the 6rd-BR. This anycast address will be shared by all the 6rd-BR and will be advertised in the operator’s IPv4 serving IGP. The 6rd-CE will send the encapsulated packets to this anycast address.

IPv6 packets are delivered on the IPv6-in-IP tunnel. MTU is a common consideration for any tunnel technology. Since 6rd is a stateless technology, the tunnel endpoints cannot perform fragmentation. The simplest solution is to increase default MTU size larger than 1500 bytes in the access network. More discussion can be found in [RFC5969].

Hosts behind the 6rd-CE may not be able to dynamically learn any DNS server via SLAAC, so they may query DNS from a DNS server in the IPv4 network. The DNS server in the IPv4 network should be configured process AAAA records.

2.1.1.3. Operation Impact

6rd is a stateless technology. It greatly simplifies the network design for scalability and high availability. Traffic engineering of the tunnels is not explicitly required since the 6rd-BRs are known via an IGP (or IGP assisted path). Operators can add or remove 6rd-BR in the network without transferring service states from one 6rd-BR to another 6rd-BR. Operators also need not assign any particular 6rd-BR to a 6rd-CE. 6rd-CE will rely on routing to find the closest 6rd-BR.

6rd is similar to VPN technology. 6rd packets are encapsulated and transparent to the network. Operator can operate, monitor and troubleshoot the 6rd network independently.

Considerations for 6rd include any in-line service or network device that monitors, controls or assists with traffic flows. Since 6rd sends IPv6 packets insider an IPv4 tunnel, all such systems must be 6rd aware to continue to supply the same functions for this new traffic type. Additionally, if an operator has enabled dynamic Q0S within their access network, the overall detection, policy and enforcement infrastructure will need to be able to manage the control of IPv6 flows within an IPv4 tunnel.

2.1.1.4. CPE Impact

CPE is required to implement the 6rd-CE specification. 6rd-CE must be the first device connecting to the cable modem and is responsible for learning the 6rd prefix and construct the 6rd delegated prefix. The CPE is also responsible to advertise the 6rd delegated prefix to hosts behind the CPE. If the CPE implements SLAAC, the hosts behind the CPE learns the prefix and default gateway via Router Advertisement. As with the network portion, any service information, including QoS, will need to be carefully managed to support the IPv6-in-IP function.

2.1.1.5. Application Impact

Applications will have dual-stack and should behave identically as of running on a native dual-stack host Application which are served via IPv6 will add additional load to BRs within the network. The operator may want to take this under consideration if they are planning to deploy high bandwidth services over IPv6. The operator may choose to offer some services over IPv4 in this case to lower the load on the BRs and allow for more efficient traffic delivery inside the network (since the BR and application systems may not share network locations).

2.1.2. MPLS

TBD

2.2. Native Dual-Stack Use Case

Providing native dual-stack to user may be the simplest for transition to IPv6, but it requires operators to upgrade the network, provisioning systems, and supporting systems to give production grade service to users. In this memo, native dual-stack means to provision a public IPv4 address, a global IPv6 address, and a global IPv6

prefix to a user.

2.2.1. IPv6 Address Design

In general, most of the IPv4 address architecture rules still apply to the IPv6 address architecture. For example: each service (e.g. VoIP vs. IPTV) should use different prefixes. Also, operators should use two separate prefixes for network infrastructure and customer services.

Due to the high utilization and the allocation policies of IPv4 prefixes, the result is each organization got many discontinuous blocks of prefixes rather than a large aggregate. The drawback is a fairly large Internet routing table. The overall IPv6 address pool is 128-bit long. Operators are normally given a prefix that contains an enormous number of addresses. If an operator carefully plans for address allocation and aggregation, it should only advertise the provider's prefix to the IPv6 Internet routing table. For example: each regional network should be a suffix of the overall provider's prefix. The result should be a smaller and more organized Internet routing table. In contrast, bad IPv6 address design may result a divided routing table and unnecessarily bubble its size.

2.2.2. Provisioning

TBD

2.2.3. Advertising Customer's Prefixes to the Access Network

Apart from an IPv6 address assignment to the CPE, the network will also delegate a prefix to the CPE for the hosts behind the CPE. This prefix is normally assigned by a DHCP server. The access network will need to learn the prefix and the associated cable modem and CPE. [I-D.droms-dhc-dhcpv6-agentopt-delegate] suggests that the DHCP Relay Agent which is the CMTS can query the DHCP server and learn the prefix. Then, it installs the prefix into its routing table. Another way is the DHCP Relay Agent inspects the DHCP IA_PD reply from the DHCP server and installs the prefix to the routing table. This topic remains open and more development is coming.

2.2.4. Benefits of Native Dual Stack

Utilizing a native dual stack option for IPv4 and IPv6 connectivity includes the overall integration ease for the provider. Although this option requires the deployment of IPv6, it is the more understood and support option. Other then standard IPv6 functionality within the network providers space and in the CPE, no new options are necessarily needed. Many inline services will need

to support IPv6, but are likely to support IPv6 native before newer connectivity options which includes DS-lite, 6rd and other such tunneling modes.

2.3. Native Dual-Stack with Shared IPv4 Addresses Use Case

This use case is an extension of the previous native dual stack option. In this particular case, all the IPv6 deployment considerations are made with an added complexity of shared IPv4 access. Shared IPv4 connectivity with a provider controlled NAT44 function may be required for dual stack deployments after IPv4 exhaustion. This option provides many of the same advantages as the native dual stack option which includes in the clear IPv4 and IPv6 flows. The provider can still utilize existing systems that support native IPv4 and IPv6 flows, but will need to add in network functionally related to the NAT44 function.

3. Offer Dual-Stack on IPv6-only Access Network

When the access network is IPv6-only, IPv6 traffic can be delivered natively over IPv6. So, there is no new requirement to enable IPv6. However, the access network will not be able to deliver IPv4 services. We provide two use cases to give dual-stack to users in an IPv6-only access network.

3.1. Shared IPv4 Address Use Case

When IPv4 addresses are limited, operators may consider multiplexing IPv4 addresses among internal users. Users will not be provisioned with a public IPv4 address. Instead, users will share a pool of public IPv4 addresses in the network.

DS-lite [I-D.ietf-softwire-dual-stack-lite] is a technology that provides IPv4 access over an IPv6-only access network. This also provides NAT44 functionality in the operator’s network to multiplex a pool of public IPv4 addresses amongst users.

3.1.1. DS-lite

3.1.1.1. Deployment Requirements

DS-lite is composed of two elements: B4 element and AFTR element. B4 element initiates an IP-in-IPv6 tunnel to the AFTR. AFTR terminates the tunnel and performs NAT44. B4 element can be implemented in a CPE or in a host. For this use case, we only discuss the CPE B4 element model.

An operator is required to deploy B4 to user premises. B4 will replace the existing CPE and must be the first network device in front of the cable modem. The operator will provision an IPv6 address to the B4 element. It will not provision any IPv4 address to the B4. Operator will also provision an IPv6 Prefix to the B4 and B4 will advertise this IPv6 prefix to the hosts behind it so that IPv6-capable hosts will have native IPv6 services.

B4 will run as DHCP server to the hosts behind it. It also acts as IPv4 default gateway and DNS proxy to the hosts. IPv4 packets will be delivered over the IP-in-IPv6 tunnel between the B4 and AFTR. From the host perspective, it will be provisioned with dual-stack and the applications running on the host can decide to use IPv4 or IPv6.

An operator is required to deploy a set of AFTR elements in the network. The AFTR should be dual-stack to terminate the IP-in-IPv6 tunnel from B4 elements and deliver NAT-ed packets to IPv4 Internet.

3.1.1.2. Network Impact

DS-lite requires the access network to support IPv6. This requires the CMTS and cable modem to be IPv6 enabled. It also requires to deploy a set of AFTR elements in the operator network. AFTR is a stateful network device, it inherits the cost to manage a stateful network device inside the network.

IPv4 packets are delivered on the IP-in-IPv6 tunnel. This reduces the effective MTU side. Neither hosts behind the B4 element nor services in front of the AFTR are aware of the tunnel. The operator can increase the MTU size in the access network. However, many cable modem implementations do not support MTU larger than default 1500 bytes, so the B4 and AFTR elements must handle fragmentation caused by the tunnel overhead.

The AFTR owns the NAT pool, it will be the aggregation point of the IPv4 addresses defined in the NAT pool. AFTR must advertise the NAT pool prefix to the IPv4 Internet. In contrast, the IPv6 tunnel interface should stay only inside the operator’s IGP and should not be advertised to the IPv6 Internet.

3.1.1.3. Operation Impact

DS-lite identifies a user by IPv6 address. Operators should be trained to understand how to map a user from an IPv6 address in the AFTR. AFTR is a NAT device, operator should maintain the NAT binding information to satisfy the government regulations. This is standard procedure for operating any NAT44 device.

DS-Lite introduces the operational mode where historical IPv4 connectivity (as experienced) is now totally dependent on IPv6. This significant change in operating conditions must be well understood by the operator. If DS-lite is introduced during deployment infancy in the operators IPv6 network, it will require careful attention to operational practices and capabilities to maintain the IPv6 network.

AFTR is critical to continuously offer IPv4 access in IPv6-only access network. Operator should scale AFTR to provide non-interruptive access to users. Operators should closely monitor two AFTR's resources: (1) Network Capacity and (2) Port Utilization. When network capacity is reached, the operator should decide to upgrade the AFTR to higher network capacity or to deploy a new AFTR to balance the workload. When port utilization is high, the operator should increase the NAT pool size.

AFTR is stateful, it will complicate the high-availability (HA) design. Operators should apply the standard HA design (e.g. cold or hot) which best fits to their network operations.

3.1.1.4. CPE Impact

CPE is required to implement the B4 element specification. Also, port-forwarding and UPnP IGD protocol will no longer function. IETF PCP Working Group was formed to address the port-forwarding and UPnP IGD issues.

CPE must know the IPv6 address of the AFTR tunnel interface. This information can be obtained from DHCP. Since there is only IPv6 access to the B4 element. Any IPv4 network service learned from DHCP must be proxy by the B4 element.

If the operator cannot increase the access network MTU size, the B4 element must handle fragmentation to ensure IPv4 service using maximum MTU size won't be affected by the tunnel overhead.

3.1.1.5. Application Impact

3.1.1.5.1. Egress Connection

Since hosts behind B4 are provisioned with dual-stack, the application can decide to use IPv4 or IPv6. If the external service is also dual-stack, the host will automatically prefer IPv6 over IPv4 if the host O/S has implemented [RFC3484]. If the host prefers IPv4 due to application logic, it will use the private IPv4 address provisioned by the B4 element. For applications expecting to use specific source port will be impacted because the AFTR inside the network won't be able to allocate a specific source port.

Applications use random source port will continue to function without modification.

3.1.1.5.2. Ingress Connection

Similar to traditional NAT, ingress connection will be blocked by default. The current techniques such as port-forwarding and UPnP IGD are required modification. Technically this could be done. But this will requires some changes in user's procedure to enable the service. It also adds cost to operators to offer port-forwarding service.

3.2. Public IPv4 Address Use Case

Some applications requires specific source port and some applications requires ingress connection. Users using those applications may want to be provisioned with a public IPv4 address to ease the potential challenges caused by NAT in the network. IPv4-over-IPv6 (4over6) [I-D.cui-softwire-host-4over6] is a simple technology to provision a public IPv4 address to a user and provide IPv4 access over an IPv6-only network.

3.2.1. IPv4 Over IPv6

3.2.1.1. Deployment Requirements

4over6 consists of two elements: 4over6 Initiator and 4over6 Tunnel Concentrator (TC). 4over6 is similar to DS-lite except two features: (1) Unlike B4 element, 4over6 Initiator will be provisioned with a public IPv4 address. (1) 4over6 TC only terminates the IP-in-IPv6 tunnel and won't perform any NAT44 function.

4over6 supports both host and CPE models. We will only discuss the 6over6 CPE model.

An operator is required to deploy 4over6 Initiator in premises. The 4over6 initiator will replace the existing CPE and must be the first network device in front of the cable modem. The operator will provide an IPv6 address and an IPv6 prefix to the CPE. The procedure is similar to Native IPv6 use case and DS-lite use case.

4over6 Initiator is very similar to the B4 element. It serves as DHCP server, IPv4 default gateway and DNS server to hosts behind it. The only difference is 4over6 will be provisioned with a public IPv4 address while B4 element will not. Once 4over6 Initiator discovers the 4over6 TC, it will issue standard DHCP request over the tunnel to the 4over6 TC. The 4over6 TC either relays the DHCP request to a centralized DHCP server or replies to the request if it is the authoritative DHCP server for the 4over6 service. Once the CPE

acquires the public IPv4 address, the user can run all his legacy IPv4 applications similar to what he is doing with a regular IPv4 home gateway.

3.2.1.2. Network Impact

Similar to DS-lite, the access network must support IPv6. This requires the CMTS and cable modem must be IPv6 enabled. It also requires the operator to deploy a set of 4over6 TC in the network.

Despite no NAT in the 6over4 TC, 6over4 TC is required to maintain the 4over6 Initiate IPv6 address (tunnel-id) and IPv4 address binding. Also, the 4over6 TC must advertise the IPv4 prefix to the Internet. It is the aggregation point of the IPv4 address prefix.

4over6 suffers the same MTU limitation which is common to any tunnel protocols. Please refer to Section 3.1.1.2 for details.

3.2.1.3. Operation Impact

Since each user will be assigned a public IPv4 address, it doesn’t require operator to log any binding. Operator should be able to identify a user by either IPv4 or IPv6 address.

Similar to AFTR, network capacity and IPv4 address utilization are critical resources to 4over6 TC. Operator must closely monitor the resources to ensure continuous IPv4 access.

Operators also need to coordinate the IPv4 address space in the DHCP server and the 4over6 Initiator which manages the space. This requires careful coordination and management.

3.2.1.4. CPE Impact

CPE is required to implement the 4over6 TC specification. Unlike B4 element, port-forwarding and the UPnP IGD will work without modification.

3.2.1.5. Application Impact

Applications will have dual-stack and should behave identically as of running on a native dual-stack host.

4. Security Considerations

TBD

5. Acknowledgements

TBD

6. IANA Considerations

This memo includes no request to IANA.

7. References

7.1. Normative References

[I-D.arkko-ipv6-transition-guidelines]
Arkko, J. and F. Baker, "Guidelines for Using IPv6 Transition Mechanisms during IPv6 Deployment", draft-arkko-ipv6-transition-guidelines-06 (work in progress), August 2010.

[I-D.cui-softwire-host-4over6]
Cui, Y., Wu, J., and P. Wu, "Host 4over6 for IPv6 host connecting IPv4 Internet", draft-cui-softwire-host-4over6-01 (work in progress), July 2010.

[I-D.ietf-softwire-dual-stack-lite]
Durand, A., Droms, R., Woodyatt, J., and Y. Lee, "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", draft-ietf-softwire-dual-stack-lite-06 (work in progress), August 2010.

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

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

7.2. Informative References

[I-D.droms-dhc-dhcpv6-agentopt-delegate]
Droms, R., "DHCP Relay Agent Assignment Notification Option", draft-droms-dhc-dhcpv6-agentopt-delegate-00 (work in progress), November 2005.

[RFC1918] Rekhter, Y., Moskowitz, R., Karrenberg, D., Groot, G., and E. Lear, "Address Allocation for Private Internets",

BCP 5, RFC 1918, February 1996.

[RFC3056] Carpenter, B. and K. Moore, "Connection of IPv6 Domains via IPv4 Clouds", RFC 3056, February 2001.

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

[RFC5569] Despres, R., "IPv6 Rapid Deployment on IPv4 Infrastructures (6rd)", RFC 5569, January 2010.

Authors' Addresses

Yiu L. Lee
Comcast
One Comcast Center
Philadelphia, PA 19103
U.S.A.

Email: yiu_lee@cable.comcast.com
URI: http://www.comcast.com

Victor Kuarsingh
Rogers Communications
8200 Dixie Road
Brampton, Ontario L6T 0C1
Canada

Email: victor.kuarsingh@rci.rogers.com
URI: http://www.rogers.com

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: April 25, 2011

J. Livingood
Comcast

October 22, 2010

IPv6 AAAA DNS Whitelisting Implications
draft-livingood-dns-whitelisting-implications-01

Abstract

The objective of this document is to describe what whitelisting of DNS AAAA resource records is, or DNS whitelisting for short, as well as what the implications of this emerging practice are and what alternatives may exist. The audience for this document is the Internet community generally, including the IETF and IPv6 implementers.

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, 2011.

Copyright Notice

Table of Contents

1. Introduction

[EDITORIAL: This is a rough first -00 draft. Some sections have not yet been completed but will be soon. Suggestions on all parts of this document are eagerly solicited.]

This document describes the emerging practice of whitelisting of DNS AAAA resource records (RRs), or DNS whitelisting for short. It also explores the implications of this emerging practice are and what alternatives may exist.

The practice of DNS whitelisting appears to have first been used by major web content sites. These web site operators observed that when they added AAAA RRs to their authoritative DNS servers that a small fraction of end users had slow or otherwise impaired access to a given web site with both AAAA and A RRs. The fraction of users with such impaired access has been estimated to be roughly 0.078% of total Internet users [IETF 77 DNSOP WG Presentation] [Network World Article on IETF 77 DNSOP WG Presentation]. Thus, in an example Internet Service Provider (ISP) network of 10 million users, approximately 7,800 of those users may experience such impaired access.

As a result of this impairment affecting end users of a given domain, a few large web site operators have begun to either implement DNS whitelisting or strongly consider the implementation of DNS whitelisting [Network World Article on DNS Whitelisting]. When implemented, DNS whitelisting in practice means that a domain's authoritative DNS will return a AAAA RR to DNS recursive resolvers [RFC1035] on the whitelist, while returning no AAAA RRs to DNS resolvers which are not on the whitelist. It is important to note that these web site operators are motivated to maintain a high-quality user experience for all of their users, and that they are attempting to shield users with impaired access from the symptoms of these impairments that would negatively affect their access to certain websites and related Internet resources.

[EDITORIAL: change web site operators --> domain operators?]

However, critics of this emerging practice of DNS whitelisting have articulated several concerns. Among these are that this is a very different behavior from the current practice concerning the publishing of IPv4 address records, that it may create a two-tiered Internet, that policies concerning whitelisting and de-whitelisting are opaque, that DNS whitelisting reduces interest in the deployment of IPv6, that new operational and management burdens are created, and that the costs and negative implications of DNS whitelisting outweigh the perceived benefits as compared to fixing underlying impairments.

This document explores the reasons and motivations for DNS whitelisting. It also explores the concerns regarding this emerging practice. As a result, readers can hopefully better understand what DNS whitelisting is, why some parties are implementing it, and why other parties are critical of the practice.

2. How DNS Whitelisting Works

DNS whitelisting is implemented in authoritative DNS servers, where those servers implement IP address-based restrictions on AAAA query responses, which contain IPv6 addresses. In practice DNS whitelisting has been primarily implemented by web server operators. For a given operator of the website www.example.com, that operator essentially applies an access control list (ACL) on their authoritative DNS servers, which are authoritative for the domain example.com. The ACL is then configured with the IPv4 and/or IPv6 addresses of DNS recursive resolvers on the Internet, which have been authorized to be added to the ACL and to therefore receive AAAA RR responses. These DNS recursive resolvers are operated by other parties, such as ISPs, universities, governments, businesses, individual end users, etc. If a DNS recursive resolver IS NOT on the ACL, then NO AAAA RRs with IPv6 addresses will be sent in response to a query for a given hostname in the example.com domain. However, if a DNS recursive resolver IS on the ACL, then AAAA RRs with IPv6 addresses will be sent in response to a query for a given hostname in the example.com domain.

In practice this generally means that a very small fraction of the DNS recursive resolvers on the Internet can receive AAAA responses with IPv6 addresses, which means that the large majority of DNS resolvers on the Internet will receive only A RRs with IPv4 addresses. Thus, quite simply, the authoritative server hands out different answers depending upon who is asking; with IPv4 and IPv6 records for some on the authorized whitelist, and only IPv4 records for everyone else. See Figure 1 and Figure 2 for two different visual descriptions of how this works in practice.

Finally, DNS whitelisting can be deployed in two primary ways: universally on a global basis, or on an ad hoc basis. These two potential deployment models are described in Section 5.

1: The authoritative DNS server for example.com receives a DNS query for www.example.com, for which both A (IPv4) and AAAA (IPv6) address records exist.
2: The authoritative DNS server examines the IP address of the DNS recursive resolver sending the query.
3: The authoritative DNS server checks this IP address against the access control list (ACL) that is the DNS whitelist.
4: If the DNS recursive resolver’s IP address IS listed in the ACL, then the response to that specific DNS recursive resolver can contain both A (IPv4) and AAAA (IPv6) address records.
5: If the DNS recursive resolver’s IP address IS NOT listed in the ACL, then the response to that specific DNS recursive resolver can contain only A (IPv4) address records and therefore cannot contain AAAA (IPv6) address records.

Figure 1: DNS Whitelisting - System Logic

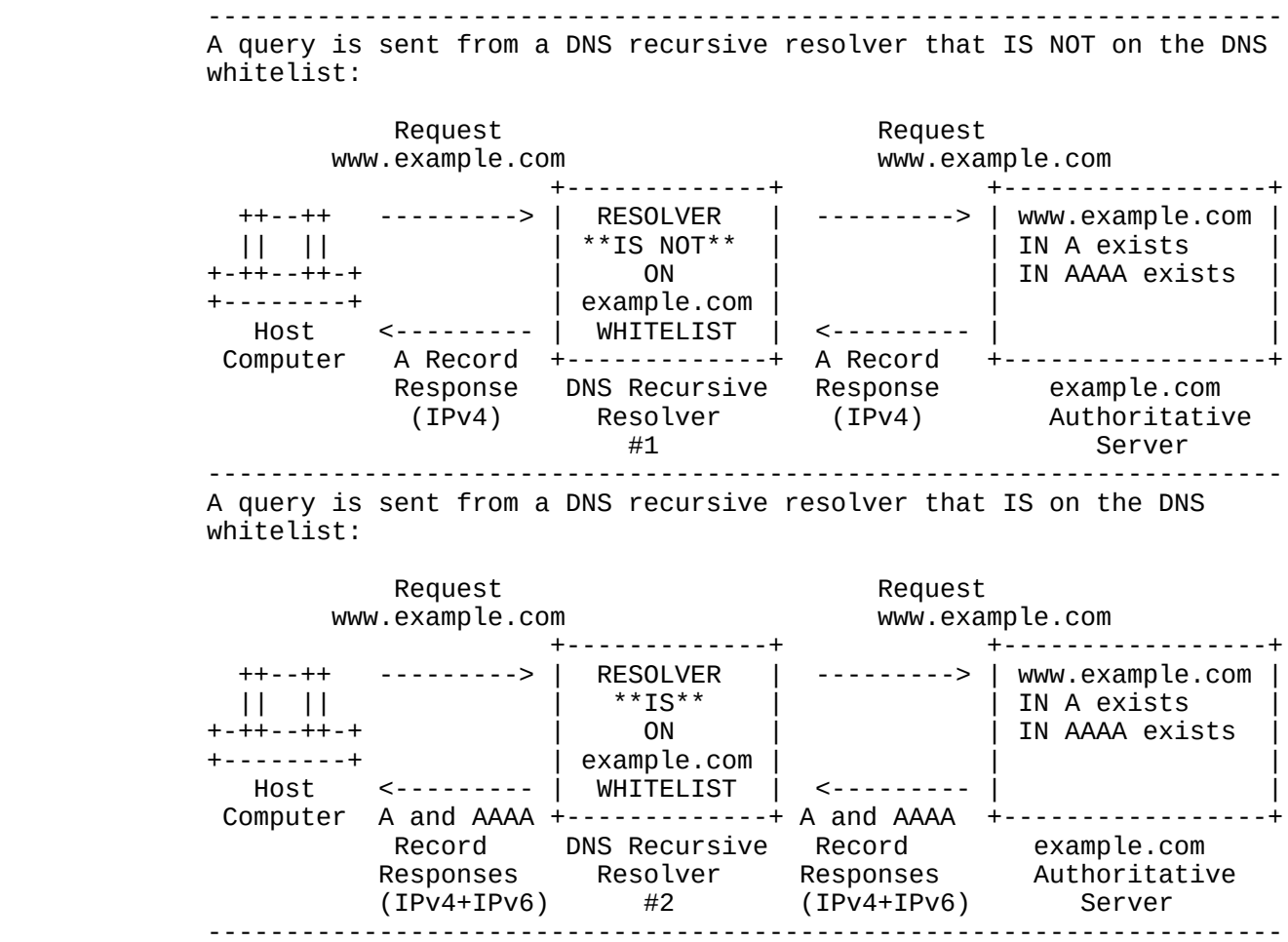


Figure 2: DNS Whitelisting - Functional Diagram

3. Concerns Regarding DNS Whitelisting

There are a number of potential implications relating to DNS whitelisting, which have raised various concerns in some parts of the Internet community. Many of those potential implications are described in Section 7.

Some parties in the Internet community are concerned that this emerging practice of DNS whitelisting for IPv6 address records could represent a departure from the generally accepted practices regarding IPv4 address records in the DNS on the Internet. These parties

explain their belief that for A records, containing IPv4 addresses, once an authoritative server operator adds the A record to the DNS, then any DNS recursive resolver on the Internet can receive that A record in response to a query. By extension, this means that any of the hosts connected to any of these DNS recursive resolvers can receive the IPv4 address records for a given FQDN. This enables new server hosts which are connected to the Internet, and for which a fully qualified domain name (FQDN) such as www.example.com has been added to the DNS with an IPv4 address record, to be almost immediately reachable by any host on the Internet. In this case, these new servers hosts become more and more widely accessible as new networks and new end user hosts connect to the Internet over time [EDITORIAL: consider reference to network effects]. It also means that the new server hosts do not need to know about these new networks and new end user hosts in order to make their content and applications available to them, in essence that each end in this end-to-end model is responsible for connecting to the Internet and once they have done so they can connect to each other without additional impediments or middle networks or intervening networks or servers knowing about these end points and whether one is allowed to contact the other.

In contrast, these parties are concerned that DNS whitelisting may fundamentally change this model. As a result, in this altered end-to-end model, one end (where the end user is located) cannot readily connect to the other end (where the content is located), without parts of the middle used by one end being known by the other end and approved for access to that end. Thus, as new networks connect to the Internet over time, those networks need to contact any and all domains which have implemented DNS whitelisting in order to apply to be added to their DNS whitelist, in the hopes of making the content and applications residing on named server hosts in those domains accessible by the end user hosts on that new network. Furthermore, this same need to contact all domains implementing DNS whitelisting also applies to all existing networks connected to the Internet.

Therefore, these concerned parties explain, whereas in the current IPv4 Internet when a new server host is added to the Internet it is widely available to all end user hosts and networks, when DNS whitelisting of IPv6 records is used then these new server hosts are not accessible to any end user hosts or networks until such time as the operator of the authoritative DNS servers for those new server hosts expressly authorizes access to those new server hosts by adding DNS recursive resolvers around the Internet to the ACL. This could represent a significant change in reachability of content and applications by end users and networks as these end user hosts and networks transition to IPv6. Therefore, a concern expressed is that if much of the content that end users are most interested in is not

accessible as a result, then end users and/or networks may resist adoption of IPv6 or actively seek alternatives to it, such as using multi-layer network address translation (NAT) techniques like NAT444 [I-D.shirasaki-nat444] on a long-term basis. There is also concern that this practice also could disrupt the continued increase in Internet adoption by end users if they cannot simply access new content and applications but must instead contact the operator of their DNS recursive resolver, such as their ISP or another third party, to have their DNS recursive resolver authorized for access to the content or applications that interests them. Meanwhile, these parties say, over 99.9% of all other end users that are also using that same network or DNS recursive resolver are unable to access the IPv6-based content, despite their experience being a positive one.

[EDITORIAL: Are there additional concerns to add here?]

4. Similarities to Split DNS

DNS whitelisting as described herein is in some ways similar to so-called split DNS, which is briefly described in Section 3.8 of [RFC2775]. When split DNS is used, the authoritative DNS server returns different responses depending upon what host has sent the query. While [RFC2775] notes the typical use of split DNS is to provide one answer to hosts on an Intranet and a different answer to hosts on the Internet, the essence is that different answers are provided to hosts on different networks. This is basically the way that DNS whitelisting works, in so far as hosts of different networks, which use different DNS recursive resolvers, receive different answers if one DNS recursive resolver is on the whitelist and the other is not. Thus, in a way, DNS whitelisting could in some ways be considered split DNS on the public Internet, though with some differences.

In [RFC2956], Internet transparency and Internet fragmentation concerns regarding split DNS are detailed in Section 2.1. [RFC2956] further notes in Section 2.7, concerns regarding split DNS and that it "makes the use of Fully Qualified Domain Names (FQDNs) as endpoint identifiers more complex." Section 3.5 of [RFC2956] further recommends that maintaining a stable approach to DNS operations is key during transitions such as the one to IPv6 that is underway now, stating that "Operational stability of DNS is paramount, especially during a transition of the network layer, and both IPv6 and some network address translation techniques place a heavier burden on DNS."

5. Likely Deployment Scenarios

In considering how DNS whitelisting may emerge more widely, there are two likely deployment scenarios, which are explored below.

5.1. Deploying DNS Whitelisting Universally

The least likely deployment scenario is one where DNS whitelisting becomes a standardized process across all authoritative DNS servers, across the entire Internet. While this scenario is the least likely, due to some parties not sharing the concerns that have so far motivated the use of DNS whitelisting, it is nonetheless conceivable that it could be one of the ways in which DNS whitelisting may be deployed.

In order for this deployment scenario to occur, it is likely that DNS whitelisting functionality would need to be built into all authoritative DNS server software, and that all operators of authoritative DNS servers would have to upgrade their software and enable this functionality. Furthermore, it is likely that new Internet Draft documents would need to be developed which describe how to properly configure, deploy, and maintain DNS whitelisting. As a result, it is unlikely that DNS whitelisting would, at least in the next several years, become universally deployed. Furthermore, these DNS whitelists are likely to vary on a domain-by-domain basis, depending upon a variety of factors. Such factors may include the motivation of each domain owner, the location of the DNS recursive resolvers in relation to the source content, as well as various other parameters that may be transitory in nature, or unique to a specific end user host type. Thus, it is probably unlikely that a single clearinghouse for managing whitelisting is possible; it will more likely be unique to the source content owners and/or domains which implement DNS whitelists.

While this scenario may be unlikely, it may carry some benefits. First, parties performing troubleshooting would not have to determine whether or not DNS whitelisting was being used, as it always would be in use. In addition, if universally deployed, it is possible that the criteria for being added to or removed from a DNS whitelist could be standardized across the entire Internet. Nevertheless, even if uniform DNS whitelisting policies were not standardized, is also possible that a central registry of these policies could be developed and deployed in order to make it easier to discover them, a key part of achieving transparency regarding DNS whitelisting.

[EDITORIAL: Are there additional benefits or challenges to add here?]

5.2. Deploying DNS Whitelisting On An Ad Hoc Basis

This is the most likely deployment scenario for DNS whitelisting, as it seems today, is where some interested parties engage in DNS whitelisting but many or most others do not do so. What can make this scenario challenging from the standpoint of a DNS recursive resolver operator is determining which domains implement DNS whitelisting, particularly since a domain may not do so as they initially transition to IPv6, and may instead do so later. Thus, a DNS recursive resolver operator may initially believe that they can receive AAAA responses with IPv6 addresses as a domain adopts IPv6, but then notices via end user reports that they no longer receive AAAA responses due to that site adopting DNS whitelisting.

Thus, in contrast to universal deployment of DNS whitelisting, deployment on an ad hoc basis is likely to be significantly more challenging from an operational, monitoring, and troubleshooting standpoint. In this scenario, a DNS recursive resolver operator will have no way to systematically determine whether DNS whitelisting is or is not implemented for a domain, since the absence of AAAA records with IPv6 addresses may simply be indicative that the domain has not yet added IPv6 addressing for the domain, not that they have done so but have restricted query access via DNS whitelisting. As a result, discovering which domains implement DNS whitelisting, in order to differentiate them from those that do not, is likely to be challenging.

On the other hand, one benefit of DNS whitelisting being deployed on an ad hoc basis is that only the domains that are interested in doing so would have to upgrade their authoritative DNS servers in order to implement the ACLs necessary to perform DNS whitelisting.

[EDITORIAL: Additional benefits or challenges to add?]

6. What Problems Are DNS Whitelisting Implementers Trying To Solve?

As noted in Section 1, domains which implement DNS whitelisting are attempting to protect a few users of their domain, which happen to have impaired IPv6 access, from having a negative end user experience. While it is outside the scope of this document to explore the various reasons why a particular user may experience impaired IPv6 access, for the users which experience this it is a very real effect and would of course affect access to all or most IPv4 and IPv6 dual stack servers. This negative end user experience can range from someone slower than usual (as compared to native IPv4-based access), to extremely slow, to no access to the domain whatsoever.

Thus, parties which implement DNS whitelisting are attempting to provide a good experience to these end users. While one can debate whether DNS whitelisting is the optimal solution, it is quite clear that DNS whitelisting implementers are extremely interested in the performance of their services for end users as a primary motivation.

[EDITORIAL 1: More motivations to add?]

[EDITORIAL 2:Any good external references to consider adding?]

7. Implications of DNS Whitelisting

There are many potential implications of DNS whitelisting. In the sections below, the key potential implications are listed in some detail.

7.1. Architectural Implications

DNS whitelisting could be perceived as somewhat modifying the end-to-end model that prevails on the IPv4 Internet today. This approach moves additional access control information and policies into the middle of the network on the IPv6-addressed Internet, which did not exist before on the IPv4-addressed Internet. This could raise some risks noted in [RFC3724], which in explaining the history of the end-to-end principle [RFC1958] explains that one of the goals is to minimize the state, policies, and other functions needed in the middle of the network in order to enable end-to-end communications on the Internet.

It is also possible that DNS whitelisting could place at risk some of the benefits of the end-to-end principle, as listed in Section 4.1 of [RFC3724], such as protection of innovation. Further, while [RFC3234] details issues and concerns regarding so-called middleboxes, there may be parallels to DNS whitelisting, especially concerning modified DNS servers noted in Section 2.16 of [RFC3234], and more general concerns noted in Section 1.2 of [RFC3234] about the introduction of new failure modes, that configuration is no longer limited to two ends of a session, and that diagnosis of failures and misconfigurations is more complex.

In [Tussle in Cyberspace], the authors note concerns regarding the introduction of new control points, as well as "kludges" to the DNS, as risks to the goal of network transparency in the end-to-end model. Some parties concerned with the emerging use of DNS whitelisting have shared similar concerns, which may make [Tussle in Cyberspace] an interesting and relevant document. In addition, [Rethinking the design of the Internet] reviews similar issues that may be of

interest to readers of this document.

In order to explore and better understand these high-level architectural implications and concerns in more detail, the following sections explore more specific potential implications.

7.2. Public IPv6 Address Reachability Implications

The predominant experience of end user hosts and servers on the IPv4-addressed Internet today is that, very generally speaking, when a new server with a public IPv4 address is added, that it is then globally accessible by IPv4-addressed hosts. For the purposes of this document, that concept can be considered "pervasive reachability". It has so far been assumed that the same expectations of reachability would exist in the IPv6-addressed Internet. However, if DNS whitelisting is deployed, this will not be the case since only end user hosts using DNS recursive resolvers which have been added to the ACL of a given domain using DNS whitelisting would be able to reach new servers in that given domain via IPv6 addresses.

Thus, the expectation of any end user host being able to connect to any server (essentially both hosts, just at either end of the network), defined here as "pervasive reachability", will change to "restricted reachability" with IPv6.

[EDITORIAL: Additional implications?]

7.3. Operational Implications

This section explores some of the operationally related implications which may occur as a result of, related to, or necessary when engaging in the practice of DNS whitelisting.

7.3.1. De-Whitelisting May Occur

If it is possible for a DNS recursive resolver to be added to a whitelist, then it is also possible for that resolver to be removed from the whitelist, also known as de-whitelisting. Since de-whitelisting can occur, whether through a decision by the authoritative server operator or the domain owner, or even due to a technical error, an operator of a DNS recursive resolver will have new operational and monitoring requirements and/or needs as noted in Section 7.3.3, Section 7.3.4, Section 7.3.5, and Section 7.5.

7.3.2. Authoritative DNS Server Operational Implications

Operators of authoritative servers may need to maintain an ACL a server-wide basis affecting all domains, on a domain-by-domain basis,

as well as on a combination of the two. As a result, operational practices and software capabilities may need to be developed in order to support such functionality. In addition, processes may need to be put in place to protect against inadvertently adding or removing IP addresses, as well as systems and/or processes to respond to such incidents if and when they occur. For example, a system may be needed to record DNS whitelisting requests, report on their status along a workflow, add IP addresses when whitelisting has been approved, remove IP addresses when they have been de-whitelisted, log the personnel involved and timing of changes, schedule changes to occur in the future, and to roll back any inadvertent changes.

Such operators may also need implement new forms of monitoring in order to apply change control, as noted briefly in Section 7.3.4.

[EDITORIAL: Additional implications?]

7.3.3. DNS Recursive Resolver Server Operational Implications

Operators of DNS recursive resolvers, which may include ISPs, enterprises, universities, governments, individual end users, and many other parties, are likely to need to implement new forms of monitoring, as noted briefly in Section 7.3.4. But more critically, such operators may need to add people, processes, and systems in order to manage countless DNS whitelisting applications, for all domains that the end users of such servers are interested in now or in which they may be interested in the future. As such anticipation of interesting domains is likely infeasible, it is more likely that such operators may either choose to only apply to be whitelisted for a domain based upon one or more end user requests, or that they will attempt to do so for all domains.

When such operators apply for DNS whitelisting for all domains, that may mean doing so for all registered domains. Thus, some system would have to be developed to discover whether each domain has been whitelisted or not, which is touched on in Section 5 and may vary depending upon whether DNS whitelisting is universally deployed or is deployed on an ad hoc basis.

Furthermore, these operators will need to develop processes and systems to track the status of all DNS whitelisting applications, respond to requests for additional information related to these applications, determine when and if applications have been denied, manage appeals, and track any de-whitelisting actions. Given the incredible number of domains in existence, the ease with which a new domain can be added, and the continued strong growth in the numbers of new domains, readers should not underestimate the potential significance in personnel and expense that this could represent for

such operators. In addition, it is likely that systems and personnel may also be needed to handle new end user requests for domains for which to apply for DNS whitelisting, and/or inquiries into the status of a whitelisting application, reports of de-whitelisting incidents, general inquiries related to DNS whitelisting, and requests for DNS whitelisting-related troubleshooting by these end users.

[EDITORIAL: Additional implications?]

7.3.4. Monitoring Implications

Once a DNS recursive resolver has been whitelisted for a particular domain, then the operator of that DNS recursive resolver may need to implement monitoring in order to detect the possible loss of whitelisting status in the future. This DNS recursive resolver operator could configure a monitor to check for a AAAA response in the whitelisted domain, as a check to validate continued status on the DNS whitelist. The monitor could then trigger an alert if at some point the AAAA responses were no longer received, so that operations personnel could begin troubleshooting, as outlined in Section 7.3.5.

Also, authoritative DNS server operators are likely to need to implement new forms of monitoring. In this case, they may desire to monitor for significant changes in the size of the whitelist within a certain period of time, which might be indicative of a technical error such as the entire ACL being removed. These operators may also wish to monitor their workflow process for reviewing and acting upon DNS whitelisting applications and appeals, potentially measuring and reporting on service level commitments regarding the time an application or appeal can remain at each step of the process, regardless of whether or not such information is shared with parties other than that authoritative DNS server operator.

These are but a few examples of the types of monitoring that may be called for as a result of DNS whitelisting, among what are likely many other types and variations.

[EDITORIAL: Additional implications?]

7.3.5. Troubleshooting Implications

The implications of DNS whitelisted present many challenges, which have been detailed in Section 7. These challenges may negatively affect the end users' ability to troubleshoot, as well as that of DNS recursive resolver operators, ISPs, content providers, domain owners (where they may be different from the operator of the authoritative DNS server for their domain), and other third parties. This may make

the process of determining why a server is not reachable significantly more complex.

[SECTION INCOMPLETE - MIGHT LIKE TO ADD SOME EXAMPLES HERE]

[EDITORIAL: Additional implications?]

7.3.6. Additional Implications If Deployed On An Ad Hoc Basis

[SECTION INCOMPLETE - IS THIS NEEDED? - PLACEHOLDER FOR NOW]

[EDITORIAL: Additional implications?]

7.4. Homogeneity May Be Encouraged

A broad trend which has existed on the Internet appears to be a move towards increasing levels of heterogeneity. One manifestation of this is in an increasing number, variety, and customization of end user hosts, including home network, operating systems, client software, home network devices, and personal computing devices. This trend appears to have had a positive effect on the development and growth of the Internet. A key facet of this that has evolved is the ability of the end user to connect any technically compliant device or use any technically compatible software to connect to the Internet. Not only does this trend towards greater heterogeneity reduce the control which is exerted in the middle of the network, described in positive terms in [Tussle in Cyberspace], [Rethinking the design of the Internet], and [RFC3724], but it can also help to enable greater and more rapid innovation at the edges.

An unfortunate implication of the adoption of DNS whitelisting may be the encouragement of a reversal of this trend, which would be a move back towards greater levels of homogeneity. In this case, a domain owner which has implemented DNS whitelisting may prefer greater levels of control be exerted over end user hosts (which broadly includes all types of end user software and hardware) in order to attempt to enforce technical standards relating to establishing certain IPv6 capabilities or the enforcing the elimination of or restriction of certain end user hosts. While the domain operator is attempting to protect, maintain, and/or optimize the end user experience for their domain, the collective result of many domains implementing DNS whitelisting, or even a few important domains implementing DNS whitelisting, may be to encourage a return to more homogenous and/or controlled end user hosts. Unfortunately, this could have unintended side effects on and counter-productive implications for future innovation at the edges of the network.

7.5. Technology Policy Implications

A key technology policy implication concerns the policies relating to the process of reviewing an application for DNS whitelisting, and the decision-making process regarding whitelisting for a domain. Important questions may include whether these policies have been fully and transparently disclosed, are non-discriminatory, and are not anti-competitive. A related implication is whether and what the process for appeals is, when a domain decides not to add a DNS recursive resolver to the whitelist. Key questions here may include whether appeals are allowed, what the process is, what the expected turn around time is, and whether the appeal will be handled by an independent third party or other entity/group.

A further implications arises when de-whitelisting occurs. Questions that may naturally be raised in such a case include whether the criteria for de-whitelisting have been fully and transparently disclosed, are non-discriminatory, and are not anti-competitive. Additionally, the question of whether or not there was a cure period available prior to de-whitelisting, during which troubleshooting activities, complaint response work, and corrective actions may be attempted, and whether this cure period was a reasonable amount of time.

It is also conceivable that whitelisting and de-whitelisting decisions could be quite sensitive to concerned parties beyond the operator of the domain which has implemented DNS whitelisting and the operator of the DNS recursive resolver, including end users, application developers, content providers, advertisers, public policy groups, governments, and other entities, which may also seek to become involved in or express opinions concerning whitelisting and/or de-whitelisting decisions. Lastly, it is conceivable that any of these interested parties or other related stakeholders may seek redress outside of the process a domain has establishing for DNS whitelisting and de-whitelisting.

A final concern is that decisions relating to whitelisting and de-whitelisting may occur as an expression of other commercial, governmental, and/or cultural conflicts, given the new control point which has be established with DNS whitelisting. For example, in one imagined scenario, it may be conceivable that one government is unhappy with a news story or book published in a particular country, and that this government may retaliate against or protest this news story or book by requiring domains operating within that government’s territory to de-whitelist commercial, governmental, or other entities involved in or related to (however tangentially) publishing the news story or book. By the same token, a news site operating in multiple territories may be unhappy with governmental policies in one

particular territory and may choose to express dissatisfaction in that territory by de-whitelisting commercial, governmental, or other entities in that territory. Thus, it seems possible that DNS whitelisting and de-whitelisting could become a vehicle for adjudicating other disputes, and that this may well have intended and unintended consequences for end users which are affected by such decisions and are unlikely to be able to express a strong voice in such decisions.

7.6. IPv6 Adoption Implications

As noted in Section 3, the implications of DNS whitelisting may drive end users and/or networks to delay, postpone, or cancel adoption of IPv6, or to actively seek alternatives to it. Such alternatives may include the use of multi-layer network address translation (NAT) techniques like NAT444 [I-D.shirasaki-nat444], which these parties may decide to pursue on a long-term basis to avoid the perceived costs and aggravations related to DNS whitelisting. This could of course come at the very time that the Internet community is trying to get these very same parties interested in IPv6 and motivated to begin the transition to IPv6. As a result, parties concerned over the negative implications of DNS whitelisting have said they are very concerned of the negative effects that this practice could have on the adoption of IPv6 if it became widespread or was adopted by key parties in the Internet ecosystem.

[EDITORIAL: Additional implications?]

8. Solutions

8.1. Implement DNS Whitelisting Universally

One obvious solution is to implement DNS whitelisted universally, and to do so using some sort of centralized registry of DNS whitelisting policies, contracts, processes, or other information. This potential solution seems unlikely at the current time.

[EDITORIAL: More to add?]

8.2. Implement DNS Whitelisting On An Ad Hoc Basis

If DNS whitelisting was to be adopted more widely, it is likely to be adopted on this ad hoc, or domain-by-domain basis. Therefore, only those domains interested in DNS whitelisting would need to adopt the practice, though as noted herein discovering that they a given domain has done so may be problematic.

[EDITORIAL: More to add?]

8.3. Do Not Implement DNS Whitelisting

As an alternative to adopting DNS whitelisting, the Internet community can instead choose to take no action whatsoever, perpetuating the current predominant authoritative DNS operational model on the Internet, and leave it up to end users with IPv6-related impairments to discover and fix those impairments.

8.3.1. Solving Current End User IPv6 Impairments

A further extension of not implementing DNS whitelisting, is to also endeavor to actually fix the underlying technical problems that have prompted the consideration of DNS whitelisting in the first place, as an alternative to trying to apply temporary workarounds to avoid the symptoms of underlying end user IPv6 impairments. A first step is obviously to identify which users have such impairments, which would appear to be possible, and then to communicate this information to end users. Such end user communication is likely to be most helpful if the end user is not only alerted to a potential problem but is given careful and detailed advice on how to resolve this on their own, or where they can seek help in doing so.

One challenge with this option is the potential difficulty of motivating members of the Internet community to work collectively towards this goal, sharing the labor, time, and costs related to such an effort. Of course, since just such a community effort is now underway for IPv6, it is possible that this would call for only a moderate amount of additional work.

[EDITORIAL: More to add?]

9. Security Considerations

There are no particular security considerations if DNS whitelisting is not adopted, as this is how the public Internet works today with A records.

However, if DNS whitelisting is adopted, organizations which apply DNS whitelisting policies in their authoritative servers should have procedures and systems which do not allow unauthorized parties to either remove whitelisted DNS resolvers from the whitelist or add non-whitelisted DNS resolvers to the whitelist. Should such unauthorized additions or removals from the whitelist can be quite damaging, and result in content providers and/or ISPs to incur substantial support costs resulting from end user and/or customer

contacts. As such, great care must be taken to control access to the whitelist for an authoritative server.

In addition, two other key security-related issues should be taken into consideration:

9.1. DNSSEC Considerations

DNS security extensions defined in [RFC4033], [RFC4034], and [RFC4035] use cryptographic digital signatures to provide origin authentication and integrity assurance for DNS data. This is done by creating signatures for DNS data on a Security-Aware Authoritative Name Server that can be used by Security-Aware Resolvers to verify the answers. Since DNS whitelisting is implemented on an authoritative server, which provides different answers depending upon which resolver server has sent a query, the DNSSEC chain of trust is not altered. Therefore there are no DNSSEC implications per se, and thus no specific DNSSEC considerations to be listed.

9.2. Authoritative DNS Response Consistency Considerations

[INCOMPLETE!!]

While Section 9.1 does not contain any specific DNSSEC considerations. However, it is certainly conceivable that security concerns may arise when end users or other parties notice that the responses sent from an authoritative DNS server appear to vary from one network or one DNS recursive resolver to another. This may give rise to concerns that, since the authoritative responses vary that there is some sort of security issue and/or some or none of the responses can be trusted.

10. IANA Considerations

There are no IANA considerations in this document.

11. Contributors

The following people made significant textual contributions to this document and/or played an important role in the development and evolution of this document:

John Brzozowski

Chris Griffiths

Tom Klieber

Yiu Lee

Rich Woundy

12. Acknowledgements

The authors and contributors also wish to acknowledge the assistance of the following individuals in helping us to develop and/or review this document:

13. References

13.1. Normative References

[RFC1035] Mockapetris, P., "Domain names - implementation and specification", STD 13, RFC 1035, November 1987.

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

[RFC2775] Carpenter, B., "Internet Transparency", RFC 2775, February 2000.

[RFC2956] Kaat, M., "Overview of 1999 IAB Network Layer Workshop", RFC 2956, October 2000.

[RFC3234] Carpenter, B. and S. Brim, "Middleboxes: Taxonomy and Issues", RFC 3234, February 2002.

[RFC3724] Kempf, J., Austein, R., and IAB, "The Rise of the Middle and the Future of End-to-End: Reflections on the Evolution of the Internet Architecture", RFC 3724, March 2004.

[RFC4033] Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "DNS Security Introduction and Requirements", RFC 4033, March 2005.

[RFC4034] Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "Resource Records for the DNS Security Extensions", RFC 4034, March 2005.

[RFC4035] Arends, R., Austein, R., Larson, M., Massey, D., and S. Rose, "Protocol Modifications for the DNS Security Extensions", RFC 4035, March 2005.

13.2. Informative References

[I-D.shirasaki-nat444]
Yamagata, I., Shirasaki, Y., Nakagawa, A., Yamaguchi, J., and H. Ashida, "NAT444", draft-shirasaki-nat444-02 (work in progress), July 2010.

[IETF 77 DNSOP WG Presentation]
Gashinsky, I., "IPv6 & recursive resolvers: How do we make the transition less painful?", IETF 77 DNS Operations Working Group, March 2010, <http://www.ietf.org/proceedings/77/slides/dnsop-7.pdf>.

[Network World Article on DNS Whitelisting]
Marsan, C., "Google, Microsoft, Netflix in talks to create shared list of IPv6 users", Network World , March 2010, <http://www.networkworld.com/news/2010/032610-dns-ipv6-whitelist.html>.

[Network World Article on IETF 77 DNSOP WG Presentation]
Marsan, C., "Yahoo proposes 'really ugly hack' to DNS", Network World , March 2010, <http://www.networkworld.com/news/2010/032610-yahoo-dns.html>.

[Rethinking the design of the Internet]
Blumenthal, M. and D. Clark, "Rethinking the design of the Internet: The end to end arguments vs. the brave new world", ACM Transactions on Internet Technology Volume 1, Number 1, Pages 70-109, August 2001, <http://dspace.mit.edu/bitstream/handle/1721.1/1519/TPRC_Clark_Blumenthal.pdf>.

[Tussle in Cyberspace]
Braden, R., Clark, D., Sollins, K., and J. Wroclawski, "Tussle in Cyberspace: Defining Tomorrow's Internet", Proceedings of ACM Sigcomm 2002, August 2002, <http://groups.csail.mit.edu/ana/Publications/PubPDFs/Tussle2002.pdf>.

Appendix A. Document Change Log

[RFC Editor: This section is to be removed before publication]

-00: First version published

-01: Updated the title of the document, to avoid confusion (based on feedback)

Appendix B. Open Issues

[RFC Editor: This section is to be removed before publication]

1. Incorporate any feedback received at IETF 79
2. Incorporate feedback from Erik Kline, received 10/1/2010
3. Incorporate feedback from Brian Carpenter, received 10/19/2010
4. Bring on new contributors: Hannes Tschofenig and Danny McPherson has so far offered to contribute.
5. Close out any EDITORIAL notes
6. Add any good references throughout the document
7. Add reviewers to the acknowledgements section
8. Ensure references are in the proper section (normative/ informative)
9. Include a number of references from RFC3724?
10. Call DNS WL something else or add note to the effect that this is unrelated to DNS WL used for email - such as www.dnswl.org

Author's Address

Jason Livingood
Comcast Cable Communications
One Comcast Center
1701 John F. Kennedy Boulevard
Philadelphia, PA 19103
US

Email: jason_livingood@cable.comcast.com
URI: http://www.comcast.com

Network Working Group
Internet-Draft
Intended status: BCP
Expires: April 15, 2011

B. Sarikaya
F. Xia
Huawei USA

October 12, 2010

DHCPv6 Prefix Delegation as IPv6 Migration Tool in Mobile Networks <draft-sarikaya-v6ops-prefix-delegation-02.txt>

Abstract

As interest on IPv6 deployment is increasing in cellular networks several migration issues are being raised and IPv6 prefix management is the one addressed in this document. Based on the idea that DHCPv6 servers can manage prefixes, we address prefix management issues such as the access router offloading delegation and release tasks of the prefixes to a DHCPv6 server using DHCPv6 PD. The access router first requests a prefix for an incoming mobile node from the DHCPv6 server. The access router may next stateless or stateful address allocation to the mobile node, e.g. with a Router Advertisement or using DHCP. We also describe prefix management using Authentication Authorization and Accounting servers.

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 15, 2011.

Copyright Notice

Table of Contents

1. Introduction

Figure 1 illustrates the key elements of a typical cellular access network. In a Long Term Evolution (LTE) network, access router is the packet data network (PDN) gateway [23.401].

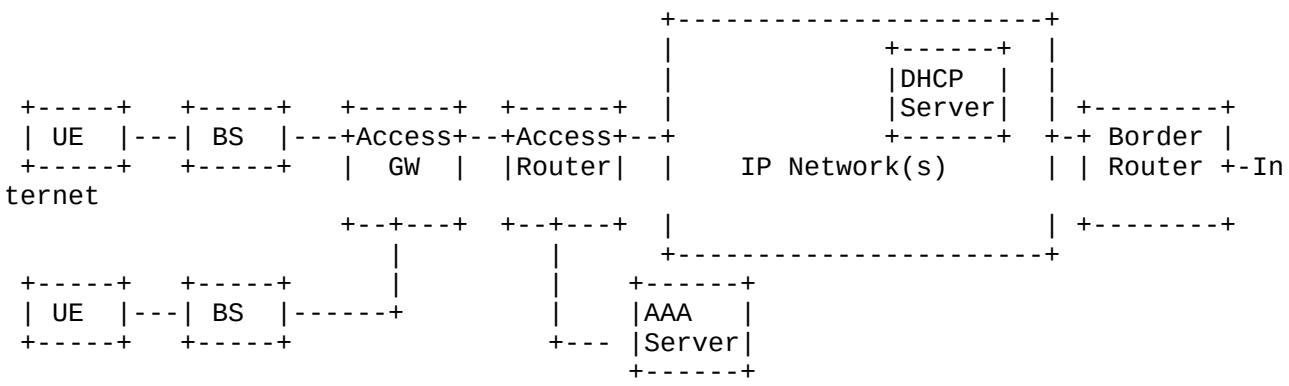


Figure 1: Key elements of a typical cellular network

User equipment (UE) attaches to a base station (BS) through LTE air interface. A BS manages connectivity of UEs and extends connections to an Access Gateway (GW), e.g. the serving gateway (S-GW) in an LTE network. The access gateway and the Access router are connected with an IP network. The access router is the first hop router of UEs and it is in charge of address/prefix management.

Access router is connected to an IP network which is owned by the operator which is connected to the public Internet via a Border Router. The network contains servers for subscriber management including Quality of Service, billing and accounting as well as DHCP server [I-D.ietf-v6ops-v6-in-mobile-networks].

As to IPv6 addressing, there are two models, shared prefix and Per-MN interface prefix. In the shared prefix model, an IPv6 prefix is shared by multiple MNs. While in the Per-MN interface prefix model, a prefix is only assigned to one interface of the MN. Different MNs can’t share a prefix, and an interface can have multiple prefixes.

[RFC4968] briefly compares the two models. Per-MN interface prefix model has some advantages, such as, no complicated duplicate address detection (DAD), fit naturally to the point-to-point links and so on. In Per-MN interface prefix model, prefix management is an issue.

When an MN attaches an AR, the AR requests one or more prefixes for the MN. When the MN detaches the AR, the prefixes should be released. When the MN becomes idle, the AR should hold the prefixes

allocated. When an operator wants to renumber its network, prefixes with different lifetime are advertised to the MN.

This document describes how to use DHCPv6 Prefix Delegation in mobile networks such as networks based on standards developed by 3GPP or WiMAX Forum. The use of prefix delegation for stateless and stateful address configuration is described in Section 3, and Section 4 is on the use of AAA protocols in prefix delegation.

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 BCP 14 [RFC2119].

This document uses the terminology defined in [RFC3315], [RFC3633] and [23.401].

3. Prefix Delegation Using DHCPv6

Access router refers to the cellular network entity that has DHCP Client. According to [23.401] DHCP Client is located in PDN Gateway. So AR is the PDN Gateway in LTE architecture.

3.1. Prefix Request Procedure for Stateless Address Configuration

There are two function modules in the AR, DHCP Client and DHCP Relay. DHCP messages should be relayed if the AR and a DHCP server are not connected directly, otherwise DHCP relay function in the AR is not necessary. Figure 2 illustrates the scenario that the AR and the DHCP Server aren’t connected directly:

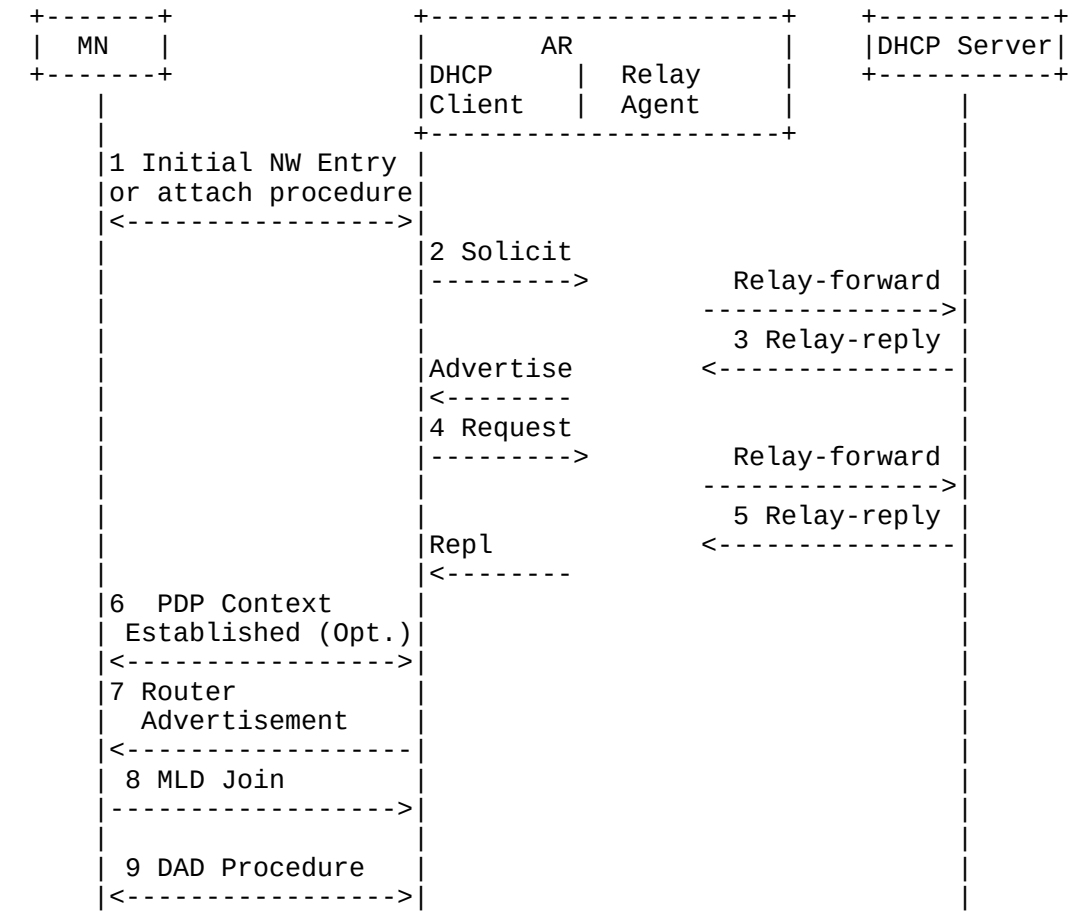


Figure 2: Prefix request

1. An MN (UE) performs initial network entry and authentication procedures, a.k.a. attach procedure.
2. On successful completion of Step 1, the AR initiates DHCP Solicit procedure to request prefixes for the MN. The DHCP Client in AR creates and transmits a Solicit message as described in sections 17.1.1, "Creation of Solicit Messages" and 17.1.2, "Transmission of Solicit Messages" of [RFC3315]. The DHCP Client in AR that supports DHCPv6 Prefix Delegation [RFC3633] creates an IA_PD and assigns it an IAID. The Client MUST include the IA_PD option in the Solicit message. Next, the Relay Agent in AR sends Relay-Forward message to the DHCP Server encapsulating Solicit message.
3. The DHCP server sends an Advertise message to the AR in the same way as described in section 17.2.2, "Creation and transmission of Advertise messages" of [RFC3315]. This message is received

encapsulated in Relay-Reply message by the Relay Agent in AR and sent as Solicit message to the DHCP Client in AR.

4. The AR (DHCP Client and Relay Agent) uses the same message exchanges as described in section 18, "DHCP Client-Initiated Configuration Exchange" of [RFC3315] and [RFC3633] to obtain or update prefixes from the DHCP server. The AR (DHCP Client and Relay Agent) and the DHCP server use the IA_PD Prefix option to exchange information about prefixes in much the same way as IA Address options are used for assigned addresses.
5. AR stores the prefix information it received in the Reply message.
6. A connection between MN and AR is established and the link becomes active. This step (called PDP Context Activation Procedure in UMTS networks) is optional and it is performed during the initial attach of Step 1 in LTE networks.
7. The AR advertises the prefixes received in IA_PD option to MN with router advertisement (RA) once the virtual link is active.
8. The MN constructs a solicited node multicast address for the corresponding Link Local Address and sends MLD Join request for the solicited node multicast address.
9. The MN starts verifying address uniqueness by sending a DAD NS on the virtual link. AR can check the address uniqueness within the virtual link scope.

4-way exchange between AR as requesting router (RR) and DHCP server as delegating router (DR) in Figure 2 MAY be reduced into a two message exchange using the Rapid Commit option [RFC3315]. DHCP Client in AR acting as RR includes a Rapid Commit option in the Solicit message. DR then sends a Reply message containing one or more prefixes.

3.2. Prefix Request Procedure for Stateful Address Configuration

After the initial attach is completed, a connection to the AR is established for the MN. DHCP Client and Relay follow the procedure shown in Figure 2 to get the new prefix for this interface of MN.

DHCPv6 client at the MN sends DHCP Request to DHCP Relay function at AR. AR MUST make sure that DHCP server assigns MN address from this prefix. AR as DHCP Relay forwards DHCP Request message in DHCP Relay Option and adds IA_PD option defined in [RFC3633] to the message. IA_PD option MUST contain the prefix. DHCP server MUST process IA_PD Option and MUST assign an address to MN using the prefix in IA_PD Option. DHCP server replies with Relay reply message. DHCP Relay sends DHCP Reply message to MN containing IA address option (IAADDR). Figure 3 shows the message sequence.

MN configures its interface with the address assigned by DHCP server

in DHCP Reply message.

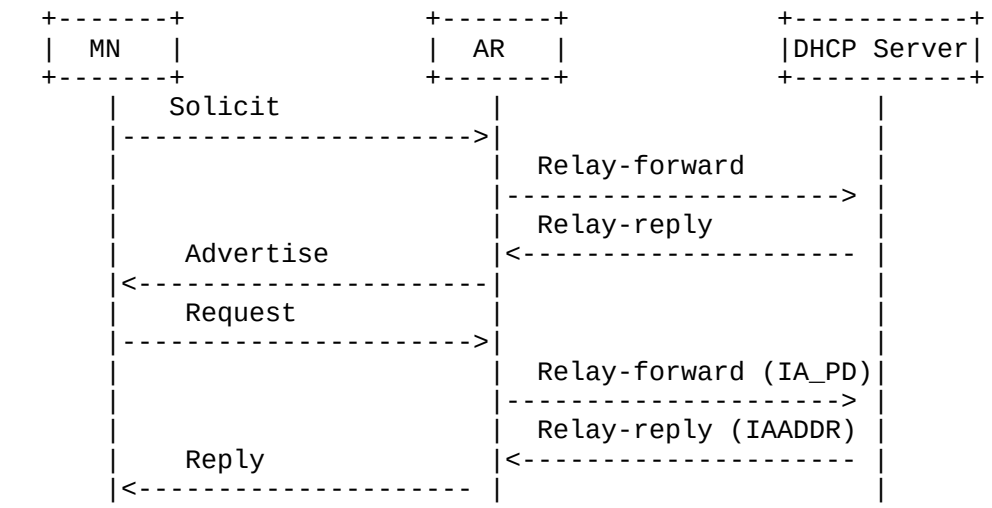


Figure 3: Stateful Address Configuration Following PD

3.3. Prefix Release Procedure

Prefixes can be released in two ways, prefix aging or DHCP release procedure. In the former way, a prefix SHOULD not be used by an MN when the prefix ages, and the DHCP Server can delegate it to another MN. A prefix lifetime is delivered from the DHCPv6 server to the MN through DHCP IA_PD Prefix option [RFC3633] and RA Prefix Information option [RFC4861]. Figure 4 illustrates how the AR releases prefixes to an DHCP Server which isn't connected directly:

1. An MN detachment signaling, such as switch-off or handover, triggers prefix release procedure.
2. The AR initiates a Release message to give back the prefixes to the DHCP server.
3. The server responds with a Reply message, and then the prefixes can be reused by other MNs.

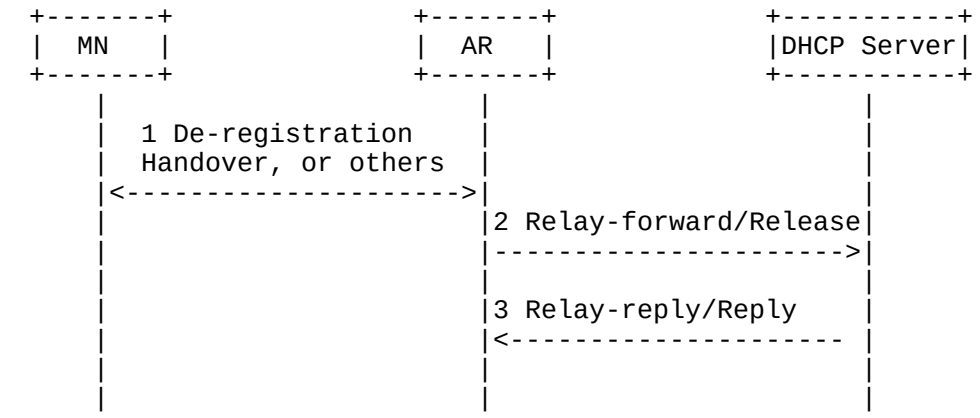


Figure 4: Prefix Release

3.4. Renumbering

3.4.1. Renumbering Through Renew Message

To extend the valid and preferred lifetimes for the prefixes associated with an MN, the AR sends a Renew message to the DHCP server. The server determines new lifetimes for the prefixes and returns the prefix to AR in a Reply message. The DHCP server MAY add new prefixes to the MN for renumbering in its Reply message. For a more detailed description of these message, refer to [RFC3633] and of renumbering, refer to [RFC4192].

3.4.2. Server Initiated Reconfiguration

DHCP server initiates a configuration message exchange with the AR by sending a Reconfigure message. The reconfigure message triggers the AR to send Renew message as described in Section 3.4.1.

3.5. Miscellaneous Considerations

3.5.1. Triggers for an AR to Initiate Prefix Request Procedure

There are some other triggers for an AR to initiate prefix request procedure besides network entry and authentication, such as, when an AR receives handover initiate (HI) message in FMIPv6 [RFC5568], or other handover signaling. After getting an incoming MN’s necessary information (such as MAC address), the AR SHOULD initiate prefix request procedure as soon as possible.

3.5.2. How to Generate IAID

IAID is 4 bytes in length and should be unique in an AR scope. Prefix table SHOULD be maintained. Prefix table contains IAID, MAC address and the prefix(es) assigned to MN. In LTE networks, International Mobile Subscriber Identity (IMNI) corrsponds to the MAC address. MAC address of the interface SHOULD be stored in the prefix table and this field is used as the key for searching the table.

IAID SHOULD be set to Start_IAID, an integer of 4 octets. The following IAID generation algorithm is used:

1. Set this IAID value in IA_PD Prefix Option. Request prefix for this MN as in Section 3.1 or Section 3.2.
2. Store IAID, MAC address and the prefix(es) received in the next entry of the prefix table.
3. Increment IAID.

Prefix table entry for an MN that handsover to another AR MUST be removed. IAID value is released to be reused.

3.5.3. Policy to Delegate Prefixes

AR should broadcast the prefix(es) dynamically upstream as the route information of all the MNs connected to this AR. In point-to-point links, this causes high routing protocol traffic (IGMP, OSPF, etc.) due to Per-MN interface prefixes. To solve the problem, route aggregation SHOULD be used. For example, each AR can be assigned a /48 or /32 prefix (aggregate prefix, aka service provider common prefix) while each interface of MN can be assigned a /64 prefix. The /64 prefix is an extension of /48 one, for example, an AR's /48 prefix is 3FFE:FFFF:0::/48, an interface of MN is assigned 3FFE:FFFF:0:2::/64 prefix. The AR only broadcasts it's /48 or /32 prefix information to Internet.

This policy can be enforced as follows: DHCP Relay MUST set the IPv6 Prefix field in IA_PD Prefix option in IA_PD option in the Relay Forward message to the aggregate prefix (/48, /32, or /16 prefix assigned to the AR).

4. Prefix Delegation Using RADIUS and Diameter

In the initial network entry procedure Figure 2, AR as RADIUS client sends Access-Request message with MN information to RADIUS server. If the MN passes the authentication, the RADIUS server may send Access-Accept message with prefix information to the AR using Framed-IPv6-Prefix attribute. AAA server also provides routing information

to be configured for MN on the AR using Framed-IPv6-Route attribute. Using such a process AR can handle initial prefix assignments to MNs but managing lifetime of the prefixes is totally left to the AR. Framed-IPv6-Prefix is not designed to support delegation of IPv6 prefixes. For this Delegated-IPv6-Prefix attribute can be used which is discussed next.

[RFC4818] defines a RADIUS attribute Delegated-IPv6-Prefix that carries an IPv6 prefix to be delegated. This attribute is usable within either RADIUS or Diameter. [RFC4818] recommends the delegating router to use AAA server to receive the prefixes to be delegated using Delegated-IPv6-Prefix attribute/AVP.

DHCP server as the delegating router in Figure 2 MAY send an Access-Request packet containing Delegated-IPv6-Prefix attribute to the RADIUS server to request prefixes. In the Access-Request message, the delegating router MAY provide a hint that it would prefer a prefix, for example, a /48 prefix. The RADIUS server MAY delegate a /64 prefix which is an extension of the /48 prefix in an Access-Accept message containing Delegated-IPv6-Prefix attribute. The attribute can appear multiple times when RADIUS server delegates multiple prefixes to the delegating router. The delegating router sends the prefixes to the requesting router using IA_PD Option and AR as RR uses them for MN’s as described in .

When Diameter is used, DHCP server as the delegating router in Figure 2 sends AA-Request message. AA-Request message MAY contain Delegated-IPv6-Prefix AVP. Diameter server replies with AA-Answer message. AA-Answer message MAY contain Delegated-IPv6-Prefix AVP. The AVP can appear multiple times when Diameter server assigns multiple prefixes to MN. The Delegated-IPv6-Prefix AVP MAY appear in an AA-Request packet as a hint by the AR to the Diameter server that it would prefer a prefix, for example, a /48 prefix. Diameter server MAY delegate in an AA-Answer message with a /64 prefix which is an extension of the /48 prefix. As in the case of RADIUS, the delegating router sends the prefixes to the requesting router using IA_PD Option and AR as RR uses them for MN’s as described in .

5. Security Considerations

This draft introduces no additional messages. Comparing to [RFC3633], [RFC2865] and [RFC3588] there is no additional threats to be introduced. DHCPv6, RADIUS and Diameter security procedures apply.

6. Protocol Variables

Start_IAID 4 octet integer value.

It can be initialized to ZERO.

7. IANA Considerations

None.

8. Acknowledgements

We are grateful to Suresh Krishnan, Hemant Singh, Qiang Zhao and others who provided in depth reviews of this document that have led to several imrpovements.

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.

[RFC2865] Rigney, C., Willens, S., Rubens, A., and W. Simpson, "Remote Authentication Dial In User Service (RADIUS)", RFC 2865, June 2000.

[RFC2866] Rigney, C., "RADIUS Accounting", RFC 2866, June 2000.

[RFC3046] Patrick, M., "DHCP Relay Agent Information Option", RFC 3046, January 2001.

[RFC3314] Wasserman, M., "Recommendations for IPv6 in Third Generation Partnership Project (3GPP) Standards", RFC 3314, September 2002.

[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.

[RFC3588] Calhoun, P., Loughney, J., Guttman, E., Zorn, G., and J. Arkko, "Diameter Base Protocol", RFC 3588, September 2003.

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

December 2003.

[RFC3963] Devarapalli, V., Wakikawa, R., Petrescu, A., and P. Thubert, "Network Mobility (NEMO) Basic Support Protocol", RFC 3963, January 2005.

[RFC4818] Salowey, J. and R. Droms, "RADIUS Delegated-IPv6-Prefix Attribute", RFC 4818, April 2007.

[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.

[RFC5213] Gundavelli, S., Leung, K., Devarapalli, V., Chowdhury, K., and B. Patil, "Proxy Mobile IPv6", RFC 5213, August 2008.

[RFC5568] Koodli, R., "Mobile IPv6 Fast Handovers", RFC 5568, July 2009.

9.2. Informative References

[23.401] "3GPP TS 23.401 V9.3.0, General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access (Release 9).", 2009.

[23.812] "3GPP TR 23.812 V9.0.0, Feasibility Study on the Security Aspects of Remote Provisioning and Change of Subscription for M2M Equipment; (Release 9).", 2009.

[I-D.ietf-v6ops-v6-in-mobile-networks] Koodli, R., "Mobile Networks Considerations for IPv6 Deployment", draft-ietf-v6ops-v6-in-mobile-networks-01 (work in progress), July 2010.

[RFC4192] Baker, F., Lear, E., and R. Droms, "Procedures for Renumbering an IPv6 Network without a Flag Day", RFC 4192, September 2005.

[RFC4968] Madanapalli, S., "Analysis of IPv6 Link Models for 802.16 Based Networks", RFC 4968, August 2007.

Authors' Addresses

Behcet Sarikaya
Huawei USA
1700 Alma Dr. Suite 500
Plano, TX 75075

Email: sarikaya@ieee.org

Frank Xia
Huawei USA
1700 Alma Dr. Suite 500
Plano, TX 75075

Phone: +1 972-509-5599
Email: xiayangsong@huawei.com

Network Working Group
Internet-Draft
Intended status: Informational
Expires: August 13, 2012

B. Sarikaya
F. Xia
Huawei USA
T. Lemon
Nominum
February 10, 2012

DHCPv6 Prefix Delegation in Long Term Evolution (LTE) Networks
draft-sarikaya-v6ops-prefix-delegation-11.txt

Abstract

As interest on IPv6 deployment is increasing in cellular networks several migration issues are being raised and IPv6 prefix management is the one addressed in this document. Based on the idea that DHCPv6 servers can manage prefixes, we address prefix management issues such as the access router offloading delegation and release tasks of the prefixes to a DHCPv6 server using DHCPv6 Prefix Delegation. The access router first requests a prefix for an incoming mobile node from the DHCPv6 server. The access router may next do stateless or stateful address allocation to the mobile node, e.g. with a Router Advertisement or using DHCP. We also describe prefix management using Authentication Authorization and Accounting servers.

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 13, 2012.

Copyright Notice

Table of Contents

1. Introduction

Figure 1 illustrates the key elements of a typical cellular access network. In a Long Term Evolution (LTE) network, access router is the packet data network (PDN) gateway [ThreeGPP23401].

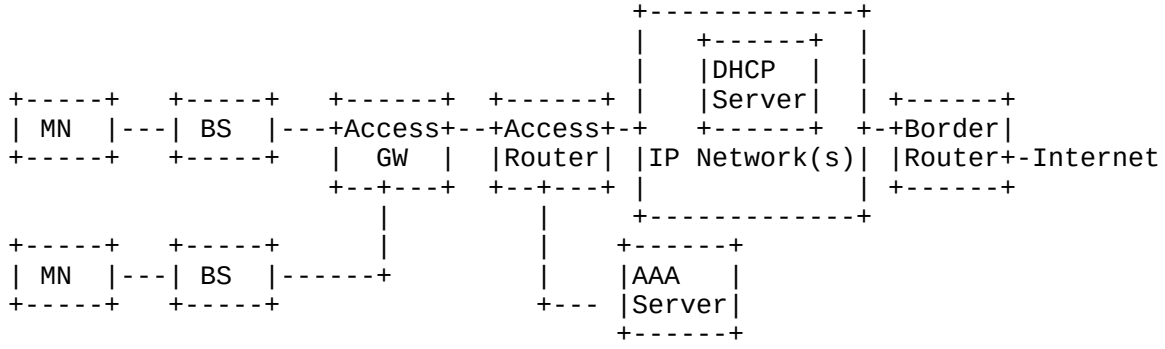


Figure 1: Key elements of a typical cellular network

Mobile node (MN) attaches to a base station (BS) through LTE air interface. A BS manages connectivity of UEs and extends connections to an Access Gateway (GW), e.g. the serving gateway (S-GW) in an LTE network. The access gateway and the Access Router (AR) are connected with an IP network. The access router is the first hop router of MNs and it is in charge of address/prefix management.

Access router is connected to an IP network which is owned by the operator which is connected to the public Internet via a Border Router. The network contains servers for subscriber management including Quality of Service, billing and accounting as well as Dynamic Host Configuration Protocol (DHCP) server [RFC6342].

As to IPv6 addressing, because mobile network links are point-to-point (p2p) Per-MN interface prefix model is used [RFC3314], [RFC3316]. In Per-MN interface prefix model, prefix management is an issue.

When an MN attaches an AR, the AR requests one or more prefixes for the MN. When the MN detaches the AR, the prefixes should be released. When the MN becomes idle, the AR should hold the prefixes allocated.

This document describes how to use DHCPv6 Prefix Delegation (PD) in mobile networks such as networks based on standards developed by the 3rd Generation Partnership Project (3GPP) and it could easily be adopted to Worldwide Interoperability for Microwave Access (WiMAX)

Forum as well. In view of migration to IPv6, the number of mobile nodes connected to the network at a given time may become very high. Traditional techniques such as prefix pools are not scalable. In such cases DHCPv6 PD becomes the viable approach to take.

The techniques described in this document have not been approved either by the IETF or by 3GPP, except what is described below in Section 3.3. This document is not a standard or best current practice. This document is published only as a possibility for consideration by operators.

This document is useful when address space needs to be managed by DHCPv6-PD. There are obviously other means of managing address space, including having the AR track internally what address space is used by what mobile.

2. Terminology and Acronyms

3GPP 3rd Generation Partnership Project

AAA Authentication Authorization and Accounting

AR Access Router

BS Base Station

DHCP Dynamic Host Control Protocol

E-UTRAN Evolved Universal Terrestrial Radio Access Network

GPRS General Packet Radio Service

LTE Long Term Evolution

MN Mobile node

PDN Packet data network

PD Prefix Delegation

p2p Point-to-point

Serving Gateway S-GW

WiMAX Worldwide Interoperability for Microwave Access

3. Prefix Delegation Using DHCPv6

Access router refers to the cellular network entity that has DHCP Client. According to [ThreeGPP23401] DHCP Client is located in PDN Gateway. So AR is the PDN Gateway in LTE architecture.

3.1. Prefix Request Procedure for Stateless Address Configuration

There are two function modules in the AR, DHCP Client and DHCP Relay. DHCP messages should be relayed if the AR and a DHCP server are not connected directly, otherwise DHCP relay function in the AR is not necessary. Figure 2 illustrates the scenario that the AR and the DHCP Server aren't connected directly:

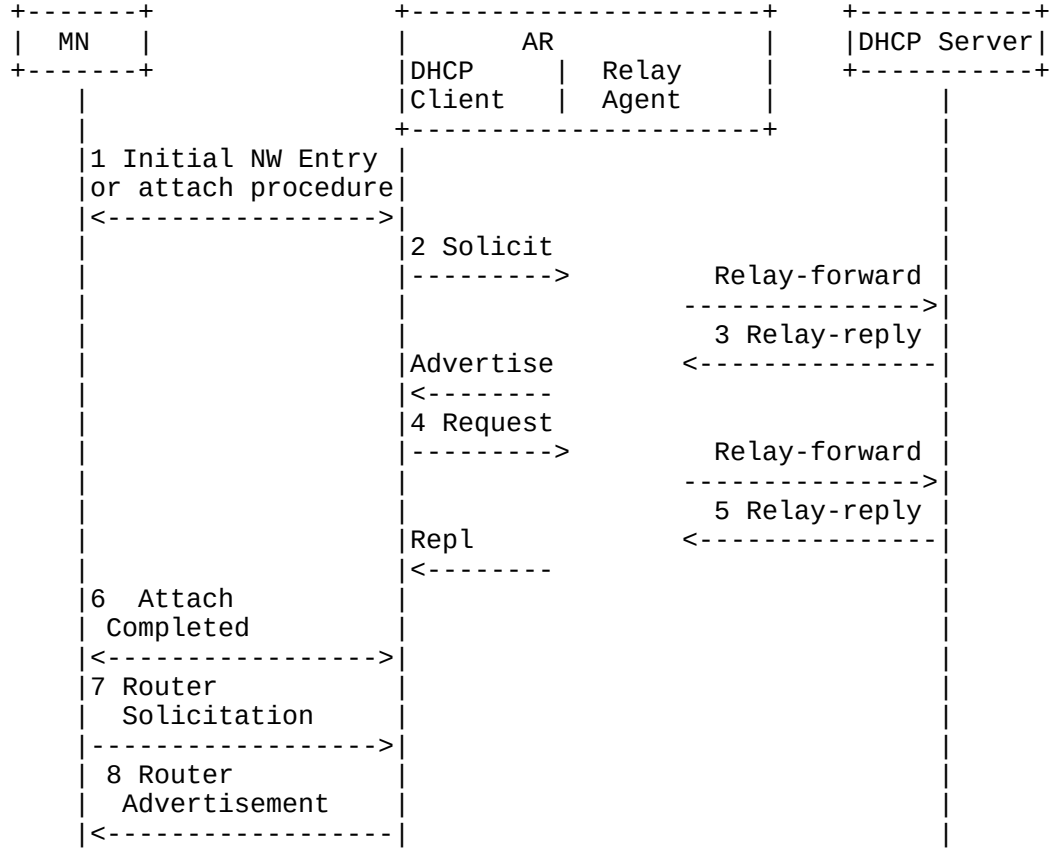


Figure 2: Prefix request

1. An MN (UE=User Equipment in 3GPP) performs initial network entry and authentication procedures, a.k.a. attach procedure.
2. On successful completion of Step 1, the AR initiates DHCP Solicit procedure to request prefixes for the MN. The DHCP Client in AR creates and transmits a Solicit message as described in sections 17.1.1, "Creation of Solicit Messages" and 17.1.2, "Transmission of Solicit Messages" of [RFC3315]. The DHCP Client in AR that supports DHCPv6 Prefix Delegation [RFC3633] creates an Identity Association for Prefix Delegation (IA_PD) and assigns it an Identity Association IDentifier (IAID). The client must include the IA_PD option in the Solicit message. DHCP Client as Requesting Router must set prefix-length field to a value less than, e.g. 48 or equal to 64 to request a /64 prefix. Next, the Relay Agent in AR sends Relay-Forward message to the DHCP Server encapsulating Solicit message.
3. The DHCP server sends an Advertise message to the AR in the same way as described in section 17.2.2, "Creation and transmission of Advertise messages" of [RFC3315]. Advertise message with IA_PD shows that the DHCP server is capable of delegating prefixes. This message is received encapsulated in Relay-Reply message by the Relay Agent in AR and sent as Advertise message to the DHCP Client in AR.
4. The AR (DHCP Client and Relay Agent) uses the same message exchanges as described in section 18, "DHCP Client-Initiated Configuration Exchange" of [RFC3315] and [RFC3633] to obtain or update prefixes from the DHCP server. The AR (DHCP Client and Relay Agent) and the DHCP server use the IA_PD Prefix option to exchange information about prefixes in much the same way as IA Address options are used for assigned addresses. This is accomplished by the AR sending a DHCP Request message and the DHCP server sending a DHCP Reply message.
5. AR stores the prefix information it received in the Reply message.
6. A connection between MN and AR is established and the link becomes active. This step completes the PDP Context Activation Procedure in UMTS and PDN connection establishment in LTE networks.
7. The MN may send a Router Solicitation message to solicit the AR to send a Router Advertisement message.
8. The AR advertises the prefixes received in IA_PD option to MN with router advertisement (RA) once the PDP Context/PDN connection is established or in response to Router Solicitation message sent from the MN.

4-way exchange between AR as requesting router (RR) and DHCP server as delegating router (DR) in Figure 2 may be reduced into a two message exchange using the Rapid Commit option [RFC3315]. DHCP Client in AR acting as RR includes a Rapid Commit option in the

Solicit message. DR then sends a Reply message containing one or more prefixes.

3.2. Prefix Request Procedure for Stateful Address Configuration

Stateful address configuration requires a different architecture than shown in Figure 2. There are two function modules in the AR, DHCP Server and DHCP Client.

After the initial attach is completed, a connection to the AR is established for the MN. DHCP Client function at the AR as requesting router and DHCP server as delegating router follow Steps 2 through 5 of the procedure shown in Figure 2 to get the new prefix for this interface of MN from IA_PD Option exchange defined in [RFC3633].

DHCPv6 client at the MN sends DHCP Request to AR. DHCP Server function at the AR must use the IA_PD option received in DHCP PD exchange to assign an address to MN. IA_PD option must contain the prefix. AR sends DHCP Reply message to MN containing IA address option (IAADDR). Figure 3 shows the message sequence.

MN configures its interface with the address assigned by DHCP server in DHCP Reply message.

In Figure 3 AR may be the home gateway of a fixed network to which MN gets connected during MN’s handover.

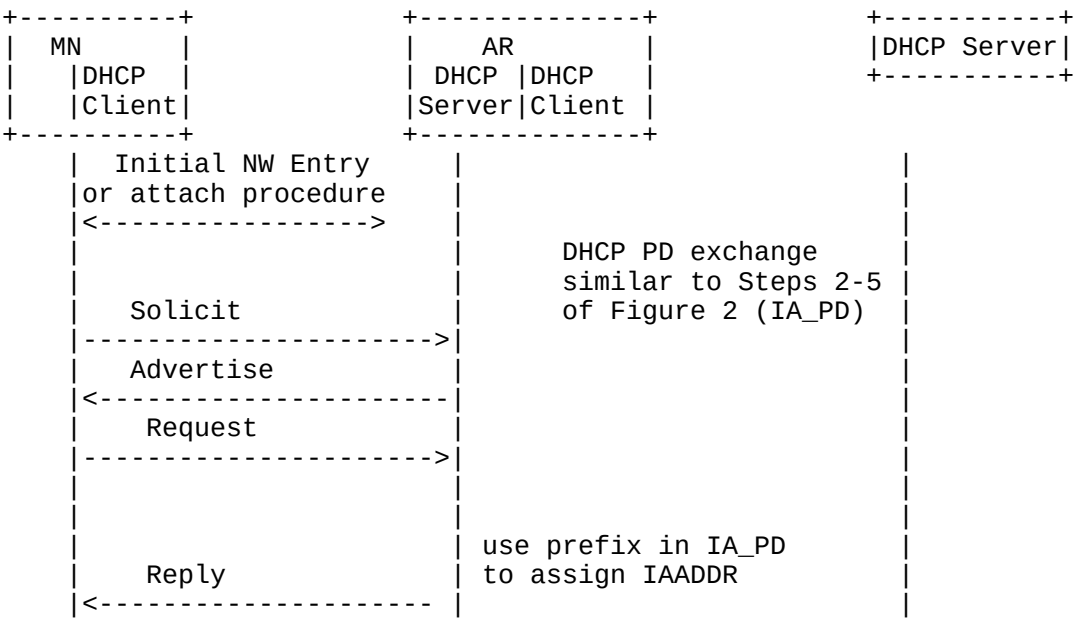


Figure 3: Stateful Address Configuration Following PD

3.3. MN as Requesting Router in Prefix Delegation

AR may use DHCPv6 prefix delegation exchange to get a delegated prefix shorter than /64 by setting prefix-length field to a value less than 64, e.g. 56 to get a /56 prefix. Each newly attaching MN first goes through the steps in Figure 2 in which AR requests a shorter prefix to establish a default connection with the MN.

MN may next request additional prefixes (/64 or shorter) from the AR using DHCPv6 prefix delegation where MN is the requesting router and AR is the delegating router [RFC6459], Section 5.3.1.2.6 in [ThreeGPP23401]. In this case the call flow is similar to Figure 3. Solicit message must include the IA_PD option with prefix-length field set to 64. MN may request more than one /64 prefixes. AR as delegating router must delegate these prefixes excluding the prefix assigned to the default connection.

3.4. Prefix Release Procedure

Prefixes can be released in two ways, prefix aging or DHCP release procedure. In the former way, a prefix should not be used by an MN when the prefix ages, and the DHCP Server can delegate it to another MN. A prefix lifetime is delivered from the DHCPv6 server to the MN

through DHCP IA_PD Prefix option [RFC3633] and RA Prefix Information option [RFC4861]. Figure 4 illustrates how the AR releases prefixes to a DHCP Server which isn't connected directly:

1. An MN detachment signaling, such as switch-off or handover, triggers prefix release procedure.
2. The AR initiates a Release message to give back the prefixes to the DHCP server.
3. The server responds with a Reply message, and then the prefixes can be reused by other MNs.

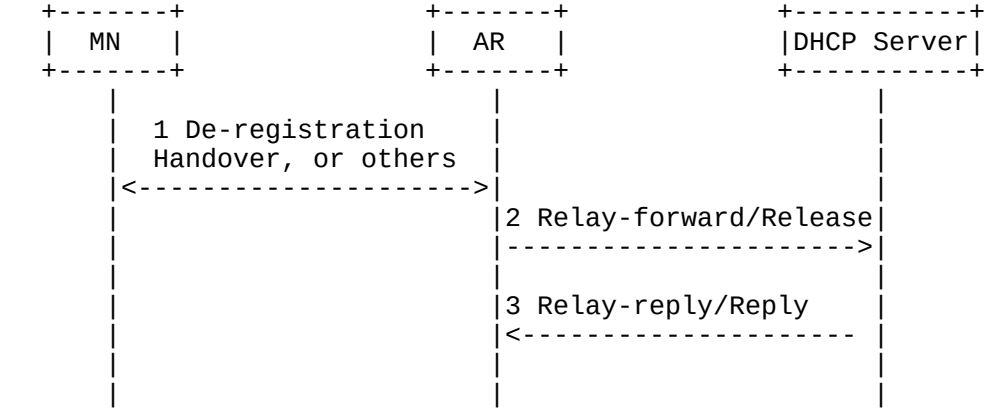


Figure 4: Prefix Release

3.5. Miscellaneous Considerations

3.5.1. How to Generate IAID

IAID is 4 bytes in length and should be unique in an AR scope. Prefix table should be maintained. Prefix table contains IAID, MAC address and the prefix(es) assigned to MN. In LTE networks, International Mobile station Equipment Identity (IMEI) uniquely identifies MN's interface and thus corresponds to the MAC address. MAC address of the interface should be stored in the prefix table and this field is used as the key for searching the table.

IAID should be set to Start_IAID, an integer of 4 octets. The following IAID generation algorithm is used:

1. Set this IAID value in IA_PD Prefix Option. Request prefix for this MN as in Section 3.1 or Section 3.2.
2. Store IAID, MAC address and the prefix(es) received in the next entry of the prefix table.

3. Increment IAID.

Prefix table entry for an MN that hands over to another AR must be removed. IAID value is released to be reused.

3.5.2. Policy to Delegate Prefixes

In point-to-point links, if /64 prefixes of all the MNs connected to one or more ARs are broadcast dynamically upstream as the route information this causes high routing protocol traffic (IGP, OSPF, etc.) due to Per-MN interface prefixes. There are two solutions this problem. One is to use static configuration, which would be preferable in many cases. No routing protocols are needed, because each AR has a known piece of address space. If the DHCP servers know this space, too, then they will assign from that space to a particular AR.

The other method is to use route aggregation. For example, each AR can be assigned a /48 or /32 prefix (aggregate prefix, aka service provider common prefix) while each interface of MN can be assigned a /64 prefix. The /64 prefix is an extension of /48 one, for example, an AR's /48 prefix is 2001:DB8:0::/48, an interface of MN is assigned 2001:DB8:0:2::/64 prefix. The border router (BR) in Figure 1 may be manually configured to broadcast only individual AR's /48 or /32 prefix information to Internet.

4. Prefix Delegation Using RADIUS and Diameter

In the initial network entry procedure Figure 2, AR as Remote Authentication Dial In User Service (RADIUS) client sends Access-Request message with MN information to RADIUS server. If the MN passes the authentication, the RADIUS server may send Access-Accept message with prefix information to the AR using Framed-IPv6-Prefix attribute. AAA server also provides routing information to be configured for MN on the AR using Framed-IPv6-Route attribute. Using such a process AR can handle initial prefix assignments to MNs but managing lifetime of the prefixes is totally left to the AR. Framed-IPv6-Prefix is not designed to support delegation of IPv6 prefixes. For this Delegated-IPv6-Prefix attribute can be used which is discussed next.

[RFC4818] defines a RADIUS attribute Delegated-IPv6-Prefix that carries an IPv6 prefix to be delegated. This attribute is usable within either RADIUS or Diameter. [RFC4818] recommends the delegating router to use AAA server to receive the prefixes to be delegated using Delegated-IPv6-Prefix attribute/AVP.

DHCP server as the delegating router in Figure 2 may send an Access-Request packet containing Delegated-IPv6-Prefix attribute to the RADIUS server to request prefixes. In the Access-Request message, the delegating router may provide a hint that it would prefer a prefix, for example, a /48 prefix. As the RADIUS server is not required to honor the hint, the server may delegate longer prefix, e.g. /56 or /64 in an Access-Accept message containing Delegated-IPv6-Prefix attribute [RFC4818]. The attribute can appear multiple times when RADIUS server delegates multiple prefixes to the delegating router. The delegating router sends the prefixes to the requesting router using IA_PD Option and AR as RR uses them for MN's as described in Section 3.

When Diameter is used, DHCP server as the delegating router in Figure 2 sends AA-Request message. AA-Request message may contain Delegated-IPv6-Prefix AVP. Diameter server replies with AA-Answer message. AA-Answer message may contain Delegated-IPv6-Prefix AVP. The AVP can appear multiple times when Diameter server assigns multiple prefixes to MN. The Delegated-IPv6-Prefix AVP may appear in an AA-Request packet as a hint by the AR to the Diameter server that it would prefer a prefix, for example, a /48 prefix. Diameter server may delegate in an AA-Answer message with a /64 prefix which is an extension of the /48 prefix. As in the case of RADIUS, the delegating router sends the prefixes to the requesting router using IA_PD Option and AR as RR uses them for MN's as described in Section 3.

5. Security Considerations

This draft introduces no additional messages. Comparing to [RFC3633], [RFC2865] and [RFC3588] there is no additional threats to be introduced. DHCPv6, RADIUS and Diameter security procedures apply.

6. IANA Considerations

None.

7. Acknowledgements

We are grateful to Suresh Krishnan, Hemant Singh, Qiang Zhao, Ole Troan, Qin Wu, Jouni Korhonen, Cameron Byrne, Brian Carpenter, Jari Arkko and Jason Lin who provided in depth reviews of this document that have led to several improvements.

8. Informative References

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

[RFC2865] Rigney, C., Willens, S., Rubens, A., and W. Simpson, "Remote Authentication Dial In User Service (RADIUS)", RFC 2865, June 2000.

[RFC3314] Wasserman, M., "Recommendations for IPv6 in Third Generation Partnership Project (3GPP) Standards", RFC 3314, September 2002.

[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.

[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.

[RFC3588] Calhoun, P., Loughney, J., Guttman, E., Zorn, G., and J. Arkko, "Diameter Base Protocol", RFC 3588, September 2003.

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

[RFC4818] Salowey, J. and R. Droms, "RADIUS Delegated-IPv6-Prefix Attribute", RFC 4818, April 2007.

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

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

[ThreeGPP23401] "3GPP TS 23.401 V11.0.0, General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access (Release 11).",

2011.

Authors' Addresses

Behcet Sarikaya
Huawei USA
5340 Legacy Dr.
Plano, TX 75074

Email: sarikaya@ieee.org

Frank Xia
Huawei USA
1700 Alma Dr. Suite 500
Plano, TX 75075

Phone: +1 972-509-5599
Email: xiayangsong@huawei.com

Ted Lemon
Nominum
2000 Seaport Blvd
Redwood City, CA 94063

Phone:
Email: mellon@nominum.com

Network Working Group
Internet Draft
Intended status: Informational
Expires: April 24, 2011

Qiong Sun
Heyu Wang
China Telecom
Xing Li
CongXiao Bao
CERNET Center/Tsinghua University
Ming Feng
China Telecom

October 25, 2010

Considerations for Stateless Translation (IVI/dIVI) in Large SP Network
draft-sunq-v6ops-ivi-sp-01.txt

Abstract

With the approaching exhaustion of IPv4 address space, large-scale SPs are now faced with the only real option to deploy IPv6 in a timely manner. In order to achieve smooth transition to IPv6, migration tools should be introduced for different deployment models. Among different IPv6 transition mechanisms, dIVI is a prefix-specific and stateless address mapping method which can directly translate IPv4 packet to IPv6 packet. This document describes the challenges and requirements for large SP to deploy IPv6 in operational network, the experimental results of dIVI in our laboratory and the considerations for dIVI deployment in large SP operational network.

This Internet-Draft will expire on April 25,2011.

Table of Contents

1. Introduction

The dramatic growth of the Internet is accelerating the exhaustion of available IPv4 addressing pool. It is widely accepted that IPv6 is the only real option on the table for the continued growth of the Internet. However, IPv6 deployment is a huge systematic project and a lot of challenges will arise especially in large SP operational network.

In order to achieve smooth transition to IPv6, migration tools should be introduced for different deployment models, among which dIVI is a stateless translation mechanism with good scalability. This document describes the challenges and requirements for large SPs in IPv6 transition period. Then, we introduce dIVI experimental results in our laboratory. And finally, the considerations for designing and deploying dIVI in operational network are discussed in terms of addressing and routing, DNS deployment requirement, AAA support and user management, network management, CPE requirement and Xlate placement.

2. Terminologies

This document uses the terminologies defined in [I-D.ietf-behave-v6v4-framework], [I-D.ietf-behave-v6v4-xlate], [I-D.ietf-behave-address-format], [I-D.ietf-behave-v6v4-xlate-stateful], and [I-D.xli-behave-ivi].

The key words MUST, MUST NOT, REQUIRED, SHALL, SHALL NOT, SHOULD, SHOULD NOT, RECOMMENDED, MAY, and OPTIONAL, when they appear in this document, are to be interpreted as described in [RFC2119].

3. Problem Statement

It is well known that the pool of public IPv4 addressing is nearing its exhaustion. The ’/8’ IANA blocks for Regional Internet Registries (RIRs) are projected to ’run-out’ towards the end of 2011. Credible estimates based on past behavior suggest that the RIRs will exhaust their remaining address space by early 2012, apart from the development of a market in IPv4 address space. Thus, it will become much more difficult to get available public IPv4 addresses. In the same time, a lot of emerging applications, e.g. Apple’s iPad, Motion’s BlackBerry, etc. will quickly deplete the available addresses. This has led to a hightened awareness among the providers to consider introducing IPv6 to keep the Internet operational.

It has been widely accepted that the end goal of IPv6 transition is to achieve an end-to-end IPv6-only network, and IPv4 can eventually

be turned off. However, since it will have impact on almost the entire world, it will take a considerable period of time to reach the ultimate goal. As a result, IPv4 and IPv6 need to coexist during the whole transition period. In this document, we mainly focus on IPv6 migration issues from large ISP's point of view. In order to facilitate smooth IPv6 migration, some factors need to be taken into consideration especially for large SPs. There are ten major requirements:

1. It should deploy in an incremental fashion and the overall transition process should be stable and operational.

2. It should largely reduce IPv4 public address consumption.

3. It should accelerate the deployment of IPv6, rather than just prolonging the lifecycle of IPv4 by introducing multiple layers of NAT.

4. There should be no perceived degradation of customer experience. As a result, IPv6 transition mechanisms should provide IPv4 service continuity.

5. It should achieve scalability, simplicity and high availability, especially for large-scale SPs.

6. It should have user management and network management ability.

7. It should support user authentication, authorization and accounting as an essential part of operational network.

8. Most ISPs need some kind of mechanisms to trace the origin of traffic in their networks. This should also be available for IPv6 traffic.

9. It should have good throughput performance and massive concurrent session support.

10. It should maintain the deployment concepts and business models which have been proven and used with existing revenue generating IPv4 services.

All existing IPv6 transition mechanisms can be widely divided into three categories: dual-stack solution, translation-based solution and tunnel-based solution. In this document, we mainly concentrate on stateless translation mechanism: dIVI. The original stateless IPv4/IPv6 translation (stateless 1:1 IVI) is scalable, [I-D.ietf-behave-v6v4-framework], [I-D.ietf-behave-v6v4-xlate], [I-D.ietf-

behave-address-format],[I-D.xli-behave-ivi]. But it cannot use the IPv4 addresses effectively. The stateless dIVI[I-D.xli-behave-divi] is a double translation mechanism which includes a 1:N stateless translator and a 1:1 Hgw translator. The 1:N stateless translator is implemented in the border between the IPv6 network and the IPv4 Internet. It translates the packets between IPv4 and IPv6 with the 1:N stateless address mapping. The 1:1 Hgw translator is implemented between an IPv6 network and user's end system. It translates the packets between IPv4 and IPv6 with 1:1 stateless address mapping. In addition, the home gateway translator maps random source port numbers to restricted port number based on the extended IPv4-translatable address format and keeps the mapping table in database for the port number mapping of the retuning packets and all the packets in the same session.

dIVI support bidirectional communication initiated from IPv4 and IPv6. It can be deployed in an IPv6-only access network, in which operational and maintenance cost can be reduced. It has very good scalability and can largely reduce IPv4 address consumption.

In this document, we firstly demonstrate the laboratory experimental results of dIVI in section 4. We can see that dIVI can support most of the current IPv4 applications in IPv6-only access network, while largely reducing IPv4 address consumption. And then dIVI deployment model and considerations in large operational network are proposed in section 5 and section 6 respectively. Some important factors need to be taken into account when introducing dIVI. Since most challenges in dIVI have no big difference compared to an IPv6-only environment, we strongly recommend that related network elements should take the corresponding modifications in order to guarantee the IPv6 transition process to be operational and manageable.

4. Laboratory experiment

We have tested dIVI using the prototype in our laboratory. The major objective listed in the following is to verify the functionality and performance of dIVI:

O Verify how to deploy dIVI in practical network.

O Verify what applications can be used in dIVI.

O Verify what Operating Systems can be supported in dIVI.

O Verify the effect of user experience with limited ports.

O Verify the performance of dIVI.

4.1. Experiment environment

We have tested dIVI in our laboratory. The network topology is depicted in Figure 1.

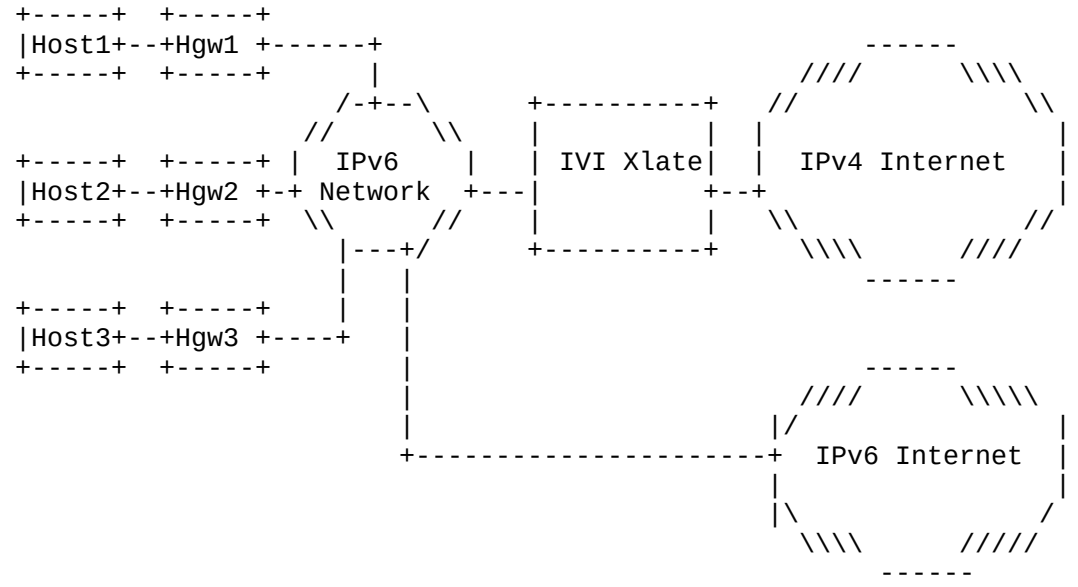


Figure 1 dIVI topology in the test

There are three key components in the test:

- The Hosts are dual-stack customers, who could run IPv4 application, IPv6 application or dual stack application.

- The Home Gateways (Hgw) are dIVI translator in user side. It translates the packets between IPv4 and IPv6 with 1:1 stateless address mapping, and maps random source port numbers to restricted port number.

- The Xlate translates the packets between IPv4 and IPv6 with the 1:N stateless address mapping.

The network between Hgw and Xlate is IPv6-only, and the network behind Hgw is dual-stack. Thus, the hosts behind Hgw can communicate with both IPv4 Internet and IPv6 Internet.

4.2. Experiment configuration

For address configuration, each host will use two IPv6 addresses: one is IVI6 address which is synthesized in Hgw with the IVI4 address and port-related information, and the other is non-IVI IPv6 address which is used for native IPv6 communication. We should notice that only non-IVI IPv6 address needs be allocated to end users. Besides, each user will get an IVI4 address from Hgw.

For routing configuration, both IVI address and non-IVI address need to be imported into the IPv6-only network.

For port configuration, since there are 65536 TCP/UDP ports for each IP address, and in fact one can use hundreds only for normal applications, so one IPv4 address can be shared by multiple customers. In our experiment, we selected ratio to be 128. That is to say, one IPv4 address is shared by 128 users, and there are 512 available ports per user.

For DNS configuration, there is no need to have additional DNS64 for dIVI. Only an IPv6 DNS server with A/AAAA records is needed and the DNS address is manually configured in Hgw. Besides, Hgw has implemented DNS Proxy and it will convert an IPv4 DNS request/response to IPv6 DNS request/response.

For ALG configuration, there is no need to deploy specific ALG for IPv4 applications in dIVI.

4.3. Experiment results

In our laboratory, we mainly have taken the following tests:

- Application test: The applications we have tested include web, email, Instant Message, ftp, telnet, SSH, video, Video Camera, P2P, online game, Voip, VPN and so on.

- Operating System test: The OS we have tested includes Win7, VISTA, windows XP.

- Performance test: We have measured the parameters of concurrent session number, throughput performance.

The experimental results are listed as follows:

Type	Experiment Result
Application test	dIVI hosts can support web, email, im, ftp , telnet, SSH, video, Video Camera, P2P, online game, voip, and so on.
Operating System test	dIVI can widely support Win7, VISTA, windows XP.
Performance test	The performance test for dIVI Xlate is carried out on a normal PC. Due to limitation of the PC hardware, the overall throughput is not quite good. However, it can still support more than one hundred million concurrent sessions.

Figure 2 dIVI test result

From the experiment, we can have the following conclusions:

1. dIVI can have good scalability. Xlate does not need to maintain any session state, and only limited session states have to been maintained in Hgw for its customer.

2. dIVI can be deployed in an incremental way. The most complicated part of dIVI is addressing and routing. The configuration for DNS and ALG is very simple.

3. dIVI can support a majority of current IPv4 applications.

4. dIVI can support a variety of Operating Systems.

5. With the ratio of 128 (512 maximum concurrent sessions), there is no perceived degradation of customer experience.

However, in the current status of equipment, e.g. BRAS, end system, etc., the support for IPv6-only function has not been fully accomplished. Therefore, there are still some limitations when deploying dIVI prototype in practical operational network:

1. Address assignment process have not been integrated with existing address allocation system.

2. Currently, IVI routing entries are configured manually.

3. Hgw has not integrated PPPoE functionality with dIVI functionality.

4. AAA system has not supported IVI-related (or IPv6-only) functions.

With regard to the listed IPv6 transition requirements in section 3, most of them can be satisfied by dIVI, except for the requirement of network management and user management. These two points should be paid special attention for large SPs, which will be further discussed in section 6.

5. dIVI Deployment Scenario

In order to investigate the ways to deploy dIVI in operational network, we firstly briefly discuss network architecture for large SP network. Then dIVI deployment model is introduced.

5.1. Network Architecture for Large SP Network

In large SPs, the generic network topology can be divided into four main parts (as depicted in Figure3): the Customer Premises Network (CPN), the Access Network (AN), the Metro Area Network (MAN), and the Backbone Network.

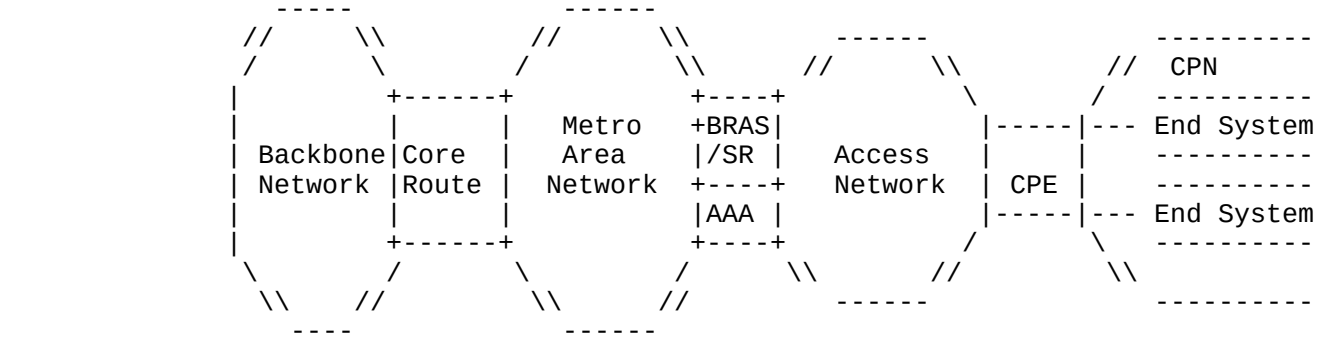


Figure 3 Network Architecture for Large SP Network

1. CPN is the part of the network by a customer when connecting to an ISP’s network which includes the CPE and the last hop link.

2. In Access Network, a very wide variety of access technologies can be used, including ADSL, Ethernet, PON, ATM, WIFI, etc.

3. MAN is the aggregated network for customers in one single metro, with the vast range of size. In most metro networks, BRAS is connected to Core Router directly, while for a small portion of large metro networks, BRAS is connected to Core Routers via aggregated routers.

4. Backbone network is to offer transit service between MANs and other ISPs.

There are typically two network models for fixed broadband access service: one is to access using PPPoE/PPPoA authentication method while the other is to use IPoE. The first one is usually applied to Residential network and SOHO networks. Subscribers in CPNs can access broadband network by PPP dial-up authentication. BRAS is the key network element which takes full responsibility of IP address assignment, user authentication, traffic aggregation, PPP session termination, etc. Then IP traffic is forwarded to Core Routers through Metro Area Network, and finally transited to external Internet via Backbone network.

The second network scenario is usually applied to large enterprise networks. Subscribers in CPNs can access broadband network by IPoE authentication. IP address is normally assigned by DHCP server, or static configuration.

5.2. dIVI Deployment Scenario in Operational Network

The deployment model of dIVI in operational network is depicted in Figure4.

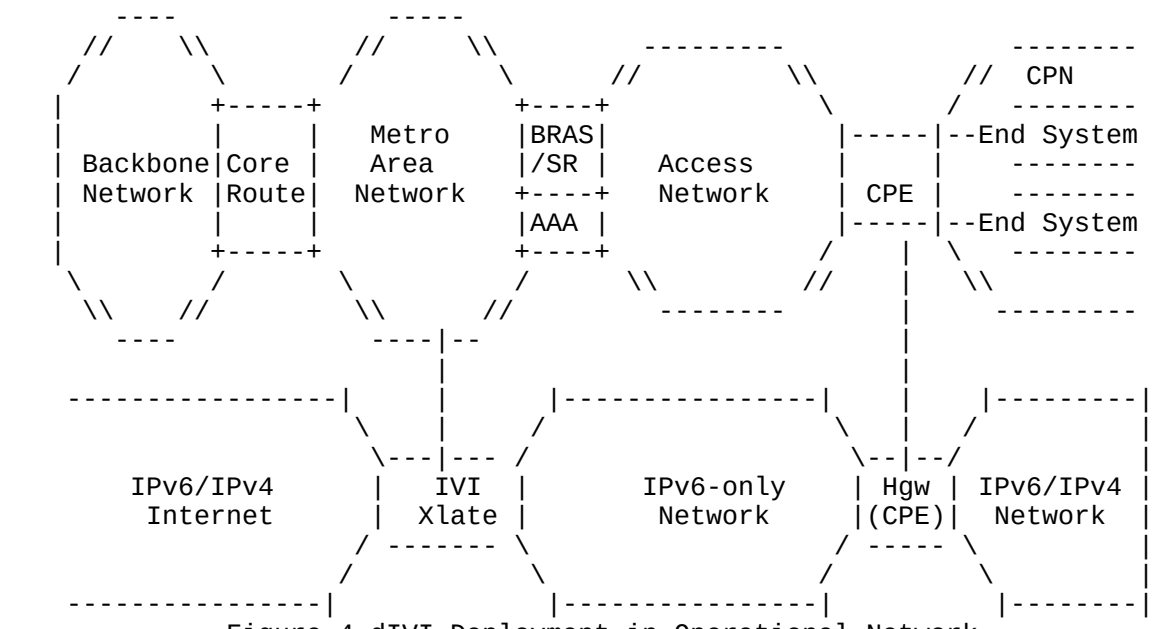


Figure 4 dIVI Deployment in Operational Network

In stateless dIVI, the network between Hgw and Xlate is an IPv6-only network, in which the operational and maintenance cost can be greatly reduced. The access network behind Hgw can either be IPv4-only or dual-stack. Thus, IPv4-only system and dual-stack system can communication with IPv4 Internet using shared IPv4 address by dIVI and the dual-stack system can also communicate with IPv6 Internet directly.

In operational network, Hgw can usually be integrated with CPE, while Xlate can be in someplace of MAN or Backbone Network. Subscribers can get IPv6 address from BRAS/SR after user authentication stage. Then, IVI-related route should be imported into the IPv6-only network between Xlate and Hgw. The detailed considerations for dIVI deployment will be discussed in section 6.

6. Considerations for dIVI deployment

This section describes the considerations for dIVI deployment in large operational network.

The major differences between dIVI deployment in laboratory and operational network lie in:

1. Operational network is a commercial network with strict user management requirement, while in laboratory it is simple and straightforward.

2. Operational network has different kinds of network equipments, e.g. BRAS/SR, CPE, Radius, etc. It would be more difficult to take modifications on all of these equipments.

3. Operational network has a large number of customers. Thus, it would be impossible to take manual configuration for all customers.

In this section, we try to outline considerations for dIVI deployment on large SP network. Some of the features are not specific to dIVI, but rather to a general requirement on all IPv6 transition techniques.

6.1. Addressing

In dIVI, there is no need to allocate IVI6 address explicitly to end users. Thus, the process of IPv6 address assignment can be integrated with existing IPv6 address allocation process. Only CPE will need to get IVI4 address, reallocate it to end user, do port-mapping and traffic translation with port-related information. Here are some basic considerations in dIVI addressing:

- Determine IVI6 prefix for each Xlate. Operators should use its own prefix as an IVI6 prefix, i.e. pref=2001:db8:a4a6::/48, in order to perform stateless translation. Address allocation process in BRAS/SR should be consistent with Xlate.

- Determine the embedded IVI4 address and port multx ratio. Operators should estimate the scale of subscribers in a certain region, evaluate the number of remaining IPv4 address, and analyze the number of concurrent ports. It is a tradeoff between multx ratio and concurrent port numbers. The bigger the multx ratio is, the more a IPv4 address can be shared by multiple subscribers and the less concurrent port number can be used per subscriber. From the above test in our laboratory, we choose multx ratio to be 128 and it is enough for current usage.

- Determine the ways to distribute the configuration profile including IVI4 address and port multx ratio to Hgw automatically, either by extended DHCP option, or other protocols.

6.2. Routing

In dIVI, IVI4(i) and IVI6(i) will be aggregated to ISP's IPv4 address and ISP's IPv6 address. They will not affect the global IPv4 and IPv6 routing tables

In dIVI deployment model, Hgws are normally configured with a default route that points to the BRAS/SR. The routers between BRAS/SR and Xlate run IPv6 dynamic routing protocols (IGP or BGP), and routers in the upper level of Xlate run IPv4 dynamic routing protocols. Therefore, the aggregated IVI6 routing directing to the upper routers will be learned/inserted by in IPv6-only domain. And the IVI6 route directing to Hgws should also be configured in BRAS/SR.

6.3. DNS

In dIVI, there is no DNS64/DNS46 needed anymore. An IPv6 DNS server is needed to process IPv6 DNS request/response, and the address of IPv6 DNS server should be passed to Hgw.

Since there is no IPv4 DNS server in IPv6-only network, it is recommended that Hgw should implement IPv4-to-IPv6 DNS Proxy to convert an IPv4 DNS request/response to IPv6 DNS request/response accordingly.

6.4. AAA and User Management

User authentication can be used to control who can have the dIVI connectivity service. This is not always required when a customer of IPv4 service automatically can have access to the dIVI service. However, it is highly recommended that IPv6-only customers should be authenticated separately. It is good for security, trouble shooting, user accounting, etc. There are some major points that AAA systems need to be taken into consideration:

- User authentication function needs to be extended to support the identification of IPv6-only subscriber, with additional dIVI-related profile for subscribers, e.g. IVI6 address, IVI4 address, non-IVI address, etc.

- User accounting and management function needs to be extended to identify dIVI user (or IPv6-only user) separately.

In summary, the major challenge of dIVI for the AAA and User Management is no big difference compared to an IPv6-only environment.

6.5. Network management

There are two issues to manage dIVI in operational network:

- Manage IPv6-only network. Operators should be able to manage IPv6-only network, including IPv6 MIB modules, Netflow Records, log information, etc.

- Manage the translation process. There are some exceptions that the MIB modules need to add dIVI related features, e.g. dIVI device management, dIVI traffic monitoring, etc.

6.6. dIVI CPE Requirements and Configuration

In dIVI, CPE is an important network element. It should perform DHCP server, integrated user authentication function, traffic translation and port mapping, DNS proxy, etc. The major operations in dIVI CPE include:

- Address assignment: dIVI CPE should support IVI4 address assignment by DHCP process to end users. It should also support IPv6 address assignment, either by stateful DHCP or stateless auto-configuration.

- Integrated user authentication function: dIVI CPE should integrate with existing user authentication function, e.g. PPPoE/PPPoA, etc.

- DNS: CPE should enable RFC 5006, well-known addresses, and DHCPv6 in order to maximize the likelihood of dIVI Hgw being able to use DNS without manual configuration. Besides, dIVI CPE should also support IPv4-to-IPv6 DNS proxy.

6.7. dIVI Xlate Placement in Large SP Network

Normally, dIVI Xlate can be deployed in "centralized model" and "distributed model".

In "centralized model", dIVI Xlate could be deployed in the place of Core Router, or even in the entrance of ICP. Since dIVI is a stateless method with better scalability than stateful ones, it can handle numerous concurrent sessions.

In "distributed model", dIVI Xlate is usually be integrated with BRAS/SR. So each Xlate should be configured with its own IVI6 prefix

and is responsible for translating the traffic of its own region. The number of subscribers is normally limited, so does the number of IVI routing entries. However, the network infrastructure should still be upgraded to dual-stack support in MAN and backbone network, and so the decreased operational cost is rather limited. Besides, since the newly emerging customers might be distributed in the whole Metro area, we have to deploy Xlate on all BRAS/SRs. This will cost a lot in the initial phase of IPv6 transition period.

In summary, we strongly recommend adopting "centralized model" for dIVI. It is a cost-effective way and easy to manage.

6.8. ALG consideration

dIVI does not require ALG, this is a very important feature in the initial development phase of IPv6.

7. Security Considerations

There are no security considerations in this document.

8. IANA Considerations

This memo adds no new IANA considerations.

Note to RFC Editor: This section will have served its purpose if it correctly tells IANA that no new assignments or registries are required, or if those assignments or registries are created during the RFC publication process. From the author’s perspective, it may therefore be removed upon publication as an RFC at the RFC Editor’s discretion.

9. References

9.1. Normative References

[I-D.ietf-behave-address-format] C., Bao, Huitema, C., Bagnulo, M., Boucadair, M., and X.Li, "IPv6 Addressing of IPv4/IPv6 Translators", draft-ietf-behave-address-format-10 (work in progress), August 2009.

[I-D.ietf-behave-dns64] Bagnulo, M., Sullivan, A., Matthews, P., and I. Beijnum, "DNS64: DNS extensions for Network Address Translation from IPv6 Clients to IPv4 Servers", draft-ietf-behave-dns64-11 (work in progress),October 2009.

[I-D.ietf-behave-v6v4-framework] Baker, F., Li, X., Bao, C., and K. Yin, "Framework for IPv4/IPv6 Translation", draft-ietf-behave-v6v4-framework-10 (work in progress), October 2009.

[I-D.ietf-behave-v6v4-xlate] Li, X., Bao, C., and F. Baker, "IP/ICMP Translation Algorithm", draft-ietf-behave-v6v4-xlate-23 (work in progress), October 2009.

[I-D.ietf-behave-v6v4-xlate-stateful] Bagnulo, M., Matthews, P., I. Beijnum, "IP/ICMP Translation Algorithm", draft-ietf-behave-v6v4-xlate-12 (work in progress), October 2009.

[I-D.xli-behave-divi] Li,X., Bao, C., and Zhang, H., "Address-sharing stateless double IVI", draft-xli-behave-divi-01, April 29, 2010.

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

10. Acknowledgments

The authors would like to thank to Fred Baker for his continuous suggestion around this topic over the years. Thanks to Qian Wang, Jie Hu and Fan Shi for useful feedback.

Authors' Addresses

Qiong SUN
China Telecom Beijing Research Institute
Room 708 No.118, Xizhimenneidajie, xicheng District Beijing 100035
China

Phone: <86 10 58552636>
Email: sunqiong@ctbri.com.cn

Heyu Wang
China Telecom Beijing Research Institute
Room 708 No.118, Xizhimenneidajie, xicheng District Beijing 100035
China

Phone: <86 10 58552117>
Email: wanghy@ctbri.com.cn

Xing Li
CERNET Center/Tsinghua University
Room 225, Main Building, Tsinghua University

Phone: <86 10 62785983>
Email: xing@cernet.edu.cn

Congxiao Bao
CERNET Center/Tsinghua University
Room 225, Main Building, Tsinghua University

Phone: <86 10 62785983>
Email: congxiao@cernet.edu.cn >

Ming Feng
China Telecom
No.31, Jinrong Ave,Xicheng District,100032

Phone: <86 10 58501428>
Email: fengm@chinatelecom.com.cn

Network Working Group
Internet Draft
Intended status: Informational
Expires: April 24, 2011

Qiong Sun
Chongfeng Xie
China Telecom
Xing Li
CongXiao Bao
CERNET Center/Tsinghua University
Ming Feng
China Telecom

March 6, 2011

Considerations for Stateless Translation (IVI/dIVI) in Large SP Network
draft-sunq-v6ops-ivi-sp-02.txt

Abstract

With the approaching exhaustion of IPv4 address space, large-scale SPs are now faced with the only real option to deploy IPv6 in a timely manner. In order to achieve smooth transition to IPv6, migration tools should be introduced for different deployment models. Among different IPv6 transition mechanisms, dIVI is a prefix-specific and stateless address mapping method which can directly translate IPv4 packet to IPv6 packet. This document describes the challenges and requirements for large SP to deploy IPv6 in operational network, the experimental results of dIVI in our laboratory and the considerations for dIVI deployment in large SP operational network.

This Internet-Draft will expire on September 6,2011.

Copyright Notice

Table of Contents

1. Introduction

The dramatic growth of the Internet is accelerating the exhaustion of available IPv4 addressing pool. It is widely accepted that IPv6 is the only real option on the table for the continued growth of the Internet. However, IPv6 deployment is a huge systematic project and a lot of challenges will arise especially in large SP operational network.

In order to achieve smooth transition to IPv6, migration tools should be introduced for different deployment models, among which dIVI is a stateless translation mechanism with good scalability. This document describes the challenges and requirements for large SPs in IPv6 transition period. Then, we introduce dIVI experimental results in our laboratory. And finally, the considerations for designing and deploying dIVI in operational network are discussed in terms of addressing and routing, DNS deployment requirement, AAA support and user management, network management, CPE requirement and Xlate placement.

2. Terminologies

This document uses the terminologies defined in [I-D.ietf-behave-v6v4-framework], [I-D.ietf-behave-v6v4-xlate], [I-D.ietf-behave-address-format], [I-D.ietf-behave-v6v4-xlate-stateful], and [I-D.xli-behave-ivi].

The key words MUST, MUST NOT, REQUIRED, SHALL, SHALL NOT, SHOULD, SHOULD NOT, RECOMMENDED, MAY, and OPTIONAL, when they appear in this document, are to be interpreted as described in [RFC2119].

3. Problem Statement

It is well known that the pool of public IPv4 addressing is nearing its exhaustion. The '/8' IANA blocks for Regional Internet Registries (RIRs) are projected to 'run-out' towards the end of 2011. Credible estimates based on past behavior suggest that the RIRs will exhaust their remaining address space by early 2012, apart from the development of a market in IPv4 address space. Thus, it will become much more difficult to get available public IPv4 addresses. In the same time, a lot of emerging applications, e.g. Apple's iPad, Motion's BlackBerry, etc. will quickly deplete the available addresses. This has led to a hightened awareness among the providers to consider introducing IPv6 to keep the Internet operational.

It has been widely accepted that the end goal of IPv6 transition is to achieve an end-to-end IPv6-only network, and IPv4 can eventually

be turned off. However, since it will have impact on almost the entire world, it will take a considerable period of time to reach the ultimate goal. As a result, IPv4 and IPv6 need to coexist during the whole transition period. In this document, we mainly focus on IPv6 migration issues from large ISP's point of view. In order to facilitate smooth IPv6 migration, some factors need to be taken into consideration especially for large SPs. There are ten major requirements:

1. It should deploy in an incremental fashion and the overall transition process should be stable and operational.

2. It should largely reduce IPv4 public address consumption.

3. It should accelerate the deployment of IPv6, rather than just prolonging the lifecycle of IPv4 by introducing multiple layers of NAT.

4. There should be no perceived degradation of customer experience. As a result, IPv6 transition mechanisms should provide IPv4 service continuity.

5. It should achieve scalability, simplicity and high availability, especially for large-scale SPs.

6. It should have user management and network management ability.

7. It should support user authentication, authorization and accounting as an essential part of operational network.

8. Most ISPs need some kind of mechanisms to trace the origin of traffic in their networks. This should also be available for IPv6 traffic.

9. It should have good throughput performance and massive concurrent session support.

10. It should maintain the deployment concepts and business models which have been proven and used with existing revenue generating IPv4 services.

All existing IPv6 transition mechanisms can be widely divided into three categories: dual-stack solution, translation-based solution and tunnel-based solution. In this document, we mainly concentrate on stateless translation mechanism: dIVI. The original stateless IPv4/IPv6 translation (stateless 1:1 IVI) is scalable, [I-D.ietf-behave-v6v4-framework], [I-D.ietf-behave-v6v4-xlate], [I-D.ietf-

behave-address-format],[I-D.xli-behave-ivi]. But it cannot use the IPv4 addresses effectively. The stateless dIVI[I-D.xli-behave-divi] is a double translation mechanism which includes a 1:N stateless translator and a 1:1 Hgw translator. The 1:N stateless translator is implemented in the border between the IPv6 network and the IPv4 Internet. It translates the packets between IPv4 and IPv6 with the 1:N stateless address mapping. The 1:1 Hgw translator is implemented between an IPv6 network and user's end system. It translates the packets between IPv4 and IPv6 with 1:1 stateless address mapping. In addition, the home gateway translator maps random source port numbers to restricted port number based on the extended IPv4-translatable address format and keeps the mapping table in database for the port number mapping of the retuning packets and all the packets in the same session.

dIVI support bidirectional communication initiated from IPv4 and IPv6. It can be deployed in an IPv6-only access network, in which operational and maintenance cost can be reduced. It has very good scalability and can largely reduce IPv4 address consumption.

In this document, we firstly demonstrate the laboratory experimental results of dIVI in section 4. We can see that dIVI can support most of the current IPv4 applications in IPv6-only access network, while largely reducing IPv4 address consumption. And then dIVI deployment model and considerations in large operational network are proposed in section 5 and section 6 respectively. Some important factors need to be taken into account when introducing dIVI. Since most challenges in dIVI have no big difference compared to an IPv6-only environment, we strongly recommend that related network elements should take the corresponding modifications in order to guarantee the IPv6 transition process to be operational and manageable.

4. Laboratory experiment

We have tested dIVI using the prototype in our laboratory. The major objective listed in the following is to verify the functionality and performance of dIVI:

O Verify how to deploy dIVI in practical network.

O Verify what applications can be used in dIVI.

O Verify what Operating Systems can be supported in dIVI.

O Verify the effect of user experience with limited ports.

O Verify the performance of dIVI.

4.1. Experiment environment

We have tested dIVI in our laboratory. The network topology is depicted in Figure 1.

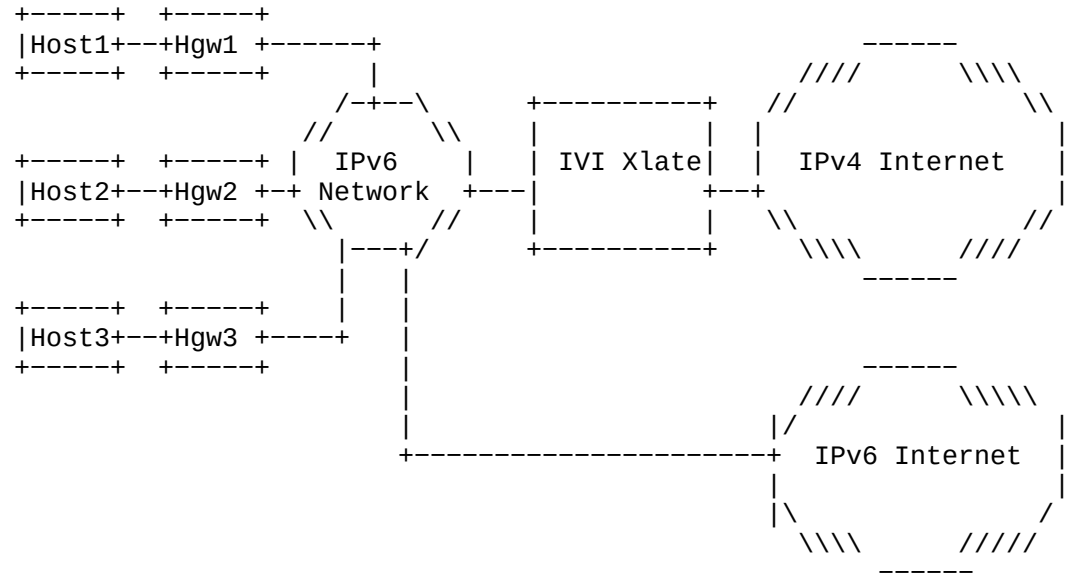


Figure 1 dIVI topology in the test

There are three key components in the test:

- The Hosts are dual-stack customers, who could run IPv4 application, IPv6 application or dual stack application.

- The Home Gateways (Hgw) are dIVI translator in user side. It translates the packets between IPv4 and IPv6 with 1:1 stateless address mapping, and maps random source port numbers to restricted port number.

- The Xlate translates the packets between IPv4 and IPv6 with the 1:N stateless address mapping.

The network between Hgw and Xlate is IPv6-only, and the network behind Hgw is dual-stack. Thus, the hosts behind Hgw can communicate with both IPv4 Internet and IPv6 Internet.

4.2. Experiment configuration

For address configuration, each host will use two IPv6 addresses: one is IVI6 address which is synthesized in Hgw with the IVI4 address and port-related information, and the other is non-IVI IPv6 address which is used for native IPv6 communication. We should notice that only non-IVI IPv6 address needs be allocated to end users. Besides, each user will get an IVI4 address from Hgw.

For routing configuration, both IVI address and non-IVI address need to be imported into the IPv6-only network.

For port configuration, since there are 65536 TCP/UDP ports for each IP address, and in fact one can use hundreds only for normal applications, so one IPv4 address can be shared by multiple customers. In our experiment, we selected ratio to be 128. That is to say, one IPv4 address is shared by 128 users, and there are 512 available ports per user.

For DNS configuration, there is no need to have additional DNS64 for dIVI. Only an IPv6 DNS server with A/AAAA records is needed and the DNS address is manually configured in Hgw. Besides, Hgw has implemented DNS Proxy and it will convert an IPv4 DNS request/response to IPv6 DNS request/response.

For ALG configuration, there is no need to deploy specific ALG for IPv4 applications in dIVI.

4.3. Experiment results

In our laboratory, we mainly have taken the following tests:

- Application test: The applications we have tested include web, email, Instant Message, ftp, telnet, SSH, video, Video Camera, P2P, online game, Voip, VPN and so on.

- Operating System test: The OS we have tested includes Win7, VISTA, windows XP.

- Performance test: We have measured the parameters of concurrent session number, throughput performance.

The experimental results are listed as follows:

Type	Experiment Result
Application test	dIVI hosts can support web, email, im, ftp , telnet, SSH, video, Video Camera, P2P, online game, voip, and so on.
Operating System test	dIVI can widely support Win7, VISTA, windows XP.
Performance test	The performance test for dIVI Xlate is carried out on a normal PC. Due to limitation of the PC hardware, the overall throughput is not quite good. However, it can still support more than one hundred million concurrent sessions.

Figure 2 dIVI test result

From the experiment, we can have the following conclusions:

1. dIVI can have good scalability. Xlate does not need to maintain any session state, and only limited session states have to been maintained in Hgw for its customer.

2. dIVI can be deployed in an incremental way. The most complicated part of dIVI is addressing and routing. The configuration for DNS and ALG is very simple.

3. dIVI can support a majority of current IPv4 applications.

4. dIVI can support a variety of Operating Systems.

5. With the ratio of 128 (512 maximum concurrent sessions), there is no perceived degradation of customer experience.

However, in the current status of equipment, e.g. BRAS, end system, etc., the support for IPv6-only function has not been fully accomplished. Therefore, there are still some limitations which would be improved in the next version of dIVI development when deploying dIVI prototype in practical operational network:

1. Address assignment process have not been integrated with existing address allocation system.

2. Currently, IVI routing entries are configured manually.

3. Hgw has not integrated PPPoE functionality with dIVI functionality.

4. AAA system has not supported IVI-related (or IPv6-only) functions.

With regard to the listed IPv6 transition requirements in section 3, most of them can be satisfied by dIVI, except for the requirement of network management and user management. These two points should be paid special attention for large SPs, which will be further discussed in section 6.

5. dIVI Deployment Scenario

In order to investigate the ways to deploy dIVI in operational network, we firstly briefly discuss network architecture for large SP network. Then dIVI deployment model is introduced.

5.1. Network Architecture for Large SP Network

In large SPs, the generic network topology can be divided into four main parts (as depicted in Figure3): the Customer Premises Network (CPN), the Access Network (AN), the Metro Area Network (MAN), and the Backbone Network.

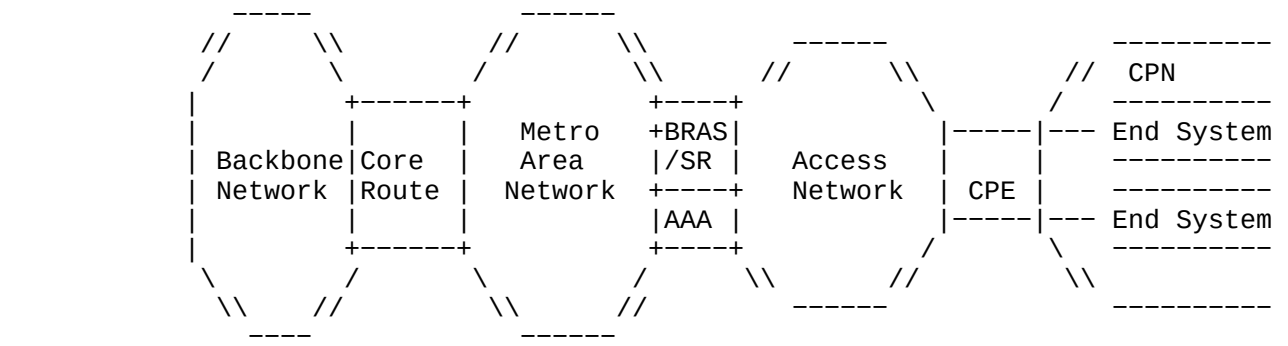


Figure 3 Network Architecture for Large SP Network

1. CPN is the part of the network by a customer when connecting to an ISP's network which includes the CPE and the last hop link.

2. In Access Network, a very wide variety of access technologies can be used, including ADSL, Ethernet, PON, ATM, WIFI, etc.

3. MAN is the aggregated network for customers in one single metro, with the vast range of size. In most metro networks, BRAS is connected to Core Router directly, while for a small portion of large metro networks, BRAS is connected to Core Routers via aggregated routers.

4. Backbone network is to offer transit service between MANs and other ISPs.

There are typically two network models for fixed broadband access service: one is to access using PPPoE/PPPoA authentication method while the other is to use IPoE. The first one is usually applied to Residential network and SOHO networks. Subscribers in CPNs can access broadband network by PPP dial-up authentication. BRAS is the key network element which takes full responsibility of IP address assignment, user authentication, traffic aggregation, PPP session termination, etc. Then IP traffic is forwarded to Core Routers through Metro Area Network, and finally transited to external Internet via Backbone network.

The second network scenario is usually applied to large enterprise networks. Subscribers in CPNs can access broadband network by IPoE authentication. IP address is normally assigned by DHCP server, or static configuration.

5.2. dIVI Deployment Scenario in Operational Network

The deployment model of dIVI in operational network is depicted in Figure4.

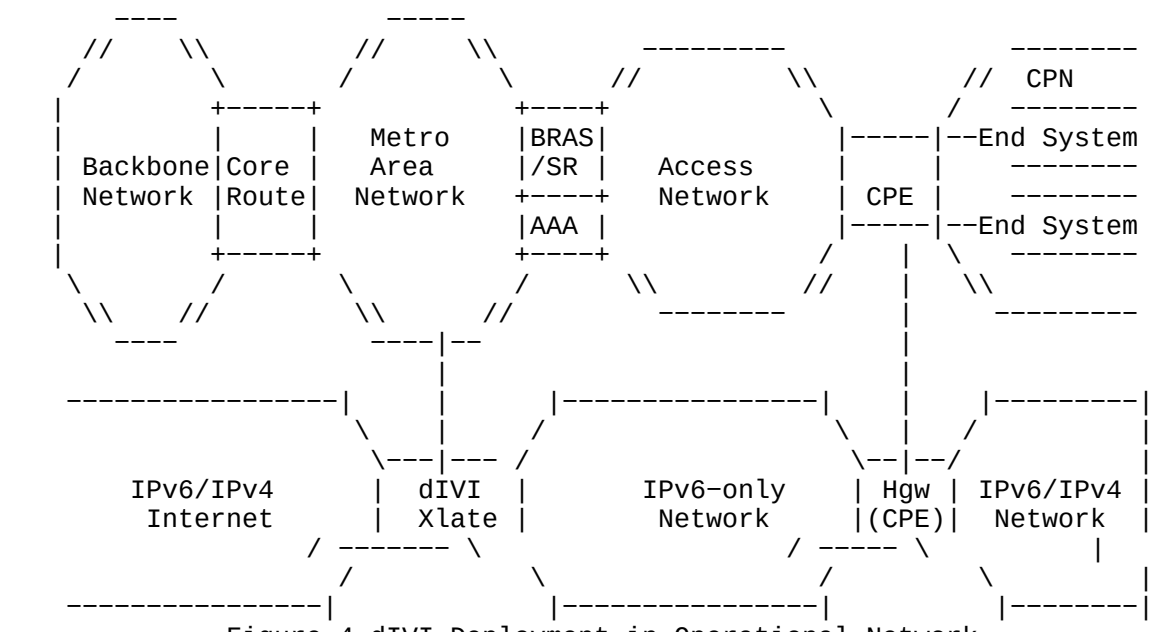


Figure 4 dIVI Deployment in Operational Network

In stateless dIVI, the network between Hgw and Xlate is an IPv6-only network, in which the operational and maintenance cost can be greatly reduced. The access network behind Hgw can either be IPv4-only or dual-stack. Thus, IPv4-only system and dual-stack system can communication with IPv4 Internet using shared IPv4 address by dIVI and the dual-stack system can also communicate with IPv6 Internet directly.

In operational network, Hgw can usually be integrated with CPE, while Xlate can be in someplace of MAN or Backbone Network. Subscribers can get IPv6 address from BRAS/SR after user authentication stage. Then, IVI-related route should be imported into the IPv6-only network between Xlate and Hgw. The detailed considerations for dIVI deployment will be discussed in section 6.

6. Considerations for dIVI deployment

This section describes the considerations for dIVI deployment in large operational network.

The major differences between dIVI deployment in laboratory and operational network lie in:

1. Operational network is a commercial network with strict user management requirement, while in laboratory it is simple and straightforward.

2. Operational network has different kinds of network equipments, e.g. BRAS/SR, CPE, Radius, etc. It would be more difficult to take modifications on all of these equipments.

3. Operational network has a large number of customers. Thus, it would be impossible to take manual configuration for all customers.

In this section, we try to outline considerations for dIVI deployment on large SP network. Some of the features are not specific to dIVI, but rather to a general requirement on all IPv6 transition techniques.

6.1. Addressing

In dIVI, there is no need to allocate IVI6 address explicitly to end users. Thus, the process of IPv6 address assignment can be integrated with existing IPv6 address allocation process. Only CPE will need to get IVI4 address, reallocate it to end user, do port-mapping and traffic translation with port-related information. Here are some basic considerations in dIVI addressing:

- Determine IVI6 prefix for each Xlate. Operators should use its own prefix as an IVI6 prefix, i.e. pref=2001:db8:a4a6::/48, in order to perform stateless translation. Address allocation process in BRAS/SR should be consistent with Xlate.

- Determine the embedded IVI4 address and port multx ratio. Operators should estimate the scale of subscribers in a certain region, evaluate the number of remaining IPv4 address, and analyze the number of concurrent ports. It is a tradeoff between multx ratio and concurrent port numbers. The bigger the multx ratio is, the more an IPv4 address can be shared by multiple subscribers and the less concurrent port number can be used per subscriber. From the above test in our laboratory, we choose multx ratio to be 128 and it is enough for current usage.

- Determine the ways to distribute the configuration profile including IVI4 address and port multx ratio to Hgw automatically, either by extended DHCP option, or other protocols.

6.2. Routing

In dIVI, IVI4(i) and IVI6(i) will be aggregated to ISP's IPv4 address and ISP's IPv6 address. They will not affect the global IPv4 and IPv6 routing tables

In dIVI deployment model, Hgws are normally configured with a default route that points to the BRAS/SR. The routers between BRAS/SR and Xlate run IPv6 dynamic routing protocols (IGP or BGP), and routers in the upper level of Xlate run IPv4 dynamic routing protocols. Therefore, the aggregated IVI6 routing directing to the upper routers will be learned/inserted by in IPv6-only domain. And the IVI6 route directing to Hgws should also be configured in BRAS/SR.

6.3. DNS

In dIVI, there is no DNS64/DNS46 needed anymore. An IPv6 DNS server is needed to process IPv6 DNS request/response, and the address of IPv6 DNS server should be passed to Hgw.

Since there is no IPv4 DNS server in IPv6-only network, it is recommended that Hgw should implement IPv4-to-IPv6 DNS Proxy to convert an IPv4 DNS request/response to IPv6 DNS request/response accordingly.

6.4. AAA and User Management

User authentication can be used to control who can have the dIVI connectivity service. This is not always required when a customer of IPv4 service automatically can have access to the dIVI service. However, it is highly recommended that IPv6-only customers should be authenticated separately. It is good for security, trouble shooting, user accounting, etc. There are some major points that AAA systems need to be taken into consideration:

- User authentication function needs to be extended to support the identification of IPv6-only subscriber, with additional dIVI-related profile for subscribers, e.g. IVI6 address, IVI4 address, non-IVI address, etc.

- User accounting and management function needs to be extended to identify dIVI user (or IPv6-only user) separately.

In summary, the major challenge of dIVI for the AAA and User Management is no big difference compared to an IPv6-only environment.

6.5. Network management

There are two issues to manage dIVI in operational network:

- Manage IPv6-only network. Operators should be able to manage IPv6-only network, including IPv6 MIB modules, Netflow Records, log information, etc.

- Manage the translation process. There are some exceptions that the MIB modules need to add dIVI related features, e.g. dIVI device management, dIVI traffic monitoring, etc.

6.6. dIVI CPE Requirements and Configuration

In dIVI, CPE is an important network element. It should perform DHCP server, integrated user authentication function, traffic translation and port mapping, DNS proxy, etc. The major operations in dIVI CPE include:

- Address assignment: dIVI CPE should support IVI4 address assignment by DHCP process to end users. It should also support IPv6 address assignment, either by stateful DHCP or stateless auto-configuration.

- Integrated user authentication function: dIVI CPE should integrate with existing user authentication function, e.g. PPPoE/PPPoA, etc.

- DNS: CPE should enable RFC 5006, well-known addresses, and DHCPv6 in order to maximize the likelihood of dIVI Hgw being able to use DNS without manual configuration. Besides, dIVI CPE should also support IPv4-to-IPv6 DNS proxy.

6.7. dIVI Xlate Placement in Large SP Network

Normally, dIVI Xlate can be deployed in "centralized model" and "distributed model".

In "centralized model", dIVI Xlate could be deployed in the place of Core Router, or even in the entrance of ICP. Since dIVI is a stateless method with better scalability than stateful ones, it can handle numerous concurrent sessions.

In "distributed model", dIVI Xlate is usually be integrated with BRAS/SR. So each Xlate should be configured with its own IVI6 prefix

and is responsible for translating the traffic of its own region. The number of subscribers is normally limited, so does the number of IVI routing entries. However, the network infrastructure should still be upgraded to dual-stack support in MAN and backbone network, and so the decreased operational cost is rather limited. Besides, since the newly emerging customers might be distributed in the whole Metro area, we have to deploy Xlate on all BRAS/SRs. This will cost a lot in the initial phase of IPv6 transition period.

In summary, we strongly recommend adopting "centralized model" for dIVI. It is a cost-effective way and easy to manage.

6.8. ALG consideration

dIVI does not require ALG, this is a very important feature in the initial development phase of IPv6.

7. Security Considerations

There are no security considerations in this document.

8. IANA Considerations

This memo adds no new IANA considerations.

Note to RFC Editor: This section will have served its purpose if it correctly tells IANA that no new assignments or registries are required, or if those assignments or registries are created during the RFC publication process. From the author's perspective, it may therefore be removed upon publication as an RFC at the RFC Editor's discretion.

9. References

9.1. Normative References

[I-D.ietf-behave-address-format] C., Bao, Huitema, C., Bagnulo, M., Boucadair, M., and X.Li, "IPv6 Addressing of IPv4/IPv6 Translators", draft-ietf-behave-address-format-10 (work in progress), August 2009.

[I-D.ietf-behave-dns64] Bagnulo, M., Sullivan, A., Matthews, P., and I. Beijnum, "DNS64: DNS extensions for Network Address Translation from IPv6 Clients to IPv4 Servers", draft-ietf-behave-dns64-11 (work in progress),October 2009.

[I-D.ietf-behave-v6v4-framework] Baker, F., Li, X., Bao, C., and K. Yin, "Framework for IPv4/IPv6 Translation", draft-ietf-behave-v6v4-framework-10 (work in progress), October 2009.

[I-D.ietf-behave-v6v4-xlate] Li, X., Bao, C., and F. Baker, "IP/ICMP Translation Algorithm", draft-ietf-behave-v6v4-xlate-23 (work in progress), October 2009.

[I-D.ietf-behave-v6v4-xlate-stateful] Bagnulo, M., Matthews, P., I. Beijnum, "IP/ICMP Translation Algorithm", draft-ietf-behave-v6v4-xlate-12 (work in progress), October 2009.

[I-D.xli-behave-divi] Li,X., Bao, C., and Zhang, H., "Address-sharing stateless double IVI", draft-xli-behave-divi-01, April 29, 2010.

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

10. Acknowledgments

The authors would like to thank to Fred Baker for his continuous suggestion around this topic over the years. Thanks to Qian Wang, Jie Hu and Fan Shi for useful feedback.

Authors' Addresses

Qiong SUN
China Telecom Beijing Research Institute
Room 708 No.118, Xizhimenneidajie, xicheng District Beijing 100035
China

Phone: <86 10 58552636>
Email: sunqiong@ctbri.com.cn

Chongfeng Xie
China Telecom Beijing Research Institute
Room 708 No.118, Xizhimenneidajie, xicheng District Beijing 100035
China

Phone: <86 10 58552116>
Email: xiechf@ctbri.com.cn

Xing Li
CERNET Center/Tsinghua University
Room 225, Main Building, Tsinghua University

Phone: <86 10 62785983>
Email: xing@cernet.edu.cn

Congxiao Bao
CERNET Center/Tsinghua University
Room 225, Main Building, Tsinghua University

Phone: <86 10 62785983>
Email: congxiao@cernet.edu.cn >

Ming Feng
China Telecom
No.31, Jinrong Ave,Xicheng District,100032

Phone: <86 10 58501428>
Email: fengm@chinatelecom.com.cn

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: January 27, 2011

O. Troan, Ed.
Cisco
D. Miles
Alcatel-Lucent
S. Matsushima
SOFTBANK TELECOM Corp.
T. Okimoto
NTT
D. Wing
Cisco
July 26, 2010

IPv6 Multihoming without Network Address Translation
draft-troan-multihoming-without-nat66-01

Abstract

Network Address and Port Translation (NAPT) works well for conserving global addresses and addressing multihoming requirements, because an IPv4 NAPT router implements three functions: source address selection, next-hop resolution and optionally DNS resolution. For IPv6 hosts one approach could be the use of IPv6 NAT. However, NAT should be avoided, if at all possible, to permit transparent host-to-host connectivity. In this document, we analyze the use cases of multihoming. We also describe functional requirements for multihoming without the use of NAT in IPv6 for hosts and small IPv6 networks that would otherwise be unable to meet minimum IPv6 allocation criteria .

Table of Contents

1. Introduction

IPv6 provides enough globally unique addresses to permit every conceivable host on the Internet to be uniquely addressed without the requirement for Network Address Port Translation (NAPT [RFC3022]) offering a renaissance in host-to-host transparent connectivity.

Unfortunately, this may not be possible due to the necessity of NAT even in IPv6, because of multihoming.

Multihoming is a blanket term to describe a host or small network that is connected to more than one upstream network. Whenever a host or small network (which does not meet minimum IPv6 allocation criteria) is connected to multiple upstream networks IPv6 addressing is assigned by each respective service provider resulting in hosts with more than one active IPv6 address. As each service provided is allocated a different address space from its Internet Registry, it in-turn assigns a different address space to the end-user network or host. For example, a remote access user may use a VPN to simultaneously connect to a remote network and retain a default route to the Internet for other purposes.

In IPv4 a common solution to the multihoming problem is to employ NAPT on a border router and use private address space for individual host addressing. The use of NAPT allows hosts to have exactly one IP address visible on the public network and the combination of NAPT with provider-specific outside addresses (one for each uplink) and destination-based routing insulates a host from the impacts of multiple upstream networks. The border router may also implement a DNS cache or DNS policy to resolve address queries from hosts.

It is our goal to avoid the IPv6 equivalent of NAT. To reach this goal, mechanisms are needed for end-user hosts to have multiple address assignments and resolve issues such as which address to use for sourcing traffic to which destination:

- If multiple routers exist on a single link the host must appropriately select next-hop for each connected network. Routing protocols that would normally be employed for router-to-router network advertisement seem inappropriate for use by individual hosts.

- Source address selection also becomes difficult whenever a host has more than one address within the same address scope. Current address selection criteria may result in hosts using an arbitrary or random address when sourcing upstream traffic. Unfortunately, for the host, the appropriate source address is a function of the upstream network for which the packet is bound for. If an

upstream service provider uses IP anti-spoofing or uRPF, it is conceivable that the packets that have inappropriate source address for the upstream network would never reach their destination.

- o In a multihomed environment, different DNS scopes or partitions may exist in each independent upstream network. A DNS query sent to an arbitrary upstream resolver may result in incorrect or poisoned responses.

In short, while IPv6 facilitates hosts having more than one address in the same address scope, the application of this causes significant issues for a host from routing, source address selection and DNS resolution perspectives. A possible consequence of assigning a host multiple identical-scoped addresses is severely impaired IP connectivity.

If a host connects to a network behind an IPv4 NAPT, the host has one private address in the local network. There is no confusion. The NAT becomes the gateway of the host and forwards the packet to an appropriate network when it is multihomed. It also operates a DNS cache server, which receives all DNS inquires, and gives a correct answer to the host.

In this document, we identify the functions present in multihomed IPv4 NAPT environments and propose requirements that address multihomed IPv6 environments without using IPv6 NAT.

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 RFC 2119 [RFC2119].

NAT66 or IPv6 NAT
: The terms "NAT66" and "IPv6 NAT" refer to [I-D.mrw-behave-nat66].

NAPT
: Network Address Port Translation as described in [RFC3022]. In other contexts, NAPT is often pronounced "NAT" or written as "NAT".

Multihomed with multi-prefix (MHMP)
: A host implementation which supports the mechanisms described in this document. Namely source address selection policy, next-hop selection and DNS selection policy.

3. IPv6 multihomed network scenarios

In this section, we classify three scenarios of the multihoming environment.

3.1. Classification of network scenarios for multihomed host

Scenario 1:

In this scenario, two or more routers are present on a single link shared with the host(s). Each router is in turn connected to a different service provider network, which provides independent address assignment and DNS resolvers. A host in this environment would be offered multiple prefixes and DNS resolvers advertised from the two different routers.

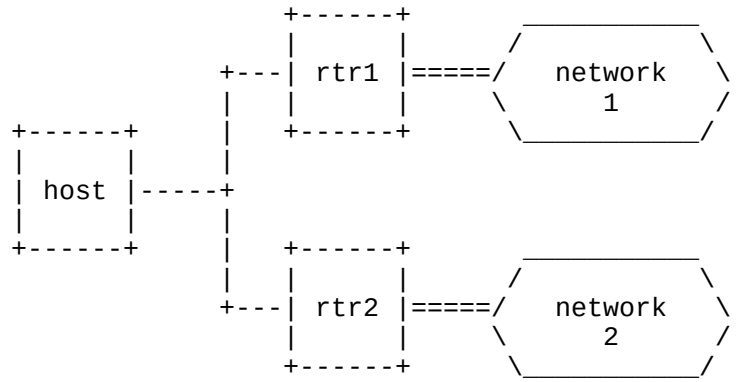


Figure 1: single uplink, multiple next-hop, multiple prefix (Scenario 1)

Figure 1 illustrates the host connecting to rtr1 and rtr2 via a shared link. Networks 1 and 2 are reachable via rtr1 and rtr2 respectively. When the host sends packets to network 1, the next-hop to network 1 is rtr1. Similarly, rtr2 is the next-hop to network 2.

- e.g., broadband service (Internet, VoIP, IPTV, etc.)

Scenario 2:

In this scenario, a single gateway router connects the host to two or more upstream service provider networks. This gateway router would receive prefix delegations from each independent service provider network and a different set of DNS resolvers. The gateway in turn advertises the provider prefixes to the host, and for DNS, may either

act as a lightweight DNS resolver/cache or may advertise the complete set of service provider DNS resolvers to the hosts.

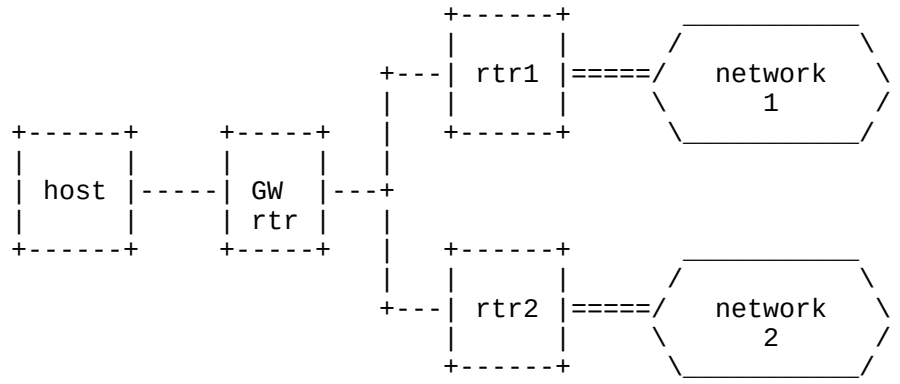


Figure 2: single uplink, single next-hop, multiple prefix (Scenario 2)

Figure 2 illustrates the host connected to GW rtr. GW rtr connects to networks 1 and 2 via rtr1 and rtr2, respectively. When the host sends packets to either network 1 or 2, the next-hop is GW rtr. When the packets are sent to network 1 (network 2), GW rtr forwards the packets to rtr1 (rtr2).

- e.g, Internet + VPN/ASP

Scenario 3:

In this scenario, a host has more than one active interfaces that connects to different routers and service provider networks. Each router provides the host with a different address prefix and set of DNS resolvers, resulting in a host with a unique address per link/ interface.

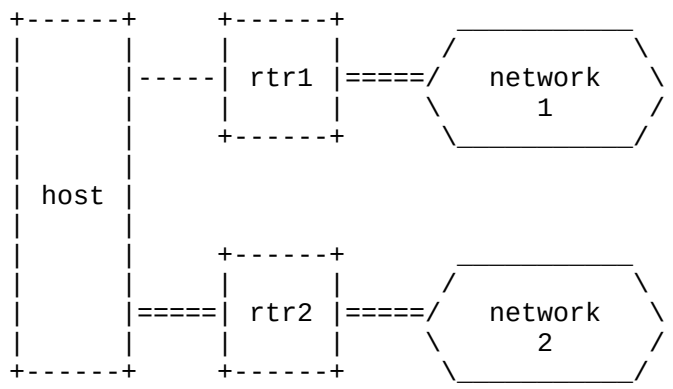


Figure 3: Multiple uplink, multiple next-hop, multiple prefix (Scenario 3)

Figure 3 illustrates the host connecting to rtr1 and rtr2 via a direct connection or a virtual link. When the host sends packets network 1, the next-hop to network 1 is rtr1. Similarly, rtr2 is the next-hop to network 2.

- e.g., Mobile Wifi + 3G, ISP A + ISP B

3.2. Multihomed network environment

In an IPv6 multihomed network, a host is assigned two or more IPv6 addresses and DNS resolvers from independent service provider networks. When this multihomed host attempts to connect with other hosts, it may incorrectly resolve the next-hop router, use an inappropriate source address, or use a DNS response from an incorrect service provider that may result in impaired IP connectivity.

Multihomed networks in IPv4 have been commonly implemented through the use of a gateway router with NAPT function (scenario 2 with NAPT). An analysis of the current IPv4 NAPT and DNS functions within the gateway router should provide a baseline set of requirements for IPv6 multihomed environments. A destination prefix/route is often used on the gateway router to separate traffic between the networks.

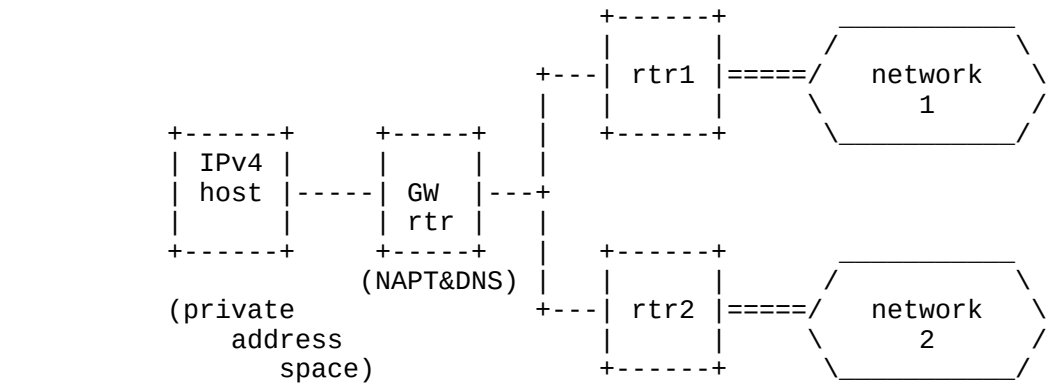


Figure 4: IPv4 Multihomed environment with Gateway Router performing NAPT

3.3. Multihomed Problem Statement

A multihomed IPv6 host has one or more assigned IPv6 addresses and DNS resolvers from each upstream service provider, resulting in the host having multiple valid IPv6 addresses and DNS resolvers. The host must be able to resolve the appropriate next-hop, the correct source address and DNS resolver to use based on the destination prefix. To prevent IP spoofing, operators will often implement IP filters and uRPF to discard traffic with an inappropriate source address, making it essential for the host to correctly resolve these three criteria before sourcing the first packet.

IPv6 has mechanisms for the provision of multiple routers on a single link and multiple address assignments to a single host. However, when these mechanisms are applied to the three scenarios in Section 3.1 a number of connectivity issues are identified:

Scenario 1:

The host has been assigned an address from each router and recognizes both rtr1 and rtr2 as valid default routers (in the default routers list).

- The source address selection policy on the host does not deterministically resolve a source address. Upstream uRPF or filter policies will discard traffic with source addresses that the operator did not assign.

- The host will select one of the two routers as the active default router. No traffic is sent to the other router.

Scenario 2:

The host has been assigned two different addresses from the single gateway router. The gateway router is the only default router on the link.

- The source address selection policy on the host does not deterministically resolve a source address. Upstream uRPF or filter policies will discard traffic with source addresses that the operator did not assign.

- The gateway router does not have a mechanism for determining which traffic should be sent to which network. If the gateway router is implementing host functions (ie, processing RA) then two valid default routers may be recognized.

Scenario 3:

A host has two separate interfaces and on each interface a different address is assigned. Each link has its own router.

- The host does not have enough information for determining which traffic should be sent to which upstream routers. The host will select one of the two routers as the active default router, and no traffic is sent to the other router.

- The default address selection rules select the address assigned to the outgoing interface as the source address. So, if a host has an appropriate routing table, an appropriate source address will be selected.

All scenarios:

- The host may use an incorrect DNS resolver for DNS queries.

4. Problem statement and analysis

The problems described in Section 3 can be classified into these three types:

- Wrong source address selection

- Wrong next-hop selection

- Wrong DNS server selection

This section reviews the problem statements presented above and the

proposed functional requirements to resolve the issues without employing IPv6 NAT.

4.1. Source address selection

A multihomed IPv6 host will typically have different addresses assigned from each service provider either on the same link (scenarios 1 & 2) or different links (scenario 3). When the host wishes to send a packet to any given destination, the current source address selection rules [RFC3484] may not deterministically resolve the correct source address when the host addressing was via RA or DHCPv6. [I-D.ietf-6man-addr-select-sol] describes the use of the policy table [RFC3484] to resolve this problem, but there is no mechanism defined to disseminate the policy table information to a host. A proposal is in [I-D.fujisaki-dhc-addr-select-opt] to provide a DHCPv6 mechanism for host policy table management.

Again, by employing DHCPv6, the server could restrict address assignment (of additional prefixes) only to hosts that support policy table management.

Scenario 1: "Host" needs to support the solution for this problem

Scenario 2: "Host" needs to support the solution for this problem

Scenario 3: If "Host" support the next-hop selection solution, there is no need to support the address selection functionality on the host.

4.2. Next-hop selection

A multihomed IPv6 host or gateway may have multiple uplinks to different service providers. Here each router would use Router Advertisements [RFC4861] for distributing default route/next-hop information to the host or gateway router.

In this case, the host or gateway router may select any valid default router from the default routers list, resulting in traffic being sent to the wrong router and discarded by the upstream service provider. Using the above scenarios as an example, whenever the host wishes to reach a destination in network 2 and there is no connectivity between networks 1 and 2 (as is the case for a walled-garden or closed service), the host or gateway router does not know whether to forward traffic to rtr1 or rtr2 to reach a destination in network 2. The host or gateway router may choose rtr1 as the default router, and traffic fails to reach the destination server. The host or gateway router requires route information for each upstream service provider, but the use of a routing protocol between a host and router causes

both configuration and scaling issues. For IPv4 hosts, the gateway router is often pre-configured with static route information or uses of Classless Static Route Options [RFC3442] for DHCPv4. Extensions to Router Advertisements through Default Router Preference and More-Specific Routes [RFC4191] provides for link-specific preferences but does not address per-host configuration in a multi-access topology because of its reliance on Router Advertisements. A DHCPv6 option, such as that in [I-D.dec-dhcpv6-route-option], is preferred for host-specific configuration. By employing a DHCPv6 solution, a DHCPv6 server could restrict address assignment (of additional prefixes) only to hosts that support more advanced next-hop and address selection requirements.

Scenario 1: "Host" needs to support the solution for this problem

Scenario 2: "GW rtr" needs to support the solution for this problem

Scenario 3: "Host" needs to support the solution for this problem

4.3. DNS server selection

A multihomed IPv6 host or gateway router may be provided multiple DNS resolvers through DHCPv6 or the experimental [RFC5006]. When the host or gateway router sends a DNS query, it would normally choose one of the available DNS resolvers for the query.

In the IPv6 gateway router scenario, the Broadband Forum [TR124] required that the query be sent to all DNS resolvers, and the gateway waits for the first reply. In IPv6, given our use of specific destination-based policy for both routing and source address selection, it is desirable to extend a policy-based concept to DNS resolver selection. Doing so can minimize DNS resolver load and avoid issues where DNS resolvers in different networks have connectivity issues, or the DNS resolvers are not publicly accessible. In the worst case, a DNS query may be unanswered if sent towards an incorrect resolver, resulting in a lack of connectivity.

An IPv6 multihomed host or gateway router should have the ability to select appropriate DNS resolvers for each service based on the domain space for the destination, and each service should provide rules specific to that network. [I-D.savolainen-mif-dns-server-selection] proposes a solution for DNS server selection policy enforcement solution with a DHCPv6 option.

Scenario 1: "Host" needs to support the solution for this problem

Scenario 2: "GW rtr" needs to support the solution for this problem

Scenario 3: "Host" needs to support the solution for this problem

5. Requirements

This section describes requirements that any solution multi-address and multi-uplink architectures need to meet.

5.1. End-to-End transparency

End-to-end transparency is a basic concept of the Internet. [RFC4966] states, "One of the major design goals for IPv6 is to restore the end-to-end transparency of the Internet. Therefore, because IPv6 is expected to remove the need for NATs and similar impediments to transparency, developers creating applications to work with IPv6 may be tempted to assume that the complex mechanisms employed by an application to work in a 'NATted' IPv4 environment are not required." The IPv6 multihoming solution SHOULD guarantee end-to-end transparency by avoiding IPv6 NAT.

5.2. Policy enforcement

The solution SHOULD have a function to enforce a policy on sites/ nodes. In particular, in a managed environment such as enterprise networks, an administrator has to control all nodes in his or her network.

The enforcement mechanisms should have:

- a function to distribute policies to nodes dynamically to update their behavior. When the network environment changes and the nodes' behavior has to be changed, a network administrator can modify the behavior of the nodes.

- a function to control every node centrally. A site administrator or a service provider could determine or could have an effect on the behavior at their users' hosts.

- a function to control node-specific behavior. Even when multiple nodes are on the same subnet, the mechanism should be able to provide a method for the network administrator to make nodes behave differently. For example, each node may have a different set of assigned prefixes. In such a case, the appropriate behavior may be different.

5.3. Scalability

The solution will have to be able to manage a large number of sites/ nodes. In services for residential users, provider edge devices have to manage thousands of sites. In such environments, sending packets periodically to each site may affect edge system performance.

6. Implementation approach

As mentioned in Section 4, in the multi-prefix environment, we have three problems in source address selection, next-hop selection, and DNS resolver selection. In this section, possible solution mechanisms for each problem are introduced and evaluated against the requirements in Section 5.

6.1. Source address selection

Possible solutions and their evaluation are summarized in [I-D.ietf-6man-addr-select-sol]. When those solutions are examined against the requirements in Section 5, the proactive approaches, such as the policy table distribution mechanism and the routing system assistance mechanism, are more appropriate in that they can propagate the network administrator’s policy directly. The policy distribution mechanism has an advantage with regard to the host’s protocol stack impact and the staticness of the assumed target network environment.

6.2. Next-hop selection

As for the source address selection problem, both a policy-based approach and a non policy-based approach are possible with regard to the next-hop selection problem. Because of the same requirements, the policy propagation-based solution mechanism, whatever the policy, should be more appropriate.

Routing information is a typical example of policy related to next-hop selection. If we assume source address-based routing at hosts or intermediate routers, the pairs of source prefixes and next-hops can be another example of next-hop selection policy.

The routing information-based approach has a clear advantage in implementation and is already commonly used.

The existing proposed or standardized routing information distribution mechanisms are routing protocols, such as RIPng and OSPFv3, the router advertisement (RA) extension option defined in [RFC4191], the DHCPv6 route information option proposed in [I-D.dec-dhcpv6-route-option], and the [TR069] standardized at BBF.

The RA-based mechanism has difficulty in per-host routing information distribution. The dynamic routing protocols such as RIPng are not usually used between the residential users and ISP networks because of their scalability implications. The DHCPv6 mechanism does not have these difficulties and has the advantages of its relaying functionality. It is commonly used and is thus easy to deploy.

[TR069], mentioned above, is a possible solution mechanism for routing information distribution to customer-premises equipment (CPE). It assumes, however, IP reachability to the Auto Configuration Server (ACS) is established. Therefore, if the CPE requires routing information to reach the ACS, [TR069] cannot be used to distribute this information.

6.3. DNS resolver selection

As in the above two problems, a policy-based approach and non policy-based approach are possible. In a non policy-based approach, a host or a home gateway router is assumed to send DNS queries to several DNS servers at once or to select one of the available servers.

In the non policy-based approach, by making a query to a resolver in a different service provider to that which hosts the service, a user could be directed to unexpected IP address or receive an invalid response, and thus cannot connect to the service provider’s private and legitimate service. For example, some DNS servers reply with different answers depending on the source address of the DNS query, which is sometimes called split-horizon. When the host mistakenly makes a query to a different provider’s DNS to resolve a FQDN of another provider’s private service, and the DNS resolver adopts the split-horizon configuration, the queried server returns an IP address of the non-private side of the service. Another problem with this approach is that it causes unnecessary DNS traffic to the DNS resolvers that are visible to the users.

The alternative of a policy-based approach is documented in [I-D.savolainen-mif-dns-server-selection], where several pairs of DNS resolver addresses and DNS domain suffixes are defined as part of a policy and conveyed to hosts in a new DHCP option. In an environment where there is a home gateway router, that router can act as a DNS proxy, interpret this option and distribute DNS queries to the appropriate DNS servers according to the policy.

7. Considerations for host without multi-prefix support

This section presents an alternative approach to mitigate the problem in a multihomed network. This approach will help IPv6 hosts that are

not capable of the enhancements for the source address selection policy, next-hop selection policy, and DNS selection policy described in Section 6.

7.1. IPv6 NAT

In a typical IPv4 multihomed network deployment, IPv4 NAPT is practically used and it can eventually avoid assigning multiple addresses to the hosts and solve the next-hop selection problem. In a similar fashion, IPv6 NAT can be used as a last resort for IPv6 multihomed network deployments where one needs to assign a single IPv6 address to a host.

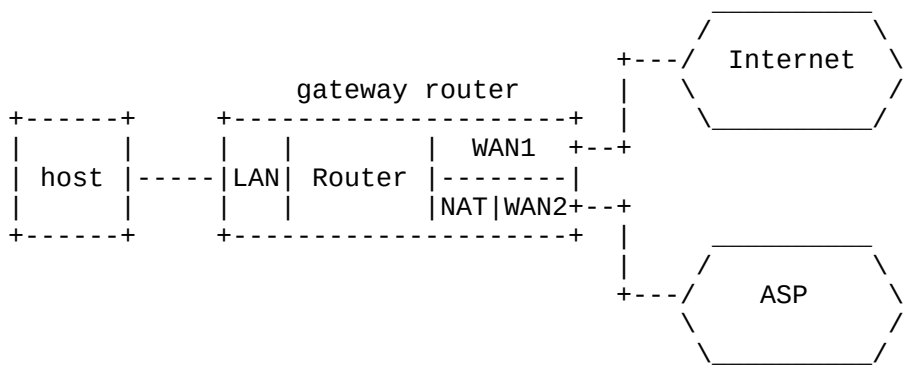


Figure 5: Legacy Host

The gateway router also has to support the two features, next-hop selection and DNS server selection, shown in Section 6.

The implementation and issues of IPv6 NAT are out of the scope of this document. They may be covered by another document under discussion [I-D.mrw-behave-nat66].

7.2. Co-exisitence consideration

The above scenario relies on the assumption that only hosts without multi-prefix support are connected to the GW rtr in scenario 2. To allow the coexistence of non-MHMP hosts and MHMP hosts(i.e. hosts supporting multi-prefix with the enhancements for the source address selection), GW-rtr may need to treat those hosts separately.

An idea to achieve this is that GW-rtr identifies the hosts, and then assigns single prefix to non-MHMP hosts and assigns multiple prefix to MHMP hosts. In this case, GW-rtr can perform IPv6 NAT only for

the traffic from MHMP hosts if its source address is not appropriate.

Another idea is that GW-rtr assigns multiple prefix to the both hosts, and it performs IPv6 NAT for the traffic from non-MHMP hosts if its source address is not appropriate.

In scenario 1 and 3, the non-MHMP hosts can be placed behind the NAT box. In this case, non-MHMP host can access the service through the NAT box.

The implementation of identifying non-MHMP hosts and NAT policy is outside the scope of this document.

8. Security Considerations

This document does not define any new mechanisms. Each solution mechanisms should consider security risks independently. Security risks that occur as a result of combining solution mechanisms should be considered in another document.

9. IANA Considerations

This document has no IANA actions.

10. Contributors

The following people contributed to this document: Akiko Hattori, Arifumi Matsumoto, Frank Brockners, Fred Baker, Tomohiro Fujisaki, Jun-ya Kato, Shigeru Akiyama, Seiichi Morikawa, Mark Townsley, Wojciech Dec, Yasuo Kashimura, Yuji Yamazaki

11. References

11.1. Normative References

[I-D.dec-dhcpv6-route-option]
Dec, W. and R. Johnson, "DHCPv6 Route Option", draft-dec-dhcpv6-route-option-03 (work in progress), March 2010.

[I-D.fujisaki-dhc-addr-select-opt]
Fujisaki, T., Matsumoto, A., and R. Hiromi, "Distributing Address Selection Policy using DHCPv6", draft-fujisaki-dhc-addr-select-opt-09 (work in progress),

March 2010.

[I-D.ietf-6man-addr-select-sol]
Matsumoto, A., Fujisaki, T., and R. Hiromi, "Solution approaches for address-selection problems", draft-ietf-6man-addr-select-sol-03 (work in progress), March 2010.

[I-D.mrw-behave-nat66]
Wasserman, M. and F. Baker, "IPv6-to-IPv6 Network Address Translation (NAT66)", draft-mrw-behave-nat66-02 (work in progress), March 2009.

[I-D.savolainen-mif-dns-server-selection]
Savolainen, T., "DNS Server Selection on Multi-Homed Hosts", draft-savolainen-mif-dns-server-selection-02 (work in progress), February 2010.

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

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

[RFC4191] Draves, R. and D. Thaler, "Default Router Preferences and More-Specific Routes", RFC 4191, November 2005.

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

11.2. Informative References

[RFC3022] Srisuresh, P. and K. Egevang, "Traditional IP Network Address Translator (Traditional NAT)", RFC 3022, January 2001.

[RFC3442] Lemon, T., Cheshire, S., and B. Volz, "The Classless Static Route Option for Dynamic Host Configuration Protocol (DHCP) version 4", RFC 3442, December 2002.

[RFC4966] Aoun, C. and E. Davies, "Reasons to Move the Network Address Translator - Protocol Translator (NAT-PT) to Historic Status", RFC 4966, July 2007.

[RFC5006] Jeong, J., Park, S., Beloeil, L., and S. Madanapalli, "IPv6 Router Advertisement Option for DNS Configuration", RFC 5006, September 2007.

[TR069] The BroadBand Forum, "TR-069, CPE WAN Management Protocol v1.1, Version: Issue 1 Amendment 2", December 2007.

[TR124] The BroadBand Forum, "TR-124i2, Functional Requirements for Broadband Residential Gateway Devices (work in progress)", May 2010.

Authors' Addresses

Ole Troan (editor)
Cisco
Bergen
Norway

Email: ot@cisco.com

David Miles
Alcatel-Lucent
Melbourne
Australia

Email: david.miles@alcatel-lucent.com

Satoru Matsushima
SOFTBANK TELECOM Corp.
Tokyo
Japan

Email: satoru.matsushima@tm.softbank.co.jp

Tadahisa Okimoto
NTT
Tokyo
Japan

Email: t.okimoto@hco.ntt.co.jp

Dan Wing
Cisco
170 West Tasman Drive
San Jose
USA

Email: dwing@cisco.com

Internet Engineering Task Force T. Tsou, Ed.
Internet-Draft T. Taylor
Intended status: Informational Huawei Technologies
Expires: April 22, 2011 October 19, 2010

IPv6 Transition Guide For A Large Mobile Operator
draft-tsou-v6ops-mobile-transition-guide-00

Abstract

This document provides a transition guide for a large-scale mobile network operator migrating its network from IPv4 to 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 22, 2011.

Table of Contents

1. Introduction

The use case for migration of a large mobile network to IPv6 is described in [ID_mobile_use_case]. That document provides an introduction to the network architecture and looks at possible strategies and tools for migration. 3GPP has decided to use Gateway-Initiated DS-lite [ID_GI_DS_lite] as the primary tool for subscriber migration. Details as far as they have been thought out are provided in [3GPP_TR_23_975]. However, they cover only a small part of the total problem.

1.1. Requirements Language

This document contains no requirements language.

2. Steps In the Transition Strategy

The transition to IPv6 must be carried out over a number of different segments within the total network:

- the operator’s operational support systems, including AAA;
- the operator’s private IP network;
- the portal to the operator’s proprietary content and applications;
- the user devices;
- portals to the partner content and application providers.

These are listed in what is probably the preferred order for making the transition, but in fact the operator will want to carry out a number of activities in parallel.

2.1. Migration Of Operational Support Systems and AAA

The transition to IPv6 may have a number of consequences for AAA and other support systems. First and most obvious, systems must be set up for the provisioning of IPv6 addresses and associated configuration data. AAA may be affected if the subscriber’s IPv4 address has been used as a correlator for accounting records. Device provisioning and configuration within the network to handle IPv6 has to be tracked so that the engineering department can monitor the progress of the necessary network upgrades and maintenance has the information it needs to carry out routine testing and restoration procedures. New maintenance procedures have to be developed.

Given the time it takes to develop new support systems or modify old ones, it is to be hoped that the most critical areas of the effort have already been identified and are well on the way to being implemented before any of the other activities begin.

2.2. Transition of the Private Internal Network

The private internal network includes signalling links between network devices, but also the IP layer over which tunnels carry user packets.

Conversion of the private internal network to pure IPv6 operation should be an early objective, for at least two reasons. In the first place, it provides the operator with experience that will be helpful when making the transition in other segments of the network. In addition, it relieves the operator of one source of demand for IPv4 addresses, at least some of which must be public to allow communication with other operators.

Unfortunately, the need for IPv4 addresses will not go away immediately. While the transition is in progress, until upgrades are completed, a transition mechanism is needed to allow the upgraded equipment to interoperate with the equipment that has not yet been upgraded. The most obvious mechanism is to use dual-stack operation with the devices being configured to use IPv6 whenever possible. It may be possible to do the upgrades in blocks of devices, where relatively few of the devices in a block need to communicate with devices outside the block. These boundary devices will continue to need IPv4 addresses until the other blocks with which they communicate have been upgraded, but communications in the interior of the block can use IPv6 and so interior devices need no IPv4 address.

Because IPv6 usage in the private network will build up as quickly as the operator can upgrade the network equipment, an IPv6 version of the internal DNS system will be needed early on. It seems likely, in fact, that the most efficient mode of operation will be a dual-stack DNS containing both A and AAAA records.

> A key use of this DNS system is to allow the Serving Gateway to locate a PDN Gateway providing access to the core network that the subscriber will use. This may be operator’s own core network or the subscriber’s home network.

2.3. Migrating the Portal to the Operator’s Proprietary Content and Applications

The operator needs to build up the availabilty of IPv6-accessible applications and content as quickly as possible, to reduce the IPv4

traffic on the network and thereby reduce the demand for public IPv4 addresses. One way to do this is to make the operator's own content and applications IPv6-accessible early on in the transition period.

As usual, because of the time it will take to transition all users, IPv4 access to the content and applications must continue to be provided, until the last Windows XP computer ceases to be tethered to a mobile device for Internet access. In the early days, until the content can be fully converted, protocol translation may be used to allow access to IPv6 users. Once the content and applications have been converted to IPv6, dual stack operation will be possible and the protocol converter (NAT64) can be removed.

2.4. User Devices

3GPP specifications for mobile devices have required dual stack support for at least a year (i.e., as of Release 8.) The operator can help things along by requiring that new devices connecting to the operator's network conform to this requirement. It will still take two or three years until the large majority of devices are capable of dual stack operation.

The transition strategies considered in [3GPP_TR_23_975] relate specifically to how user traffic is carried. That document offers four different scenarios, or strategies, for achieving transition. In the early stages, before a large portion of the content and applications accessed by users can be reached by IPv6, the most likely strategy will be to interpose Gateway-Initiated DS-lite [ID_GI_DS_lite] using a minimal set of private IPv4 addresses at the user devices and sharing public IPv4 addresses between multiple users using some system of block port allocation as proposed in [ID_natx4-log-reduction].

When the operator converts a given area to Gateway-Initiated DS-lite access, a number of public IPv4 addresses are freed because of the introduction of address sharing. This suggests that one strategy may be to introduce Gateway-Initiated DS-lite initially in high-growth areas, using the addresses thus freed to handle demand in lower-growth-rate areas until they can be converted.

DNS access is an issue with Gateway-Initiated DS-lite. The original DS-lite proposal had a point (the B4) at which all DNS queries could be intercepted and sent to an IPv6 DNS. From here IPv4 queries could be forwarded to an IPv4 DNS, or the IPv6 DNS could maintain both AAAA and A records. It is not so obvious that such interception can be carried out at the Gateway in Gateway-Initiated DS-lite, since the Gateway is essentially performing a layer 2 operation. [ID_GI_DS_lite] does not mention the issue.

2.5. Portals to the Partner Content and Application Providers

As mentioned above, content conversion is the key to building up IPv6 traffic and thereby relieving pressure on the supply of public IPv4 addresses. To some extent the operator may be able to solve this at a business level, through negotiations to encourage the content provider to convert. However, technical solutions will also be necessary.

One possible solution is to provide IPv6 access to the content provider's site by installing protocol translation for traffic between that site and IPv6 users. This would have to be accompanied by the installation of AAAA records in DNS giving the address IPv6 address via which the site is reachable.

3. Conclusions

To be completed after review.

4. Acknowledgements

5. IANA Considerations

This memo includes no request to IANA.

6. Security Considerations

To be completed.

7. Informative References

[3GPP_TR_23_975]
3rd Generation Partnership Project, "Technical Specification Group Services and System Aspects; IPv6 Migration Guidelines (Release 10)", TR 23.975, May 2010.

[ID_GI_DS_lite]
Brockners, F., Gundavelli, S., Speicher, S., and D. Ward, "Gateway Initiated Dual-Stack lite Deployment (Work in progress)", May 2010.

[ID_mobile_use_case]
Zhou, C. and T. Taylor, "IPv6 Transition Use Case For a

Large Mobile Network (Work in progress)", September 2010.

[ID_natx4-log-reduction]
Huang, J., "Port Management To Reduce Logging In Large-Scale NATs (Work in progress)", August 2010.

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

Authors' Addresses

Tina Tsou (editor)
Huawei Technologies
Bantian, Longgang District
Shenzhen 518129
P.R. China

Phone:
Email: tena@huawei.com

Tom Taylor
Huawei Technologies
1852 Lorraine Ave.
Ottawa K1H 6Z8
Canada

Phone:
Email: tom111.taylor@bell.net

v6ops Working Group
Internet-Draft
Intended status: Informational
Expires: March 4, 2011

G. Van de Velde
O. Troan
Cisco Systems
T. Chown
University of Southampton
August 31, 2010

Non-Managed IPv6 Tunnels considered Harmful
<draft-vandevelde-v6ops-harmful-tunnels-01.txt>

Abstract

IPv6 is ongoing and natively being deployed by a growing community and it is important that the quality perception and traffic flows are as optimal as possible. Ideally it would be as good as the IPv4 perceptive experience.

This paper looks into a set of transitional technologies where the actual user has IPv6 connectivity through the means of IPv6-in-IPv4 tunnels. A subset of the available tunnels has the property of being non-managed (i.e. 6to4 [RFC3056] and Teredo [RFC4380]).

While native IPv6 deployments will keep growing it is uncertain or even expected that non-managed IPv6 tunnels will be providing the same user experience and operational quality as managed tunnels or native IPv6 connectivity.

This paper will detail some considerations around non-managed tunnels and will document the harmful element of these for the future growth of networks and 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 March 4, 2011.

Table of Contents

1. Introduction

While the Internet and networks continue to grow, it is found that the deployment of IPv6 within these networks is an ongoing activity due to global IPv4 address pool depletion. An important aspect is that the quality, availability and security of the IPv6 connectivity is as good as possible, and when possible even more advanced as the IPv4 connectivity.

Historically IETF has been facilitating a variety of technologies and procedures to deploy IPv6 successfully in addition to existing IPv4 connectivity. In general and for the sake of this draft these procedures and technologies can be divided into three major groups: (1) native (dual-stack) IPv6, (2) Tunnelled IPv6 and (3) Translation. While native IPv6 deployments has been steadily growing, the value and the drawbacks of some tunnelling mechanisms can be investigated. Translational techniques provide a total different aspect of considerations and applicability and is beyond the scope of this paper. Transition techniques have been and still are in many cases important for the bootstrapping of IPv6, this paper will look into a range of property aspects of non-managed IPv6 tunnelling techniques. Areas of perverse traffic paths, security considerations, lack of business incentives to run tunnel relays/gateways, black holing and ownership of supportability will be analysed. Finally the paper will conclude that for the growth of IP connectivity, non-managed tunnelling techniques are considered harmful especially for those that want to access applications over the network through pervasive IPv6 connectivityand have no particular interrest on how connectivity to the applications is established (IPv4, translation, IPv6, etc...)

2. Managed Tunnelling Properties

A managed tunnel is a tunnel has a few properties supporting the ownership and quality of the tunnel.

When using a managed service, there tends to be an administrative entity which provides quality assurance and can take action if users of the service are experiencing a degraded service. An example would be 6rd tunnels [RFC5969]

In addition there is a general trust awareness and agreement between the user of the managed tunnel service and the provider of the managed tunnel service.

3. Tunnel User Experience Views

The tunnel experience can be divided into three distinct segments: (1) the End-user view, (2) the Enterprise View and (3) the Service Provider View.

The End-user view exists mainly out of two different user profiles. The technical power user and the general user mainly trying to reach their favourite application on the network. The technical power user may have a particular interrest to run IPv6 as a transport mechanism, and if his upstream service provider has no native IPv6 connectivity available, then non-managed tunneling mechanisms may provide a solution satisfying to the immediate needs of the technical power user. Alternatively, the general user trying to reach his favourite network application, may have no interest or awareness of his IPv6 usage, particulary when non-managed tunnels are utilized.

The Enterprise View is a more traffic flows and network oriented possitioning. When the upstream service provider does not have an IPv6 offer, then the enterprise may start to rely upon a technology as 6to4 [RFC3056]. However this technology has the potential of creating quite perverse traffic paths when user want to reach applications on the Internet. When user would like to reach other 6to4 [RFC3056] users, then more optimized traffic paths, generally following the IPv4 traffic paths are realized

The final view is how a Internet service provider looks into non-managed tunnel usage. A service provider may decide to deploy a 6to4 relay to increase the IPv6 quality of their customers. This a service which require resources (money, maintenance, etc...). Often the 6to4 relay service is not just (always) restricted to only the service providers customers, which as result provides often results in a demotivation to provide quality tunnel relay devices. From a content service provider perspective the usage of non-managed tunnel often results in measurable differences in RTT and reliability in some cases, and hence are reluctant to bring all services to mainstream IPv6 for all users ’just yet’.

4. Why do non-managed tunnels exist?

Non-managed tunnels exist due to a variety of reasons.

Early adopters: people and organisations with a desire to use new and potentially market disrupting technologies and applications may have a desire to use the latest IP even when the upstream provider doesn’t have an available service offering.

Lock-step process to implement IPv6: It is not trivial to move a system or an organisation in lock-step towards IPv6 and the aid of tunnels help in this process.

The utilisation of tunnels aid in providing a de-coupling between infrastructure readiness and application readiness, and hence contribute to the development of both elements.

5. Non-Managed Tunnelling Properties

The properties of Non-managed tunnels span many different areas. In this section the properties are analysed and segmented within different areas of impact. In each case the comparison is made between native IPv6 connectivity and connectivity through a non-managed tunnel. A common property of non-managed tunnels is that they often use well-known anycast addresses or other well known addresses and anticipate upon the goodwill of middleware (typically a relay or gateway) device to serve as a tunnel termination point. In some cases, for example a 6to4 relay can be provided by a connected responsible service provider, and hence good quality operation can be guaranteed.

Non-managed tunnels often have asymmetric behaviour. There is an outbound and an inbound connectivity behaviour from the tunnel initiator. It is possible to influence the good quality tunnel behaviour of the outbound connectivity (e.g. by explicit setting of the 6to4 relay), however, influencing good inbound connectivity is often an issue.

5.1. Performance

Deploying a tunnelling mechanism mostly results in encapsulation and de-capsulation efforts. Often this activity has a performance impact on the device, especially when the device does not use hardware acceleration for this functionality. If the performance impact is scoped into the device its lifetime through performance capacity management then the actual impact is predictive. Non-deterministic tunnels tend to have a non-predictive behaviour for capacity, and hence application and network performance is non-predictive. The key reason for this is the decoupling of the capacity management of the tunnel aggregation devices from the capacity desired by users of the aggregation devices.

During initial IPv6 deployment there have been mainly technical savvy people that have been using non-managed tunnel technologies and it has for many been working well. However, if non-managed tunnelling would be deployed in mass and especially when enabled by default by

CPE vendors or host vendors then those aggregation points could become overloaded and result in bad performance. There are a few measures that can be taken, i.e. upgrade the CPU power of the aggregation device or its bandwidth available, however this may not happen without the right motivation for the operator of the aggregation device (i.e. cash flows, reputation, competitive reasons, etc...).

5.2. Topological Considerations

Due to non-managed IPv6 tunnels the traffic flows may result in sub-optimal flows through the network topology between two communicating devices. The impact for example can cause increase of the RTT and packet loss, especially considering the availability (or better non-availability) of tunnel aggregation/de-aggregation points of certain topological areas or realms. The increase of non-managed tunnel usage would amplify the negative impact on good quality connectivity. For many operators of tunnel aggregation/de-aggregation devices there is little motivation to increase the quality and number of available devices within a topological area or logistical realm.

5.3. Operational Provisioning

Some elements regarding provisioning of both managed and non-managed tunnels can be controlled, while others are beyond control or influence of people and applications using tunnels. To make applications highly reliable and performing, all elements within the traffic path must provide an expected quality service and performance. For managed tunnels, the user or provider of the tunnel can exercise a degree of operational management and hence influence good quality behaviour upon the tunnel especially upon the aggregation and de-aggregation devices. In some cases even the traffic path between both aggregation and de-aggregation can be controlled. Non-managed tunnels however have less good quality behaviour of both tunnel aggregation and de-aggregation devices because often good quality behaviour is beyond the control or influence of the tunnel user. For non-managed tunnels the tunnel aggregator and/or tunnel de-aggregator are operated by a 3rd party which may have a conflicting interest with the user of the non-managed tunnel. An exception is where the use of the tunnel mechanism is all within one ISP, or ISPs who are ’well coupled’, e.g. as happens between many NRENs.

5.4. Operational Troubleshooting

When one is using non-managed tunnels, then these tunnels may get aggregated or de-aggregated by a 3rd party or a device outside the control of a contracted service provider. Troubleshooting these

devices these devices will be pretty hard for the tunnel user or to work around the issue.

Also some tools like traceroute don’t work too well on asymmetric paths. Another aspect is that tunnels show as one hop in a traceroute, not indicating where problems may be.

5.5. Security

For an aggregating or de-aggregating tunnel device it is a non-trivial issue to separate the valid traffic from non-valid traffic because it is from the aggregation device perspective almost impossible to know -from- and -towards- about the tunnel traffic. This imposes potential attacks on the available resources of the aggregating/de-aggregating router. A detailed security analysis for 6to4 tunnels can be found in [RFC3964].

For the user of the non-managed IPv6 tunnel there is an underlying trust that the aggregating/de-aggregating device is a trustworthy device. However, some of the devices used are run by anonymous 3rd parties outside the trusted infrastructure from the user perspective, which is not an ideal situation. The usage of non-managed tunnels increases the risk of rogue aggregation/de-aggregation devices and may be open to malicious packet analyses or manipulation.

From the operator perspective, managing the aggregating/ de-aggregating tunnel device, there is a trust assumption that no-one abuses the service. Abuse may impact preset or assumed service quality levels, and hence the quality provided can be impacted

There is also an impact caused by ipv4 firewalling upon non-managed tunnels. Common firewall policies recommend to block tunnels, especially non-managed tunnels, because there is no trust that the traffic within the tunnel is not of mallicious intend. This restricts the applicability of some non-managed tunnel mechanisms (e.g. 6to4). Other tunnel mechanisms have found manners to avoid traditional firewall filtering (e.g. Teredo) and open the local network infrastructure for mallicious influence (e.g. virus, worms, infrastructure attacs, etc..).

5.6. Content Services

When providing content services a very important related aspect is that these services are accessible with high reliability, are trustworthy and have a high performance. Using non-managed tunnels makes this a much harder equation and can result in all three elements to suffer negatively, without the ability to uniquely identify and resolve the root cause. The statistical impact of non-

mnaged tunnels has been measured by some Internet Content providers and is often an additional delay of O(100msec) (need to add reference here)

This reduces the interest of content providers to provide content services over IPv6 when non-managed tunnels are used.

6. Conclusion

Non-managed tunnels have properties impacting the growth of networks and the Internet in a negative way. Consequences regarding black-holing, perverse traffic paths, lack of business incentive and operational management influence and security issues are a real pragmatic concern, while universal supportability for the tunnel relay services appear to be non-trivial. Due to these elements the usage of non-managed tunnelling can be considered harmful for the growth of networks and the Internet.

7. IANA Considerations

There are no extra IANA consideration for this document.

8. Security Considerations

There are no extra Security consideration for this document.

9. Acknowledgements

10. References

10.1. Normative References

10.2. Informative References

[RFC3056] Carpenter, B. and K. Moore, "Connection of IPv6 Domains via IPv4 Clouds", RFC 3056, February 2001.

[RFC3964] Savola, P. and C. Patel, "Security Considerations for 6to4", RFC 3964, December 2004.

[RFC4380] Huitema, C., "Teredo: Tunneling IPv6 over UDP through Network Address Translations (NATs)", RFC 4380, February 2006.

[RFC4798] De Clercq, J., Ooms, D., Prevost, S., and F. Le Faucheur, "Connecting IPv6 Islands over IPv4 MPLS Using IPv6 Provider Edge Routers (6PE)", RFC 4798, February 2007.

[RFC5214] Templin, F., Gleeson, T., and D. Thaler, "Intra-Site Automatic Tunnel Addressing Protocol (ISATAP)", RFC 5214, March 2008.

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

Authors' Addresses

Gunter Van de Velde
Cisco Systems
De Kleetlaan 6a
Diegem 1831
Belgium

Phone: +32 2704 5473
Email: gvandeve@cisco.com

Ole Troan
Cisco Systems
Folldalslia 17B
Bergen N-5239
Norway

Phone: +47 917 38519
Email: ot@cisco.com

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

Phone: +44 23 8059 3257
Email: tjc@ecs.soton.ac.uk

v6ops Working Group
Internet-Draft
Intended status: Informational
Expires: February 17, 2011

G. Van de Velde
C. Popoviciu
T. Hain
S. Venaas
Cisco Systems
k. Chittimaneni
Google Inc

August 16, 2010

Network signaling for IPv4/IPv6 protocol selection for end-systems <draft-vandevelde-v6ops-pref-ps-00.txt>

Abstract

Within an administrative realm, especially during an IPv6 implementation period, the network operator has interest to have control on the IP protocol version (IPv4 or IPv6) used by the end systems and network devices. This document provides a problem statement about both protocol preference and protocol selection many network operators face when implementing IPv6 in a controlled process.

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 17, 2011.

Table of Contents

1. Introduction

With IPv6 TCP/IP stacks preferring IPv6 over IPv4 for forwarding, network administrators don't have control over which protocol end hosts use. Such lack of control has the potential to negatively impact the end host, especially in cases where the network is dual stacked well before the backend systems and/or applications are. It is thus preferable to have control over the device preference or selection of IP4 or IPv6. This control will allow network administrators to seamlessly implement IPv6 on the network with the ability to carefully integrate IPv6 into production as and when all other critical non-network components are found to be working as expected.

This document describes a problem statement to control and potentially communicate the IPv4/IPv6 protocol preference for devices. The document outlays the various considerations for protocol preference selection. This capability improves the ability of hosts to pick an appropriate protocol (IPv4 or IPv6) for off-link and on-link destinations.

Note that this procedure is applicable to end-systems and their applications only; the forwarding algorithm used by routers is not affected.

2. Components

2.1. IPv6 deployment model

When a network operator is in process of deploying a new technology on the network, the network operator will likely include a set of fallback mechanisms and will try to place as much control as possible in each of the deployment steps. An element of control during the integration of IPv6 is the management of IPv6 use by the end-systems and the applications running on those systems. The protocol preference management is done thorough a signaling mechanism. This signaling allows the network operator to introduce IPv6 on the routers and other network infrastructure elements without impacting existing IPv4 behavior of the end-systems. Once the network operator decides to activate IPv6 for end-systems, in order to allow each end-system to include IPv6 as valid communication protocol following RFC3484 address selection.

This operational sequence will help the enablement of IPv6 in a controlled manner once the network infrastructure is found correctly working according the expectations of the network operator.

2.2. Policy table

The policy table references to an information database defining the expected behavior regarding IPv4/IPv6 protocol preference and selection behavior for various parts of the administrative domain.

3. Considerations for IPv4 and IPv6 selection on end-systems

This section will detail considerations for an end-system with respect to IPv4/IPv6 protocol preference.

3.1. Dynamics of end-system configuration

An end-system is usually configured in three possible ways:

(a) Preset configuration: these end-systems have a configuration which has been defined during the manufacturing of the device; (b) Manual configuration: In this case the end-systems are configured by a set of parameters and settings which are individually configured on the device through human interaction; (c) Dynamic configuration: Some end-systems download through the network infrastructure a set of parameters i.e. IP and DNS addresses through DHCP

3.2. Hosts with multiple interfaces

Lots of end-systems are connected to the network infrastructure with only a single interface. For these systems, the IPv4 and IPv6 preference can be quite simply defined, either using or not-using IPv6. However, there are many end-systems connected through two or more interfaces to the network infrastructure. These systems require a protocol preference to be defined for each interface independently. These additional considerations also include aspects of conflicting information received through the different interfaces regarding the IPv6 protocol preference.

3.3. Backward compatibility

A solution for the IPv4/IPv6 preference shouldn’t have an impact on end-systems not capable to understand this functionality.

3.4. Network based learning

The IPv4/IPv6 protocol preference on an end-system should be signaled by the ’network’ (network devices or other infrastructure components such as DHCP) to automate the end-systems behavior, however the IP4/ IPv6 protocol preference solution should not exclude manual configuration on end-systems.

3.5. Impact on RFC3484

A solution for IPv4/IPv6 protocol preference may influence the availability or better the non-availability of IPv6 parameters within the RFC3484 end-system address selection algorithm. This influence must be understood very clearly for end-systems with single and with multiple interfaces attached to the network infrastructure.

3.6. Influence of IPv4/IPv6 protocol preference on applications

The IPv4/IPv6 protocol preference should be propagated by the end-system towards the applications running on the end-system. It should not be excluded that a protocol preference solution may have more specific information per application of importance to the end-systems. As consequence the end system could use this information for IPv4/IPv6 protocol preference per application or session for example.

3.7. Dynamic tunneling

Some end systems make usage of dynamic tunnels for IPv6 even when the network infrastructure does not support IPv6 as a native protocol. The IPv4/IPv6 preference signal could influence the creation of these tunnels based upon the signaled IPv4/IPv6 protocol preference.

4. Considerations for IPv4 and IPv6 selection on network infrastructure elements

This section will detail considerations for network infrastructure devices with respect to IPv4/IPv6 protocol preference.

4.1. Signaling IPv4/IPv6 preference

The IPv4/IPv6 preference signal can be sent by four methods.

(a) The end-system is configured during manufacturing of the system; (b) A network operator configures the end-system by the console of the end-system; (c) The network infrastructure could signal the end-systems the IPv4/IPv6 preference through existing or new link-local packets; (d) The network infrastructure signals the end-system that there are elements that influence protocol selection, and that the end-system may want to request the network infrastructure what these elements exactly are.

4.2. Influence of IPv4/IPv6 preference on network infrastructure elements

The IPv4/IPv6 preference selection should only have impact on the end-systems and the network infrastructure devices should be ignoring the preference signal.

4.3. IPv4/IPv6 policy table location

The IPv4/IPv6 protocol preference needs to be stored somewhere within the network. This could be done either centrally or distributed. Crucial is that the network infrastructure device is directly connected to the end-system it wants to signal IPv4/IPv6 preference, so that link-local communication between the end-system and the network infrastructure device can be used. Between the network infrastructure device and the policy table location non-link local addresses may be utilized.

4.4. Backward compatibility

There should be no impact on either the network infrastructure when end-systems do not understand the IPv4/IPv6 protocol preference solution.

5. IANA Considerations

There are no extra IANA consideration for this document.

6. Security Considerations

7. Acknowledgements

8. References

8.1. Normative References

8.2. Informative References

Authors' Addresses

Gunter Van de Velde
Cisco Systems
De Kleetlaan 6a
Diegem 1831
Belgium

Phone: +32 476 476 022
Email: gvandeve@cisco.com

Ciprian Popoviciu
Cisco Systems
7025-6 Kit Creek Road
Research Triangle Park, North Carolina NC 27709-4987
United States

Phone: +1 919 392-3723
Email: cpopovic@cisco.com

Tony Hain
Cisco Systems
500 108th Ave. NE
Bellevue, Wa.
USA

Email: alh-ietf@tndh.net

Stig Venaas
Cisco Systems
Tasman Drive
San Jose, CA 95134
USA

Phone:
Email: stig@cisco.com

Kiran Kumar Chittimaneni
Google Inc
1600 Amphitheatre Parkway
Mountain View, CA 94043
USA

Phone: +1 650 253 6185
Email: kk@google.com

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: April 28, 2011

H. Singh
W. Beebee
Cisco Systems, Inc.
C. Donley
CableLabs
B. Stark
AT&T
O. Troan, Ed.
Cisco Systems, Inc.
October 25, 2010

Advanced Requirements for IPv6 Customer Edge Routers
draft-wbeebee-v6ops-ipv6-cpe-router-bis-04

Abstract

This document continues the work undertaken by the IPv6 CE Router Phase I work in the IETF v6ops Working Group. Advanced requirements or Phase II work is covered 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 April 28, 2011.

Copyright Notice

Table of Contents

1. Introduction

This document defines Advanced IPv6 features for a residential or small office router referred to as an IPv6 CE router. Typically these routers also support IPv4. The IPv6 End-user Network Architecture for such a router is described in [I-D.ietf-v6ops-ipv6-cpe-router]. This version of the document includes the requirements for Advanced features.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

2. Terminology

End-user Network	one or more links attached to the IPv6 CE router that connect IPv6 hosts.
IPv6 Customer Edge router	a node intended for home or small office use which forwards IPv6 packets not explicitly addressed to itself. The IPv6 CE router connects the end-user network to a service provider network.
IPv6 host	any device implementing an IPv6 stack receiving IPv6 connectivity through the IPv6 CE router
LAN interface	an IPv6 CE router’s attachment to a link in the end-user network. Examples are Ethernets (simple or bridged), 802.11 wireless or other LAN technologies. An IPv6 CE router may have one or more network layer LAN Interfaces.
Service Provider	an entity that provides access to the Internet. In this document, a Service Provider specifically offers Internet access using IPv6, and may also offer IPv4 Internet access. The Service Provider can provide such access over a variety of different transport methods such as DSL, cable, wireless, and others.

WAN interface	an IPv6 CE router's attachment to a link used to provide connectivity to the Service Provider network; example link technologies include Ethernets (simple or bridged), PPP links, Frame Relay, or ATM networks as well as Internet-layer (or higher-layer) "tunnels", such as tunnels over IPv4 or IPv6 itself.

3. Conceptual Configuration Variables

The CE Router maintains such a list of conceptual optional configuration variables.

1. Enable an IGP on the LAN.

4. Architecture

This document extends the architecture described in [I-D.ietf-v6ops-ipv6-cpe-router] to cover a strictly larger set of operational scenarios. In particular, QoS, multicast, DNS, routed network in the home, transition technologies, and conceptual configuration variables. This document also extends the model described in [I-D.ietf-v6ops-ipv6-cpe-router] to a two router topology where the two routers are connected back-to-back (the LAN of one router is connected to the WAN of the other router). This topology is depicted below:

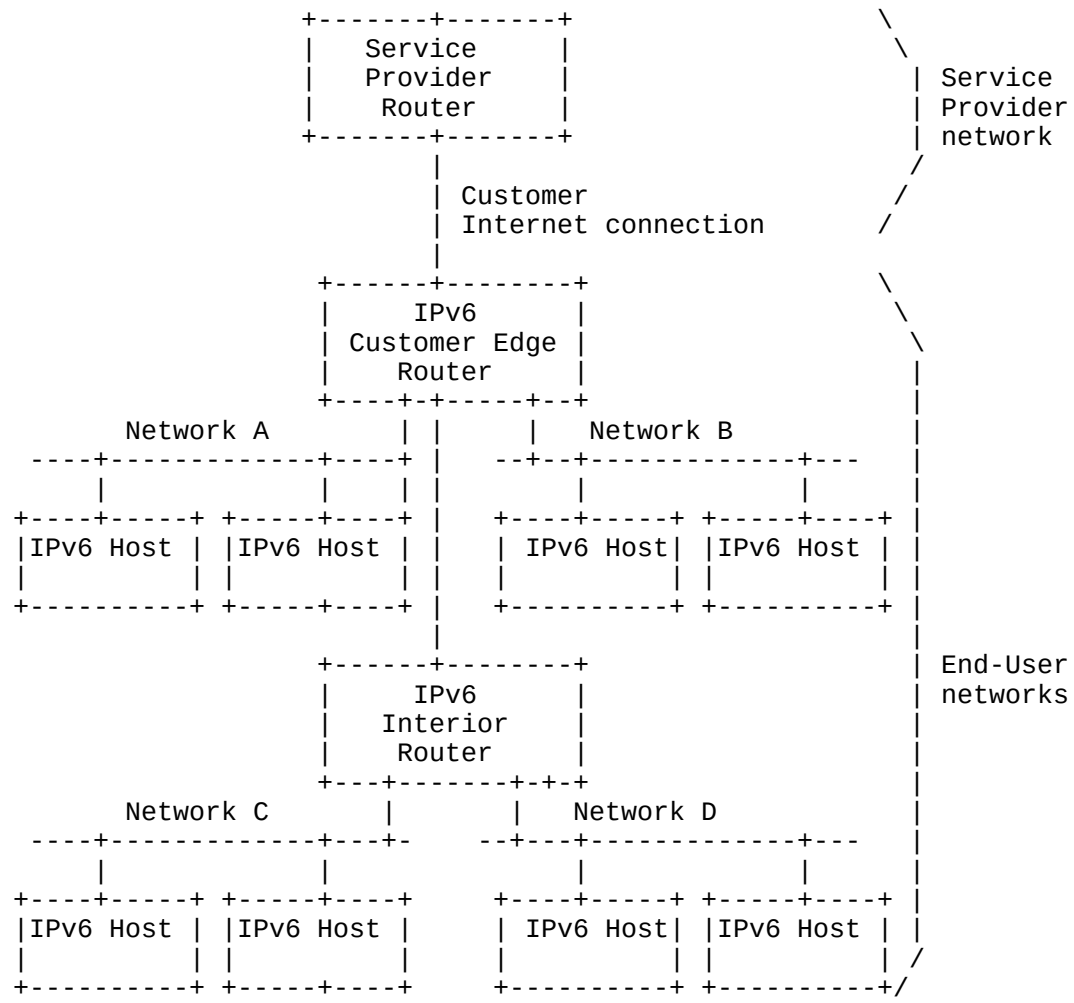


Figure 1.

For DNS, the operational expectation is that the end-user would be able to access home hosts from the home using DNS names instead of more cumbersome IPv6 addresses. Note that this is distinct from the requirement to access home hosts from outside the home.

End-users are expected to be able to receive multicast video in the home without requiring the CE router to include the cost of supporting full multicast routing protocols.

5. Advanced Features and Feature Requirements

The IPv6 CE router will need to support connectivity to one or more access network architectures. This document describes an IPv6 CE router that is not specific to any particular architecture or Service Provider, and supports all commonly used architectures.

5.1. DNS

D-1: For local DNS queries for configuration, the CE Router may include a DNS server to handle local queries. Non-local queries can be forwarded unchanged to a DNS server specified in the DNS server DHCPv6 option. The CE Router may also include DNS64 functionality which is specified in [I-D.bagnulo-behave-dns64].

D-2: The local DNS server MAY also handle renumbering from the Service Provider provided prefix for local names used exclusively inside the home (the local AAAA and PTR records are updated). This capability provides connectivity using local DNS names in the home after a Service Provider renumbering. A CE Router MAY add local DNS entries based on dynamic requests from the LAN segment(s). The protocol to carry such requests from hosts to the CE Router is yet to be described.

5.2. Multicast Behavior

This section is only applicable to a CE Router with at least one LAN interface. A host in the home is expected to receive multicast video. Note the CE Router resides at edge of the home and the Service Provider, and the CE Router has at least one WAN connection for multiple LAN connections. In such a multiple LAN to a WAN toplogy at the CE Router edge, it is not necessary to run a multicast routing protocol and thus MLD Proxy as specified in [RFC4605] can be used. The CE Router discovers the hosts via a MLDv2 Router implementation on a LAN interface. A WAN interface of the CE Router interacts with the Service Provider router by sending MLD Reports and replying to MLD queries for multicast Group memberships for hosts in the home.

The CE router SHOULD implement MLD Proxy as specified in [RFC4605]. For the routed topology shown in Figure 1, each router implements a MLD Proxy. If the CE router implements MLD Proxy, the requirements on the CE Router for MLD Proxy are listed below.

WAN requirements, MLD Proxy:

WMLD-1: Consistent with [RFC4605], the CE router MUST NOT implement the router portion of MLDv2 for the WAN interface.

LAN requirements, MLD Proxy:

LMMLD-1: The CPE Router MUST follow the model described for MLD Proxy in [RFC4605] to implement multicast.

LMMLD-2: Consistent with [RFC4605], the LAN interfaces on the CPE router MUST NOT implement an MLDv2 Multicast Listener.

LAN requirements:

LM-1: If the CE Router has bridging configured between the LAN interfaces, then the LAN interfaces MUST support snooping of MLD [RFC3810] messages.

5.3. ND Proxy

LAN requirements:

LNDP-1: If the CE Router has only one /64 prefix to be used across multiple LAN interfaces and the CE Router supports any two LAN interfaces that cannot bridge data between them because the two interfaces have disparate MAC layers, then the CE Router MUST support Proxying Neighbor Advertisements as specified in Section 7.2.8 of [RFC4861]. If any two LAN interfaces support bridging between the interfaces, then Proxying Neighbor Advertisements is not necessary between the two interfaces. Legacy 3GPP networks have the following requirements:

1. No DHCPv6 prefix is delegated to the CE Router.

2. Only one /64 is available on the WAN link.

3. The link types between the WAN interface and LAN interface(s) are disparate and, therefore, can’t be bridged.

4. No NAT66 is to be used.

5. Each LAN interface needs global connectivity.

6. Uses SLAAC to configure LAN interface addresses.

For these legacy 3GPP networks, the CPE Router MUST support ND Proxy between the WAN and LAN interface(s). If a CE

Router will never be deployed in an environment with these characteristics, then ND Proxy is not necessary.

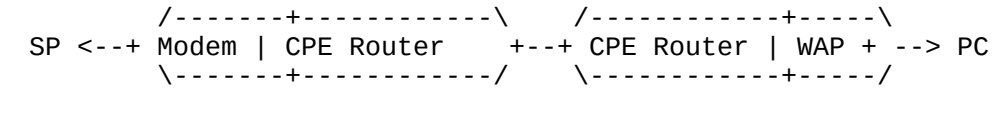
5.4. Prefix Delegation on LAN interface(s) (More details are TBD)

This section is only applicable to a CE Router with at least one LAN interface. The LAN interface(s) are delegated prefixes subnetted from the delegated prefix acquired by the WAN interface and the ULA prefix. After the CE router has assigned prefixes for all of its internally defined needs (its interfaces and any other purposes defined in its internal logic), any leftover prefixes are available for delegation. Any automated prefix delegation mechanism is TBD.

5.5. Routed network behavior(General Cases TBD)

CPE Router Behavior in a routed network:

R-1: One example of the CPE Router use in the home is shown below. The home has a broadband modem combined with a CPE Router, all in one device. The LAN interface of the device is connected to another standalone CPE Router that supports a wireless access point. To support such a network, this document recommends using prefix delegation of the prefix obtained either via IA_PD from WAN interface or a ULA from the LAN interface . The network interface of the downstream router may obtain an IA_PD via stateful DHCPv6. If the CPE router supports the routed network through automatic prefix delegation, the CPE router MUST support a DHCPv6 server or DHCPv6 relay agent. Further, if an IA_PD is used, the Service Provider or user MUST allocate an IA_PD or ULA prefix short enough to be delegated and subsequently used for SLAAC. Therefore, a prefix length shorter than /64 is needed. The CPE Router MAY support and IGP in the home network.



WAP = Wireless Access Point

Figure 2.

5.6. Transition Technologies Support

5.6.1. Dual-Stack(DS)-Lite

Even as users migrate from IPv4 to IPv6 addressing, a significant percentage of Internet resources and content will remain accessible only through IPv4. Also, many end-user devices will only support IPv4. As a consequence, Service Providers require mechanisms to allow customers to continue to access content and resources using IPv4 even after the last IPv4 allocations have been fully depleted. One technology that can be used for IPv4 address extension is DS-Lite.

DS-Lite enables a Service Provider to share IPv4 addresses among multiple customers by combining two well-known technologies: IP in IP (IPv4-in-IPv6) tunneling and Carrier Grade NAT. More specifically, Dual-Stack-Lite encapsulates IPv4 traffic inside an IPv6 tunnel at the IPv6 CE Router and sends it to a Service Provider Address Family Translation Router (AFTR). Configuration of the IPv6 CE Router to support IPv4 LAN traffic is outside the scope of this document.

The IPv6 CE Router SHOULD implement DS-Lite functionality as specified in [I-D.ietf-softwire-dual-stack-lite].

WAN requirements:

DLW-1: To facilitate IPv4 extension over an IPv6 network, if the CE Router supports DS-Lite functionality, the CE Router WAN interface MUST implement a B4 Interface as specified in [I-D.ietf-softwire-dual-stack-lite].

DLW-2: If the IPv6 CE Router implements DS-Lite functionality, the CE Router MUST support using a DS-Lite DHCPv6 option [I-D.ietf-softwire-ds-lite-tunnel-option] to configure the DS-Lite tunnel. The IPv6 CE Router MAY use other mechanisms to configure DS-Lite parameters. Such mechanisms are outside the scope of this document.

DLW-3: IPv6 CE Router MUST NOT perform IPv4 Network Address Translation (NAT) on IPv4 traffic encapsulated using DS-Lite.

DLW-4: If the IPv6 CE Router is configured with an non-RFC1918 IPv4 address on its WAN interface, the IPv6 CE Router MUST disable the DS-Lite B4 element.

DLW-5: If DS-Lite is operational on the IPv6 CE Router, multicast data MUST NOT be sent on any DS-Lite tunnel.

5.6.2. 6rd

The IPv6 CE Router can be used to offer IPv6 service to a LAN, even when the WAN access network only supports IPv4. One technology that supports IPv6 service over an IPv4 network is IPv6 Rapid Deployment (6rd). 6rd encapsulates IPv6 traffic from the end user LAN inside IPv4 at the IPv6 CE Router and sends it to a Service Provider Border Relay (BR). The IPv6 CE Router calculates a 6rd delegated IPv6 prefix during 6rd configuration, and sub-delegates the 6rd delegated prefix to devices in the LAN.

The IPv6 CE Router SHOULD implement 6rd functionality as specified in [RFC5969].

6rd requirements:

6RD-1: If the IPv6 CE Router implements 6rd functionality, the CE Router WAN interface MUST support at least one 6rd Virtual Interface and 6rd CE functionality as specified in [RFC5969].

6RD-2: If the IPv6 CE Router implements 6rd CE functionality, it MUST support using the 6rd DHCPv4 Option (212) for 6rd configuration. The IPv6 CE Router MAY use other mechanisms to configure 6rd parameters. Such mechanisms are outside the scope of this document.

6RD-3: If 6rd is operational on the IPv6 CE Router, multicast data MUST NOT be sent on any 6rd tunnel.

5.6.3. Transition Technologies Coexistence

Run the following four in parallel to provision CPE router connectivity to the Service Provider:

1. Initiate IPv4 address acquisition.

2. Initiate IPv6 address acquisition as specified by [I-D.ietf-v6ops-ipv6-cpe-router].

3. If 6rd is provisioned, initiate 6rd.

4. If DS-Lite is provisioned, initiate DS-Lite.

The default route for IPv6 through the native physical interface should have preference over the 6rd tunnel interface. The default

route for IPv4 through the native physical interface should have preference over the DS-Lite tunnel interface.

5.7. Quality Of Service

Q-1: The CPE router MAY support differentiated services [RFC2474].

5.8. Unicast Data Forwarding

The null route introduced by the WPD-6 requirement in [I-D.ietf-v6ops-ipv6-cpe-router] has lower precedence than other routes except for the default route.

5.9. ZeroConf

The CE Router MAY support manual configuration via the web using a URL string like http://router.local as per multicast DNS (mDNS). Zero-configuration is vendor-dependent.

6. Security Considerations

None.

7. Acknowledgements

Thanks to the following people (in alphabetical order) for their guidance and feedback:

Mikael Abrahamsson, Merete Asak, Scott Beuker, Mohamed Boucadair, Rex Bullinger, Brian Carpenter, Remi Denis-Courmont, Gert Doering, Alain Durand, Katsunori Fukuoka, Tony Hain, Thomas Herbst, Kevin Johns, Stephen Kramer, Victor Kuarsingh, Francois-Xavier Le Bail, David Miles, Shin Miyakawa, Jean-Francois Mule, Michael Newbery, Carlos Pignataro, John Pomeroy, Antonio Querubin, Teemu Savolainen, Matt Schmitt, Hiroki Sato, Mark Townsley, Bernie Volz, James Woodyatt, Dan Wing and Cor Zwart

This draft is based in part on CableLabs' eRouter specification. The authors wish to acknowledge the additional contributors from the eRouter team:

Ben Bekele, Amol Bhagwat, Ralph Brown, Eduardo Cardona, Margo Dolas, Toerless Eckert, Doc Evans, Roger Fish, Michelle Kuska, Diego Mazzola, John McQueen, Harsh Parandekar, Michael Patrick, Saifur Rahman, Lakshmi Raman, Ryan Ross, Ron da Silva, Madhu Sudan, Dan Torbet and Greg White.

8. Contributors

The following people have participated as co-authors or provided substantial contributions to this document: Ralph Droms, Kirk Erichsen, Fred Baker, Jason Weil, Lee Howard, Jean-Francois Tremblay, Yiu Lee, John Jason Brzozowski and Heather Kirksey.

9. IANA Considerations

This memo includes no request to IANA.

10. References

10.1. Normative References

[I-D.bagnulo-behave-dns64]
Bagnulo, M., Sullivan, A., Matthews, P., Beijnum, I., and M. Endo, "DNS64: DNS extensions for Network Address Translation from IPv6 Clients to IPv4 Servers", draft-bagnulo-behave-dns64-02 (work in progress), March 2009.

[I-D.ietf-softwire-ds-lite-tunnel-option]
Hankins, D. and T. Mrugalski, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6) Option for Dual- Stack Lite", draft-ietf-softwire-ds-lite-tunnel-option-05 (work in progress), September 2010.

[I-D.ietf-softwire-dual-stack-lite]
Durand, A., Droms, R., Woodyatt, J., and Y. Lee, "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", draft-ietf-softwire-dual-stack-lite-06 (work in progress), August 2010.

[I-D.ietf-v6ops-ipv6-cpe-router]
Singh, H., Beebee, W., Donley, C., Stark, B., and O. Troan, "Basic Requirements for IPv6 Customer Edge Routers", draft-ietf-v6ops-ipv6-cpe-router-07 (work in progress), August 2010.

[I-D.vyncke-advanced-ipv6-security]
Vyncke, E. and M. Townsley, "Advanced Security for IPv6 CPE", draft-vyncke-advanced-ipv6-security-01 (work in progress), March 2010.

[RFC1122] Braden, R., "Requirements for Internet Hosts -

Communication Layers", STD 3, RFC 1122, October 1989.

[RFC2080] Malkin, G. and R. Minnear, "RIPng for IPv6", RFC 2080, January 1997.

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

[RFC2464] Crawford, M., "Transmission of IPv6 Packets over Ethernet Networks", RFC 2464, December 1998.

[RFC2474] Nichols, K., Blake, S., Baker, F., and D. Black, "Definition of the Differentiated Services Field (DS Field) in the IPv4 and IPv6 Headers", RFC 2474, December 1998.

[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.

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

[RFC3646] Droms, R., "DNS Configuration options for Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", RFC 3646, December 2003.

[RFC3736] Droms, R., "Stateless Dynamic Host Configuration Protocol (DHCP) Service for IPv6", RFC 3736, April 2004.

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

[RFC4075] Kalusivalingam, V., "Simple Network Time Protocol (SNTP) Configuration Option for DHCPv6", RFC 4075, May 2005.

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

[RFC4242] Venaas, S., Chown, T., and B. Volz, "Information Refresh Time Option for Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", RFC 4242, November 2005.

[RFC4294] Loughney, J., "IPv6 Node Requirements", RFC 4294, April 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.

[RFC4541] Christensen, M., Kimball, K., and F. Solensky, "Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches", RFC 4541, May 2006.

[RFC4605] Fenner, B., He, H., Haberman, B., and H. Sandick, "Internet Group Management Protocol (IGMP) / Multicast Listener Discovery (MLD)-Based Multicast Forwarding ("IGMP/MLD Proxying")", RFC 4605, August 2006.

[RFC4632] Fuller, V. and T. Li, "Classless Inter-domain Routing (CIDR): The Internet Address Assignment and Aggregation Plan", BCP 122, RFC 4632, August 2006.

[RFC4779] Asadullah, S., Ahmed, A., Popoviciu, C., Savola, P., and J. Palet, "ISP IPv6 Deployment Scenarios in Broadband Access Networks", RFC 4779, January 2007.

[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.

[RFC4864] Van de Velde, G., Hain, T., Droms, R., Carpenter, B., and E. Klein, "Local Network Protection for IPv6", RFC 4864, May 2007.

[RFC5072] S.Varada, Haskins, D., and E. Allen, "IP Version 6 over PPP", RFC 5072, September 2007.

[RFC5571] Storer, B., Pignataro, C., Dos Santos, M., Stevant, B., Toutain, L., and J. Tremblay, "Softwire Hub and Spoke Deployment Framework with Layer Two Tunneling Protocol Version 2 (L2TPv2)", RFC 5571, June 2009.

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

10.2. Informative References

[I-D.ietf-behave-v6v4-framework]
Baker, F., Li, X., Bao, C., and K. Yin, "Framework for IPv4/IPv6 Translation", draft-ietf-behave-v6v4-framework-10 (work in progress), August 2010.

[UPnP-IGD]
UPnP Forum, "Universal Plug and Play (UPnP) Internet Gateway Device (IGD)", November 2001, <http://www.upnp.org/standardizeddcps/igd.asp>.

Authors' Addresses

Hemant Singh
Cisco Systems, Inc.
1414 Massachusetts Ave.
Boxborough, MA 01719
USA

Phone: +1 978 936 1622
Email: shemant@cisco.com
URI: http://www.cisco.com/

Wes Beebee
Cisco Systems, Inc.
1414 Massachusetts Ave.
Boxborough, MA 01719
USA

Phone: +1 978 936 2030
Email: wbeebee@cisco.com
URI: http://www.cisco.com/

Chris Donley
CableLabs
858 Coal Creek Circle
Louisville, CO 80027
USA

Email: c.donley@cablelabs.com

Barbara Stark
AT&T
725 W Peachtree St
Atlanta, GA 30308
USA

Email: barbara.stark@att.com

Ole Troan (editor)
Cisco Systems, Inc.
Veversmauet 8
N-5017 BERGEN,
Norway

Email: ot@cisco.com

Network Working Group
Internet-Draft
Intended status: Informational
Expires: May 15, 2011

J. Weil
Cox Communications
V. Kuarsingh
Rogers Communications
C. Donley
CableLabs
C. LILJENSTOLPE
Telstra Corp
M. Azinger
Frontier Communications

November 11, 2010

IANA Reserved IPv4 Prefix for Shared Transition Space
draft-weil-shared-transition-space-request-01

Abstract

This document requests a reserved IANA IPv4 address allocation as Shared Transition Space to support the deployment of IPv4 address sharing technologies post IPv4 exhaustion.

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 May 15, 2011.

Table of Contents

1. Introduction

Many operators are currently implimenting their IPv6 transition plans. During the transition, continued support for heritage IPv4 only devices will be required. While most operators are well aware of the limitations of NAT444 [I-D.shirasaki-nat444] (see [I-D.donley-nat444-impacts]), it is the transition mechnism that has the least customer impact for many carriers.

To deal with some of the NAT444 limitations, it becomes necessary for a provider to utilize address space in the NAT444 infrastructure that will not conflict with it’s customer space.

This document requests that IANA reserve a portion of the remaining unallocated space as Shared Transition Space for the enablement of a clean transition strategy in provider networks.

2. 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].

3. Motivation

The Internet community is rapidly consuming the remaining supply of unallocated IPv4 addresses. During the transition period to IPv6, it is imperative that Service Providers maintain IPv4 service for devices and networks that are currently incapable of upgrading to IPv6.

In order to provide IPv4 service to customers and/or devices once the IPv4 address space is exhausted, Service Providers must multiplex several subscribers behind a single IPv4 address using one of several techniques including NAT444 . Providers need sufficient non-[RFC1918] address space to deploy such technologies and avoid overlap with customer use of private address space.

Many CPE router devices used to provide residential or small-medium business services have been optimized for IPv4 operation, and typically require replacement in order to fully support the transition to IPv6 (either natively or via one of many transition technologies). In addition, various consumer devices including IP-enabled televisions, gaming consoles, medical and family monitoring devices, etc. are IPv4-only, and cannot be upgraded. While these will eventually be replaced with dual-stack or IPv6 capable devices, this transition will take many years. As these are typically consumer-owned devices, service providers do not have control over the speed of their replacement cycle. However, consumers have an expectation that they will continue to receive IPv4 service, and that such devices will continue to have IPv4 Internet connectivity after the IPv4 pool is exhausted, even if the customer contracts for new service with a new provider.

Until such customers replace their Home Gateways and all IPv4-only CPE devices with IPv6-capable devices, Service Providers will be required to continue to offer IPv4 services through the use of an IPv4 address sharing technology such as NAT444 [I-D.shirasaki-nat444]. The challenges associated with these deployments are identified in [I-D.shirasaki-nat444-isp-shared-addr], [I-D.donley-nat444-impacts], and [I-D.ietf-intarea-shared-addressing-issues].

Addressing solutions for dealing with the depletion of the IPv4 public address space and the lack of available private addresses within large providers are presented in [I-D.azinger-additional-private-ipv4-space-issues] as well as [I-D.shirasaki-nat444-isp-shared-addr]. For infrastructure providers whose customers are already using [RFC1918] space, the preferred method for addressing the problems presented in both documents is to direct IANA to reserve address space from its unassigned IPv4 address

pool for Shared Transition Space.

4. Shared Transition Space

This document proposes the assignment of the equivalent of a /10 as Shared Transition Space. This block could be composed of one contiguous assignment, or several discontiguous assignments. Shared Transition Space is IPv4 address space reserved for Infrastructure provider use with the purpose of facilitating IPv6 transition and IPv4 coexistence deployment. The requested block SHOULD NOT be utilized for any purpose other than IPv4 to IPv6 transition infrastructure. Network equipment manufacturers MUST NOT use the assigned block in default or example device configurations.

Because Shared Transition addresses have no meaning outside of the Infrastructure Provider, routing information about shared transition space networks MUST NOT be propagated on interdomain links, and packets with shared transition source or destination addresses SHOULD NOT be forwarded across such links. Internet service providers SHOULD filter out routing information about shared transition space networks on ingress links.

5. Problems using Future Use Space

[I-D.fuller-240space] and [I-D.wilson-class-e] suggest that 240.0.0.0/4 space could be used as Shared Transition Space. However, as discussed in [I-D.azinger-additional-private-ipv4-space-issues], some existing network equipment does not support addresses in the 240.0.0.0/4 range. In particular, [CISCO] states that "no addresses are allowed with the highest-order bits set to 1111". It is likely that many home routers will not support this range, either. In order to use this range, equipment vendors would need to update software code for existing routers and end users would need to upgrade their home devices. As many older home routers do not support automatic updates, it is unlikely that enough end users would upgrade to make the 240.0.0.0/4 range viable for Shared Transition Space use.

6. Security Considerations

This memo does not define any protocol, and raises no security issues. Any addresses allocated as Shared Transition Space would not be routable on the Internet.

7. IANA Considerations

IANA is asked to reserve an IPv4 /10 from its remaining pool of unallocated IPv4 addresses for use as Shared Transition Space.

8. Informative References

[CISCO] Cisco Systems, "TCP/IP Overview", <http://www.cisco.com/univercd/cc/td/doc/product/rtrmgmt/cwhubs/starvwug/83428.htm#xtocid74886>.

[I-D.azinger-additional-private-ipv4-space-issues] Azinger, M. and L. Vegoda, "Additional Private IPv4 Space Issues", draft-azinger-additional-private-ipv4-space-issues-04 (work in progress), April 2010.

[I-D.donley-nat444-impacts] Donley, C., Howard, L., Kuarsingh, V., Chandrasekaran, A., and V. Ganti, "Assessing the Impact of NAT444 on Network Applications", draft-donley-nat444-impacts-01 (work in progress), October 2010.

[I-D.fuller-240space] Fuller, V., "Reclassifying 240/4 as usable unicast address space", draft-fuller-240space-02 (work in progress), March 2008.

[I-D.ietf-intarea-shared-addressing-issues] Ford, M., Boucadair, M., Durand, A., Levis, P., and P. Roberts, "Issues with IP Address Sharing", draft-ietf-intarea-shared-addressing-issues-02 (work in progress), October 2010.

[I-D.shirasaki-nat444] Yamagata, I., Shirasaki, Y., Nakagawa, A., Yamaguchi, J., and H. Ashida, "NAT444", draft-shirasaki-nat444-02 (work in progress), July 2010.

[I-D.shirasaki-nat444-isp-shared-addr] Shirasaki, Y., Miyakawa, S., Nakagawa, A., Yamaguchi, J., and H. Ashida, "NAT444 addressing models", draft-shirasaki-nat444-isp-shared-addr-04 (work in progress), July 2010.

[I-D.wilson-class-e] Wilson, P., Michaelson, G., and G. Huston, "Redesignation of 240/4 from "Future Use" to "Private Use"", draft-wilson-class-e-02 (work in progress), September 2008.

[RFC1918] Rekhter, Y., Moskowitz, R., Karrenberg, D., Groot, G., and E. Lear, "Address Allocation for Private Internets",

BCP 5, RFC 1918, February 1996.

[RFC2119] "".

Appendix A. Acknowledgements

Thanks to the following people (in alphabetical order) for their guidance and feedback:

John Brzozowski

Isaiah Connell

Greg Davies

Kirk Erichsen

Wes George

Tony Hain

Philip Matthews

John Pomeroy

Barbara Stark

Jean-Francois Tremblay

Leo Vegoda

Steven Wright

Ikuhei Yamagata

Authors' Addresses

Jason Weil
Cox Communications
1400 Lake Hearn Drive
Atlanta, GA 30319
USA

Email: jason.weil@cox.com

Victor Kuarsingh
Rogers Communications
8200 Dixie Road
Brampton, ON L6T 0C1
Canada

Email: victor.kuarsingh@rci.rogers.com

Chris Donley
CableLabs
858 Coal Creek Circle
Louisville, CO 80027
USA

Email: c.donley@cablelabs.com

Christopher Liljenstolpe
Telstra Corp
7/242 Exhibition Street
Melbourne, VIC 316
AU

Phone: +61 3 8647 6389
Fax:
Email: cdl@asgaard.org
URI:

Marla Azinger
Frontier Communications
Vancouver, WA
US

Phone: +1.360.513.2293
Fax:
Email: marla.azinger@frontiercorp.com
URI:

Network Working Group
Internet-Draft
Updates: 5735 (if approved)
Intended status: BCP
Expires: August 19, 2012

J. Weil
Time Warner Cable
V. Kuarsingh
Rogers Communications
C. Donley
CableLabs
C. Liljenstolpe
Telstra Corp
M. Azinger
Frontier Communications

February 16, 2012

IANA Reserved IPv4 Prefix for Shared Address Space
draft-weil-shared-transition-space-request-15

Abstract

This document requests the allocation of an IPv4 /10 address block to be used as Shared Address Space to accommodate the needs of Carrier Grade Network Address Translation (CGN) devices. It is anticipated that Service Providers will use this Shared Address Space to number the interfaces that connect CGN devices to Customer Premise Equipment (CPE).

Shared Address Space is distinct from RFC1918 private address space because it is intended for use on Service Provider networks. However, it may be used in a manner similar to RFC 1918 private address space on routing equipment that is able to do address translation across router interfaces when the addresses are identical on two different interfaces. Details are provided in the text of this document.

As this document proposes the allocation of an additional special-use IPv4 address block, it updates RFC 5735.

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 19, 2012.

Copyright Notice

Table of Contents

1. Introduction

IPv4 address space is nearly exhausted. However, ISPs must continue to support IPv4 growth until IPv6 is fully deployed. To that end, many ISPs will deploy Carrier Grade NAT (CGN) such as that described in [RFC6264]. Because CGNs are used on networks where public address space is expected, and currently available private address space causes operational issues when used in this context, ISPs require a new IPv4 /10 address block. This address block will be called the Shared Address Space and will be used to number the interfaces that connect CGN devices to Customer Premise Equipment (CPE).

Shared Address Space is similar to [RFC1918] private address space in that it is not global routable address space and can be used by multiple pieces of equipment. However, Shared Address Space has limitations in its use that the current [RFC1918] private address space does not have. In particular, Shared Address Space can only be used in Service Provider networks or on routing equipment that is able to do address translation across router interfaces when the addresses are identical on two different interfaces.

This document requests the allocation of an IPv4 /10 address block to be used as Shared Address Space. In conversations with many ISPs, a /10 is the smallest block that will allow them to deploy CGNs on a regional basis without requiring nested CGNs. For Instance, as described in [I-D.shirasaki-isp-shared-addr], a /10 is sufficient to service Points of Presence in the Tokyo area.

As this document proposes the allocation of an additional special-use IPv4 address block, it updates [RFC5735].

2. 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].

3. Alternatives to Shared Address Space

The interfaces that connect CGN devices to CPE might conceivably be numbered from any of the following address spaces:

- legitimately assigned globally unique address space
- usurped globally unique address space (i.e., squat space)
- [RFC1918] space
- Shared Address Space

A Service Provider can number the interfaces in question from legitimately assigned globally unique address space. While this solution poses the fewest problems, it is impractical because globally unique IPv4 address space is in short supply. While the Regional Internet Registries (RIR) have enough address space to allocate a single /10 to be shared by all Service Providers, they do not have enough address space to make a unique assignment to each Service Provider.

Service Providers MUST NOT number the interfaces in question from usurped globally unique address space (i.e., squat space). If a Service Provider leaks advertisements for squat space into the global Internet, the legitimate holders of that address space may be adversely impacted, as would those wishing to communicate with them. Even if the Service Provider did not leak advertisements for squat space, the Service Provider and its subscribers might lose connectivity to the legitimate holders of that address space.

A Service Provider can number the interfaces in question from [RFC1918] space if either of the following conditions are true:

- The Service Provider knows that the CPE/NAT works correctly when the same [RFC1918] address block is used both on its inside and outside interfaces.
- The Service Provider knows that the [RFC1918] address block that it uses to number interfaces between the CGN and CPE is not used on the subscriber side of the CPE.

Unless at least one of the conditions above is true, the Service Provider cannot safely use [RFC1918] address space and must resort to Shared Address Space. This is typically the case in an unmanaged service, where subscribers provide their own CPE and number their own internal network.

4. Use of Shared CGN Space

Shared Address Space is IPv4 address space designated for Service Provider use with the purpose of facilitating CGN deployment. Also, Shared Address Space can be used as additional non-globally routable space on routing equipment that is able to do address translation across router interfaces when the addresses are identical on two different interfaces.

Devices MUST be capable of performing address translation when identical Shared Address Space ranges are used on two different interfaces.

Packets with Shared Address Space source or destination addresses MUST NOT be forwarded across Service Provider boundaries. Service Providers MUST filter such packets on ingress links. As above, one exception to the above proscriptions is in the case of business relationships such as hosted CGN service.

When running a single DNS infrastructure, Service Providers MUST NOT include Shared Address Space in zone files. When running a split DNS infrastructure, Service Providers MUST NOT include Shared Address Space in external-facing zone files.

Reverse DNS queries for Shared Address Space addresses MUST NOT be forwarded to the global DNS infrastructure. DNS Providers SHOULD filter requests for Shared Address Space reverse DNS queries on recursive nameservers. This is done to avoid having to set up something similar to AS112.net for RFC 1918 private address space that a host has incorrectly sent for a DNS reverse-mapping queries on the public Internet [RFC6304].

Because CGN service requires non-overlapping address space on each side of the home NAT and CGN, entities using Shared Address Space for purposes other than for CGN service, as described in this document, are likely to experience problems implementing or connecting to CGN service at such time as they exhaust their supply of public IPv4 addresses.

5. Risk

5.1. Analysis

Some existing applications discover the outside address of their local CPE, determine whether the address is reserved for special-use, and behave differently based on that determination. If a new IPv4 address block is reserved for special-use and that block is used to number CPE outside interfaces, some of the above-mentioned applications may fail.

For example, assume that an application requires its peer (or some other device) to initiate an incoming connection directly with its CPE outside address. That application discovers the outside address of its CPE and determines whether that address is reserved for special-use. If the address is reserved for special-use, the application rightly concludes the that address is not reachable from the global Internet and behaves in one manner. If the address is not reserved for special-use, the application assumes that the address is reachable from the global Internet and behaves in another manner.

While the assumption that a non-special-use address is reachable from the global Internet is generally safe, it is not always true (e.g., when the CPE outside interface is numbered from globally unique address space but that address is not advertised to the global Internet as when it is behind a CGN). Such an assumption could cause certain applications to behave incorrectly in those cases.

5.2. Empirical Data

The primary motivation for the allocation of Shared Address Space is as address space for CGNs; the use and impact of CGNs has been previously described in [RFC6269] and [I-D.donley-nat444-impacts]. Some of the services adversely impacted by CGNs are:

1. Console gaming - some games fail when two subscribers using the same outside public IPv4 address try to connect to each other.

2. Video streaming - performance is impacted when using one of several popular video streaming technologies to deliver multiple video streams to users behind particular CPE routers.

3. Peer-to-peer - some peer-to-peer applications cannot seed content due to the inability to open incoming ports through the CGN. Likewise, some SIP client implementations cannot receive incoming calls unless they first initiate outgoing traffic or open an incoming port through the CGN using [I-D.ietf-pcp-base] or similar mechanism.

4. Geo-location - geo-location systems identify the location of the CGN server, not the end host.

5. Simultaneous logins - some websites (particularly banking and social networking websites) restrict the number of simultaneous logins per outside public IPv4 address.

6. 6to4 - 6to4 requires globally reachable addresses, and will not work in networks that employ addresses with limited topological span such as those employing CGNs.

Based on testing documented in [I-D.donley-nat444-impacts], the CGN impacts on 1-5 are comparable regardless of whether globally unique, Shared Address Space, or [RFC1918] addresses are used. There is, however, a difference between the three alternatives in the treatment of 6to4.

As described in [RFC6343], CPE routers do not attempt to initialize 6to4 tunnels when they are configured with [RFC1918] or [RFC5735] WAN addresses. When configured with globally unique or Shared Address Space addresses, such devices may attempt to initiate 6to4, which would fail. Service Providers can mitigate this issue using 6to4-PMT [I-D.kuarsingh-v6ops-6to4-provider-managed-tunnel] or blocking the route to 192.88.99.1 and generating an IPv4 'destination unreachable' message [RFC6343]. When the address range is well-defined, as with Shared Address Space, CPE router vendors can include Shared Address Space in their list of special-use addresses (e.g., [RFC5735]) and treat Shared Address Space similarly to [RFC1918] space. When the CGN-CPE address range is not well-defined, as in the case of globally unique space, it will be more difficult for CPE router vendors to mitigate against this issue.

Thus, when comparing the use of [RFC1918] and Shared Address Space, Shared Address Space poses an additional impact on 6to4 connectivity, which can be mitigated by Service Provider or CPE router vendor action. On the other hand, the use of [RFC1918] address space poses more of a challenge vis-a-vis Shared Address Space when the subscriber and Service Provider use overlapping [RFC1918] space, which will be outside the Service Provider's control in the case of unmanaged service. Service Providers have indicated that it is more challenging to mitigate the possibility of overlapping [RFC1918] address space on both sides of the CPE router than it is to mitigate the 6to4 impacts of Shared Address Space.

6. Security Considerations

Similar to other [RFC5735] special use IPv4 addresses, Shared Address Space does not directly raise security issues. However, the Internet does not inherently protect against abuse of these addresses. Attacks have been mounted that depend on the unexpected use of similar special-use addresses. Network operators are encouraged to review this document and determine what security policies should be associated with this address block within their specific operating environments and should consider including Shared Address Space in Ingress Filter lists [RFC3704] unless their Internet service incorporates a CGN.

To mitigate against potential misuse of Shared Address Space, except where required for hosted CGN service or similar business relationship,

- Routing information about Shared Address Space networks MUST NOT be propagated across Service Provider boundaries. Service Providers MUST filter incoming advertisements regarding Shared Address Space.

- Packets with Shared Address Space source or destination addresses MUST NOT be forwarded across Service Provider boundaries. Service Providers MUST filter such packets on ingress links.

- Service Providers MUST NOT include Shared Address Space in external-facing DNS zone files.

- Reverse DNS queries for Shared Address Space addresses MUST NOT be forwarded to the global DNS infrastructure.

- DNS Providers SHOULD filter requests for Shared Address Space reverse DNS queries on recursive nameservers.

7. IANA Considerations

IANA is asked to record the allocation of an IPv4 /10 for use as Shared Address Space.

The Shared Address Space address range is: x.x.0.0/10. [Note to RFC Editor: this address range to be added before publication]

8. References

8.1. Normative References

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

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

[RFC5735] Cotton, M. and L. Vegoda, "Special Use IPv4 Addresses", BCP 153, RFC 5735, January 2010.

8.2. Informative References

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

[I-D.ietf-pcp-base] Wing, D., Cheshire, S., Boucadair, M., Penno, R., and P. Selkirk, "Port Control Protocol (PCP)", draft-ietf-pcp-base-13 (work in progress), July 2011.

[I-D.kuarsingh-v6ops-6to4-provider-managed-tunnel] Kuarsingh, V., Lee, Y., and O. Vautrin, "6to4 Provider Managed Tunnels", draft-kuarsingh-v6ops-6to4-provider-managed-tunnel-03 (work in progress), September 2011.

[I-D.shirasaki-isp-shared-addr] Yamagata, I., Miyakawa, S., Nakagawa, A., Yamaguchi, J., and H. Ashida, "ISP Shared Address", draft-shirasaki-isp-shared-addr-06 (work in progress), July 2011.

[RFC3704] Baker, F. and P. Savola, "Ingress Filtering for Multihomed Networks", BCP 84, RFC 3704, March 2004.

[RFC6264] Jiang, S., Guo, D., and B. Carpenter, "An Incremental Carrier-Grade NAT (CGN) for IPv6 Transition", RFC 6264, June 2011.

[RFC6269] Ford, M., Boucadair, M., Durand, A., Levis, P., and P. Roberts, "Issues with IP Address Sharing", RFC 6269,

June 2011.

[RFC6304] Abley, J. and W. Maton, "AS112 Nameserver Operations", RFC 6304, July 2011.

[RFC6343] Carpenter, B., "Advisory Guidelines for 6to4 Deployment", RFC 6343, August 2011.

Appendix A. Acknowledgments

Thanks to the following people (in alphabetical order) for their guidance and feedback:

Stan Barber

John Brzozowski

Isaiah Connell

Greg Davies

Owen DeLong

Kirk Erichsen

Wes George

Chris Grundemann

Tony Hain

Philip Matthews

John Pomeroy

Barbara Stark

Jean-Francois Tremblay

Leo Vegoda

Steven Wright

Ikuhei Yamagata

Authors' Addresses

Jason Weil
Time Warner Cable
13820 Sunrise Valley Drive
Herndon, VA 20171
USA

Email: jason.weil@twcable.com

Victor Kuarsingh
Rogers Communications
8200 Dixie Road
Brampton, ON L6T 0C1
Canada

Email: victor.kuarsingh@gmail.com

Chris Donley
CableLabs
858 Coal Creek Circle
Louisville, CO 80027
USA

Email: c.donley@cablelabs.com

Christopher Liljenstolpe
Telstra Corp
7/242 Exhibition Street
Melbourne, VIC 316
Australia

Phone: +61 3 8647 6389
Email: cdl@asgaard.org

Marla Azinger
Frontier Communications
Vancouver, WA
USA

Phone: +1.360.513.2293
Email: marla.azinger@frontiercorp.com

v6ops
Internet-Draft
Intended status: Standards Track
Expires: April 28, 2011

D. Wing
A. Yourtchenko
Cisco

October 25, 2010

Happy Eyeballs: Trending Towards Success with Dual-Stack Hosts
draft-wing-v6ops-happy-eyeballs-ipv6-01

Abstract

This document describes how a dual-stack client can determine the functioning path to a dual-stack server. This provides a seemless user experience during initial deployment of dual-stack networks and during outages of IPv4 or outages of IPv6.

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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."

The list of current Internet-Drafts can be accessed at http://www.ietf.org/ietf/1id-abstracts.txt.

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html.

This Internet-Draft will expire on April 28, 2011.

Copyright Notice

Table of Contents

1. Introduction

In order to use HTTP successfully over IPv6, it is necessary that the user enjoys nearly identical performance as compared to IPv4. A combination of today's applications, IPv6 tunneling and IPv6 service providers, and some of today's content providers all cause the user experience to suffer (Section 3). For IPv6, Google ensures a positive user experience by using a DNS white list of IPv6 service providers who peer directly with Google [whitelist]. However, this is not scalable to all service providers worldwide, nor is it scalable for other content providers to operate their own DNS white list.

Instead, this document suggests a mechanism for applications to quickly determine if IPv6 or IPv4 is the most optimal to connect to a server. The suggestions in this document provide a user experience which is superior to connecting to ordered IP addresses which is helpful during the IPv6/IPv4 transition with dual stack hosts.

Following the procedures in this document, once a certain address family is successful, the application trends towards preferring that address family. Thus, repeated use of the application DOES NOT cause repeated probes over both address families.

While the application recommendations in this document are described in the context of HTTP clients ("web browsers"), but is useful and applicable to other time-sensitive applications.

2. Notational Conventions

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].

3. Problem Statement

As discussed in more detail in Section 3.1, it is important that the same URI and hostname be used for IPv4 and IPv6. Using separate namespaces causes namespace fragmentation and reduces the ability for users to share URIs and hostnames, and complicates printed material that includes the URI or hostname.

As discussed in more detail in Section 3.2, IPv6 connectivity is sometimes broken entirely or slower than native IPv4 connectivity.

3.1. URIs and hostnames

URIs are often used between users to exchange pointers to content -- such as on Facebook, email, instant messaging, or other systems. Thus, production URIs and production hostnames containing references to IPv4 or IPv6 will only function if the other party is also using an application, OS, and a network that can access the URI or the hostname.

3.2. IPv6

When IPv6 connectivity is impaired, today's IPv6-capable web browsers incur many seconds of delay before falling back to IPv4. This harms the user's experience with IPv6, which will slow the acceptance of IPv6, because IPv6 is frequently disabled in its entirety on the end systems to improve the user experience.

Reasons for such failure include no connection to the IPv6 Internet, broken 6to4 or Teredo tunnels, and broken IPv6 peering.

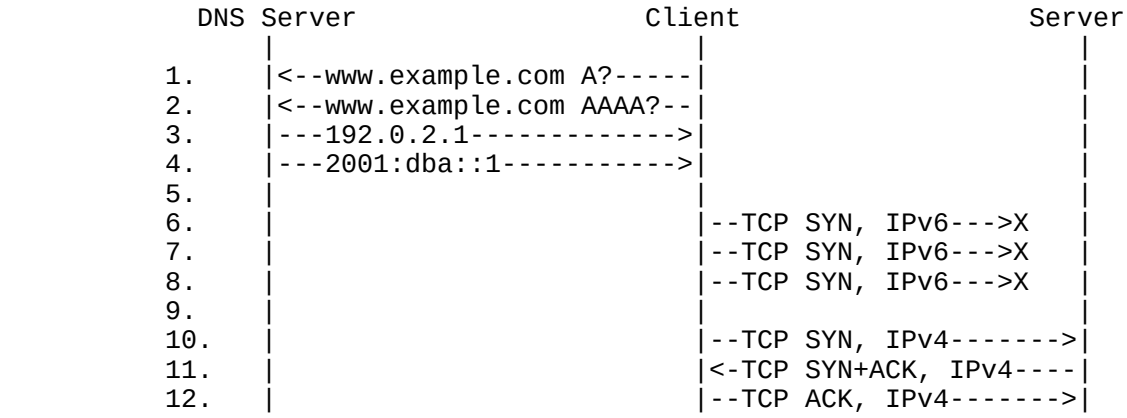


Figure 1: Existing behavior message flow

The client obtains the IPv4 and IPv6 records for the server (1-4). The client attempts to connect using IPv6 to the server, but the IPv6 path is broken (6-8), which consumes several seconds of time. Eventually, the client attempts to connect using IPv4 (10) which succeeds.

4. Client Recommendations

To provide fast connections for users, clients should make connections quickly over various technologies, automatically tune

itself to avoid flooding the network with unnecessary connections (i.e., for technologies that have not made successful connections), and occasionally flush its self-tuning.

4.1. IPv6

If a TCP client supports IPv6 and IPv4 and is connected to IPv4 and IPv6 networks, it can perform the procedures described in this section.

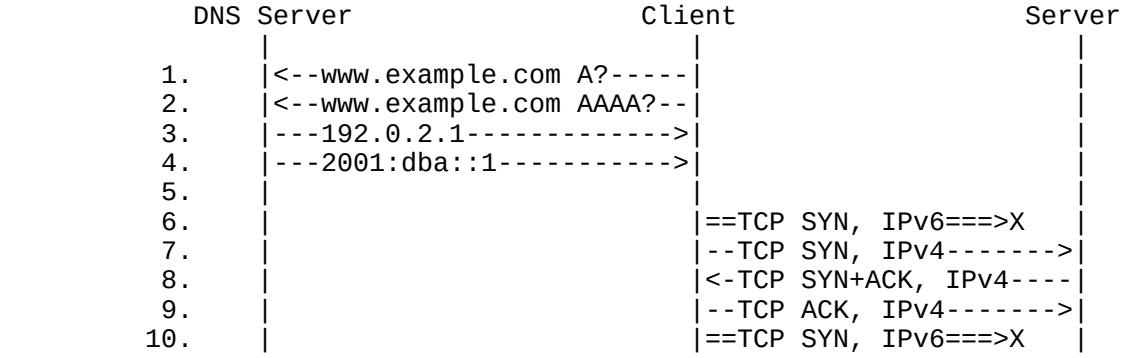


Figure 2: Happy Eyeballs flow 1, IPv6 broken

In diagram above, the client sends two TCP SYNs at the same time over IPv6 (6) and IPv4 (7). In the diagram, the IPv6 path is broken but has little impact to the user because there is no long delay before using IPv4. The IPv6 path is retried until the application gives up (10).

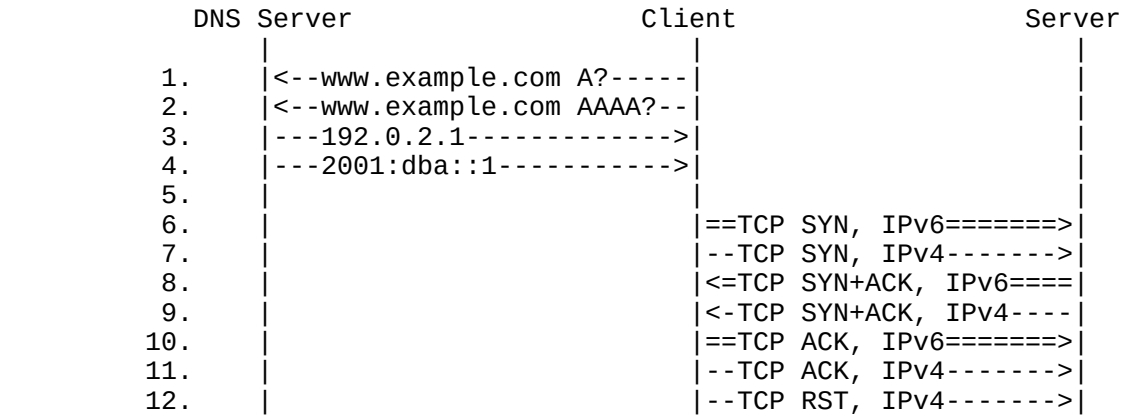


Figure 3: Happy Eyeballs flow 2, IPv6 working

The diagram above shows a case where both IPv6 and IPv4 are working, and IPv4 is abandoned (12).

This section details how to provide robust dual stack service for both IPv6 and IPv4, so that the user perceives very fast application response.

The TCP client application is configured with one value, P. A positive value indicates a preference for IPv6 and a negative value indicates a preference for IPv4. A value of 0 indicates equal weight, which means the A and AAAA queries and associated connection attempts will be sent as quickly as possible. The absolute value of P is the measure of a delay before initiating a connection attempt on the other address family. There are two P values maintained: one is application-wide and the other is specific per each destination (hostname and port).

The algorithm attempts to delay the DNS query until it expects that address family will be necessary; that is, if the preference is towards IPv6, then AAAA will be queried immediately and the A query will be delayed.

The TCP client application starts two threads in order to minimize the user-noticeable delay ("dead time") during the connection attempts:

thread 1: (IPv6)

* If P<0, wait for absolute value of p*10 milliseconds
* send DNS query for AAAA
* wait until DNS response is received
* Attempt to connect over IPv6 using TCP

thread 2: (IPv4)

* if P>0, wait for p*10 milliseconds
* send DNS query for A
* wait until DNS response is received
* Attempt to connect over IPv4 using TCP

The first thread that succeeds returns the completed connection to the parent code and aborts the other thread (Section 4.2.2).

After a connection is successful, we want to adjust the application-wide preference and the per-destination preference. The value of P is incremented (decremented) each time an IPv6 (IPv4) connection is successfully made. When a connection using the less-preferred address family is successful, it indicates the wrong address family was used and the P is halved:

- If P>0 (indicating IPv6 is preferred over IPv4) and the first thread to finish was the IPv6 thread it indicates the IPv6 preference is correct and we need to re-enforce this by increasing the application-wide P value by 1. However, if the first thread to finish was the IPv4 thread it indicates an IPv6 connection problem occurred and we need to aggressively prefer IPv4 more by halving P and rounding towards 0.

- If P<0 (indicating IPv4 is preferred over IPv6) and the first thread to finish was the IPv4 thread it indicates the preference is correct and we need to re-enforce this gently by decreasing the application-wide P value by 1. However, if the first thread to finish was the IPv6 thread it indicates an IPv4 connection problem and we need to aggressively avoid IPv4 by halving P and rounding towards 0.

- If P=0 (indicating equal preference), P is incremented if the first thread to complete was the IPv6 thread, or decremented if the first thread to complete was the IPv4 thread.

After adjusting P, it should never be larger than 4 seconds -- which is similar to the value used by many IPv6-capable TCP client applications to switch to an alternate A or AAAA record.

> Note: Proof of concept tests on fast networks show that even smaller value (around 0.5 seconds) is practical. More extensive testing would be useful to find the best upper boundary that still ensures a good user experience.

4.2. Additional Considerations

This section discusses considerations and requirements that are common to new technology deployment.

4.2.1. Additional Network and Host Traffic

Additional network traffic and additional server load is created due to these recommendations and mitigated by application-wide and per-destination timer adjustments. The procedures described in this document retain a quality user experience while transitioning from IPv4-only to dual stack. The quality user experience benefits the

user but to the detriment of the network and server that are serving the user.

4.2.2. Abandon Non-Winning Connections

It is RECOMMENDED that the non-winning connections be abandoned, even though they could be used to download content. This is because some web sites provide HTTP clients with cookies (after logging in) that incorporate the client's IP address, or use IP addresses to identify users. If some connections from the same HTTP client are arriving from different IP addresses, such HTTP applications will break.

4.2.3. Flush or Expire Cache

Because every network has different characteristics (e.g., working or broken IPv6 connectivity) the IPv6/IPv4 preference value (P) SHOULD be reset to its default whenever the host is connected to a new network ([cx-osx], [cx-win]). However, in some instances the application and the host are unaware the network connectivity has changed so it is RECOMMENDED that per-destination values expire after 10 minutes of inactivity.

4.2.4. Determining Address Type

[[[IS THIS SECTION NECESSARY ??

For some transitional technologies such as a dual-stack host, it is easy for the application to recognize the native IPv6 address (learned via a AAAA query) and the native IPv4 address (learned via an A query). For other transitional technologies [RFC2766] it is impossible for the host to differentiate a transitional technology IPv6 address from a native IPv6 address (see Section 4.1 of [RFC4966]). Replacement transitional technologies are attempting to bridge this gap. It is necessary for applications to distinguish between native and transitional addresses in order to provide the most seamless user experience.

]]]

4.2.5. Debugging and Troubleshooting

This mechanism is aimed to help the user experience in case of connectivity problems. However, this precise reason also makes it tougher to use these applications as a means of the verification that the problems are fixed. To assist in that regard, the applications implementing the proposal in this document SHOULD also provide a mechanism to temporarily use only one address family.

4.2.6. DNS Behavior

Unique to DNS AAAA queries are the problems described in [RFC4074] which, if they still persist, require applications to perform an A query before the AAAA query.

[[Editor's Note: It is believed these defective DNS servers have long since been upgraded. If so, we can remove this section.]]

4.2.7. Thread safe DNS resolvers

Some applications and some OSs do not have thread safe DNS resolvers, which complicates implementation of simultaneous A and AAAA queries for IPv4/IPv6.

4.2.8. Middlebox Issues

Some devices are known to exhibit what amounts to a bug, when the A and AAAA requests are sent back-to-back over the same 4-tuple, and drop one of the requests or replies [DNS-middlebox]. However, in some cases fixing this behaviour may not be possible either due to the architectural limitations or due to the administrative constraints (location of the faulty device is unknown to the end hosts or not controlled by the end hosts). The algorithm described in this draft, in the case of this erroneous behaviour will eventually pace the queries such that this issue is will be avoided. The algorithm described in this draft also avoids calling the operating system's getaddrinfo() with "any", which should prevent the operating system from sending the A and AAAA queries on the same port.

4.2.9. Multiple Interfaces

Interaction of the suggestions in this document with multiple interfaces, and interaction with the MIF working group, is for further study.

4.3. Content Provider Recommendations

Content providers SHOULD provide both AAAA and A records for servers using the same DNS name for both IPv4 and IPv6.

4.4. Security Considerations

[[Placeholder.]]

See Section 4.2.2.

4.5. Acknowledgements

The mechanism described in this paper was inspired by Stuart Cheshire's discussion at the IAB Plenary at IETF72, the author's understanding of Safari's operation with SRV records, Interactive Connectivity Establishment (ICE [RFC5245]), and the current IPv4/IPv6 behavior of SMTP mail transfer agents.

Thanks to Fred Baker, Jeff Kinzli, Christian Kuhtz, and Iljitsch van Beijnum for fostering the creation of this document.

Thanks to Scott Brim and Stig Venaas for providing feedback on the document.

4.6. IANA Considerations

This document has no IANA actions.

5. References

5.1. Normative References

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

5.2. Informational References

[DNS-middlebox] Various, "DNS middlebox behavior with multiple queries over same source port", June 2009, <https://bugzilla.redhat.com/show_bug.cgi?id=505105>.

[RFC2766] Tsirtsis, G. and P. Srisuresh, "Network Address Translation - Protocol Translation (NAT-PT)", RFC 2766, February 2000.

[RFC4074] Morishita, Y. and T. Jinmei, "Common Misbehavior Against DNS Queries for IPv6 Addresses", RFC 4074, May 2005.

[RFC4966] Aoun, C. and E. Davies, "Reasons to Move the Network Address Translator - Protocol Translator (NAT-PT) to Historic Status", RFC 4966, July 2007.

[RFC5245] Rosenberg, J., "Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator (NAT) Traversal for Offer/Answer Protocols", RFC 5245, April 2010.

[cx-osx] Adium, "AIHostReachabilityMonitor", June 2009, <https://bugzilla.redhat.com/show_bug.cgi?id=505105>.

[cx-win] Microsoft, "NetworkChange.NetworkAvailabilityChanged Event", June 2009, <http://msdn.microsoft.com/en-us/library/system.net.networkinformation.networkchange.networkavailabilitychanged.aspx>.

[whitelist] Google, "Google IPv6 DNS Whitelist", March 2008, <http://www.google.com/intl/en/ipv6>.

Authors' Addresses

Dan Wing
Cisco Systems, Inc.
170 West Tasman Drive
San Jose, CA 95134
USA

Email: dwing@cisco.com

Andrew Yourtchenko
Cisco Systems, Inc.
De Kleetlaan, 7
San Jose, Diegem B-1831
Belgium

Email: ayourtch@cisco.com

Internet Engineering Task Force G. Yang, Ed.
Internet-Draft L. Hu
Intended status: Informational J. Lin
Expires: April 23, 2011 China Telecom
October 20, 2010

IPv6 Transition Guide For A Large-scale Broadband Network
draft-yang-v6ops-v4v6tran-bb-transition-guide-01

Abstract

This document discusses about different IPv6 migrating solutions for each part of the Large-scale broadband network with layer 2 access infrastructure. They are based on the requirements of providing existing broadband services in v4v6-coexisting or IPv6-only scenarios. The document provides analysis of different solutions and also describes the suitable scenarios that each solution may be deployed in.

Table of Contents

1. Introduction

As we known, broadband subscriber is one of the largest parts of the Internet participants. It is significant to migrate them to IPv6, which will seem as an important step on IPv6 development. This document describes an IPv6 transition guide for a large-scale broadband network with Layer 2 accessing. And it will focus on the cases that the network infrastructure is large and widely covered, and the new subscriber's number is very large and is still increasing very fast.

In some cases, the broadband network is serving several dozen millions of subscribers with more than 20% annual increases in next few years. It is predicted that after the IPv4 addresses allocated by IANA are exhausted, the broadband users in these cases will still keep a high increasing rate, which will bring unprecedented pressure to not only the development of broadband services, but also the development of Internet.

Due to IPv4 addresses shortage, the network infrastructure and Internet services will no doubt to migrate to IPv6 eventually. And it is also our final goal. However, IPv6-based new services and applications are few and far between.

During the IPv6 transition, large-scale broadband network basically should take a smooth transition strategy because of the inactive IPv6 industrial chain. The first rule could be customer-oriented which means any changes to the network infrastructure should guarantee the users' experience. At the same time, the transition technology and strategy should be consistent with the future direction in order to protect the investments and maintain the network stability. And the technologies and solutions should be compatible with the existing broadband service access method and provisioning method.

This document is aimed to identify the pros and cons of all possible solutions in every part of the broadband network with considering its features. And it also provides the applicable scenarios for each solution.

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. Terminologies

Backbone network: Backbone network interconnects various regional broadband networks, providing a path for the exchange of information between different networks.

Regional Broadband Network: Regional broadband network interconnects the central offices in a geographical area.

L2 Access Network: L2 Access Network is the broadband access infrastructure which is a Layer 2 network.

Customer Premises Network: Customer Premises Network will contain one or more terminal equipment devices possibly interconnected by a customer premises network.

POP: Internet point of presence, POP, is the access point to the backbone network for regional broadband network.

MPLS PE routers: Provider edge router in a MPLS backbone network.

BRAS: Broadband Remote Access Server, BRAS, is the aggregation point for the subscriber traffic. It provides aggregation capabilities between the Access Network and the Metro Network. Beyond aggregation, it is also the injection point for access authentication, policy management and IP QoS.

CR: Core Router in a regional broadband network is the egress router of the regional broadband network and connecting to a POP of the backbone in upstream and connecting to BRASs for downstream.

SR: Service Router, SR, is the access nodes for different services providers (e.g. Internet Contents Provider).

AR: Aggregation Router is connected to CRs and provide traffic aggregation for BRASs in a large-scale regional broadband network.

DSLAM: Digital Subscriber Line Access Multiplexer, DSLAM, is the access node for Digital Subscriber Lines (DSL) subscribers.

OLT: An optical line termination (OLT) is a device which serves as the endpoint of a passive optical network (PON).

CPE: Customer Premises Equipment, CPE, is the edge of Customer Premises Network. In this document, there are two types of CPEs, Routing mode CPE and Bridging mode CPE.

User End: In this document, we consider the user end is a PC with a popular operating system like Windows, MAC OS, or Linux.

3. High Level Architecture

In this section, a High Level Broadband Architecture with layer 2 (L2) access network is shown in Figure 1. There are basically five parts in this architecture, Customer Premises Network, Layer 2 Access Network, Regional Broadband Network, Backbone and the Internet. We don’t discuss the physical layer infrastructures in this document. (e.g.Main Distribution Frame (MDF), and Optical Network Terminal (ONT)).

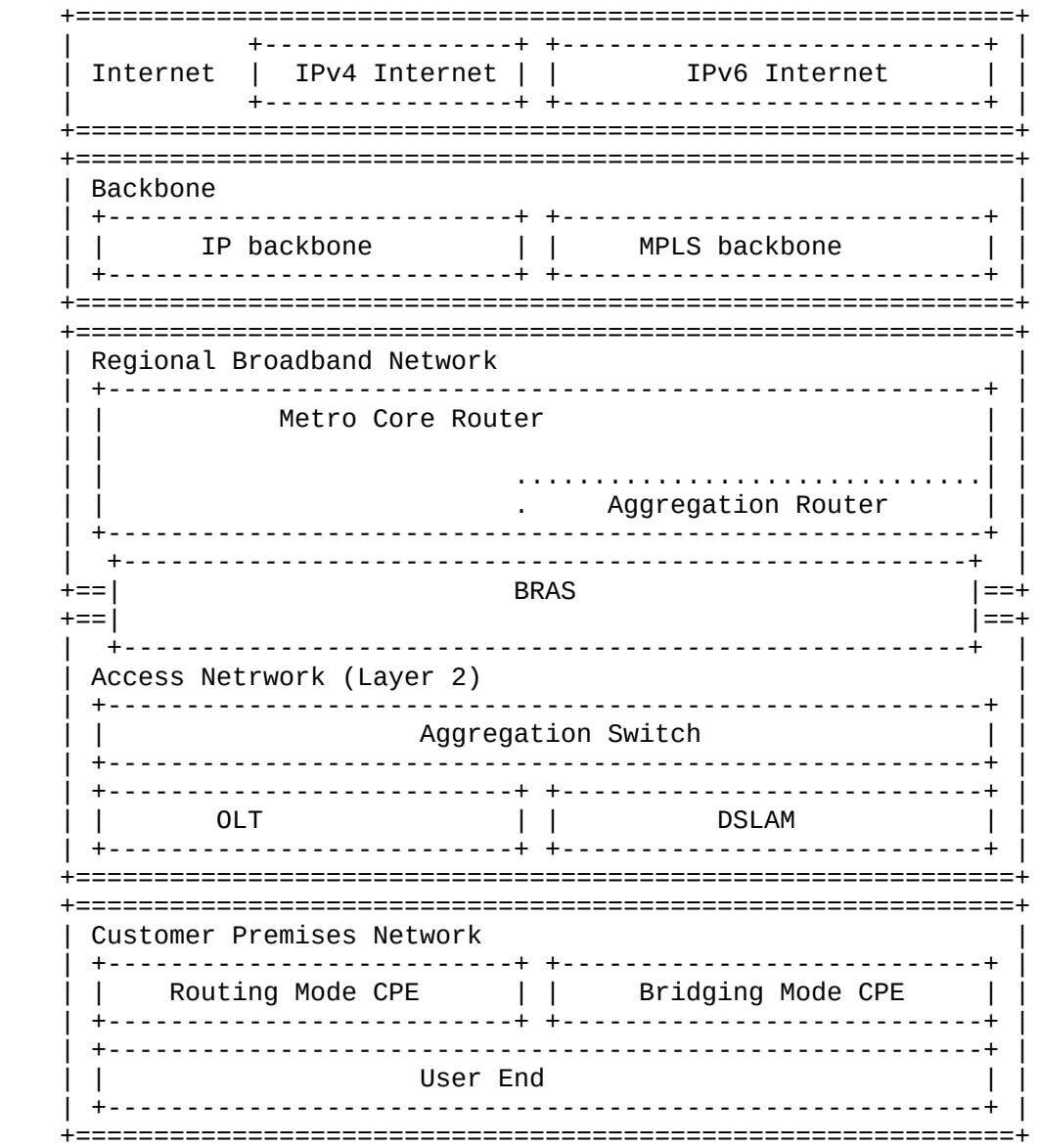


Figure 1: High Level Broadband Architecture with L2 Access network

In general, IP backbone and MPLS-enabled Layer 3 IP backbone (to simplify discuss, we call it "MPLS backbone" in this document) [RFC3031] are two major types of backbone for long distance transmission in the large scale broadband network. We classified two

types of MPLS backbone here, the combined backbone forwarding both the non-labeled packets by IP switching and the labeled packets by label switching (MPLS+IP backbone), or a MPLS backbone with label switching only.

In the Regional Broadband Network, Metro Core Router (CR) is connected to IP backbone, MPLS+IP backbone or both IP backbone and MPLS backbone. In most situations, Broadband Remote Access Server (BRAS) is directly connected to CR, while in some cases, BRAS could be connected to an Aggregation Router (AR) that connected to CR. Service Router (SR) is basically the access nodes for different services. As this document is focus on the IPv6 transition of broadband service, it does not discuss about the transition of SR.

BRAS acts as the aggregation point for the subscriber traffic. It provides aggregation capabilities between the L2 Access Network and the Regional Broadband Network. Beyond aggregation, it is also the injection point for access authentication, policy management and IP QoS.

The access network in this architecture is a L2 network. It is from the BRAS to the Customer Premises Equipment (CPE) located at the edge of customer premises network. The most popular access method in this architecture currently could be PPPoE [RFC2516]. In theory, the devices in the access network have no needs to be IPvX protocol aware.

Note that although the IPoE access method may be using the same L2 access network, the discussion of IPv6 transition with IPoE is out-of-scope in this document. And it does not consider the transition from a layer 2 access network to a layer 3 one as well.

This document will focus on the IPv6 transition of the Backbone Network and the Regional Broadband Network in the architecture above [Figure 1]. And it will also discuss how to provide IPv6 service for broadband subscribers in such kind of architecture.

4. Overview of Solutions

This document describes the IPv6 transition solutions and related technologies for the Use Case for Large-Scale Broadband Network. [I-D.huang-v6ops-v4v6tran-bb-usecase] By analysing the features of the case, The following factors make the networks' and services' IPv6 transition complicated and difficult:

- Large number of broadband subscribers and their terminals with diverse IPv6 capabilities;

- Large number of network devices with diverse IPv6 capabilities;
- Various types of the Internet services and applications have diverse IPv6 capabilities and they will not migrated to IPv6 synchronized.

During the IPv6 transition, the user experience should be guaranteed. So, it is important to take the terminal into consideration besides the networking issues.

Moreover, in order to migrate both of the network infrastructures and the Internet services smoothly, it is significant to select the proper technologies at each point of time on the IPv6 roadmap according to different network scenarios.

Because the global IPv4 addresses is depleting, and most of the Internet contents and applications are not ready yet, carrier graded NAT44 technologies and Large Scale NAT devices (LSN) may be deployed in the network during the initial stage of the IPv6 transition. So, the issues that are brought from NAT44 technology itself and the interoperating with other IPv6 transition technologies should be considered.

5. Transition For the Backbone Network

According to the architecture in Figure 1 above, there are three possible backbones in a large-scale broadband network:

- There is an IP backbone only;
- There is a MPLS+IP combined backbone;
- There is an IP backbone and a MPLS backbone.

The discussion on backbone will focus onthe scenario that there is an IP backbone and a MPLS backbone separately. When there is an IP backbone only, it can use the Dual-stack solution or the IPv6 solutions. And when there is a MPLS+IP combined backbone, the solutions are similar. It can use the Dual-stack solution or the IPv6 solutions with IP-based switching; alternatively, it use the 6PE solution with labeled switching for the IPv6 traffic.

Basically, there are three main ways for the transition of backbone network:

- Dual-stack IP Backbone

- o IPv6-Only Backbone
- o 6PE in MPLS Backbone

5.1. Dual-stack IP Backbone

The Dual-stack IP Backbone Solution includes two implementation options, building a completely new Dual-Stack IP backbone or upgrading the existing routers in the IP backbone to Dual-Stack by enabling IPv6. Technically, the main idea is carrying both the IPv6 and IPv4 traffic on one backbone. Except the management issue and engineering cost, there may be little difference on technical aspect. The upgrade implementation could be incrementally to reduce the risk on each Internet point of presence (POP), but it can not avoid the risk on provider (P) routers. In the new-built implementation, the new POPs and P routers are sperately with the existing backbone. Therefore, the risk of upgrading the P routers, which is considerable high, will be avoided. The upgrade implementation will be much more complicated, and this section will focus on it.

Upgrading the existing backbone to a Dual-stack IP backbone requires enable IPv6 on the all the routers and support both IPv4 and IPv6 routing protocols. In some cases, the routers may upgrade to a new software and hardware version to support a better performance and functionality. So, the changes will contains two parts, devices upgrade and reconfiguration. Figure 2 presents the architecture after the upgrade implementation.

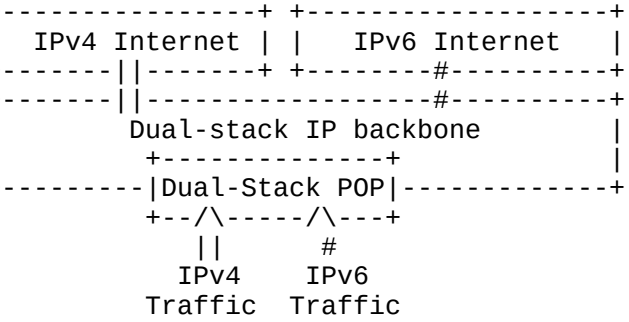


Figure 2: Dual-stack IP Backbone

Generally, the pros and cons of this solution are:

Pros:

- o According the the use CASE [I-D.huang-v6ops-v4v6tran-bb-usecase], most of existing routers have IPv6 capability already (but not

enabled). To support IPv6, it only needs to make some configurations or upgrade the software version. It does not need the extra engineering cost for the new infrastructure. It is no need to build or expand the existing facilities like power, air conditioning and transmission infrastructures, which may take a long engineering period.

- The existing IP backbone is usually covering a widely area already. So, this solution is flexible and can conduct a rapid deployment of IPv6 services anywhere it covered.

- The upgraded IP backbone is compatible with the existing IPv4 traffic and the new IPv6 traffic. There is no extra new backbone which will lead a large amount of extra management cost.

- The upgraded dual-stack backbone can be upgraded to IPv6-only by turning off the IPv4 after the IPv4 traffic is disappeared.

Cons:

- Until now, the routers in the IP backbone that supported Dual-stack will usually route IPv4 and IPv6 traffic separately based on the IPv4 routing table and the IPv6 one respectively. The router may have challenges for the performance and stability after they are upgraded to dual-stack, for example, the size of routing table, routing lookup/forwarding capability and routing convergence capability due to the sharing of resources.

- There is lack of technical and management experience of large-scale changing in a high volume traffic backbone, even though the change is very little.

- There is a high risk to upgrade the P router to dual-stack because of its high volume traffic. Any fault (hardware or software) may lead a significant impact to the existing services. And these impacts is difficult to predict.

5.2. IPv6-Only Backbone

It seems impossible to upgrade the existing IPv4 backbone to a native IPv6 backbone. This section will discuss a solution of building a new IPv6-only backbone network and keeping the original IPv4 infrastructure unchanged. There is only IPv6 routing in the new backbone, and the IPv4 traffic will be kept going through the legacy IPv4 backbone. Figure 3 presents the architecture of co-existing of IPv4 backbone and IPv6 backbone.

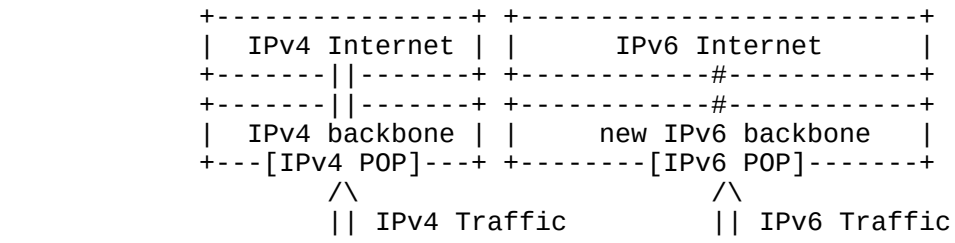


Figure 3: New IPv6 Backbone

Pros:

- o In line with the future network; It is a one-step solution and no need the second step which will also brings the risks (e.g. dual-stack backbone upgrade to IPv6 only);
- o Nearly with no impact on the existing IPv4 backbone and the services on it;
- o Simple to maintain two physically separated infrastructures compared with a complex dual-stack network with two logical network.

Cons:

- o The cost of building a new backbone is considerable high and the engineering cycle could be very long.
- o If the IPv6 services are still very few, the subscribers' IPv4 traffic will not be forwarded to the new IPv6 backbone. The inefficiency of the new IPv6 backbone could be a waste. Moreover, there may be a separate network operation and management cost for the new backbone.

5.3. 6PE on MPLS Backbone

The Multiprotocol Label Switching (MPLS) [RFC3031] is a popular networking technology that forwards packets by label switching instead of by IP switching. In this solution, the provider edge (PE) routers are dual-stack. The egress routers of the regional broadband network are connected to the PE router via a normal interface. The IPv6 routing distribution between two IPv6 enabled PE routers is done via Multiprotocol iBGP (MP-iBGP). The iBGP sessions distribute the IPv6 prefixes and the associated MPLS label. This is known as IPv6+ label and is encoded according to [RFC3107]. The communication of IPv6 is achieved by the label switched path (LSP) among PE

routers.Figure 4 presents the architecture of 6PE solution over a MPLS backbone.

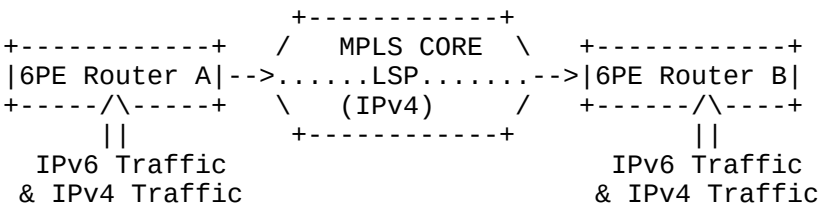


Figure 4: 6PE in MPLS backbone

Pros:

- 6PE technology [RFC4798] is relatively mature compared to other tunnel technology in backbone.

- There is no need to make changes on P routers which the LSP goes through. Only the PE router connecting to IPv6-enabled networks needs to implement Dual-stack and make some corresponding configurations for 6PE. The reengineering cost and risk of this kind of changes is comparable low.

- Little impact to the existing services: There is little impact to the existing services. It is similar to the existing MPLS network is carrying a new service with a new label.

- According to the use case [I-D.huang-v6ops-v4v6tran-bb-usecase], the MPLS backbone covers a widely area already which means it can provide IPv6 servces with a rapid deployment when there is an IPv6 demand in some regional broadband networks. Therefore, this solution is flexible and supporting incremental deployment.

Cons:

- This solution changes the original designed purpose of the MPLS network which is normally used to carry VPN traffic and usually light load. 6PE brings the public traffic in to the MPLS infrastructure. When the this kind of traffic grows, there may be significant to the existing services.

- Unable to deploy QoS policy for IPv6 traffic.

- It may bring some inconvenient for troubleshooting.

5.4. Conclusion

The 6PE solution may be applicable in the IPv6 initial stage while the most traffic is still IPv4 in the backbone and there is a demand for the rapid deployment of IPv6 service.

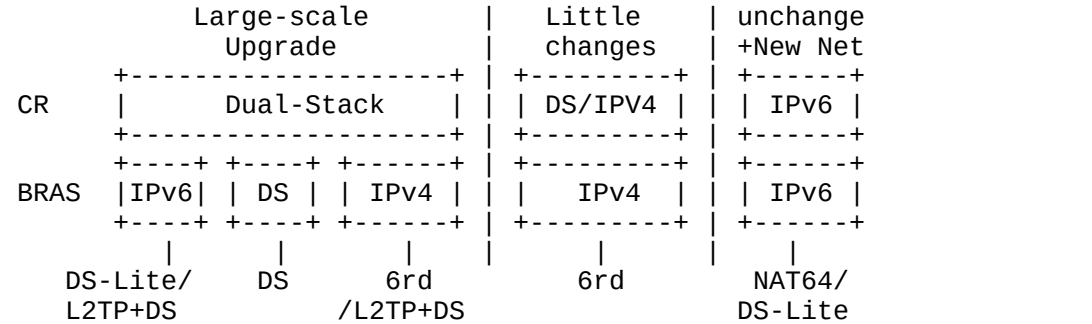
The Dual-stack solution may be applicable to the intermediate stage when IPv6 traffic is relatively large. And for the network devices, the Dual-stack capability, performance, and stability need to be reasonable high enough to support two IP stacks.

The native IPv6 solution may be suitable to the latter phase of IPv6 transition with most of the services being IPv6 capable. It can also be upgraded from the Dual-stack backbone by turning off the IPv4 after the IPv4 traffic is disappeared.

6. Transition of Regional IP Network

According to the use case [I-D.huang-v6ops-v4v6tran-bb-usecase], The Overview of the solutions in the Regional Broadband Network can be summarized into the following three types:

- Providing IPv6 service by Large-scale upgrading the existing regional broadband network infrastructure;
- Providing IPv6 service by little changes on the existing regional broadband network infrastructure;
- Providing IPv6 service by building a completely new IPv6 regional broadband network infrastructure;



Note: "DS" stands for "Dual-Stack".

Figure 5: Overview of Solutions in Regional Broadband Network

In each transition solution of regional broadband network, it can connect to one of the following backbone described in section 3.1??:

- Connect to the Dual-stack IP backbone;
- Connect to the IPv6-only backbone for IPv6 traffic and to the existing IP backbone for IPv4 traffic if it has;
- Connect to the 6PE on the MPLS backbone for IPv6 traffic and to the existing IP backbone for IPv4 traffic if it has.

Some less possible transition solutions haven't been listed above:

- Upgrade the existing regional broadband network to IPv6-only; It will lead to a huge influence to existing network and services.
- Create a new regional broadband network with native IPv6 CRs and Dual-Stack BRASs; It has very low possibilities because if we create a new regional broadband network to provide dual-stack service with new dual-stack BRAS, the simplest solution will be let the new CRs to be dual-stack too. If the new CRs are IPv6-only, they need other transition technologies working together which seem to be more complicated.
- Create a new regional broadband network with Dual-stack CRs and native IPv6 BRASs; It also has very low possibilities and the reason is same as the above one.

In the following sections, the technical solutions based on the scenarios in Figure 5 are discussed. Although there may be many technical options in each scenario, the discussion will focus on one of them.

The possible solutions referred to Figure 5 that we will discuss:

- Solution 1: Dual-Stack and L2TP
- Solution 2: Dual-Stack over IPv6 - DS-lite
- Solution 3: Dual-Stack over IPv4 - 6rd
- Solution 4: IPv6 and NAT64

In this document, it is considering that the CPE is basically purchased by customers. In the PPPoE dial-up cases, most users dial-up from PC, but there is some deployed a Home Gateway (e.g.

WLAN AP) by themselves and set up an automatically dial-up from it. Until now, most terminals, including PCs and CPEs, will still be IPv4-only. Although most PC operating system (OS) declared that they already supported IPv6, there is still a problem on supporting PPPoE with IPv6. Not only the most widely used OS, Windows(TM) XP, doesn't support PPPoE with IPv6, but also nearly all CPEs in the market does not support this function. This problem will be a significant bottleneck of the development of IPv6 broadband with PPPoE access method.

6.1. Dual-Stack and L2TP

In this solution, both the CRs and BRASs will be transition to Dual-stack by upgrading or replacing the existing devices. However, there are so many different BRASs with diverse IPv6 capability in a large-scale broadband network. So there is a possibility that some BRASs cannot upgrade to Dual-stack and support PPPoE with IPv6.

In the Figure 6 below, there are two scenarios in this solution.

- Scenario 1: A Dual-stack or IPv6 terminal accessing to a Dual-stack BRAS.
- Scenario 2: A Dual-stack or IPv6 terminal accessing to a legacy IPv4-only BRAS.

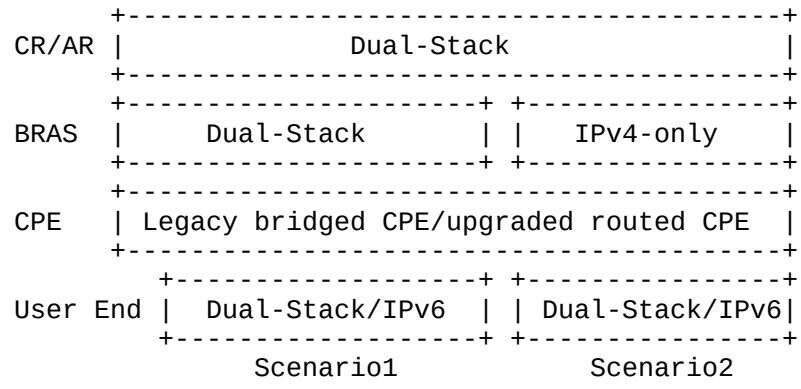


Figure 6: Dual-stack Transition Solution

The Scenario 1 is very simple. But the routing CPE at the edge of customer premises network need to be upgraded to support IPv6 and PPPoE with IPv6. And the PC operation system (OS) also need to

support PPPoE with IPv6.

The Scenario 2 is a little bit complicated. The BRAS which the subscriber is connecting to is not support IPv6 and PPPoE with IPv6. So, one possible solution could be terminating the point-to-point protocol (PPP) [RFC1661] link at a remote Dual-stack BRAS. A tunnel technology like Layer Two Tunneling Protocol (L2TP) [RFC2661] can be used in this scenario. Other technologies could be an alternative. But considering the device capabilities and the maturity of the technology, the following discussion will focus on the solution that Dual-stack network with L2TP to provide Dual-stack services. [SeeFigure 7]

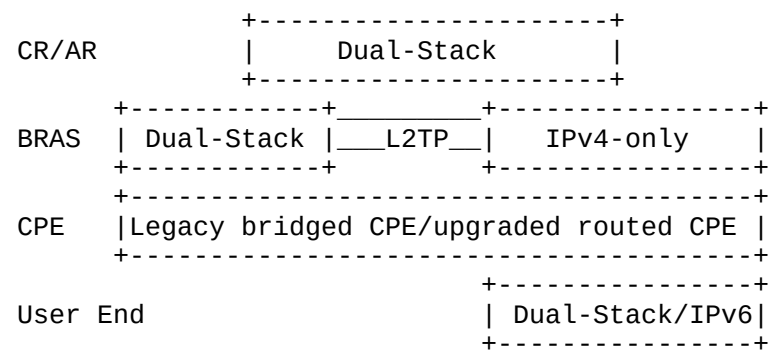


Figure 7: The L2TP Solution in partly Dual-Stack network

Although tunnel technologies can solve this problem, it is considered as a temporary solution. The legacy IPv4-only BRASs will be replaced eventually.

For the Dual-stack service, IPv4 address is still need to allocate to terminal. After the IPv4 addresses exhaustion, Dual-stack BRASs could allocate private IPv4 addresses for broadband subscribers, and a NAT44 Large Scale NAT (LSN) [I-D.kuarsingh-lsn-deployment] device will be deployed to provide IPv4 NAT services for subscribers who are using private IPv4 addresses.

The operating system (OS) of the new subscriber is recommended to support PPPoE with IPv6. Third-party dial-up software could be provided if the OS is not support PPPoE with IPv6.

The routing mode CPE of new subscriber is required to support PPPoE with IPv6. Otherwise, they are required to turn off the auto-dialup function, and initial the PPPoE dial-up session from the host that supports PPPoE with IPv6.

The legacy subscribers are recommended to upgrade their OSs and CPEs, but not required. They can still access by IPv4-only. Third-party dial-up software could also be provided to support PPPoE with IPv6.

The benefits and drawbacks for this solution could be:

Pros:

- L2TP can be a temporary solution to provide Dual-stack services with fast and incremental deployment. The BRASs that are unable to upgraded to Dual-stack can be replaced at the end of its lifecycle.

- When the IPv4 traffic disappears in the future, the network could be migrated to native IPv6 network gradually.

- There is no need to change the existing access method. Although many existing routing mode CPEs are not supporting auto-dialup via PPPoE with IPv6, they can be replaced by existing subscribers smoothly or waiting for new technologies because the network still providing IPv4 service. When the IPv6 contents are abundant enough, legacy subscribers would like to replace or upgrade their CPEs and OSs to gain more services.

- Although NAT44 technology is needed after the IPv4 address exhaustion, NAT[RFC3022] is relatively mature compared with IPv4/IPv6 Translation (e.g. stateful NAT64 [I-D.ietf-behave-v6v4-xlate-stateful], or stateless NAT64 [I-D.xli-behave-ivi]. The major existing applications, such as Instant Messengers, E-mail Terminals and P2P downloaders are already supporting NAT traverse. Transitioning with providing Dual-stack service is much smoother than providing IPv6-only service, especially at the initial stage of the IPv6 transition.

- There is no extra cost on the network operation and management. There is still only one physical network in each regional area and the operation and management team can use the original one, though some compulsory training is needed.

Cons:

- The requirment is that there is at least one Dual-Stack BRAS in the network. And if the BRASs that cannnot support Dual-stack is the majority, when the Dual-stack/IPv6 subscribers grows there could be many L2TP tunnels across the network and the traffic load among BRASs is not balanced which may impact the customer experience. Incremental replacement of the old BRAS should be considered.

o Although it is the same issue for any protocol translation technology, some existing applications which do not consider NAT44 traverse may have some problems after the deployment of NAT44 LSN. For example, after the deployment of NAT44 LSN, the service of PPTP VPN has malfunction. Parts of P2P users have worse experience.

Applicable scenarios: This solution could be suitable for the initial stage or the intermediate stage of the IPv6 transition when the IPv4 traffic is still very large in the network. And the broadband network is going to provide Dual-stack services with incremental deployment. It is also suitable when the number of subscribers is increasing very fast, and there is a large amount of CPEs and OSs that do not support PPPoE with IPv6.

6.2. Dual-Stack over IPv6 - DS-lite

In this solution, the CRs in the regional broadband network are Dual-stack or IPv6-only and the BRASs are IPv6-only. This network is providing a Dual-stack service or an IPv6-only service for subscribers. [See Figure 8]

o A Dual-stack or IPv6 terminal accessing to an IPv6-only BRAS.

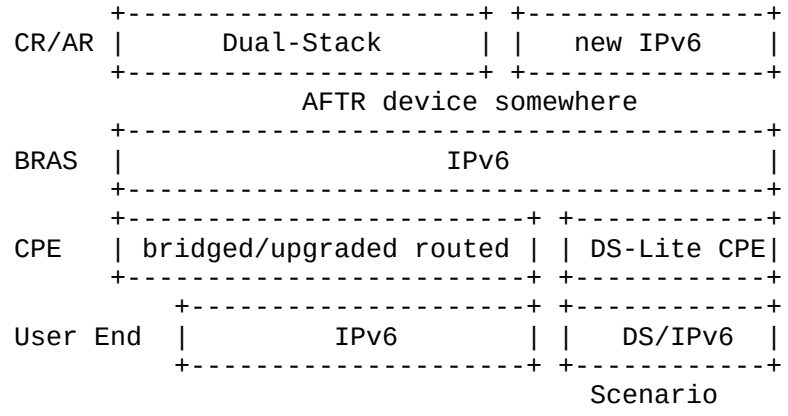


Figure 8: The DS-Lite Solution in IPv6 Infrastructure

The Scenario for IPv6-only subscriber accessing a IPv6 BRAS is very simple. But the routed CPE at the edge of customer premises network needs upgrade to support IPv6 and PPPoE with IPv6. And the PC operation system (OS) also needs to support PPPoE with IPv6 with a

bridged CPE. This scenario will exist when the IPv6 traffic is already dominant in the network. The little IPv4 traffic will be translated by a NAT64 device located at the edge of IPv6 Ocean. This situation is similar to the solution in Section 6.4

The Scenario for Dual-stack subscriber is a little bit complicated. It provides Dual-stack service over an IPv6-only infrastructure. The technologies like DS-Lite [I-D.ietf-softwire-dual-stack-lite] can be deployed in this scenario. This section will focus on this technology.

DS-Lite is a tunnel technology with a point-to-multipoint IPv4-in-IPv6 tunnel between B4 element and AFTR. According to the definition in [I-D.ietf-softwire-dual-stack-lite], the B4 element is a function implemented on a dual-stack capable node, either a directly connected device or a CPE, which creates a tunnel to an DS-Lite Address Family Translation Router (AFTR) deployed somewhere in the regional broadband network.

The operating system (OS) of the new subscriber is recommended to support PPPoE with IPv6. Third-party dial-up software could be provided if the OS is not support PPPoE with IPv6. Subscribers need to replace the existing CPEs for DS-Lite services.

The pros and cons of the DS-Lite solution will be:

Pros:

- We are assuming a completely IPv6 scenario, it is a one-step solution of IPv6 transition. The network infrastructure does not need to upgrade to native IPv6 network in the future.

- AFTR is performing a NAT [RFC3022] behavior, which is relatively mature compared with IPv4/IPv6 Translation. The major existing applications, such as Instant Messengers, E-mail Terminals and P2P downloaders are already supporting NAT traverse. Transitioning with providing Dual-stack service is much smoother than providing IPv6-only service, especially at the initial stage of the IPv6 transition.

- The IPv4 address used in DS-lite does not need to planned. It is easy for operation and management.

Cons:

- DS-Lite solution seems not suitable for the transition of the existing network that contains thousands of IPv4-only BRASs. Because the first step should be upgrade the existing BRAS to at

least Dual-stack. if the BRASs is upgraded to Dual-stack, DS-Lite does not need any more. Moreover, if the CRs are Dual-stack, and it is planned to add new BRASs to provide Dual-stack service, the simplist solution is let the new BRASs to be Dual-stack. So, DS-Lite solution is only suitable for the new build IPv6 network scenario.

- o DS-lite CPEs are needed to provide to subscribers, and there is no mature product until now. It may change the original access method and service provisioning method. And the extra installment cost is unacceptable when the number of subscribers is huge.

- o DS-Lite has not been deployed and verified in any large-scale commercial trail. In the regional broadband network with a large number of dual-stack subscribers, a number of DS-Lite AFTR devices with high performance are needed. The reengineering cost for this solution may be very high.

Applicable scenarios: This solution is suitable for the scenarios that providing dual-stack services over an IPv6-only network infrastructure when IPv6 traffic is already dominant in the network. It is also suitable when the broadband subscribers are increasing slowly and the CPEs are provided by operators. Some IPv6-only BRASs could be added for these new subscribers. It is also suitable for the new services which may using IPv6-only, for example, IPTV and Machine-to-machine (M2M) services.

6.3. Dual-Stack over IPv4 - 6rd

In this solution, the CRs in the regional broadband network are IPv4-only or Dual-stack and the BRASs are IPv4-only. It provides a Dual-stack service or an IPv6-only service with a completely/partly IPv4 infrastructure for subscribers.

The discussion will focus on scenario in Figure 9.

- o An IPv6 or Dual-stack terminal accessing to an IPv4-only BRAS.

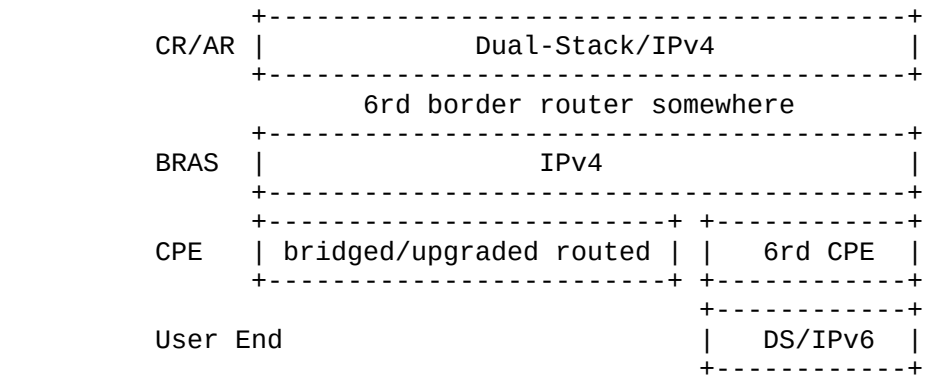


Figure 9: The 6rd Solution in an IPv4 infrastructure

A possible technical solution for this scenario is IPv6 Rapid Deployment (6rd) [RFC5969]. There are two components in this solution. 6rd CPEs support IPv6 on their customer premise side and support 6rd on the provider side. 6rd gateway (a.k.a 6rd border router or 6rd relay) is operated at the border between IPv4 infrastructure and the IPv6 Internet. The 6rd mechanism operates statelessly, which ensures simplicity and scalability. The IPv4 address in the IPv4 infrastructure could be a private address, 6rd mechanism can support the private IPv4 address.

The pros and cons of the 6rd solution will be:

Pros:

- o 6rd solution does not need to upgrade the IPv4 infrastructure. It can deploy incrementally by adding some 6rd gateways in the network. Therefore the engineering complexity and cost is low compared with other solutions.

- o Although NAT44 technology is needed after the IPv4 address exhaustion, NAT[RFC3022] is relatively mature compared with IPv4/ IPv6 Translation. The major existing applications, such as Instant Messengers, E-mail Terminals and P2P downloaders are already supporting NAT traverse. Transitioning with providing Dual-stack service is much smoother than providing IPv6-only service, especially at the initial stage of the IPv6 transition.

- o There is no extra cost on the network operation and management. There is still only one physical network in each regional area and the operation and management team can use the original one, though some compulsory training is needed.

Cons:

- When the network infrastructure is transitioning to IPv6 in the future, the access method may need to be changed again. 6rd is not applicable when the infrastructure is IPv6-only in the future. When the BRASs are removed by nature, the new BRASs for replacement will enable IPv6 and the 6rd may be useless.

- 6rd CPE need to be provided to subscribers, which will lead to a huge amount of devices cost and installment cost. And when the IPv6 traffic is extremely high, each regional broadband network may need a number of 6RD border routers to ensure the performance.

Applicable scenarios: This solution may be suitable for the initial stage of the IPv6 transition, providing IPv6 services with rapid deployment.

6.4. IPv6 and NAT64

This solution is for the IPv6-only subscribers that are accessing to the new built IPv6-only regional broadband network. Basically, only IPv6 address is allocated to the subscribers. And for the requirement ofaccessing IPv4 applications and contents, it needs to deploy a IPv6/IPv4 Translation device to solve the intercommunication problem between IPv6 and IPv4 for the IPv6-only subscribers.

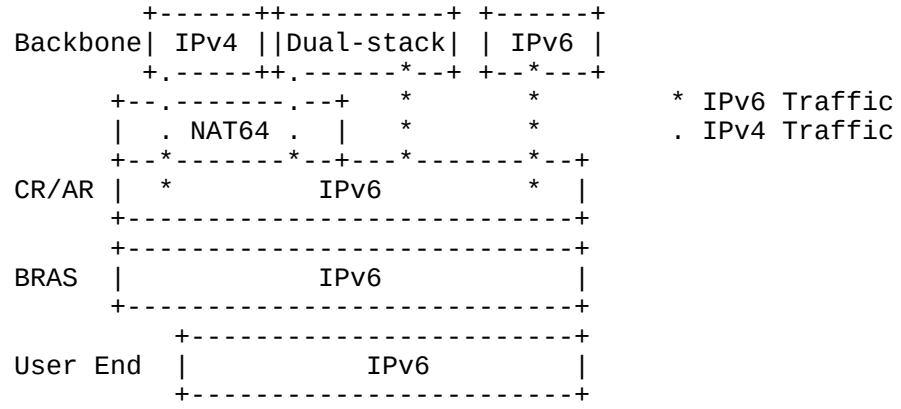


Figure 10: The NAT64 Solution in an IPv6 infrastructure - Scenario 1

The operating system (OS) is required to support PPPoE with IPv6. Third-party dial-up software could be provided if the OS of the new hosts is unable to support PPPoE with IPv6.

The routing mode CPE that is purchased by subscribers is also required to support PPPoE with IPv6 as well, or to turn off the auto-dialup function, and initial the PPPoE with IPv6 dial-up session from the host.

Pros:

- This solution is usually building a new IPv6 network which could avoid the risk from changing the existing regional broadband network.
- Building a completely new network is much easier than upgrade from the existing one. That is because there is no subscriber with no traffic on the new network.
- It is no need to allocate IPv4 address for subscribers.
- It is benefit the development of IPv6 and push the IPv6 transition of ICPs.

Cons:

- All the user terminals including CPEs have to support IPv6 which is unpractical at the initial stage of IPv6.
- NAT64 mechanisms have not been deployed and verified in large scale commercial trails. And the NAT64 technologies are still immature, the users experience with this solution could be worse.
- The requirement of NAT64 device performance is very high, especially when there is a large amount of subscribers. The situation could be much worse because the IPv4 traffic is still the majority at the IPv6 initial stage.
- The cost is huge and the investment is duplicated with the existing one. The existing network infrastructure is usually kept up-to-date. Building a new network may lead to an early end of the lifecycle of the existing network infrastructure, which will lead to investment loss.
- Building a completely new regional broadband network is usually along with building a new transmission infrastructure. And the engineering period will be very long.
- After the new network is built, there are two networks in the same area, which will lead an extra operational cost.

Applicable scenarios: This solution is suitable for the last-step of

the IPv6 transition. The IPv6 contents and services are already very popular and abundant. Besides, IPv6/IPv4 Translation technology is relatively mature and the IPv4 traffic is little. In this case, we can disable the IPv4 protocol stack of the Dual-stack devices, and NAT64 devices can also be deployed at several sites to meet the requirement of IPv6-only users who are visiting the historical IPv4 services.

7. Backwards Compatibility

8. Conclusions

TBD...

9. Acknowledgements

The authors would like to acknowledge the discussions on this topic with Dave Thaler, Denis Alexeitsev, Jari Arkko, R"|mi Despr"|s, Tina TSOU, Tom Taylor and Tony Li.

10. IANA Considerations

This memo includes no request to IANA.

11. Security Considerations

The IETF is specifying security considerations for the solutions that it is providing for IPv6 migration. However, it is possible that additional considerations arise due to the interoperation of these solutions, and the fact that the network is in a transitional state.

12. References

12.1. Normative References

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

[min_ref] authSurName, authInitials., "Minimal Reference", 2006.

12.2. Informative References

[I-D.huang-v6ops-v4v6tran-bb-usecase]
Huang, C., Li, X., and L. Hu, "Use Case For IPv6 Transition For a Large-Scale Broadband network", draft-huang-v6ops-v4v6tran-bb-usecase-00 (work in progress), September 2010.

[I-D.ietf-behave-v6v4-xlate-stateful]
Bagnulo, M., Matthews, P., and I. Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", draft-ietf-behave-v6v4-xlate-stateful-12 (work in progress), July 2010.

[I-D.ietf-softwire-dual-stack-lite]
Durand, A., Droms, R., Woodyatt, J., and Y. Lee, "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", draft-ietf-softwire-dual-stack-lite-06 (work in progress), August 2010.

[I-D.kuarsingh-lsn-deployment]
Kuarsingh, V. and J. Cianfarani, "NAT44/LSN Deployment Options and Experiences", draft-kuarsingh-lsn-deployment-00 (work in progress), July 2010.

[I-D.xli-behave-ivi]
Li, X., Bao, C., Chen, M., Zhang, H., and J. Wu, "The CERNET IVI Translation Design and Deployment for the IPv4/IPv6 Coexistence and Transition", draft-xli-behave-ivi-07 (work in progress), January 2010.

[I-D.zhou-softwire-ds-lite-p2p]
Zhou, C., ZOU), T., Lee, Y., and G. Yang, "Deployment DS-lite in Point-to-Point Access Network", draft-zhou-softwire-ds-lite-p2p-02 (work in progress), July 2010.

[RFC1661] Simpson, W., "The Point-to-Point Protocol (PPP)", STD 51, RFC 1661, July 1994.

[RFC2516] Mamakos, L., Lidl, K., Evarts, J., Carrel, D., Simone, D., and R. Wheeler, "A Method for Transmitting PPP Over Ethernet (PPPoE)", RFC 2516, February 1999.

[RFC2629] Rose, M., "Writing I-Ds and RFCs using XML", RFC 2629, June 1999.

[RFC2637] Hamzeh, K., Pall, G., Verthein, W., Taarud, J., Little, W., and G. Zorn, "Point-to-Point Tunneling Protocol", RFC 2637, July 1999.

[RFC2661] Townsley, W., Valencia, A., Rubens, A., Pall, G., Zorn, G., and B. Palter, "Layer Two Tunneling Protocol "L2TP"", RFC 2661, August 1999.

[RFC3022] Srisuresh, P. and K. Egevang, "Traditional IP Network Address Translator (Traditional NAT)", RFC 3022, January 2001.

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.

[RFC3107] Rekhter, Y. and E. Rosen, "Carrying Label Information in BGP-4", RFC 3107, May 2001.

[RFC3376] Cain, B., Deering, S., Kouvelas, I., Fenner, B., and A. Thyagarajan, "Internet Group Management Protocol, Version 3", RFC 3376, October 2002.

[RFC3552] Rescorla, E. and B. Korver, "Guidelines for Writing RFC Text on Security Considerations", BCP 72, RFC 3552, July 2003.

[RFC3697] Rajahalme, J., Conta, A., Carpenter, B., and S. Deering, "IPv6 Flow Label Specification", RFC 3697, March 2004.

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

[RFC4798] De Clercq, J., Ooms, D., Prevost, S., and F. Le Faucheur, "Connecting IPv6 Islands over IPv4 MPLS Using IPv6 Provider Edge Routers (6PE)", RFC 4798, February 2007.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.

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

Authors' Addresses

GuoLiang Yang (editor)
China Telecom
109, Zhongshan Ave. West,
Guangzhou, Tianhe District 510630
P.R. China

Phone:
Email: yanggl@gsta.com

LeMing Hu
China Telecom
109, Zhongshan Ave. West,
Guangzhou, Tianhe District 510630
P.R. China

Phone:
Email: hulm@gsta.com

JinYan Lin
China Telecom
109, Zhongshan Ave. West,
Guangzhou, Tianhe District 510630
P.R. China

Phone:
Email: jasonlin.gz@gmail.com

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: April 22, 2011

C. Zhou, Ed.
T. Taylor
Huawei Technologies

October 19, 2010

IPv6 Transition Use Case For a Large Mobile Network
draft-zhou-v6ops-mobile-use-case-00

Abstract

This document describes an use case for migrating from IPv4 to IPv6 in a very large mobile network. Its purpose is to enhance general understanding of the challenges associated with the migration of such a network to IPv6 operation.

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 22, 2011.

Table of Contents

1. Introduction

Consider a major mobile network operator, with hundreds of millions of subscribers, currently growing at a rate in the order of 1% per month. To assure continued revenue growth as market penetration approaches its limit, the operator has been deploying 3GPP technology. Currently about 1.5 percent of the operator's subscribers use the third and fourth generation technology discussed in this document.

The operator is looking to mobile data services for future revenue gains. Mobile data services currently include mobile payments, music downloads, mobile reading (book downloads), streaming and broadcast video, and internet search services in partnership with content providers such as news agencies. By their nature, these services require communication between the mobile subscriber and a third party application or content provider. Given the importance of these third parties to the operator's business, the operator's IPv6 transition plans have to ensure continuity of service to the third party servers regardless of the IP version they run.

At the subscriber end, mobile handsets are typically replaced within two to three years after purchase, apparently putting an upper limit on how long it will take to make IPv6 the preferred protocol for the majority of subscribers. However, in some markets the most popular use of mobile data access is to provide access for personal computers attached to the mobile terminal. This means the transition period at the subscriber end depends to an important extent on the rate at which personal computer operating systems and applications evolve to support IPv6.

As a further complication for the migration to IPv6, the operator is facing a major upgrade of its access networks from the older 3GPP technology to LTE ("Long Term Evolution"). LTE flattens out the access network by bringing the IP edge closer to the user equipment. LTE will provide higher data rates, opening up the possibilities for improved services and increased revenue from them.

1.1. Requirements Language

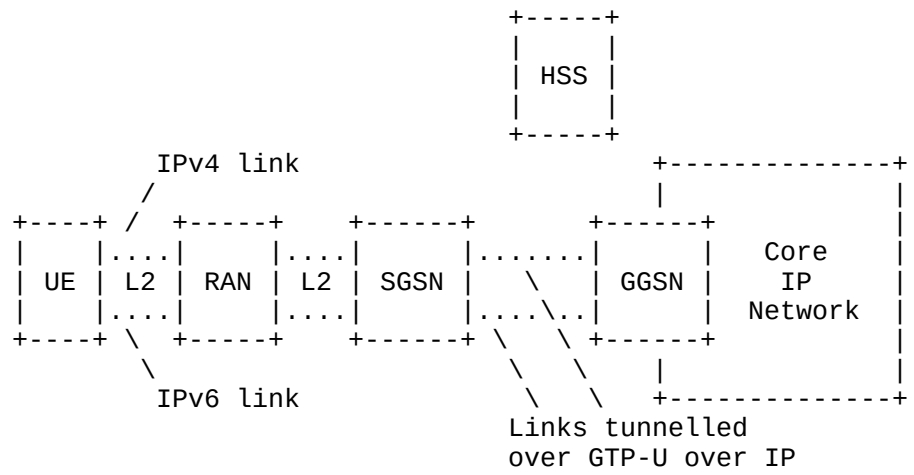
This document is descriptive, and as such, contains no requirements language.

2. Overview of the Mobile Network Architecture

3GPP has specified layer 2 access for IPv6 in the legacy 3GPP architecture and in the LTE ("Long Term Evolution") version. The

third generation architecture is shown in Figure 1. Only the user data paths are shown. The IP edge (of the core IP network) is located at the GGSN (Gateway GPRS Support Node, where GPRS itself stands for General Packet Radio Service). Within this system, IPv6 support requires separate bearers (i.e., links) for IPv4 and IPv6 respectively, extending from the User Equipment through the radio network and the SGSN (Serving GPRS Support Node, at which the User Equipment is registered) and, finally, to the GGSN.

The bearers are actually propagated from the SGSN to the GGSN using GTP-U, a 3GPP-specified tunneling protocol running over IP. The IP addresses assigned to the SGSN and GGSN for this purpose are not visible to the mobile station. The SGSN locates the GGSN using a DNS service that is also not visible to the User Equipment.



UE = User Equipment
RAN = Radio access network
SGSN = Serving GPRS Support Node
GGSN = Gateway GPRS Support Node
HSS = Home Subscriber Server (holds subscriber profile)
GTP = GPRS Tunneling Protocol

Figure 1: Third Generation GPRS Mobile Access Network

The use of IPv4 and/or IPv6 is controlled by the combination of subscriber profile and core operator preference. Address allocation to the User Equipment (UE) differs between IPv4 and IPv6. For IPv4, addresses can be assigned in a number of ways. From the point of view of the UE, the choice is between allocation during bearer

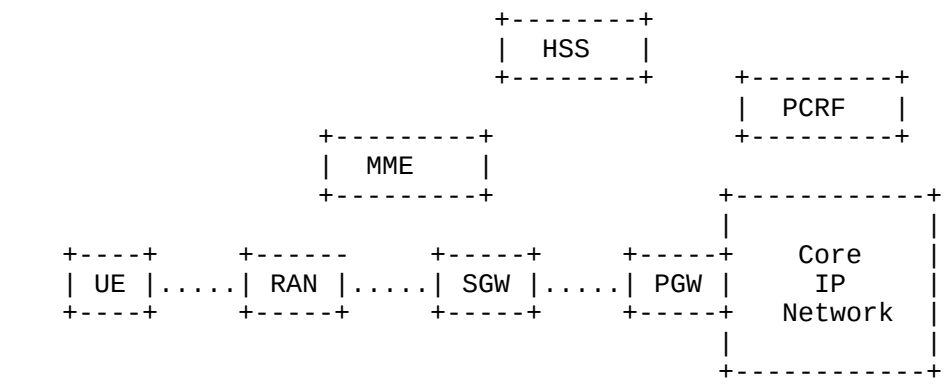
activation vs. DHCPv4 exchanges with the core network following bearer activation. From the point of view of the network, the choices are between static and dynamic address allocation. For dynamic allocation, the choice is between:

- o allocation by the access network (with the GGSN responsible for managing the address pool);
- o allocation by the core network (with the GGSN acting as DHCPv4 or RADIUS client and passing the configuration data on to the UE during bearer setup; or
- o via DHCPv4 to the UE after bearer setup, as already mentioned.

IPv6 prefix allocation is done by stateless address autoconfiguration (SLAAC), with the GGSN responsible for sending Router Announcements in response to the UE's Router Solicitation. For further details see Sections 9.2.1 and 9.2.1.1 of [3GPP_TR_23_060].

The fourth generation mobile access architecture is described in [3GPP_TR_23_401] and [3GPP_TR_23_402] and shown in Figure 2. The SGSN has been split into a control part, the Mobile Management Entity (MME), and a forwarding part, the Serving Gateway (SGW). The GGSN has been replaced by the Packet Data Network Gateway (PDN Gateway, or PGW).

The user data path to the core network changes in several ways with LTE. First, it becomes possible to carry both IPv4 and IPv6 over the same bearer. Thus the figure shows only a single link from the User Equipment to the PDN Gateway. The second change is that the layer 2 protocol used in the third generation architecture to carry the link between the edge of the Radio Access Network and the SGSN is replaced by a tunnel over IP, the same protocol stack (User packets/GTP-U/IP/...) used between the SGSN and the GGSN. Finally, the operator has the option to use PMIPv6 [RFC5213] instead of GTP-U tunneling between the SGW and the PDN Gateway. The SGW acts in the role of Mobility Access Gateway (MAG), while the PDN Gateway acts in the role of Local Mobility Anchor (LMA). PMIP is used only when the core IP network to which the UE is connected has the same operator as the access network.



UE = User Equipment
RAN = Radio Access Network
SGW = Serving Gateway
PGW = Packet Data Network (PDN) Gateway
MME = Mobility Management Entity
PCRF = Policy and Charging Rules Function
HSS = Home Subscriber Server

Figure 2: Long Term Evolution (LTE) Mobile Access Network

Essentially the same options are available in LTE for address configuration of the User Equipment, except that the PDN Gateway replaces the GGSN as the entity responsible for obtaining and propagating that information.

To ease the upgrade from third generation access to LTE, it is possible to mix equipment types in the same access network. Traffic from the SGSN is forwarded to the SGW, which relays it to the PDN. If the Radio Access Network control is upgraded to use GTP tunneling, it is possible to tunnel traffic directly between the Radio Access Network and the SGW. The SGSN retains a control function. The final step is to replace the SGSN by an MME to carry out the control function.

To achieve service continuity during handover, legacy mobile devices support MIPv4 [RFC3344]. The GGSN provides the Foreign Agent functionality. More recent devices support DSMIPv6 [RFC5555]. DSMIPv6 assumes that both the UE and the Home Agent are dual-stack.

3. Approaches To IPv4 / IPv6 Coexistence

This section discusses the coexistence of user-plane IPv4 and IPv6 traffic. The operator is also faced with the upgrade of the network equipment to use IPv6 for control signalling and for the IP wrapper

for PMIP and GTP-U tunnels carrying user data. This upgrade can proceed independently of the work on user-plane traffic, but presents its own coexistence problem. In the immediate case this is because it will take time to reconfigure all of the equipment in one network. Over the longer term the problem arises because different networks will upgrade at different times, and they must interoperate to support mobile roaming in the meantime.

3.1. IPv6 Coexistence Strategy 1: Dual-Stack Connectivity With Limited Public IPv4 Address Pools

In this IPv6 transition scenario, the UEs operate in dual stack mode and are assigned both an IPv6 prefix and an IPv4 address in order to allow them to utilise both IPv4- and IPv6-capable applications. Dual-stack UEs are able to support parallel IPv4 and IPv6 connectivity to a single PDN. As popular services start to support IPv6, IPv4 traffic will gradually be offloaded into the IPv6 domain. Services owned and deployed by the operator may be IPv6-enabled first (while retaining IPv4 capability) and hence will be accessible to dual-stack capable MSs running IPv6.

Issue: In dual stack mode, every UE still needs an IPv4 address. As the number of subscribers grows, the lack of additional public IPv4 addresses will force the use of private rather than public IPv4 addresses for some UEs. Aside from anything else, this will complicate the operation of mobile IP.

3.2. IPv6Coexistence Strategy 2: Dual Stack Connectivity With Limited Private IPv4 Address Pools

In this scenario, UEs operate in dual stack mode and are assigned both an IPv6 prefix and a private IPv4 address in order to allow them to utilise both IPv4- and IPv6-capable applications. The IPv4 addresses assigned to UEs are taken from one of the private address ranges specified in RFC 1918. NAT is performed on the interface between the PDN Gateway and the core IP network.

Issue : The number of private IP addresses is limited. If more than 16 million UEs are active in the same network simultaneously, the network could run out of private IPv4 addresses to assign.

The problem could be worse than this, depending on how many addresses are temporarily unused because they haven‘t been reclaimed after the UE roamed out of the network or shut down.

3.3. IPv6 Coexistence Strategy 3: UEs With IPv6-only Transport and Applications Using IPv6

In this scenario, the operator decides to assign only IPv6 prefixes to the UEs due to, e.g., shortage of IPv4 addresses or because the UEs support only IPv6.

Issue: UEs with IPv6-only connectivity running applications using IPv6 should be able to access both IPv4- or IPv6-enabled services. NAT64 and DNS64 should be performed to let the IPv6-only UEs have access to IPv4 services.

3.4. IPv6 Coexistence Strategy 4: IPv4 Applications Running On a Dual-Stack Host With an Assigned IPv6 Prefix and a Shared IPv4 Address

In this scenario an IPv4 application running on a dual-stack UE needs to access IPv4 services without the operator having to allocate a unique non-shared (private or public) IPv4 address to the UE. The dual-stack UE running these applications uses an IPv4 address that is shared amongst many other UEs, and uses an IPv6 prefix.

Issue: The obvious issue arises, that IPv4 services need to be able to distinguish between UEs using the same IPv4 address. The source port may be used for this purpose.

4. Consideration of IPv4 / IPv6 Coexistence Solutions

The different strategies described above require support to make them work.

4.1. Gateway-Initiated Dual-Stack Lite

Dual-stack lite (DS-lite) [I-D.softwire-dual-stack-lite-06] transports IPv4 traffic from the user device in an IPv6 tunnel across an IPv6 provider network to a NAT44, where sharing of IPv4 public addresses can be implemented. To prevent user DNS queries from going through the NAT44, all queries are intercepted and sent to an IPv6 DNS server. Gateway-initiated DS-lite [I-D.softwire-gateway-init-ds-lite-00] makes explicit use of the tunnel set up through the access network to carry user packets to the IP network. The gateway at the IP end of this tunnel maintains a single tunnel between itself and the NAT44, to which it forwards packets from all of the user devices connected to it. The inner source IP address of the individual packets is actually a 32-bit context identifier used with other information to retrieve forwarding state at the gateway and the NAT44. Gateway-initiated DS-lite also reduces or in some configurations eliminates the need for a unique

IPv4 address, public or private, at the UE.

As applied to the architectures described in Section 2, the gateway is represented by the PDN Gateway or GGSN. The NAT44 is located beyond the gateway, in the core IP network. The access tunnels are GTP over IP, and indeed the tunnel IP version may be either IPv4 or IPv6 without affecting the operation of gateway-initiated DS lite. Section 5.6.1.2 of [3GPP_TR_23_060] suggests that GTP tunnels run between the core nodes too, so the core network may run IPv4 or IPv6 independently of what is used in the access network.

By making the IPv4 address provisioned at the UE almost irrelevant, gateway-initiated DS-lite supports any of the strategies described in Section 3 except the all-IPv6 approach. One issue is that, at the beginning when most traffic is IPv4, a large investment in NAT44 capacity will be needed. As traffic migrates to IPv6, this investment becomes stranded and not reusable. Another issue is that all IPv4 traffic suffers the quality of service penalty imposed by the use of NAT.

One issue that has to be resolved is how to handle DNS queries for IPv4 addresses. [I-D.softwire-dual-stack-lite-06] requires that all DNS queries be sent to an IPv6 DNS server. Discussion on the softwires list leading up to this suggested that from there, A record requests could be forwarded to an IPv4 server. Alternatively, the IPv6 server could maintain both AAAA and A records. A third possibility was that the address of an IPv4 DNS server could be configured manually at the UE. This is messy, both on grounds of operational cost and because it would push DNS queries through the NAT44, greatly increasing its workload.

4.2. Protocol Translation

NAT-PT was first described in [RFC2766]. [RFC4966] summarized a number of issues identified with NAT-PT in the intervening years. As a result, [RFC2766] was deprecated and given Historic status.

The IETF has made a new attempt at solving the problem. [ID_v6v4_framework] explores a number of different translation scenarios. Any particular application must identify which of the scenarios it is dealing with before it can choose stateless versus stateful translation.

[ID_IVI] reports on an early implementation of stateless translation, operating in the two directions between an IPv6 network and the IPv4 Internet. This has been deployed in the China Education and Research Network (CERNET). IVI predates the official IETF work in this area, references to which are given both in [ID_v6v4_framework] and

[ID_IVI].

5. IANA Considerations

This memo includes no request to IANA.

6. Security Considerations

This memo does not in itself introduce any security issues.

7. informative References

[3GPP_TR_23_060]
3rd Generation Partnership Project, "Technical Specification Group Services and System Aspects; General Packet Radio Service (GPRS); Service description; Stage 2 (Release 9)", TR 23.060, June 2010.

[3GPP_TR_23_401]
3rd Generation Partnership Project, "Technical Specification Group Services and System Aspects; General Packet Radio Service (GPRS) enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access (Release 9)", TR 23.401, March 2010.

[3GPP_TR_23_402]
3rd Generation Partnership Project, "Technical Specification Group Services and System Aspects; Architecture enhancements for non-3GPP accesses (Release 9)", TR 23.402, June 2010.

[I-D.huang-behave-pnat-01]
Huang, B., "Prefix NAT: Host based IPv6 translation", February 2010.

[I-D.softwire-dual-stack-lite-06]
Durand, A., "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", August 2010.

[I-D.softwire-gateway-init-ds-lite-00]
Brockners, F., "Gateway Initiated Dual-Stack Lite Deployment", May 2010.

[ID_IVI] Li, X., Bao, C., Chen, M., Zhang, H., and J. Wu, "", January 2010.

[ID_v6v4_framework]
Baker, F., Li, X., Bao, C., and K. Yin, "Framework for IPv4/IPv6 Translation (Work in progress)", August 2010.

[RFC2766] Tsirtsis, G. and P. Srisuresh, "Network Address Translation - Protocol Translation (NAT-PT)", RFC 2766, February 2000.

[RFC3344] Perkins, C., "IP Mobility Support for IPv4", RFC 3344, August 2002.

[RFC4966] Aoun, C. and E. Davies, "Reasons to Move the Network Address Translator - Protocol Translator (NAT-PT) to Historic Status", RFC 4966, July 2007.

[RFC5213] Gundavelli, S., Leung, K., Devarapalli, V., Chowdhury, K., and B. Patil, "Proxy Mobile IPv6", RFC 5213, August 2008.

[RFC5555] Soliman, H., "Mobile IPv6 Support for Dual Stack Hosts and Routers", RFC 5555, June 2009.

Authors' Addresses

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

Phone:
Email: cathyzhou@huawei.com

Tom Taylor
Huawei Technologies
1852 Lorraine Ave.
Ottawa K1H 6Z8
Canada

Phone:
Email: tom111.taylor@bell.net