Network Working Group
Internet Draft
Intended status: Informational
Expires: July 4, 2011

S. Jiang
D. Guo
Huawei Technologies Co., Ltd
B. Carpenter
University of Auckland
January 4, 2011

An Incremental Carrier-Grade NAT (CGN) for IPv6 Transition
draft-ietf-v6ops-incremental-cgn-03.txt

Status of this Memo

Copyright Notice

Abstract

Global IPv6 deployment was slower than originally expected. As IPv4 address exhaustion approaches, IPv4 to IPv6 transition issues become more critical and less tractable. Host-based transition mechanisms

used in dual stack environments cannot meet all transition requirements. Most end users are not sufficiently expert to configure or maintain host-based transition mechanisms. Carrier-Grade NAT (CGN) devices with integrated transition mechanisms can reduce the operational changes 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 hosts and IPv4 access services for IPv4 hosts, while leaving much of a legacy ISP network unchanged during the initial stage of IPv4 to IPv6 migration. Unlike CGN alone, incremental CGN also supports and encourages smooth transition towards dual-stack or IPv6-only ISP networks. An integrated configurable CGN device and an adaptive Home Gateway (HG) device are described. Both are re-usable during different transition phases, avoiding multiple upgrades. This enables IPv6 migration to be incrementally achieved according to real user requirements.

Table of Contents

1. Introduction.....3
2. An Incremental CGN Approach.....4
   2.1. Incremental CGN Approach Overview.....4
   2.2. Choice of tunneling technology.....5
   2.3. Behavior of Dual-stack Home Gateway.....6
   2.4. Behavior of Dual-stack CGN.....7
   2.5. Impact for existing hosts and unchanged networks.....7
   2.6. IPv4/IPv6 intercommunication.....7
   2.7. Discussion.....8
3. Smooth transition towards IPv6 infrastructure.....9
4. Security Considerations.....10
5. IANA Considerations.....11
6. Acknowledgements.....11
7. Change Log [RFC Editor please remove].....11
8. References.....12
   8.1. Normative References.....12
   8.2. Informative References.....12
Author's Addresses.....15

1. Introduction

Global IPv6 deployment did not happen as was forecast 10 years ago. Network providers were hesitant to make the first move while IPv4 was and is still working well. However, IPv4 address exhaustion is imminent. The dynamically-updated IPv4 Address Report [IPUSAGE] has analyzed this issue. It predicts early 2011 for IANA unallocated address pool exhaustion and middle 2012 for RIR unallocated address pool exhaustion. Based on this fact, the Internet industry appears to have reached consensus that global IPv6 deployment is inevitable and has to be done expeditiously.

IPv4 to IPv6 transition issues therefore become more critical and complicated for the approaching global IPv6 deployment. Host-based transition mechanisms alone are not able to meet the requirements in all cases. Therefore, network-based supporting functions and/or new transition mechanisms with simple user-side operation are needed.

Carrier-Grade NAT (CGN) [I-D.nishitani-cgn], also called NAT444 CGN or Large Scale NAT, compounds IPv4 operational problems when used alone, but does nothing to encourage IPv4 to IPv6 transition. Deployment of NAT444 CGN allows ISPs to delay the transition, and therefore causes double transition costs (once to add CGN, and again to support IPv6).

CGN deployments that integrate multiple transition mechanisms can simplify the operation of end user services during the IPv4 to IPv6 migration and coexistence periods. CGNs are deployed on the network side and managed/maintained by professionals. On the user side, new Home Gateway (HG) devices may be needed too. They may be provided by network providers, depending on the specific business model. Dual-stack lite [I-D.ietf-softwire-dual-stack-lite], also called DS-Lite, is a CGN-based solution that supports transition, but it requires the ISP to upgrade its network to IPv6 immediately. Many ISPs hesitate to do this as the first step. Theoretically, DS-Lite can be used with double encapsulation (IPv4-in-IPv6-in-IPv4) but this seems even less likely to be accepted by an ISP and is not discussed in this document.

This document proposes an incremental CGN approach for IPv6 transition. It does not define any new protocols or mechanisms, but describes how to combine existing proposals in an incremental deployment. The approach is similar to DS-Lite, but the other way around. It mainly combines v4-v4 NAT with v6-over-v4 tunneling functions. It can provide IPv6 access services for IPv6-enabled hosts and IPv4 access services for IPv4 hosts, while leaving most of legacy

IPv4 ISP networks unchanged. The deployment of this technology does not affect legacy IPv4 hosts with global IPv4 addresses at all. It is suitable for the initial stage of IPv4 to IPv6 migration. It also supports transition towards dual-stack or IPv6-only ISP networks.

A smooth transition mechanism is also described in this document. It introduces an integrated configurable CGN device and an adaptive HG device. Both CGN and HG are re-usable devices during different transition periods, so they do not need to be replaced as the transition proceeds. This enables IPv6 migration to be incrementally achieved according to the real user requirements.

2. An Incremental CGN Approach

Today, most consumers primarily use IPv4. Network providers are starting to provide IPv6 access services for end users. At the initial stage of IPv4 to IPv6 migration, IPv4 connectivity and traffic would continue to represent the majority of traffic for most ISP networks. ISPs would like to minimize the changes to their IPv4 networks. Switching the whole ISP network into IPv6-only would be considered as a radical strategy. Switching the whole ISP network to dual stack is less radical, but introduces operational costs and complications. Although some ISPs have successfully deployed dual stack networks, others prefer not to do this as their first step in IPv6. However, they currently face two urgent pressures - to compensate for an immediate shortage of IPv4 addresses by deploying some method of address sharing, and to prepare actively for the use of deployment of IPv6 address space and services. ISPs facing only one pressure out of two could adopt either CGN (for shortage of IPv4 addresses) or 6rd (to provide IPv6 connectivity services). The approach described in this draft is intended to address both of these pressures at the same time by combining v4-v4 CGN with v6-over-v4 tunneling technologies.

2.1. Incremental CGN Approach Overview

The incremental CGN approach we propose is illustrated as the following figure.

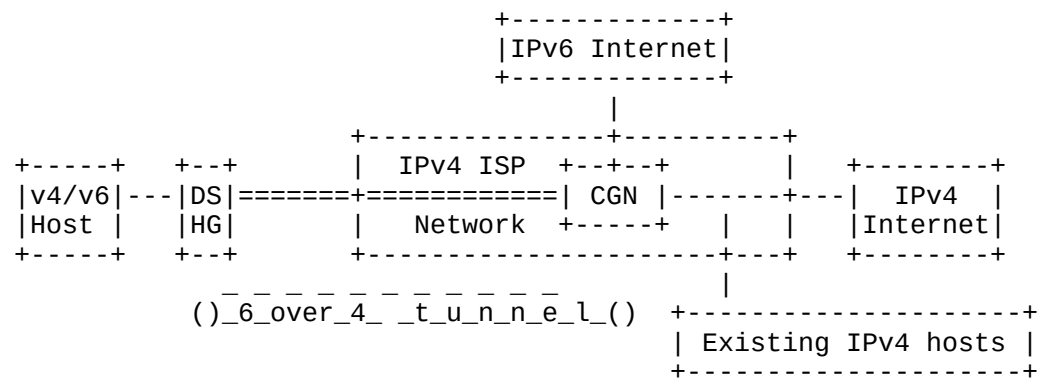


Figure 1: incremental CGN approach with IPv4 ISP network

DS HG = Dual-Stack Home Gateway (CPE - Customer Premises Equipment).

As shown in the above figure, the ISP has not significantly changed its IPv4 network. This approach enables IPv4 hosts to access the IPv4 Internet and IPv6 hosts to access the IPv6 Internet. A dual stack host is treated as an IPv4 host when it uses IPv4 access service and as an IPv6 host when it uses an IPv6 access service. In order to enable IPv4 hosts to access the IPv6 Internet and IPv6 hosts to access IPv4 Internet, NAT64 can be integrated with the CGN; see Section 2.6 for details regarding IPv4/IPv6 intercommunication. The integration of such mechanisms is out of scope for this document.

Two types of device need to be deployed in this approach: a dual-stack home gateway (HG), and a dual-stack CGN. The dual-stack home gateway integrates both IPv6 and IPv4 forwarding and v6-over-v4 tunneling functions. It should follow the requirements of [I-D.ietf-v6ops-ipv6-cpe-router], including IPv6 prefix delegation. It may integrate v4-v4 NAT functionality, too. The dual-stack CGN integrates v6-over-v4 tunneling and v4-v4 CGN functions, as well as standard IPv6 and IPv4 routing

The approach does not require any new mechanisms for IP packet forwarding or encapsulation or decapsulation at tunnel end points. The following sections describe how the HG and the incremental CGN interact.

2.2. Choice of tunneling technology

In principle, this model will work with any form of tunnel between the dual-stack HG and the dual-stack CGN. However, tunnels that require individual configuration are clearly undesirable because of

their operational cost. Configured tunnels based directly on [RFC4213] are therefore not suitable. A tunnel broker according to [RFC3053] would also have high operational costs and be unsuitable for home users.

6rd [RFC5569, RFC5969] technology appears suitable to support v6-over-v4 tunneling with low operational cost. GRE [RFC2784] with an additional auto-configuration mechanism is also able to support v6-over-v4 tunneling. Other tunneling mechanisms such as 6over4 [RFC2529], 6to4 [RFC3056], the Intra-Site Automatic Tunnel Addressing Protocol (ISATAP) [RFC5214] or Virtual Enterprise Traversal (VET) [RFC5558] could be considered. If the ISP has an entirely MPLS infrastructure between the HG and the dual-stack CGN, it would also be possible to use a 6PE [RFC4798] tunnel directly over MPLS. This would, however, only be suitable for an advanced HG that is unlikely to be found as a consumer device, and is not further discussed here. To simplify the discussion, we assume the use of 6rd.

2.3. Behavior of Dual-stack Home Gateway

When a dual-stack home gateway receives a data packet from a host, it will determine whether the packet is an IPv4 or IPv6 packet. The packet will be handled by an IPv4 or IPv6 stack as appropriate. For IPv4, and in the absence of v4-v4 NAT on the HG, the stack will simply forward the packet to the CGN, which will normally be the IPv4 default router. If v4-v4 NAT is enabled, the HG translates the packet source address from a HG-scope private IPv4 address into a CGN-scope IPv4 address, performs port mapping if needed, and then forwards the packet towards the CGN. The HG records the v4-v4 address and port mapping information for inbound packets, like any other NAT.

For IPv6, the HG needs to encapsulate the data into an IPv4 tunnel packet, which has the dual-stack CGN as the IPv4 destination. The HG sends the new IPv4 packet towards the CGN, for example using 6rd.

If the HG is linked to more than one CGN, it will record the mapping information between the tunnel and the source IPv6 address for inbound packets. Detailed considerations for the use of multiple CGNs by one HG are for further study.

IPv4 packets from the CGN, and encapsulated IPv6 packets from the CGN, will be translated or decapsulated according to the stored mapping information and forwarded to the customer side of the HG.

2.4. Behavior of Dual-stack CGN

When a dual-stack CGN receives an IPv4 data packet from a dual-stack home gateway, it will determine whether the packet is a normal IPv4 packet, which is non-encapsulated, or a v6-over-v4 tunnel packet addressed to a tunnel end point within the CGN. For a normal IPv4 packet, the CGN translates the packet source address from a CGN-scope IPv4 address into a public IPv4 address, performs port mapping if necessary, and then forwards it normally to the IPv4 Internet. The CGN records the v4-v4 address and port mapping information for inbound packets, just like a normal NAT does. For a v6-over-v4 tunnel packet, the tunnel end point within the CGN will decapsulate it into the original IPv6 packet and then forward it to the IPv6 Internet. The CGN records the mapping information between the tunnel and the source IPv6 address for inbound packets.

Depending on the deployed location of the CGN, it may use a further v6-over-v4 tunnel to connect to the IPv6 Internet.

Packets from the IPv4 Internet will be appropriately translated by the CGN and forwarded to the HG, and packets from the IPv6 Internet will be tunneled to the appropriate HG, using the stored mapping information as necessary.

2.5. Impact for existing hosts and unchanged networks

This approach does not affect the unchanged parts of ISP networks at all. Legacy IPv4 ISP networks and their IPv4 devices remain in use. The existing IPv4 hosts, shown as the lower right box in Figure 1, either having global IPv4 addresses or behind v4-v4 NAT, can connect to the IPv4 Internet as it is now. These hosts, if they are upgraded to become dual-stack hosts, can access the IPv6 Internet through the IPv4 ISP network by using IPv6-over-IPv4 tunnel technologies. (See section 2.7 for a comment on MTU size.)

2.6. IPv4/IPv6 intercommunication

IPv6-only public services are not expected as long as there is significant IPv4-only customer base in the world, for obvious commercial reasons. However, IPv4/IPv6 intercommunication may become issues in many scenarios.

The IETF is expected to standardize a recommended IPv4/IPv6 translation algorithm, sometimes referred to as NAT64. It is specified in

- "Framework for IPv4/IPv6 Translation" [I-D.ietf-behave-v6v4-framework]
- "IPv6 Addressing of IPv4/IPv6 Translators" [RFC6052]
- "DNS64: DNS extensions for Network Address Translation from IPv6 Clients to IPv4 Servers" [I-D.ietf-behave-dns64]
- "IP/ICMP Translation Algorithm" [I-D.ietf-behave-v6v4-xlate]
- "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers" [I-D.ietf-behave-v6v4-xlate-stateful]
- "An FTP ALG for IPv6-to-IPv4 translation" [I-D.ietf-behave-ftp64]

The service, as described in the IETF's "Guidelines for Using IPv6 Transition Mechanisms during IPv6 Deployment" [I-D.arkko-ipv6-transition-guidelines], provides for stateless translation between hosts in an IPv4-only domain or which offer an IPv4-only service and hosts with an IPv4-embedded IPv6 address in an IPv6-only domain. It additionally provide access from IPv6 hosts with general format addresses to hosts in an IPv4-only domain or which offer an IPv4-only service. It does not provide any-to-any translation. One result is that client-only hosts in the IPv6 domain gain access to IPv4 services through stateful translation. Another result is that the IPv6 network operator has the option of moving servers into the IPv6-only domain while retaining accessibility for IPv4-only clients, through stateless translation of an IPv4-embedded IPv6 address.

Also, "Architectural Implications of NAT" [RFC2993] applies across the service just as in IPv4/IPv4 translation: apart from the fact that a system with an IPv4-embedded IPv6 address is reachable across the NAT, which is unlike IPv4, any assumption on the application's part that a local address is meaningful to a remote peer, and any use of an IP address literal in the application payload, is a source of service issues. In general, the recommended mitigation for this is

- Ideally, applications should use DNS names rather than IP address literals in URLs, URIs, and referrals, and in general be network layer agnostic.
- If they do not, the network may provide a relay or proxy that straddles the domains. For example, an SMTP MTA with both IPv4 and IPv6 connectivity handles IPv4/IPv6 translation cleanly at the application layer.

2.7. Discussion

For IPv4 traffic, the incremental CGN approach inherits all the problems of CGN address sharing techniques [I-D.ietf-intarea-shared-addressing-issues] (e.g., scaling, and the

difficulty of supporting well-known ports for inbound traffic). Application layer problems created by double NAT are beyond the scope of this document.

For IPv6 traffic, a user behind the DS HG will see normal IPv6 service. We observe that an IPv6 tunnel MTU of at least 1500 bytes would ensure that the mechanism does not cause excessive fragmentation of IPv6 traffic nor excessive IPv6 path MTU discovery interactions. This, and the absence of NAT problems for IPv6, will create an incentive for users and application service providers to prefer IPv6.

ICMP filtering [RFC4890] may be included as part of CGN functions.

3. Smooth transition towards IPv6 infrastructure

Transition from pure NAT444 CGN or 6rd to the incremental CGN approach is straightforward. The HG and CGN devices and their locations do not have to be changed; only software upgrading may be needed. In the ideal model, described below, even software upgrading is not necessary; reconfiguration of the devices is enough. NAT444 CGN solves the public address shortage issues in the current IPv4 infrastructure. However, it does not contribute towards IPv6 deployment at all. The incremental CGN approach can inherit NAT444 CGN functions while providing overlay IPv6 services. 6rd mechanisms can also transform smoothly into this incremental CGN model. However, the home gateways need to be upgraded correspondingly to perform the steps described below

The incremental CGN can also easily be transitioned to an IPv6-enabled infrastructure, in which the ISP network is either dual-stack or IPv6-only.

If the ISP prefers to move to dual-stack routing, the HG should simply switch off its v6-over-v4 function when it observes native IPv6 RA or DHCPv6 traffic, and then forward both IPv6 and IPv4 traffic directly, while the dual-stack CGN keeps only its v4-v4 NAT function.

However, we expect ISPs to choose the approach described as incremental CGN here because they intend to avoid dual-stack routing, and to move incrementally from IPv4-only routing to IPv6-only routing. In this case, the ideal model for the incremental CGN approach is that of an integrated configurable CGN device and an adaptive HG device. The integrated CGN device will be able to support multiple functions, including NAT444 CGN, 6rd router (or an alternative tunneling mechanism), DS-Lite, and dual-stack forwarding.

The HG has to integrate the corresponding functions, and be able to detect relevant incremental changes on the CGN side. Typically the HG will occasionally poll the CGN to discover which features are operational. For example, starting from a pure IPv4-only scenario (in which the HG treats the CGN only as an IPv4 default router), the HG would discover by infrequent polling when 6rd became available. The home user would then acquire an IPv6 prefix. At a later stage, the HG would observe the appearance of native IPv6 Route Advertisement messages or DHCPv6 messages to discover the availability of an IPv6 service including DS-Lite. Thus, when an ISP decides to switch from incremental CGN to DS-Lite CGN, only a configuration change or a minor software update is needed on the CGNs. The home gateway would detect this change and switch automatically to DS-Lite mode. The only impact on the home user will be to receive a different IPv6 prefix.

In the smooth transition model, both CGN and HG are re-usable devices during different transition periods. This avoids potential multiple upgrades. Different regions of the same ISP network may be at different stages of the incremental process, using identical equipment but with different configurations of the incremental CGN devices in each region. Thus, IPv6 migration may be incrementally achieved according to the real ISP and customer requirements.

4. Security Considerations

Security issues associated with NAT have been documented in [RFC2663] and [RFC2993]. Security issues for large-scale address sharing, including CGN, are discussed in [I-D.ietf-intarea-shared-addressing-issues]. The present specification does not introduce any new features to CGN itself, and hence no new security considerations. Security issues for 6rd are documented in [RFC5569, RFC5969] and those for DS-Lite in [I-D.ietf-softwire-dual-stack-lite].

Since the tunnels proposed here exist entirely within a single ISP network, between the HG/CPE and the CGN, the threat model is relatively simple. [RFC4891] describes how to protect tunnels using IPsec, but an ISP could reasonably deem its infrastructure to provide adequate security without the additional protection and overhead of IPsec. The intrinsic risks of tunnels are described in [I-D.ietf-v6ops-tunnel-security-concerns], which recommends that tunneled traffic should not cross border routers. The incremental CGN approach respects this recommendation. To avoid other risks caused by tunnels, it is important that any security mechanisms based on packet inspection, and any implementation of ingress filtering, are applied to IPv6 packets after they have been decapsulated by the CGN. The dual-stack home gateway will need to provide basic security

functionality for IPv6 [I-D.ietf-v6ops-cpe-simple-security]. Other aspects are described in [RFC4864].

5. IANA Considerations

This draft does not request any IANA action.

6. Acknowledgements

Useful comments were made by Fred Baker, Dan Wing, Fred Templin, Seiichi Kawamura, Remi Despres, Janos Mohacsi, Mohamed Boucadair, Shin Miyakawa, Joel Jaeggli, Jari Arkko, Tim Polk, Sean Turner and other members of the IETF V6OPS working group.

7. Change Log [RFC Editor please remove]

draft-jiang-incremental-cgn-00, original version, 2009-02-27

draft-jiang-v6ops-incremental-cgn-00, revised after comments at IETF74, 2009-04-23

draft-jiang-v6ops-incremental-cgn-01, revised after comments at v6ops mailing list, 2009-06-30

draft-jiang-v6ops-incremental-cgn-02, remove normative parts (to be documented in other WGs), 2009-07-06

draft-jiang-v6ops-incremental-cgn-03, revised after comments at v6ops mailing list, 2009-09-24

draft-ietf-v6ops-incremental-cgn-00, accepted as v6ops wg document, 2009-11-17

draft-ietf-v6ops-incremental-cgn-01, revised after comments at v6ops mailing list, 2010-06-21

draft-ietf-v6ops-incremental-cgn-02, revised after comments at v6ops WGLC, 2010-10-11

draft-ietf-v6ops-incremental-cgn-03, revised according comments from IESG, 2011-1-4

8. References

8.1. Normative References

[RFC2529] B. Carpenter, and C. Jung, "Transmission of IPv6 over IPv4 Domains without Explicit Tunnels", RFC2529, March 1999.

[RFC2784] D. Farinacci, T. Li, S. Hanks, D. Meyer and P. Traina, "Generic Routing Encapsulation (GRE)", RFC 2784, March 2000.

[RFC5569] R. Despres, "IPv6 Rapid Deployment on IPv4 infrastructures (6rd)", RFC 5569, January 2010.

[RFC5969] W. Townsley and O. Troan, "IPv6 via IPv4 Service Provider Networks '6rd'", RFC5969, May 2010.

8.2. Informative References

[RFC2663] P. Srisuresh and M. Holdrege, "IP Network Address Translator (NAT) Terminology and Considerations", RFC 2663, August 1999.

[RFC2993] T. Hain, "Architectural Implications of NAT", RFC 2993, November 2000.

[RFC3053] A. Durand, P. Fasano, I. Guardini and D. Lento, "IPv6 Tunnel Broker", RFC 3053, January 2001.

[RFC3056] B. Carpenter and K. Moore, "Connection of IPv6 Domains via IPv4 Clouds", RFC 3056, February 2001.

[RFC4213] E. Nordmark and R. Gilligan, "Basic Transition Mechanisms for IPv6 Hosts and Routers", RFC 4213, October 2005.

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

[RFC4864] G. Van de Velde, T. Hain, R. Droms, B. Carpenter and E. Klein, "Local Network Protection for IPv6", RFC 4864, May 2007.

[RFC4890] E. Davies and J. Mohacsi, "Recommendations for Filtering ICMPv6 Messages in Firewalls", RFC 4890, May 2007.

[RFC4891] R. Graveman, "Using IPsec to Secure IPv6-in-IPv4 Tunnels", RFC4891, May 2007.

[RFC5214] F. Templin, T. Gleeson and D. Thaler, "Intra-Site Automatic Tunnel Addressing Protocol (ISATAP)", RFC 5214, March 2008.

[RFC5558] F. Templin, "Virtual Enterprise Traversal (VET)", RFC 5558, February 2010.

[RFC6052] Bao, C., Huitema, C., Bagnulo, M., Boucadair, M., and X. Li, "IPv6 Addressing of IPv4/IPv6 Translators", RFC 6052, 2010.

[IPUSAGE] G. Huston, IPv4 Address Report, March 2009, http://www.potaroo.net/tools/ipv4/index.html.

[I-D.ietf-softwire-dual-stack-lite]
A. Durand, "Dual-stack lite broadband deployments post IPv4 exhaustion", draft-ietf-softwire-dual-stack-lite, work in progress.

[I-D.ietf-v6ops-ipv6-cpe-router]
H. Singh, W. Beebee, C. Donley, B. Stark and O. Troan, "IPv6 CPE Router Recommendations", draft-ietf-v6ops-ipv6-cpe-router, work in progress.

[I-D.ietf-v6ops-cpe-simple-security]
J. Woodyatt, "Recommended Simple Security Capabilities in Customer Premises Equipment for Providing Residential IPv6 Internet Service", draft-ietf-v6ops-cpe-simple-security, work in progress.

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

[I-D.ietf-intarea-shared-addressing-issues]
M. Ford, et al, "Issues with IP Address Sharing", draft-ietf-intarea-shared-addressing-issues, work in progress.

[I-D.nishitani-cgn]
I. Yamagata, T. Nishitani, S. Miyahawa, A. nakagawa and H. Ashida, "Common requirements for IP address sharing schemes", draft-nishitani-cgn, work in progress.

[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, work in progress.

[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, work in progress.

[I-D.ietf-behave-ftp64]
Beijnum, I., "An FTP ALG for IPv6-to-IPv4 translation", draft-ietf-behave-ftp64, work in progress.

[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, work in progress.

[I-D.ietf-behave-v6v4-xlate]
Li, X., Bao, C., and F. Baker, "IP/ICMP Translation Algorithm", draft-ietf-behave-v6v4-xlate, work in progress.

Author's Addresses

Sheng Jiang
Huawei Technologies Co., Ltd
Huawei Building, No.3 Xinxi Rd.,
Shang-Di Information Industry Base, Hai-Dian District, Beijing 100085
P.R. China
Email: shengjiang@huawei.com

Dayong Guo
Huawei Technologies Co., Ltd
Huawei Building, No.3 Xinxi Rd.,
Shang-Di Information Industry Base, Hai-Dian District, Beijing 100085
P.R. China
Email: guoseu@huawei.com

Brian Carpenter
Department of Computer Science
University of Auckland
PB 92019
Auckland, 1142
New Zealand
Email: brian.e.carpenter@gmail.com

IPv6 Operations Working Group
Internet-Draft
Intended status: Informational
Expires: April 28, 2011

S. Krishnan
Ericsson
D. Thaler
Microsoft
J. Hoagland
Symantec

October 25, 2010

Security Concerns With IP Tunneling
draft-ietf-v6ops-tunnel-security-concerns-04

Abstract

A number of security concerns with IP tunnels are documented in this memo. The intended audience of this document includes network administrators and future protocol developers. The primary intent of this document is to raise the awareness level regarding the security issues with IP tunnels as deployed today.

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 3
2. Tunnels May Bypass Security 3
2.1. Network Security Bypass 3
2.2. IP Ingress and Egress Filtering Bypass 5
2.3. Source Routing After the Tunnel Client 6
3. Challenges in Inspecting and Filtering Content of Tunneled Data Packets 6
3.1. Inefficiency of Selective Network Filtering of All Tunneled Packets 7
3.2. Problems with deep packet inspection of tunneled data packets 8
4. Increased Exposure Due to Tunneling 9
4.1. NAT Holes Increase Attack Surface 9
4.2. Exposure of a NAT Hole 11
4.3. Public Tunnels Widen Holes in Restricted NATs 12
5. Tunnel Address Concerns 12
5.1. Feasibility of Guessing Tunnel Addresses 12
5.2. Profiling Targets Based on Tunnel Address 13
6. Additional Security Concerns 14
6.1. Attacks Facilitated By Changing Tunnel Server Setting . . 14
7. Mechanisms to secure the use of tunnels 16
8. Acknowledgments 17
9. Security Considerations 17
10. IANA Considerations 17
11. Informative References 17
Authors' Addresses 19

1. Introduction

With NAT devices becoming increasingly more prevalent, there have recently been many tunneling protocols developed that go through NAT devices or firewalls by tunneling over UDP or TCP. For example, Teredo [RFC4380], L2TPv2 [RFC2661], and L2TPv3 [RFC3931] all tunnel IP packets over UDP. Similarly, many SSL VPN solutions that tunnel IP packets over HTTP (and hence over TCP) are deployed today.

This document discusses security concerns with tunneling IP packets, and includes recommendations where relevant.

The primary intent of this document is to help improve security deployments using tunnel protocols. In addition, the document aims to provide information that can be used in any new or updated tunnel protocol specification. The intended audience of this document includes network administrators and future protocol developers.

2. Tunnels May Bypass Security

2.1. Network Security Bypass

2.1.1. Problem

Tunneled IP traffic may not receive the intended level of inspection or policy application by network-based security devices unless such devices are specifically tunnel-aware. This reduces defense in depth and may cause security gaps. This applies to all network-located devices and to any end-host based firewalls whose existing hooking mechanism(s) would not show them the IP packet stream after the tunnel client does decapsulation or before it does encapsulation.

2.1.2. Discussion

Evasion by tunneling is often a problem for network-based security devices such as network firewalls, intrusion detection and prevention systems, and router controls. To provide such functionality in the presence of tunnels, the developer of such devices must add support for parsing each new protocol. There is typically a significant lag between when the security developer recognizes that a tunnel will be used (or will be remotely usable) to a significant degree and when the parsing can be implemented in a product update, the update tested and released, and customers begin using the update. Late changes in the protocol specification or in the way it is implemented can cause additional delays. This becomes a significant security concern when a delay in applied coverage is occurring frequently. One way to cut down on this lag is for security developers to follow the progress of

new IETF protocols but this will still not account for any new proprietary protocols.

For example, for L2TP or Teredo, an unaware network security device would inspect or regulate the outer IP and the IP-based UDP layer as normal, but it would not recognize that there is an additional IP layer contained inside the UDP payload to which it needs to apply the same controls as it would to a native packet. (Of course, if the device discards the packet due to something in the IP or UDP header, such as referring to an unknown protocol, the embedded packet is no longer a concern.) In addition, if the tunnel does encryption, the network-based security device may not be able to do much, just as if IPsec end-to-end encryption were used without tunneling.

Network security controls being not applied must be a concern to those that set them up, since those controls are supposed to provide an additional layer of defense against external attackers. If network controls are being bypassed due to the use of tunneling, the strength of the defense (i.e. the number of layers of defense) is reduced. Since security administrators may have a significantly reduced level of confidence without this layer, this becomes a concern to them.

One implication of the security control bypass is that defense in depth has been reduced, perhaps down to zero unless a local firewall is in use as recommended in [RFC4380]. However, even if there are host-based security controls that recognize tunnels, security administrators may not have configured them with full security control parity, even if all controls that were maintained by the network are available on the host. Thus there may be gaps in desired coverage.

Compounding this is that, unlike what would be the case for native IP, some network administrators will not even be aware that their hosts are globally reachable, if the tunnel provides connectivity to/ from the Internet; for example, they may not be expecting this for hosts behind a stateful firewall. In addition, Section 3.2 discusses how it may not be efficient to find all tunneled traffic for network devices to examine.

2.1.3. Recommendations

Security administrators who do not consider tunneling an acceptable risk should disable tunnel functionality either on the end-nodes (hosts) or on the network nodes at the perimeter of their network. However, there may be an awareness gap. Thus, due to the possible negative security consequences, tunneling functionality should be easy to disable on the host and through a central management facility

if one is provided.

To minimize security exposure due to tunnels, we recommend that a tunnel be an interface of last resort, independent of IP version. Specifically, we suggest that when both native and tunneled access to a remote host is available, that the native access be used in preference to tunneled access except when the tunnel endpoint is known to not bypass security (e.g., an IPsec tunnel to a gateway provided by the security administrator of the network). This should also promote greater efficiency and reliability.

Note that although Rule 7 of [RFC3484] section 6 will prefer native connectivity over tunnels, this rule is only a tie-breaker when a choice is not made by earlier rules; hence tunneling mechanisms that are tied to a particular range of IP address space will be decided based on the prefix precedence. For example, using the prefix policy mechanism of [RFC3484] section 2.1, Teredo might have a precedence of 5 so that native IPv4 is preferred over Teredo.

2.2. IP Ingress and Egress Filtering Bypass

2.2.1. Problem

IP addresses inside tunnels are not subject to ingress and egress filtering in the network they tunnel over, unless extraordinary measures are taken. Only the tunnel endpoints can do such filtering.

2.2.2. Discussion

Ingress filtering (sanity-checking incoming destination addresses) and egress filtering (sanity-checking outgoing source addresses) are done to mitigate attacks and to make it easier to identify the source of a packet and are considered to be a good practice. e.g. ingress filtering at the network perimeter should not allow packets with a source address that belongs to the network to enter the network from the outside the network. This function is most naturally (and in the general case, by requirement) done at network boundaries. Tunneled IP traffic bypassing this network control is a specific case of Section 2.1, but is illustrative.

2.2.3. Recommendations

Tunnel servers can apply ingress and egress controls to tunneled IP addresses passing through them to and from tunnel clients.

Tunnel clients could make an effort to conduct ingress and egress filtering.

Implementations of protocols that embed an IPv4 address in a tunneled IPv6 address directly between peers should perform filtering based on checking the correspondence.

Implementations of protocols that accept tunneled packets directly from a server, relay or protocol peer do filtering the same way as it would be done on a native link with traffic from a router.

Some protocols such as 6to4 [RFC3056], Teredo, and ISATAP [RFC5214] allow both other hosts and a router over a common tunnel. To perform host-based filtering with such protocols a host would need to know the outer IP address of each router from which it could receive traffic, so that packets from hosts beyond the router will be accepted even though the source address would not embed the router's IP address. Router addresses might be learned via Secure Neighbor Discovery (SEND) [RFC3971] or some other mechanism (e.g., [RFC5214] section 8.3.2).

2.3. Source Routing After the Tunnel Client

2.3.1. Problem

If the encapsulated IP packet specifies source routing beyond the recipient tunnel client, the host may forward the IP packet to the specified next hop. This may be unexpected and contrary to administrator wishes and may have bypassed network-based source routing controls.

2.3.2. Discussion

A detailed discussion of issues related to source routing can be found in [RFC5095] and [SECA-IP].

2.3.3. Recommendations

Tunnel clients should by default discard tunneled IP packets that specify additional routing, as recommended in [RFC5095] and [SECA-IP], though they may also allow the user to configure what source routing types are allowed. All pre-existing source routing controls should be upgraded to apply these controls to tunneled IP packets as well.

3. Challenges in Inspecting and Filtering Content of Tunneled Data Packets

3.1. Inefficiency of Selective Network Filtering of All Tunneled Packets

3.1.1. Problem

There is no mechanism to both efficiently and immediately filter all tunneled packets (other than the obviously faulty method of filtering all packets). This limits the ability to prevent tunnel use on a network.

3.1.2. Discussion

Given concerns about tunnel security or a network's lack of preparedness for tunnels, a network administrator may wish to simply block all use of tunnels that bypass security policies. He or she may wish to do so using network controls; this could be either due to not having the capability to disable tunneling on all hosts attached to the network or due to wanting an extra layer of prevention.

One simple method of doing this easily for many tunnel protocols is to block outbound packets to the UDP or TCP port used (e.g., destination UDP port is 3544 for Teredo, UDP port 1701 for L2TP, etc.). This prevents a tunnel client from establishing a new tunnel. However, existing tunnels will not necessarily be affected if the blocked port is used only for initial setup. In addition, if the blocking is applied on the outside of the client's NAT device, the NAT device will retain the port mapping for the client. In some cases, however, blocking all traffic to a given outbound port (e.g., port 80) may interfere with non-tunneled traffic so this should be used with caution.

Another simple alternative, if the tunnel server addresses are well-known, is to filter out all traffic to/from such addresses.

The other approach is to find all packets to block in the same way as would be done for inspecting all packets (Section 3.2). However; this faces the difficulties in terms of efficiency of filtering, as is discussed there.

3.1.3. Recommendations

Developers of protocols that tunnel over UDP or TCP (including HTTP) to reach the Internet should disable their protocols in networks that wish to enforce security policies on the user traffic. (Windows, for example, disables Teredo by default if it detects that it is within an enterprise network that contains a Windows domain controller.)

Administrators of such networks may wish to filter all tunneled

traffic at the boundaries of their networks. It is sufficient to filter out the tunneled connection requests (if they can be identified) to stop further tunneled traffic. The easiest mechanism for this would be to filter out outgoing traffic sent to the destination port defined by the tunneling protocol, and incoming traffic with that source port. Similarly, in certain cases, it is also possible to use the IP protocol field to identify and filter tunneled packets. e.g. 6to4 [RFC3056] is a tunneling mechanism that uses the IPv4 packets to carry encapsulated IPv6 packets, and can be identified by the IPv4 protocol type 41.

3.2. Problems with deep packet inspection of tunneled data packets

3.2.1. Problem

There is no efficient mechanism for network-based devices, which are not the tunnel endpoint, to inspect the contents of all tunneled data packets, the way they can for native packets. This makes it difficult to apply the same controls as they do to native IP.

3.2.2. Discussion

Some tunnel protocols are easy to identify, such as if all data packets are encapsulated using a well-known UDP or TCP port that is unique to the protocol.

Other protocols, however, either use dynamic ports for data traffic, or else share ports with other protocols (e.g., tunnels over HTTP).

The implication of this is that network-based devices that wish to passively inspect (and perhaps selectively apply policy to) all encapsulated traffic must inspect all TCP or UDP packets (or at least all packets not part of a session that is known not to be a tunnel). This is imperfect since a heuristic must then be applied to determine if a packet is indeed part of a tunnel. This may be too slow to make use of in practice, especially if it means that all TCP or UDP packets must be taken off of the device’s "fast path".

One heuristic that can be used on packets to determine if they are tunnel-related or not is as follows. For each known tunnel protocol, attempt parsing the packet as if it were a packet of that protocol, destined to the local host (i.e., where the local host has the destination address in the inner IP header, if any). If all syntax checks pass, up to and including the inner IP header (if the tunnel doesn’t use encryption), then treat the packet as if it is a tunneled packet of that protocol.

It is possible that non-tunnel packets will match as tunneled using

this heuristic, but tunneled packets (of the known types of tunnels) should not escape inspection, absent implementation bugs.

For some protocols, it may be possible to monitor setup exchanges to know to expect that data will be exchanged on certain ports later. (Note that this does not necessarily apply to Teredo, for example, since communicating with another Teredo client behind a cone NAT [RFC5389] device does not require such signaling. In such cases this control will not work. However, deprecation of the cone bit as discussed in [RFC5991] means this technique may indeed work with updated Teredo implementations.)

3.2.3. Recommendations

As illustrated above, it should be clear that inspecting the contents of tunneled data packets is highly complex and often impractical. For this reason, if a network wishes to monitor IP traffic, tunneling across, as opposed to tunneling to, the security boundary is not recommended. For example, to provide an IPv6 transition solution, the network should provide native IPv6 connectivity or a tunnel solution (e.g., ISATAP or 6over4) that encapsulates data packets between hosts and a router within the network.

4. Increased Exposure Due to Tunneling

4.1. NAT Holes Increase Attack Surface

4.1.1. Problem

If the tunnel allows inbound access from the public Internet, the opening created in a NAT device due to a tunnel client increases its Internet attack surface area. If vulnerabilities are present, this increased exposure can be used by attackers and their programs.

If the tunnel allows inbound access only from a private network (e.g., a remote network to which one has VPN’ed), the opening created in the NAT device still increases its attack surface area, although not as much as in the public Internet case.

4.1.2. Discussion

When a tunnel is active, a mapped port is maintained on the NAT device through which remote hosts can send packets and perhaps establish connections. The following sequence is intended to sketch out the processing on the tunnel client host that can be reached through this mapped port; the actual processing for a given host may be somewhat different.

1. Link-layer protocol processing
2. (Outer) IP host firewall processing
3. (Outer) IP processing by stack
4. UDP/TCP processing by stack
5. Tunnel client processing
6. (Inner) IP host firewall processing
7. (Inner) IP processing by stack
8. Various upper layer processing may follow

The inner firewall (and other security) processing may or may not be present, but if it is, some of the inner IP processing may be filtered. (For example, [RFC4380] section 7.1 recommends that an IPv6 host firewall be used on all Teredo clients.)

(By the virtue of the tunnel being active, we can infer that the inner host firewall is unlikely to do any filtering based on the outer IP.) Any of this processing may expose vulnerabilities an attacker can exploit; similarly these may expose information to an attacker. Thus, even if firewall filtering is in place (as is prudent) and filters all incoming packets, the exposed area is larger than if a native IP Internet connection were in place, due to the processing that takes place before the inner IP is reached (specifically, the UDP/TCP processing, the tunnel client processing, and additional IP processing, especially if one is IPv4 and the other is IPv6).

One possibility is that a layer 3 targeted worm makes use of a vulnerability in the exposed processing. The main benefit tunneling provides to worms is enabling L3 reachability to the end host. Even a thoroughly firewalled host could be subject to a worm that spreads with a single UDP packet if the right remote code vulnerability is present.

4.1.3. Recommendations

This problem seems inherent in tunneling being active on a host, so the solution seems to be to minimize tunneling use.

For example, it can be active only when it is really needed and only for as long as needed. So, the tunnel interface can be initially not configured and only used when it is entirely the last resort. The

interface should then be deactivated (ideally, automatically) again as soon as possible. Note however that the hole will remain in the NAT device for some amount of time after this, so some processing of incoming packets is inevitable unless the client's native IP address behind the NAT device is changed.

4.2. Exposure of a NAT Hole

4.2.1. Problem

Attackers are more likely to know about a tunnel client's NAT hole than a typical hole in the NAT device. If they know about the hole, they could try to use it.

4.2.2. Discussion

There are at least three reasons why an attacker may be more likely to learn of the tunnel client's exposed port than a typical NAT exposed port:

1. The NAT mapping for a tunnel is typically held open for a significant period of time, and kept stable. This increases the chance of it being discovered.

2. In some protocols (e.g., Teredo), the external IP address and port are contained in the client's address that is used end-to-end and possibly even advertised in a name resolution system. While the tunnel protocol itself might only distribute this address in IP headers, peers, routers, and other on-path nodes still see the client's IP address. Although this point does not apply directly to protocols (e.g., L2TP) that do not construct the inner IP address based on the outer IP address, the inner IP address is still known to peers, routers, etc. and can still be reached by attackers without knowing the external IP address or port.

3. The tunnel protocol often contains more messages that are exchanged and with more parties (e.g., due to a longer path length) than without using the tunnel, offering more chance for visibility into the port and address in use.

4.2.3. Recommendations

The recommendations from Section 4.1 seem to apply here as well: minimize tunnel use.

4.3. Public Tunnels Widen Holes in Restricted NATs

4.3.1. Problem

Tunnels that allow inbound connectivity from the Internet (e.g., Teredo, tunnel brokers, etc) essentially disable the filtering behavior of the NAT for all tunnel client ports. This eliminates NAT devices filtering for such ports and may eliminate the need for an attacker to spoof an address.

4.3.2. Discussion

NATs that implement Address-Dependent or Address and Port-Dependent Filtering [RFC4787] limit the source of incoming packets to just those that are a previous destination. This poses a problem for tunnels that intend to allow inbound connectivity from the Internet.

Various protocols (e.g., Teredo, STUN [RFC5389], etc.) provide a facility for peers, upon request, to become a previous destination. This works by sending a "bubble" packet via a server, which is passed to the client, and then sent by the client (through the NAT) to the originator.

This removes any NAT-based barrier to attackers sending packets in through the client's service port. In particular, an attacker would no longer need to either be an actual previous destination or to forge its addresses as a previous destination. When forging, the attacker would have had to learn of a previous destination and then would face more challenges in seeing any returned traffic.

4.3.3. Recommendations

If the tunnel can provide connectivity to the Internet, the tunnel client should run a host firewall on the tunnel interface. Also, minimizing public tunnel use (see Section 4.1.3) would lower the attack opportunity related to this exposure.

5. Tunnel Address Concerns

5.1. Feasibility of Guessing Tunnel Addresses

5.1.1. Problem

For some types of tunneling protocols, it may be feasible to guess IP addresses assigned to tunnels, either when looking for a specific client or when looking for an arbitrary client. This is in contrast to native IPv6 addresses in general, but is no worse than for native

IPv4 addresses today.

For example, some protocols (e.g., 6to4 and Teredo) use well-defined address ranges. As another example, using well-known public servers for Teredo or tunnel brokers also implies using a well known address range.

5.2. Profiling Targets Based on Tunnel Address

5.2.1. Problem

An attacker encountering an address associated with a particular tunneling protocol or well-known tunnel server has the opportunity to infer certain relevant pieces of information that can be used to profile the host before sending any packets. This can reduce the attacker's footprint and increase the attacker's efficiency.

5.2.2. Discussion

The tunnel address reveals some information about the nature of the client.

- That a host has a tunnel address associated with a given protocol means that the client is running on some platform for which there exists a tunnel client implementation of that protocol. In addition, if some platforms have that protocol installed by default and where the host's default rules for using it make it susceptible to being in use, then it is more likely to be running on such a platform than on one where it is not used by default. For example, as of this writing, seeing a Teredo address suggests that the host it is on is probably running Windows.

- Similarly, the use of an address associated with a particular tunnel server also suggests some information. Tunnel client software is often deployed, installed, and/or configured using some degree of automation. It seems likely that the majority of the time the tunnel server that results from the initial configuration will go unchanged from the initial setting. Moreover, the server that is configured for use may be associated with a particular means of installation, which often suggests the platform. For example, if the server field in a Teredo address is one of the IPv4 addressees to which teredo.ipv6.microsoft.com resolves, it suggests that the host is running Windows.

- The external IPv4 address of a NAT device can of course be readily associated with a particular organization or at least an ISP, and hence putting this address in an IPv6 address reveals this information. However, this is no different than using a native IP

address, and hence is not new with tunneling.

- o It is also possible that external client port numbers may be more often associated with particular client software or the platform on which it is running. The usefulness of this for platform determination is, however, reduced by the different NAT port number assignment behaviors. In addition, the same observations would apply to use of UDP or TCP over native IP as well, and hence this is not new with tunneling.

The platform, tunnel client software, or organization information can be used by an attacker to target attacks more carefully. For example, an attacker may decide to attack an address only if it is likely to be associated with a particular platform or tunnel client software with a known vulnerability. (This is similar to the ability to guess some platforms based on the OUI in the EUI-64 portion of an IPv6 address generated from a MAC address, since some platforms are commonly used with interface cards from particular vendors.)

5.2.3. Recommendations

If installation programs randomized the server setting, that would reduce the extent to which they can be profiled. Similarly, administrators can choose to change the default setting to reduce the degree to which they can be profiled ahead of time.

Randomizing the tunnel client port in use would mitigate any profiling that can be done based on the external port, especially if multiple different tunnel clients did this. Further discussion on randomizing ports can be found at [TSV-PORT].

It is recommended that tunnel protocols minimize the propagation of knowledge about whether the NAT is a cone NAT.

6. Additional Security Concerns

6.1. Attacks Facilitated By Changing Tunnel Server Setting

6.1.1. Problem

If an attacker could either change a tunnel client’s server setting or change the IP addresses to which a configured host name resolves (e.g., by intercepting DNS queries) AND the tunnel is not authenticated, it would let the attacker become a man in the middle. This would allow them to at least monitor peer communication and at worst to impersonate the remote peer.

6.1.2. Discussion

A client's server has good visibility into the client's communication with IP peers. If the server were switched to one that records this information and makes it available to third parties (e.g., advertisers, competitors, spouses, etc.) then sensitive information would be disclosed, especially if the client's host prefers the tunnel over native IP. Assuming the server provides good service, the user would not have reason to suspect the change.

Full interception of IP traffic could also be arranged (including pharming) which would allow any number of deception or monitoring attacks including phishing. We illustrate this with an example phishing attack scenario.

It is often assumed that the tunnel server is a trusted entity. It may be possible for malware or a malicious user to quietly change the client's tunnel server setting and have the user be unaware their trust has been misplaced for an indefinite period of time. However, malware or a malicious user can do much worse than this, so this is not a significant concern. Hence it is only important that an attacker on the network cannot change the client's server setting.

1. A phisher sets up a malicious tunnel server (or tampers with a legitimate one). This server, for the most part, provides correct service.

2. An attacker, by some means, switches the host's tunnel server setting, or spoofs a DNS reply, to point to the above server. If neither DNS nor the tunnel setup is secured (i.e., if the client does not authenticate the information), then the attacker's tunnel server is seen as legitimate.

3. A user on the victim host types their bank's URL into his/her browser.

4. The bank's hostname resolves to one or more IP addresses and the tunnel is selected for socket connection for whatever reason (e.g., the tunnel provides IPv6 connectivity and the bank has an IPv6 address).

5. The tunnel client uses the server for help in connecting to the bank's IP address. Some tunneling protocols use a separate channel for signaling vs data, but this still allows the server to place itself in the data path by an appropriate signal to the client. For example, in Teredo, the client sends a ping request through a server which is expected to come back through a data relay, and a malicious server can simply send it back itself to

indicate that is a data relay for the communication.

6. The rest works pretty much like any normal phishing transaction, except that the attacker acts as a tunnel server (or data relay, for protocols such as Teredo) and a host with the bank's IP address.

This pharming type attack is not unique to tunneling. Switching DNS server settings to a malicious DNS server or DNS cache poisoning in a recursive DNS resolver could have a similar effect.

6.1.3. Recommendations

In general, anti-phishing and anti-fraud provisions should help with aspects of this, as well as software that specifically monitors for tunnel server changes.

Perhaps the best way to mitigate tunnel-specific attacks is to have the client either authenticate the tunnel server, or at least the means by which the tunnel server's IP address is determined. For example, SSL VPNs use https URLs and hence authenticate the server as being the expected one. Another mechanism, when IPv6 Router Advertisements are sent over the tunnel is to use SEcure Neighbor Discovery (SEND) [RFC3971] to verify that the client trusts the server.

On the host, it should require an appropriate level of privilege in order to change the tunnel server setting (as well as other non-tunnel-specific settings such as the DNS server setting, etc.). Making it easy to see the current tunnel server setting (e.g., not requiring privilege for this) should help detection of changes.

The scope of the attack can also be reduced by limiting tunneling use in general but especially in preferring native IPv4 to tunneled IPv6; this is because it is reasonable to expect that banks and similar web sites will continue to be accessible over IPv4 for as long as a significant fraction of their customers are still IPv4-only. Please refer to Section 3 of [TUNNEL-LOOPS] for a detailed description and mitigation measures for a class of attacks based on IPv6 automatic tunnels.

7. Mechanisms to secure the use of tunnels

This document described several security issues with tunnels. This does not mean that tunnels need to be avoided at any cost. On the contrary, tunnels can be very useful if deployed, operated and used properly. The threats against IP tunnels are documented here. If

the threats can be mitigated, network administrators can efficiently and securely use tunnels in their network. Several measures can be taken in order to secure the operation of IPv6 tunnels:

- Operating on-premise tunnel servers/relays so that the tunneled traffic does not cross border routers.
- Setting up internal routing to steer traffic to these servers/ relays
- Setting up of firewalls [RFC2979] to allow known and controllable tunneling mechanisms and disallow unknown tunnels.

8. Acknowledgments

The authors would like to thank Remi Denis-Courmont, Fred Templin, Jordi Palet Martinez, James Woodyatt, Christian Huitema, Brian Carpenter, Nathan Ward, Kurt Zeilenga, Joel Halpern, Erik Kline, Alfred Hoenes and Fernando Gont for reviewing earlier versions of the document and providing comments to make this document better.

9. Security Considerations

This entire document discusses security concerns with tunnels.

10. IANA Considerations

This document does not require any IANA action.

11. Informative References

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

[RFC2529] Carpenter, B. and C. Jung, "Transmission of IPv6 over IPv4 Domains without Explicit Tunnels", RFC 2529, March 1999.

[RFC2661] Townsley, W., Valencia, A., Rubens, A., Pall, G., Zorn, G., and B. Palter, "Layer Two Tunneling Protocol "L2TP"", RFC 2661, August 1999.

[RFC2979] Freed, N., "Behavior of and Requirements for Internet Firewalls", RFC 2979, October 2000.

[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.

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

[RFC3971] Arkko, J., Kempf, J., Zill, B., and P. Nikander, "SEcure Neighbor Discovery (SEND)", RFC 3971, March 2005.

[RFC4380] Huitema, C., "Teredo: Tunneling IPv6 over UDP through Network Address Translations (NATs)", RFC 4380, February 2006.

[RFC4787] Audet, F. and C. Jennings, "Network Address Translation (NAT) Behavioral Requirements for Unicast UDP", BCP 127, RFC 4787, January 2007.

[RFC5095] Abley, J., Savola, P., and G. Neville-Neil, "Deprecation of Type 0 Routing Headers in IPv6", RFC 5095, December 2007.

[RFC5214] Templin, F., Gleeson, T., and D. Thaler, "Intra-Site Automatic Tunnel Addressing Protocol (ISATAP)", RFC 5214, March 2008.

[RFC5389] Rosenberg, J., Mahy, R., Matthews, P., and D. Wing, "Session Traversal Utilities for NAT (STUN)", RFC 5389, October 2008.

[RFC5991] Thaler, D., Krishnan, S., and J. Hoagland, "Teredo Security Updates", RFC 5991, September 2010.

[SECA-IP] Gont, F., "Security Assessment of the Internet Protocol version 4", draft-ietf-opsec-ip-security-03 (work in progress), April 2010.

[TSV-PORT]
Larsen, M. and F. Gont, "Transport Protocol Port Randomization Recommendations", draft-ietf-tsvwg-port-randomization-09 (work in progress), August 2010.

[TUNNEL-LOOPS]
Nakibly, G. and F. Templin, "Routing Loop Attack using IPv6 Automatic Tunnels: Problem Statement and Proposed

Mitigations", draft-ietf-v6ops-tunnel-loops-00 (work in progress), September 2010.

Authors' Addresses

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

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

Dave Thaler
Microsoft Corporation
One Microsoft Way
Redmond, WA 98052
USA

Phone: +1 425 703 8835
Email: dthaler@microsoft.com

James Hoagland
Symantec Corporation
350 Ellis St.
Mountain View, CA 94043
US

Email: Jim_Hoagland@symantec.com
URI: http://symantec.com/

v6ops Working Group
Internet-Draft
Intended status: Informational
Expires: November 20, 2011

Rajeev. Koodli
Cisco Systems
May 19, 2011

Mobile Networks Considerations for IPv6 Deployment
draft-ietf-v6ops-v6-in-mobile-networks-05.txt

Abstract

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.

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 November 20, 2011.

Copyright Notice

Table of Contents

1. Introduction 3
2. Reference Architecture and Terminology 3
3. IPv6 Considerations 5
3.1. IPv4 Address Exhaustion 5
3.2. NAT Placement in the mobile networks 7
3.3. IPv6-only Deployment Considerations 10
3.4. Fixed - Mobile Convergence 13
4. Summary and Conclusion 15
5. Security Considerations 16
6. IANA Considerations 16
7. Acknowledgement 16
8. Informative References 16
Appendix A. Change Log 18
Author's Address 18

1. Introduction

The dramatic growth of the Mobile Internet is accelerating the exhaustion of the available IPv4 addresses. It is widely accepted that IPv6 is necessary for the continued operation and growth of the Internet in general, and that of the Mobile Internet in particular. While IPv6 brings many benefits, certain unique challenges arise when deploying it in mobile networks. This document describes such challenges and outlines the applicability of the existing IPv6 deployment solutions. As such, it can be a useful reference document for service providers as well as network designers. This document does not propose any new protocols or suggest new protocol specification work.

The primary considerations that we address in this document on IPv6 deployment in mobile networks are:

- o Public and Private IPv4 address exhaustion and implications to mobile network deployment architecture;
- o Placement of Network Address Translation (NAT) functionality and its implications;
- o IPv6-only deployment considerations and roaming implications;
- o Fixed-Mobile Convergence and implications to overall architecture.

In the following sections, we discuss each of these in detail.

For the most part, we assume the 3GPP 3G and 4G network architectures specified in [3gpp.3g] and [3gpp.4g]. However, the considerations are general enough for other mobile network architectures as well [3gpp2.ehrpd].

2. Reference Architecture and Terminology

The following is a reference architecture of a mobile network.

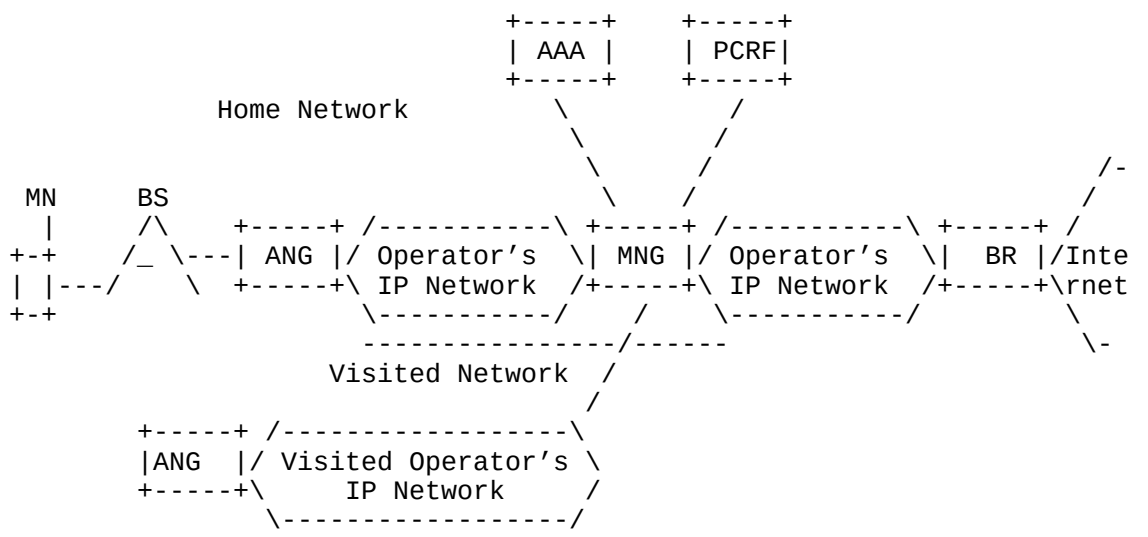


Figure 1: Mobile Network Architecture

A Mobile Node (MN) connects to the mobile network either via its Home Network or via a Visited Network when the user is roaming outside of the Home Network. In the 3GPP network architecture, a MN accesses the network by connecting to an Access Point Name (APN), which maps to a mobile gateway. Roughly speaking, an APN is similar to an SSID in wireless LAN. An APN is a logical concept which can be used to specify what kinds of services, such as Internet access, high-definition video streaming, content-rich gaming, and so on, a MN is entitled to. Each APN can specify what type of IP connectivity (i.e., IPv4, IPv6, IPv4v6) is enabled on that particular APN.

While an APN directs a MN to an appropriate gateway, the MN needs an end-to-end "link" to that gateway. In the Long-Term Evolution (LTE) networks, this link is realized through an Evolved Packet System (EPS) bearer. In the 3G UMTS networks, such a link is realized through a Packet Data Protocol (PDP) Context. The end-to-end link traverses multiple nodes which are defined below:

- o Base Station (BS): The radio Base Station provides wireless connectivity to the MN.

- o Access Network Gateway (ANG): The ANG forwards IP packets to and from the MN. Typically, this is not the MN's default router, and the ANG does not perform IP address allocation and management for the mobile nodes. The ANG is located either in the Home Network or in the Visited Network.

o The Mobile Network Gateway (MNG): The MNG is the MN's default router which provides IP address management. The MNG performs functions such as offering Quality of Service (QoS), applying subscriber-specific policy, and enabling billing and accounting; these functions are sometimes collectively referred to as "subscriber-management" operations. The mobile network architecture, as shown in the figure, defines the necessary protocol interfaces to enable subscriber management operations. The MNG is typically located in the Home Network.

o Border Router (BR): As the name implies, a BR borders the Internet for the mobile network. The BR does not perform subscriber management for the mobile network.

o Authentication, Authorization and Accounting (AAA): The general functionality of AAA is used for subscriber authentication and authorization for services, as well as for generating billing and accounting information.
In the 3GPP network environments, the subscriber authentication and the subsequent authorization for connectivity and services is provided using the "Home Location Register" (HLR)/"Home Subscriber Server" (HSS) functionality.

o Policy and Charging Rule Function (PCRF): The PCRF enables applying policy and charging rules at the MNG.

In the rest of this document, we use the terms operator, service provider or provider interchangeably.

3. IPv6 Considerations

3.1. IPv4 Address Exhaustion

It is generally agreed that the pool of public IPv4 addresses is nearing its exhaustion. The IANA has exhausted the available '/8' blocks for allocation to the Regional Internet Registries (RIRs). The RIRs themselves have either "run-out" of their blocks or are projected to exhaust them in the near future. This has led to a heightened awareness among the service providers to consider introducing technologies to keep the Internet operational. For providers, there are two simultaneous approaches to addressing the run-out problem: delaying the IPv4 address pool exhaustion (i.e., conserving their existing pool) and introducing IPv6 in operational networks. We consider both in the following.

Delaying the public IPv4 address exhaustion for providers involves assigning private IPv4 addressing for end-users, or extending an IPv4 address with the use of port ranges which requires tunneling and additional signaling. A mechanism such as the Network Address Translator (NAT) is used at the provider premises (as opposed to customer premises) to manage the private IP address assignment and access to the Internet. In the following, we primarily focus on translation based mechanisms such as NAT44 (i.e., translation from public IPv4 to private IPv4 and vice versa) and NAT64 (i.e., translation from public IPv6 to public IPv4 and vice versa). We do this because the 3GPP architecture already defines a tunneling infrastructure with the GPRS Tunneling Protocol (GTP), and the architecture allows for dual-stack and IPv6-only deployments.

In a mobile network, the IPv4 address assignment for a MN is performed by the MNG. In the 3GPP network architecture, this assignment is performed in conjunction with the PDN connectivity establishment. A PDN connection implies an end-end link (i.e., an EPS bearer in 4G LTE or a PDP context in 3G UMTS) from the MN to the MNG. There can be one or more PDN connections active at any given time for each MN. A PDN connection may support both IPv4 and IPv6 traffic (as in a dual-stack PDN in 4G LTE networks) or it may support only one of the two traffic types (as in the existing 3G UMTS networks). The IPv4 address is assigned at the time of PDN connectivity establishment, or is assigned using the DHCP protocol after the PDN connectivity is established. In order to delay the exhaustion of public IPv4 addresses, this IP address needs to be a private IPv4 address which is translated into a shared public IPv4 address. Hence, there is a need for private - public IPv4 translation mechanism in the mobile network.

In the Long-Term Evolution (LTE) 4G network, there is a requirement for an always-on PDN connection in order to reliably reach a mobile user in the All-IP network. This requirement is due to the need for supporting Voice over IP service in LTE which does not have circuit-based infrastructure. If this PDN connection were to use IPv4 addressing, a private IPv4 address is needed for every MN that attaches to the network. This could significantly affect the availability and usage of private IPv4 addresses. One way to address this is by making the always-on PDN (that requires voice service) to be IPv6. The IPv4 PDN is only established when the user needs it.

The 3GPP standards also specify a deferred IPv4 address allocation on a dual-stack IPv4v6 PDN at the time of connection establishment. This has the advantage of a single PDN for IPv6 and IPv4 along with deferring IPv4 address allocation until an application needs it. The deferred address allocation requires support for a dyncamic configuration protocol such as DHCP as well as appropriate triggers

to invoke the protocol. Such a support does not exist today in mobile phones. The newer iterations of Smartphones could provide such support. Also, the tethering of Smartphones to laptops (which typically support DHCP) could use deferred allocation depending on when a laptop attaches to the Smartphone. Until appropriate triggers and host stack support is available, the applicability of the address deferring option may be limited.

On the other hand, in the existing 3G UMTS networks, there is no requirement for an always-on connection even though many SmartPhones seldom relinquish an established PDP context. The existing so-called pre-Release-8 deployments do not support the dual-stack PDP connection. Hence, two separate PDP connections are necessary to support IPv4 and IPv6 traffic. Even though some MNs, especially the SmartPhones, in use today may have IPv6 stack, there are two remaining considerations. First, there is little operational experience and compliance testing with these existing stacks. Hence, it is expected that their use in large deployments may uncover software errors and interoperability problems which inhibit providing services based on IPv6 for such hosts. Second, only a fraction of current phones in use have such a stack. As a result, providers need to test, deploy and operationalize IPv6 as they introduce new handsets which also, continue to need, access to the predominantly IPv4 Internet.

The considerations from the preceeding paragraphs lead to the following observations. First, there is an increasing need to support private IPv4 addressing in mobile networks because of the public IPv4 address run-out problem. Correspondingly, there is a greater need for private - public IPv4 translation in the mobile networks. Second, there is support for IPv6 in both 3G and 4G LTE networks already in the form of PDP context and PDN connections. To begin with, the operators can introduce IPv6 for their own applications and services. In other words, the IETF’s recommended model of dual-stack IPv6 and IPv4 networks is readily applicable to mobile networks with the support for distinct APNs and the ability to carry IPv6 traffic on PDP/PDN connections. The IETF dual-stack model can be applied using a single IPv4v6 PDN connection in Release-8 and onwards, but requires separate PDP contexts in the earlier releases. Finally, operators can make IPv6 as the default for always-on mobile connections using either the IPv4v6 PDN or the IPv6 PDN, and use IPv4 PDNs as necessary.

3.2. NAT Placement in the mobile networks

In the previous section, we observed that the NAT44 functionality is needed in order to conserve the available pool and delay public IPv4 address exhaustion. However, the available private IPv4 pool itself

is not abundant for large networks such as mobile networks. For instance, the so-called NET10 block [RFC1918] has approximately 16.7 million private IPv4 addresses starting with 10.0.0.0. A large mobile service provider network can easily have more than 16.7 million subscribers attached to the network at a given time. Hence, the private IPv4 address pool management and the placement of NAT44 functionality becomes important.

In addition to the developments cited above, NAT placement is important for other reasons as well. Access networks generally need to produce network and service usage records for billing and accounting. This is true also for mobile networks where "subscriber management" features (i.e., QoS, Policy, and Billing and Accounting) can be fairly detailed. Since a NAT introduces a binding between two addresses, the bindings themselves become necessary information for subscriber management. For instance, the offered QoS on private IPv4 address and the (shared) public IPv4 address may need to be correlated for accounting purposes. As another example, the Application Servers within the provider network may need to treat traffic based on policy provided by the PCRF. If the IP address seen by these Application Servers is not unique, the PCRF needs to be able to inspect the NAT binding to disambiguate among the individual MNs. And, the subscriber session management information and the service usage information also need to be correlated in order to produce harmonized records. Furthermore, there may be legal requirements for storing the NAT binding records. Indeed, these problems disappear with the transition to IPv6. For now, it suffices to state that NAT introduces state which needs to be correlated and possibly stored with other routine subscriber information.

Mobile network deployments vary in their allocation of IP address pools. Some network deployments use the "centralized model" where the pool is managed by a common node, such as the PDN’s BR, and the pool shared by multiple MNGs all attached to the same BR. This model has served well in the pre-3G deployments where the number of subscribers accessing the mobile Internet at any given time has not exceeded the available address pool. However, with the advent of 3G networks and the subsequent dramatic growth in the number of users on the mobile Internet, the service providers are increasingly forced to consider their existing network design and choices. Specifically, the providers are forced to address private IPv4 pool exhaustion as well as scalable NAT solutions.

In order to tackle the private IPv4 exhaustion in the centralized model, there would be a need to support overlapped private IPv4 addresses in the common NAT functionality as well as in each of the gateways. In other words, the IP addresses used by two or more MNs (which may be attached to the same MNG) are very likely to overlap at

the centralized NAT, which needs to be able to differentiate traffic. Tunneling mechanisms such as Generic Routing Encapsulation (GRE) [RFC2784], [RFC2890], MPLS [RFC3031] VPN tunnels or even IP-in-IP encapsulation [RFC2003] which can provide a unique identifier for a NAT session can be used to separate overlapping private IPv4 traffic as described in [gi-ds-lite]. An advantage of centralizing the NAT and using the overlapped private IPv4 addressing is conserving the limited private IPv4 pool. It also enables the operator's enterprise network to use IPv6 from the MNG to the BR; this (i.e., the need for an IPv6-routed enterprise network) may be viewed as an additional requirement by some providers. The disadvantages include the need for additional protocol to correlate the NAT state (at the common node) with subscriber session information (at each of the gateways), suboptimal MN - MN communication, absence of subscriber-aware NAT (and policy) function, and of course the need for a protocol from the MNG to BR itself. Also, if the NAT function were to experience failure, all the connected gateway service will be affected. These drawbacks are not present in the "distributed" model which we discuss in the following.

In a distributed model, the private IPv4 address management is performed by the MNG which also performs the NAT functionality. In this model, each MNG has a block of 16.7 million unique addresses, which is sufficient compared to the number of mobile subscribers active on each MNG. By distributing the NAT functionality to the edge of the network, each MNG is allowed to re-use the available NET10 block which avoids the problem of overlapped private IPv4 addressing at the network core. In addition, since the MNG is where subscriber management functions are located, the NAT state correlation is readily enabled. Furthermore, an MNG already has existing interfaces to functions such as AAA and PCRF, which allows it to perform subscriber management functions with the unique private IPv4 addresses. Finally, the MNG can also pass-through certain traffic types without performing NAT to the application servers located within the service provider's domain, which allows the servers to also identify subscriber sessions with unique private IPv4 addresses. The disadvantages of the "distributed model" include the absence of centralized addressing and centralized management of NAT.

In addition to the two models described above, a hybrid model is to locate NAT in a dedicated device other than the MNG or the BR. Such a model would be similar to the distributed model if the IP pool supports unique private addressing for the mobile nodes, or it would be similar to the centralized model if it supports overlapped private IP addresses. In any case, the NAT device has to be able to provide the necessary NAT session binding information to an external entity (such as AAA or PCRF) which then needs to be able to correlate those records with the user's session state present at the MNG.

The foregoing discussion can be summarized as follows: First, the management of available private IPv4 pool has become important given the growth of the mobile Internet users. The mechanisms that enable re-use of the available pool are required. Second, in the context of private IPv4 pool management, the placement of NAT functionality has implications to the network deployment and operations. The centralized models with a common NAT have the advantages of continuing their legacy deployments and the re-use of private IPv4 addressing. However, they need additional functions to enable traffic differentiation and NAT state correlation with subscriber state management at the MNG. The distributed models also achieve private IPv4 address re-use and avoid overlapping private IPv4 traffic in the operator's core, but without the need for additional mechanisms. Since the MNG performs (unique) IPv4 address assignment and has standard interfaces to AAA and PCRF, the distributed model also enables a single point for subscriber and NAT state reporting as well as policy application. In summary, providers interested in readily integrating NAT with other subscriber management functions, as well as conserving and re-using their private IPv4 pool, may find the distributed model compelling. On the other hand, those providers interested in common management of NAT may find the cetralized model more compelling.

3.3. IPv6-only Deployment Considerations

As we observed in the previous section, the presence of NAT functionality in the network brings multiple issues which would otherwise be absent. NAT should be viewed as an interim solution until IPv6 is widely available, i.e., IPv6 is available for mobile users for all (or most) practical purposes. Whereas NATs at provider premises may slow down the exhaustion of public IPv4 addresses, expeditious and simultaneous introduction of IPv6 in the operational networks is necessary to keep the "Internet going and growing". Towards this goal, it is important to understand the considerations in deploying IPv6-only networks.

There are three dimensions to IPv6-only deployments: the network itself, the mobile nodes and the applications, represented by the 3-tuple {nw, mn, ap}. The goal is to reach the co-ordinate {IPv6, IPv6, IPv6} from {IPv4, IPv4, IPv4}. However, there are multiple paths to arrive at the goal. The classic dual-stack model would traverse the co-ordinate {IPv4v6, IPv4v6, IPv4v6}, where each dimension supports co-existence of IPv4 and IPv6. This appears to be the path of least disruption, although we are faced with the implications of supoorting large-scale NAT in the network. There is also the cost of supporting separate PDP contexts in the existing 3G UMTS networks. The other intermediate co-ordinate of interest is {IPv6, IPv6, IPv4}, where the network and the MN are IPv6-only, and

the Internet applications are recognized to be predominantly IPv4. This transition path would, ironically, require interworking between IPv6 and IPv4 in order for the IPv6-only MNs to be able to access IPv4 services and applications on the Internet. In other words, in order to disengage NAT (for IPv4 - IPv4), we need to introduce another form of NAT (i.e., IPv6 - IPv4) to expedite the adoption of IPv6.

It is interesting to consider the preceeding discussion surrounding the placement of NAT for IPv6 - IPv4 interworking. There is no overlapping private IPv4 address problem because each IPv6 address is unique and there are plenty of them available. Hence, there is also no requirement for (IPv6) address re-use, which means no protocol is necessary in the centralized model to disambiguiate NAT sessions. However, there is an additional requirement of DNS64 [dns64] functionality for IPv6 - IPv4 translation. This DNS64 functionality must ensure that the synthesized AAAA record correctly maps to the IPv6 - IPv4 translator.

The IPv6-only deployments in mobile networks need to reckon with the following considerations. First, both the network and the MNs need to be IPv6-capable. Expedited network upgrades as well as roll-out of MNs with IPv6 would greatly facilitate this. Fortunately, the 3GPP network design for LTE already requires the network nodes and the mobile nodes to support IPv6. Even though there are no requirements for the transport network to be IPv6, an operational IPv6 connectivity service can be deployed with appropriate existing tunneling mechanisms in the IPv4-only transport network. Hence a service provider may choose to enforce IPv6-only PDN and address assignment for their own subscribers in their Home Networks, see Figure 1. This is feasible for the newer MNs when the mobile network is able to provide IPv6-only PDN support and IPv6 - IPv4 interworking for Internet access. For the existing MNs however, the provider still needs to be able to support IPv4-only PDP/PDN connectivity.

Migration of applications to IPv6 in MNs with IPv6-only PDN connectivity brings challenges. The applications and services offered by the provider obviously need to be IPv6-capable. However, a MN may host other applications which also need to be IPv6-capable in IPv6-only deployments. This can be a "long-tail" phenomenon; however, when a few prominant applications start offering IPv6, there can be a strong incentive to provide application layer (e.g., socket interface) upgrades to IPv6. Also, some IPv4-only applications may be able to make use of alternative access such as WiFi when available. A related challenge in the migration of applications is the use of IPv4 literals in application layer protocols (such as XMPP) or content (as in html or xml). Some Internet applications expect their clients to supply IPv4 addresses as literals, and this

will not be possible with IPv6-only deployments. Some of these experiences and the related considerations in deploying IPv6-only network are documented in [arkko-v6]. In summary, migration of applications to IPv6 needs to be done, and such a migration is not expected to be uniform across all subsets of existing applications.

Voice over LTE (VoLTE) also brings some unique challenges. The signaling for voice is generally expected to be available for free while the actual voice call itself is typically charged on its duration. Such a separation of signaling and the payload is unique to voice, whereas an Internet connection is accounted without specifically considering application signaling and payload traffic. This model is expected to be supported even during roaming. Furthermore, the providers and the users generally require the voice service regardless of roaming whereas the Internet usage is subject to subscriber preferences and roaming agreements. This requirement to ubiquitously support voice service while providing the flexibility for Internet usage exacerbates the addressing problem, and may hasten provisioning of VoLTE using the IPv6-only PDN.

As seen earlier, roaming is unique to mobile networks and it introduces new challenges. The service providers can control their own network design but not their peer's networks which they rely on for roaming. The users expect uniformity in experience even when they are roaming. This imposes a constraint on providers interested in IPv6-only deployments to also support IPv4 addressing when their own (outbound) subscribers roam to networks which do not offer IPv6. For instance, when an LTE deployment is IPv6-only, a roamed 3G network may not offer IPv6 PDN connectivity. Since a PDN connection involves the radio base station, the ANG and the MNG (See Figure 1), it would not be possible to enable IPv6 PDN connectivity without the roamed network support. These considerations also apply when the visited network is used for offering services such as VoLTE in the so-called Local Breakout model; the roaming MN's capability as well as the roamed network capability to support VoLTE using IPv6 determine whether fallback to IPv4 would be necessary. Similarly, there are inbound roamers to an IPv6-ready provider network whose MN's are not capable of IPv6. The IPv6-ready provider network has to be able to support IPv4 PDN connectivity for such inbound roamers. There are encouraging signs that the existing deployed network nodes in the 3GPP architecture already provide support for IPv6 PDP context. It would be necessary to scale this support for a (very) large number of mobile users and offer it as a ubiquitous service which can be accounted for.

In summary, IPv6-only deployments should be encouraged along-side the dual-stack model which is the recommended IETF approach. This is relatively straightforward for an operator's own services and

applications, provisioned through an appropriate APN and the corresponding IPv6-only PDP or EPS bearer. Some providers may consider IPv6-only deployment for Internet access as well, and this would require IPv6 - IPv4 interworking. When the IPv6 - IPv4 translation mechanisms are used in IPv6-only deployments, the protocols and the associated considerations specified in [xlate-stateful] and [xlate-stateless] apply. Finally, such IPv6-only deployments can be phased-in for newer mobile nodes, while the existing ones continue to demand IPv4-only connectivity.

Roaming is important in mobile networks and roaming introduces diversity in network deployments. Until IPv6 connectivity is available in all mobile networks, IPv6-only mobile network deployments need to be prepared to support IPv4 connectivity (and NAT44) for their own outbound roaming users as well as for inbound roaming users. However, by taking the initiative to introduce IPv6-only for the newer MNs, the mobile networks can significantly reduce the demand for private IPv4 addresses.

3.4. Fixed - Mobile Convergence

Many service providers have both fixed broadband and mobile networks. Access networks are generally disparate, with some common characteristics but with enough differences to make it challenging to achieve "convergence". For instance, roaming is not a consideration in fixed access networks. An All-IP mobile network service provider is required to provide voice service, whereas this is not required for a fixed network provider. A "link" in fixed networks is generally capable of carrying IPv6 and IPv4 traffic, whereas not all mobile networks have "links" (i.e., PDP/PDN connections) capable of supporting IPv6 and IPv4. Indeed roaming makes this problem worse when a portion of the link (i.e., the Home Network in Figure 1) is capable of supporting IPv6 and the other portion of the link (i.e., the Visited Network in Figure 1) is not. Such architectural differences, as well as policy and business model differences make convergence challenging.

Nevertheless, within the same provider’s space, some common considerations may apply. For instance, IPv4 address management is a common concern for both of the access networks. This implies that the same mechanisms discussed earlier, i.e., delaying IPv4 address exhaustion and introducing IPv6 in operational networks, apply for the converged networks as well. However, the exact solutions deployed for each access network can vary for a variety of reasons. For instance:

o Tunneling of private IPv4 packets within IPv6 is feasible in fixed networks where the end-point is often a cable or DSL modem. This is not the case in mobile networks where the end-point is a MN itself.

o Encapsulation-based mechanisms such as 6rd [RFC5969] are useful where the operator is unable to provide native or direct IPv6 connectivity and a residential gateway can become a tunnel end-point for providing this service. In mobile networks, the operator could provide IPv6 connectivity using the existing mobile network tunneling mechanisms without introducing an additional layer of tunneling.

o A mobile network provider may have application servers (e.g., an email server) in its network that require unique private IPv4 addresses for MN identification, whereas a fixed network provider may not have such a requirement or the service itself.

These examples illustrate that the actual solutions used in an access network are largely determined by the requirements specific to that access network. Nevertheless, some sharing between access and core network may be possible depending on the nature of the requirement and the functionality itself. For example, when a fixed network does not require a subscriber-aware feature such as NAT, the functionality may be provided at a core router while the mobile access network continues to provide the NAT functionality at the mobile gateway. If a provider chooses to offer common subscriber management at the MNG for both fixed and wireless networks, the MNG itself becomes a convergence node that needs to support the applicable transition mechanisms for both fixed and wireless access networks.

Different access networks of a provider are more likely to share a common core network. Hence, common solutions can be more easily applied in the core network. For instance, configured tunnels or MPLS VPNs from the gateways from both mobile and fixed networks can be used to carry traffic to the core routers, until the entire core network is IPv6-enabled.

There can also be considerations due to the use of NAT in access networks. Solutions such as Femto Networks rely on a fixed Internet connection being available for the Femto Base Station to communicate with its peer on the mobile network, typically via an IPsec tunnel. When the Femto Base Station needs to use a private IPv4 address, the mobile network access through the Femto Base Station will be subject to NAT policy administration including periodic clean-up and purge of NAT state. Such policies affect the usability of the Femto Network, and has implications to the mobile network provider. Using IPv6 for

the Femto (or any other access technology) could alleviate some of these concerns if the IPv6 communication could bypass the NAT.

In summary, there is interest in fixed-mobile convergence at least among some providers. While there are benefits from harmonizing the network as much as possible, there are also idiosyncrasies of disparate access networks which influence the convergence. Perhaps greater harmonization is feasible at the higher service layers, e.g., in terms of offering unified user experience for services and applications. Some harmonization of functions across access networks into the core network may be feasible. A provider's core network appears to the place where most convergence is feasible.

4. Summary and Conclusion

IPv6 deployment in mobile networks is crucial for the mobile Internet. In this document, we discussed the considerations in deploying IPv6 in mobile networks. We summarize the discussion in the following:

- o IPv4 address exhaustion and its implications to mobile networks: As the mobile service providers begin to deploy IPv6, conserving their available IPv4 pool implies the need for network address translation in mobile networks. At the same time, providers can make use of the 3GPP architecture constructs such as the APN and PDN connectivity to introduce IPv6 without affecting the predominantly IPv4 Internet access. The IETF dual-stack model [RFC4213] can be applied to the mobile networks readily.

- o The placement of NAT functionality in mobile networks: Both the centralized and distributed models of private IPv4 address pool management have their relative merits. By enabling each MNG to manage its own NET10 pool, the distributed model achieves re-use of available private IPv4 pool and avoids the problems associated with the non-unique private IPv4 addresses for the MNs without additional protocol mechanisms. The distributed model also augments the "subscriber management" functions at an MNG, such as readily enabling NAT session correlation with the rest of the subscriber session state. On the other hand, the existing deployments which have used the centralized IP address management can continue their legacy architecture by placing the NAT at a common node. The centralized model also achieves private IPv4 address re-use, but needs additional protocol extensions to differentiate overlapping addresses at the common NAT as well as to integrate with policy and billing infrastructure.

o IPv6-only mobile network deployments: This deployment model is feasible in the LTE architecture for an operator's own services and applications. The existing MNs still expect IPv4 address assignment. And, roaming which is unique to mobile networks, requires that a provider support IPv4 connectivity when their (outbound) users roam into a mobile network that is not IPv6-enabled. Similarly, a provider needs to support IPv4 connectivity for (inbound) users whose MNs are not IPv6-capable. The IPv6 - IPv4 interworking is necessary for IPv6-only MNs to access IPv4 Internet.

o Fixed-Mobile Convergence: The examples discussed illustrate the differences in the requirements of fixed and mobile networks. While some harmonization of functions may be possible across the access networks, the service provider's core network is perhaps better-suited for converged network architecture. Similar gains in convergence are feasible in the service and application layers.

5. Security Considerations

This document does not introduce any new security vulnerabilities.

6. IANA Considerations

This document does not require any actions from IANA.

7. Acknowledgement

This document has benefitted from discussions with and reviews from Cameron Byrne, David Crowe, Hui Deng, Remi Despres, Fredrik Garneij, Jouni Korhonen, Teemu Savolainen and Dan Wing; thanks to all of them. Mohamed Boucadair provided an extensive review of individual draft version 01 of this document; many thanks Mohamed. Cameron Byrne, Kent Leung, Kathleen Moriarty and Jari Arkko provided reviews which have helped improve this document. Thanks to Nick Heatley for providing valuable review and input on VoLTE.

8. Informative References

[3gpp.3g] "General Packet Radio Service (GPRS); Service description; Stage 2, 3GPP TS 23.060, December 2006", .

[3gpp.4g] "General Packet Radio Service (GPRS);enhancements for Evolved Universal Terrestrial Radio Access Network (E-UTRAN) access", 3GPP TS 23.401 8.8.0, December 2009.", .

[3gpp2.ehrpd] "E-UTRAN - eHRPD Connectivity and Interworking: Core Network Aspects", http://www.3gpp2.org/Public_html/Misc/X.P0057-0_v0.13_E-UTRAN-eHRPD_Interworking_VV_Due_5_December-2008.pdf.

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

[RFC2003] Perkins, C., "IP Encapsulation within IP", RFC 2003, October 1996.

[RFC2784] Farinacci, D., Li, T., Hanks, S., Meyer, D., and P. Traina, "Generic Routing Encapsulation (GRE)", RFC 2784, March 2000.

[RFC2890] Dommety, G., "Key and Sequence Number Extensions to GRE", RFC 2890, September 2000.

[RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol Label Switching Architecture", RFC 3031, January 2001.

[RFC4213] Nordmark, E. and R. Gilligan, "Basic Transition Mechanisms for IPv6 Hosts and Routers", RFC 4213, October 2005.

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

[arkko-v6] Arkko, J. and A. Keranen, "Experiences from an IPv6-Only Network", draft-arkko-ipv6-only-experience-01, Jul 2010.

[dns64] Bagnulo, M., Sullivan, A., Matthews, P., and I. van Beijnum, "DNS64: DNS extensions for Network Address Translation from IPv6 Clients to IPv4 Servers", draft-ietf-behave-dns64-11, Mar 2010.

[gi-ds-lite] Brockners et al., F., "Gateway Initiated Dual-stack Lite Deployment", draft-ietf-softwire-gateway-init-ds-lite, Oct 2009.

[xlate-stateful]
Bagnulo, M., Matthews, P., and I. van Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", draft-ietf-behave-v6v4-xlate-stateful-11, Mar 2010.

[xlate-stateless]
Li, X., Bao, C., and F. Baker, "IP/ICMP Translation Algorithm", draft-ietf-behave-v6v4-xlate-20, May 2010.

Appendix A. Change Log

Revisions (from draft-koodli-**), descending chronological order

- o: More IESG reviews
- o: Addressed IESG reviews
- o: VoLTE related text
- o: FMC, Femto Networks text
- o: Dedicated NAT device model (in addition to the centralized and distributed models)
- o: IPv6-only deployment considerations: - IPv4 literals discussion and reference, - IPv6 prefix assignment clarification, - DNS64 requirement and reference
- o: Overall revisions based on comments from reviews (C. Byrne, K. Leung)
- o: Dual-stack being the recommended model, while encouraging IPv6-only deployments.
- o: Clarifications on on-demand IPv4 PDN usage, DHCP usage and on-demand IPv4 assignment.
- o: Clarifications regarding IPv6-only deployment: Roaming and Applications considerations.

Author's Address

Rajeev Koodli
Cisco Systems
USA

Email: rkoodli@cisco.com

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: June 12, 2011

S. Kawamura
NEC BIGLOBE, Ltd.
E. Jankiewicz
SRI International, Inc.

December 9, 2010

A Basic Guideline for Listing ISPs that Run IPv6
draft-kawamura-ipv6-isp-listings-01

Abstract

There are many web sites that list IPv6 enabled service providers, or attempt to categorize the IPv6 capability of ISPs. While these opinions are helpful, there is no standard criteria used by the sites, so it is difficult to compare the results. This document surveys current listings, and proposes a set of guidelines that could be taken into consideration by theses sites, or by anyone looking to evaluate an ISP's IPv6 capability. This guideline can also be used as a checklist by ISPs planning activation of IPv6 in their network.

Status of this Memo

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

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

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

This Internet-Draft will expire on June 12, 2011.

Copyright Notice

1. Introduction

There are many web sites that give listings of IPv6 enabled service providers, or rate ISPs according to their IPv6 enabledness. Appendix A gives examples of these.

There are several motivations for these listings which benefit both the ISPs and the users. It gives ISPs a goal to work for in turning up IPv6, i.e. earning a rating as "IPv6 capable". It also can be used by ISPs for publicity, a platform for telling the world that their service is ready for IPv4 address exhaustion. Listings can also be a guide for users to select the IPv6 capability they want when they choose their ISP, assuming they have a choice in their service area.

This document surveys examples of currently known listings, and proposes a set of basic guidelines that can be used in revised or new listings like this or by individuals evaluating an ISP’s capability. These guidelines would help those that intend to start such programs. It may also help in keeping one listing or rating guideline from being widely different from another, so it would not confuse users who decided to choose ISPs on the basis that the ISP is on one of these IPv6 enabled service provider listings.

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. Examples of Listing Criteria

2.1. IPv6 Enabled Program

The IPv6 enabled program (http://ipv6forum.org/ipv6_enabled/) lists ISPs at two levels: basic and advanced. At the time of this writing, the advanced level list has not been started yet. The requirements for being listed in the basic list are, to have a prefix assigned or allocated (IPv6 enabled program does not check if the prefix is an assignment or allocation), have a global AS route it, and keep reachability as much as possible.

The IPv6 Enabled Program checks the following.

2.1.1. Network Accessibility

The ISP's AS number is checked against a database to see if the AS exists and is unique.

2.1.2. Active IPv6 Address Requirement

The ISP's IPv6 prefix is checked against a database to see if the applying ISP is the rightful owner. Actual traffic to the prefix from a customer is also checked. Checking at the time of writing is done by using a script that the ISP will paste to a web site, and the script checks if it was accessed via IPv6.

2.1.3. Persistence of IPv6 Service Reachability

The check noted in the previous section is done periodically to check global reachability.

2.2. IPv6 Ripeness

IPv6 Ripeness (http://labs.ripe.net/content/ipv6-ripeness/) is part of a study conducted by RIPE NCC. Stars are given to LIRs registered in the RIPE NCC service region by checking there status in IPv6 deployment.

2.2.1. Criteria

Stars are earned by checking the following criteria.

- Have an IPv6 prefix allocated or a PI assigned.
- Prefix is visible in the Routing Information System(RIS).
- A route6 object is registered in the RIPE database.
- Reverse DNS is setup for the IPv6 prefix.

2.3. Summary of the Checking Criteria

The programs discussed in this section share these criteria in common.

- Have an IPv6 prefix allocated or a PI assigned.
- Prefix is visible in a routing database.

IPv6 Ripeness also checks if a route6 is registered (have good routing manners), and a reverse DNS is set up. IPv6 Enabled Program checks for actual traffic which requires the presence of an active web server inside the ISP.

3. Guidelines for Listing an IPv6 Enabled ISP

3.1. Scope of the Guideline

This guideline can be used to check any LIR or a PI address holder, that claims to be an ISP. The guideline is only intended to check an ISP's network accessibility. In turn, this guideline can also be used as a minimum requirement checklist by ISPs who want to newly turn up IPv6 in their network.

3.2. Levels of the Listing

We divide the listing into three levels, Experimental, Basic, and Advanced. Experimental level is what is a minimal set of capabilities for any ISP to claim that they have some form of IPv6 working and available to some subset of customers. The Experimental level will not guarantee that the ISP has a fully working or production quality IPv6 network or that IPv6 service is available to all customers. The Experimental level is what is absolutely necessary to provide service defined in [RFC5211] section 2.1 as PREP1+PREP2+PREP3 strengthened by the addition of section 2.2 "Trans1". This means that in addition to preparing for IPv6 deployment, an Experimental level ISP MUST offer IPv6-based Internet Service to at least some customers as a trial.

The Basic level will take the requirements one step further in bring the level of deployment closer to the quality of the IPv4 network. The Basic level includes what is absolutely necessary to provide service defined as MUST in [RFC5211] section 2.2 as TRANS1+TRANS2+ TRANS3 and to the extent possible the capabilities defined as SHOULD.

The requirements of the Basic level should be covered in order to provide any of the service types defined in the General Terminology section in [RFC4084].

The Advanced level will take the requirements further to bring the level of deployment and support to parity with what is generally recognized as "full production support" in the IPv4 services offered by ISPs today. This corresponds to the service level defined in [RFC5211] section 2.3 as POST1+POST2+POST3.

3.3. Experimental

The Experimental level listing checks an ISP to meet the following criteria.

- Have an IPv6 prefix allocated or a PI assigned.
- Prefix is visible in at least one routing database.
- Have at least one server with an IPv6 address where accessibility can be checked.

3.4. Basic

The Basic level listing checks an ISP to meet the following criteria.

- Reverse DNS for is set up for allocated prefixes.
- DNS cache servers are accessible via IPv6 transport.
- Path MTU discovery [RFC1981] is functional and is not filtered.
- Prefix visibility is seen in at least two routing databases belonging in different regions of the world.
- Some form of support is available to customers and to operators that want to contact the ISP on an issue that cannot be resolved within their network.
- Mail exchange(MX) servers are accessible via IPv6.

3.5. Advanced

Detailed criteria for Advanced level are difficult to specify, as they depend on the specific operational characteristic of the particular network. In general the Advanced level listing requires an ISP to meet the following criteria, essentially full parity with IPv4 level of service.

- The capabilities described in Basic level MUST be available to all customers by default.
- Full support for IPv6 services comparable to support for IPv4 services MUST be available to all customers and operators.
- All public websites provided by the ISP for customer and other operators SHOULD be accessible from an IPv6-only client.

3.6. Considerations

The listings can be made more useful if checking is done according to the target users of the ISP service. ISP for residential, ISP for ISPs (transit providers), ISP for enterprises, and ISP for data centers have different requirements. This document does not go into discussing the requirements for each type of services are. This document intends to discuss the requirements that should be common to any services provided by any ISP.

4. Security Considerations

This draft does not introduce any new Security Considerations.

5. IANA Considerations

None.

6. Acknowledgements

The author would like to thank the Task Force on IPv4 Address Exhaustion, Japan. Parts of this document was inspired from work by Brian Carpenter and Sheng Jiang. Thanks to Vesna Manojlovic for providing generous input to the draft.

7. References

7.1. Normative References

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

7.2. Informative References

[RFC1981] McCann, J., Deering, S., and J. Mogul, "Path MTU Discovery for IP version 6", RFC 1981, August 1996.

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

[RFC4084] Klensin, J., "Terminology for Describing Internet Connectivity", BCP 104, RFC 4084, May 2005.

Appendix A. Links to Listing Programs

Below are some programs that list IPv6 enabled service providers.

IPv6 Enabled Program http://ipv6forum.org/ipv6_enabled/

IPv6 Ripeness http://labs.ripe.net/content/ipv6-ripeness/

SixXS http://www.sixxs.net/wiki/IPv6_Enabled_Service_Providers

IPv6 to Standard http://www.ipv6-to-standard.org/

Hurricane Electric IPv6 Progress Report
http://bgp.he.net/ipv6-progress-report.cgi

Authors' Addresses

Seiichi Kawamura
NEC BIGLOBE, Ltd.
14-22, Shibaura 4-chome
Minatoku, Tokyo 108-8558
JAPAN

Email: kawamucho@mesh.ad.jp

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

Email: edward.jankiewicz@sri.com

6man
Internet-Draft
Intended status: Standards Track
Expires: April 11, 2011

M. Kohno
Juniper Networks, Keio University
B. Nitzan
Juniper Networks
R. Bush
Y. Matsuzaki
Internet Initiative Japan
L. Colitti
Google
T. Narten
IBM Corporation
October 8, 2010

Using 127-bit IPv6 Prefixes on Inter-Router Links
draft-kohno-ipv6-prefixlen-p2p-03.txt

Abstract

On inter-router point-to-point links, it is useful for security and other reasons, to use 127-bit IPv6 prefixes. Such a practice parallels the use of 31-bit prefixes in IPv4 [RFC3021]. This document specifies motivation and usages of 127-bit IPv6 prefix lengths on inter-router point-to-point links.

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 11, 2011.

Copyright Notice

Table of Contents

1. Conventions Used In This Document 3
2. Introduction 3
3. Scope Of This Memo 3
4. Problems identified with 127-bit prefix lengths in the past . . 4
5. Reasons for using longer prefixes 4
5.1. Ping-pong issue 4
5.2. Neighbor Cache Exhaustion issue 4
5.3. Other reasons 5
6. Recommendations 6
7. Security Considerations 6
8. IANA Considerations 6
9. Contributors 6
10. Acknowledgments 6
11. References 7
11.1. Normative References 7
11.2. Informative References 7
Authors' Addresses 7

1. Conventions Used In This Document

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. Introduction

[RFC4291] specifies that interface IDs for all unicast address, except those that start with the binary value 000, are required to be 64 bits long and to be constructed in Modified EUI-64 format. In addition, it defines the Subnet-Router anycast address, which is intended to be used for applications where a node needs to communicate with any one of the set of routers on a link.

Some operators have been using 127-bit prefixes, but this has been discouraged due to conflicts with Subnet-Router anycast [RFC3627]. However, using 64-bit prefixes creates security issues which are particularly problematic on inter-router links, and there are other valid reasons to use prefixes longer than 64 bits, in particular /127 (see Section 5).

This document provides rationale for using 127-bit prefix lengths, reevaluates the reasons why doing so was considered harmful, and specifies how /127 prefixes can be used on inter-router links configured for use as point-to-point links.

3. Scope Of This Memo

This document is applicable to cases where operators assign specific addresses on inter-router point-to-point links and do not rely on link-local addresses. Many operators assign specific addresses for purposes of network monitoring, reverse DNS resolution for traceroute and other management tools, EBGP peering sessions, and so on.

For the purposes of this document, an inter-router point-to-point link is a link to which only two routers and no hosts are attached. This may include Ethernet links which are configured to be point-to-point. In such cases, there is no need to support Neighbor Discovery for address resolution, and other general scenarios like the use of stateless address autoconfiguration are not relevant.

Links between a router and a host, or links to which both routers and hosts are attached, are out of scope of this document.

4. Problems identified with 127-bit prefix lengths in the past

[RFC3627] discourages the use of 127-bit prefix lengths due to conflicts with the Subnet-Router anycast addresses, while stating that the utility of Subnet-Router Anycast for point-to-point links is questionable.

[RFC5375] also says the usage of 127-bit prefix lengths is not valid and should be strongly discouraged, but the stated reason for doing this is to be in compliance with [RFC3627].

Though the analyses in the RFCs are correct, operational experience with IPv6 has shown that /127 prefixes can be used successfully.

5. Reasons for using longer prefixes

There are reasons network operators use IPv6 prefix lengths greater than 64, particularly 127, for inter-router point-to-point links.

5.1. Ping-pong issue

A forwarding loop may occur on a point-to-point link with a prefix length shorter than 127. This does not affect interfaces that perform Neighbor Discovery, but some point-to-point links, which uses medium such as SONET, do not use Neighbor Discovery. As a consequence, configuring any prefix length shorter than 127 bits on these links can create an attack vector in the network.

The pingpong issue happens in case of IPv4 as well. But due to the scarcity of IPv4 address space, the current practice is to assign long prefix lengths such as /30 or /31 [RFC3021] on point-to-point links, thus the problem did not come to the fore.

The latest ICMPv6 specification [RFC4443] mitigates this problem by specifying that a router receiving a packet on a point-to-point link, which is destined to an address within a subnet assigned to that same link (other than one of the receiving router’s own addresses), MUST NOT forward the packet back on that link. Instead, it SHOULD generate an ICMPv6 Destination Unreachable message code 3 in response. This check is on the forwarding processing path, so it may have performance impact.

5.2. Neighbor Cache Exhaustion issue

As described in Section 4.3.2 of [RFC3756], the use of a 64-bit prefix length on an inter-router link that uses Neighbor Discovery (e.g., Ethernet) potentially allows for denial-of-service attacks on

the routers on the link.

Consider an Ethernet link between two routers A and B to which a /64 subnet has been assigned. A packet sent to any address on the /64 (except the addresses of A and B) will cause the router attempting to forward it to create an new cache entry in state INCOMPLETE, send a Neighbor Solicitation message to be sent on the link, start a retransmit timer, and so on [RFC4861].

By sending a continuous stream of packets to a large number of the 2^64 - 3 unassigned addresses on the link (one for each router and one for Subnet-Router Anycast), an attacker can create a large number of neighbor cache entries and send a large number of Neighbor Solicitation packets which will never receive replies, thereby consuming large amounts of memory and processing resources. Sending the packets to one of the 2^24 addresses on the link which has the same Solicited-Node multicast address as one of the routers also causes the victim to spend large amounts of processing time discarding useless Neighbor Solicitation messages.

Careful implementation and rate-limiting can limit the impact of such an attack, but are unlikely to neutralize it completely. Rate-limiting neighbor solicitation messages will reduce CPU usage, and following the garbage-collection recommendations in [RFC4861] will maintain reachability, but if the link is down and neighbor cache entries have expired while the attack is ongoing, legitimate traffic (for example, BGP sessions) over the link might never be re-established because the routers cannot resolve each others' IPv6 addresses to MAC addresses.

This attack is not specific to point-to-point links, but is particularly harmful in the case of point-to-point backbone links, which may carry large amounts of traffic to many destinations over long distances.

While there are a number of ways to mitigate this kind of issue, assigning /127 subnets eliminates it completely.

5.3. Other reasons

Though address space conservation considerations are less important for IPv6 than they are in IPv4, some operators prefer not to assign /64s to individual point-to-point links. Instead, they may be able to number all of their point-to-point links out of a single (or small number of) /64s.

6. Recommendations

Routers MUST support the assignment of /127 prefixes on point-to-point inter-router links.

When assigning and using any /127 prefixes, the following considerations apply.Some addresses have special meanings, in particular addresses corresponding to reserved anycast addresses. When assigning prefixes (and addresses) to links, care should be taken to ensure that addresses reserved for such purposes aren't inadvertently assigned and used as unicast addresses. Otherwise, nodes may receive packets that they are not intended to receive. Specifically, assuming that a number of point-to-point links will be numbered out of a single /64 prefix:

a) Addresses with all zeros in the rightmost 64 bits SHOULD NOT be assigned as unicast addresses, to avoid colliding with the Subnet-Router anycast address. [RFC4291]

b) Addresses in which the rightmost 64 bits are assigned the highest 128 values SHOULD NOT be used as unicast addresses, to avoid colliding with Reserved Subnet Anycast Addresses. [RFC2526]

7. Security Considerations

Section 5.1 and 5.2 discuss about seurity related issues.

8. IANA Considerations

None.

9. Contributors

Chris Morrow, morrowc@google.com

Pekka Savola, pekkas@netcore.fi

Remi Despres, remi.despres@free.fr

Seiichi Kawamura, karamucho@mesh.ad.jp

10. Acknowledgments

We'd like to thank Ron Bonica, Pramod Srinivasan, Olivier Vautrin,

Tomoya Yoshida, Warren Kumari and Tatsuya Jinmei for their helpful inputs.

11. References

11.1. Normative References

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

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

[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

[RFC2526] Johnson, J. and S. Deering, "Reserved IPv6 Subnet Anycast Addresses", RFC 2526, March 1999.

[RFC3021] Retana, A., White, R., and V. Fuller, "Using 31-Bit Prefixes on IPv4 Point-to-Point Links", December 2000.

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

[RFC3756] Nikander, P., Kempf, J., and E. Nordmark, "IPv6 Neighbor Discovery (ND) Trust Models and Threats", RFC 3756, May 2004.

[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.

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

Authors' Addresses

Miya Kohno
Juniper Networks, Keio University
Shinjuku Park Tower, 3-7-1 Nishishinjuku
Shinjuku-ku, Tokyo 163-1035
Japan

Email: mkohno@juniper.net

Becca Nitzan
Juniper Networks
1194 North Marhilda Avenue
Sunnyvale, CA 94089
USA

Email: nitzan@juniper.net

Randy Bush
Internet Initiative Japan
5147 Crystal Springs
Bainbridge Island, WA 98110
USA

Email: randy@psg.com

Yoshinobu Matsuzaki
Internet Initiative Japan
Jinbocho Mitsui Building,
1-105 Kanda Jinbo-cho, Tokyo 101-0051
Japan

Email: maz@iij.ad.jp

Lorenzo Colitti
Google
1600 Amphitheatre Parkway,
Mountain View, CA 94043
USA

Email: lorenzo@google.com

Thomas Narten
IBM Corporation
3039 Cornwallis Ave.
PO Box 12195 - BRQA/502 Research Triangle Park, NC 27709-2195
USA

Email: narten@us.ibm.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 3
2. Terminology and Acronyms 4
3. Prefix Delegation Using DHCPv6 5
3.1. Prefix Request Procedure for Stateless Address Configuration 5
3.2. Prefix Request Procedure for Stateful Address Configuration 7
3.3. MN as Requesting Router in Prefix Delegation 8
3.4. Prefix Release Procedure 8
3.5. Miscellaneous Considerations 9
3.5.1. How to Generate IAID 9
3.5.2. Policy to Delegate Prefixes 10
4. Prefix Delegation Using RADIUS and Diameter 10
5. Security Considerations 11
6. IANA Considerations 11
7. Acknowledgements 11
8. Informative References 12
Authors' Addresses 13

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].

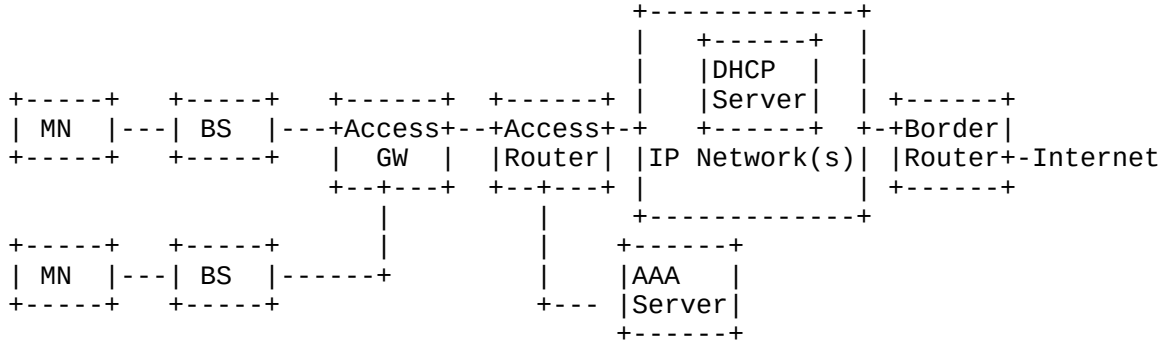


Figure 1: Key elements of a typical cellular network

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:

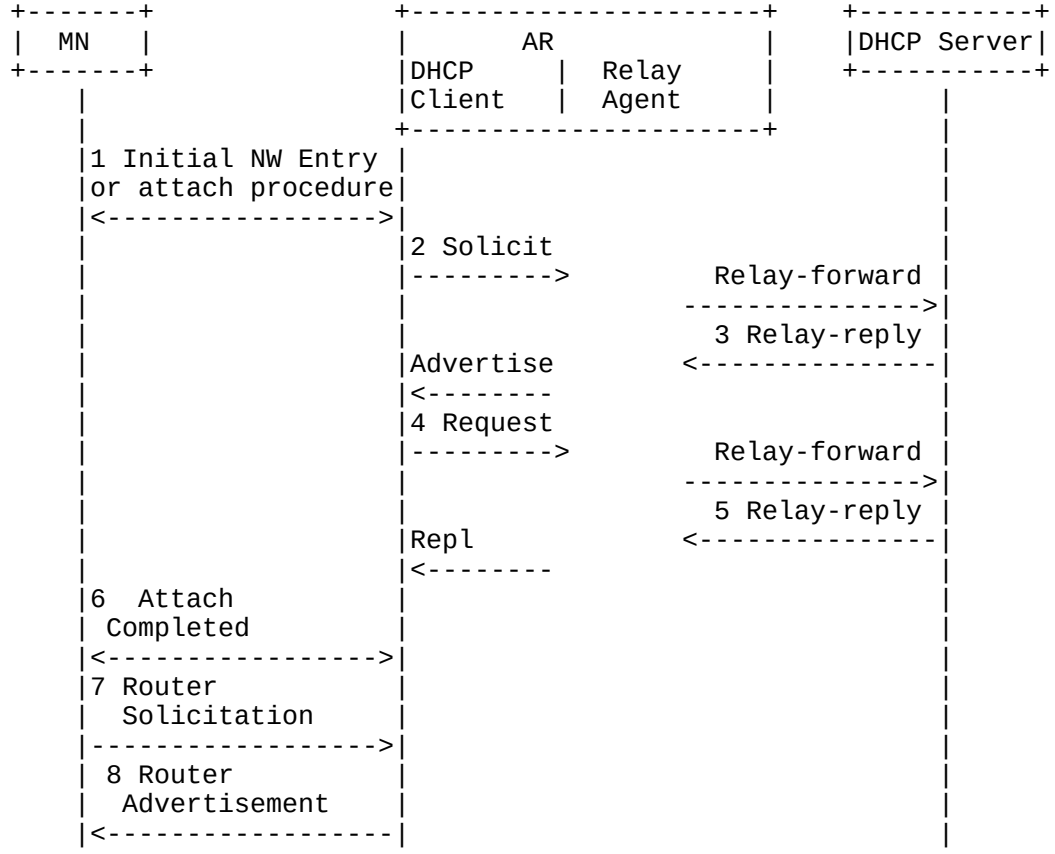


Figure 2: Prefix request

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.

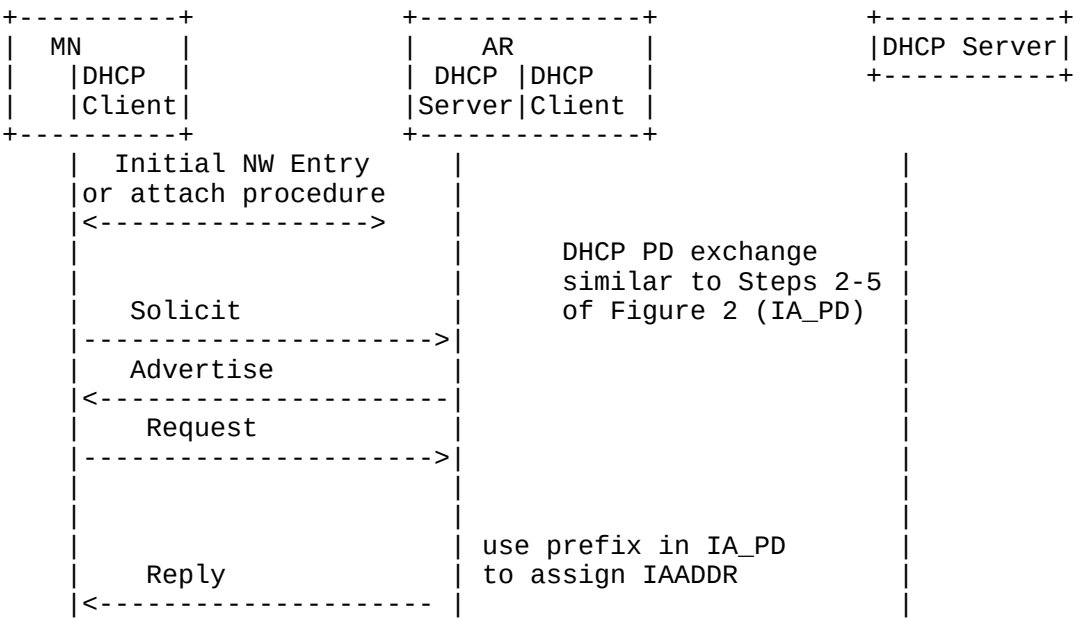


Figure 3: Stateful Address Configuration Following PD

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.

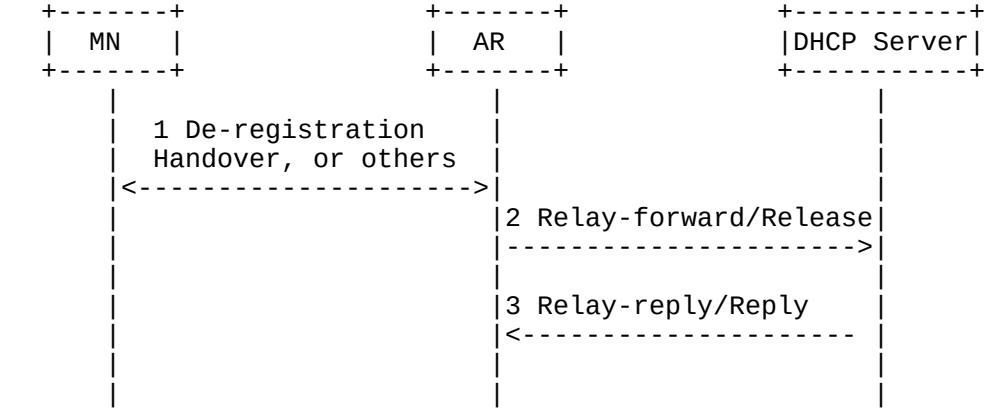


Figure 4: Prefix Release

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

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 3
2. Terminology 4
3. IPv6 multihomed network scenarios 5
3.1. Classification of network scenarios for multihomed host . 5
3.2. Multihomed network environment 7
3.3. Multihomed Problem Statement 8
4. Problem statement and analysis 9
4.1. Source address selection 10
4.2. Next-hop selection 10
4.3. DNS server selection 11
5. Requirements 12
5.1. End-to-End transparency 12
5.2. Policy enforcement 12
5.3. Scalability 13
6. Implementation approach 13
6.1. Source address selection 13
6.2. Next-hop selection 13
6.3. DNS resolver selection 14
7. Considerations for host without multi-prefix support 14
7.1. IPv6 NAT 15
7.2. Co-exisitence consideration 15
8. Security Considerations 16
9. IANA Considerations 16
10. Contributors 16
11. References 16
11.1. Normative References 16
11.2. Informative References 17
Authors' Addresses 18

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.

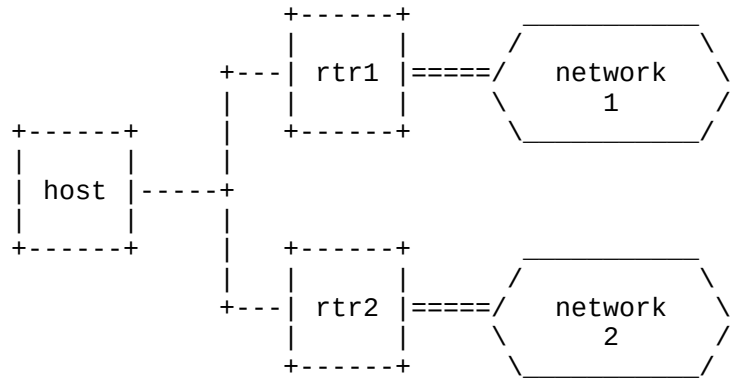


Figure 1: single uplink, multiple next-hop, multiple prefix (Scenario 1)

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.

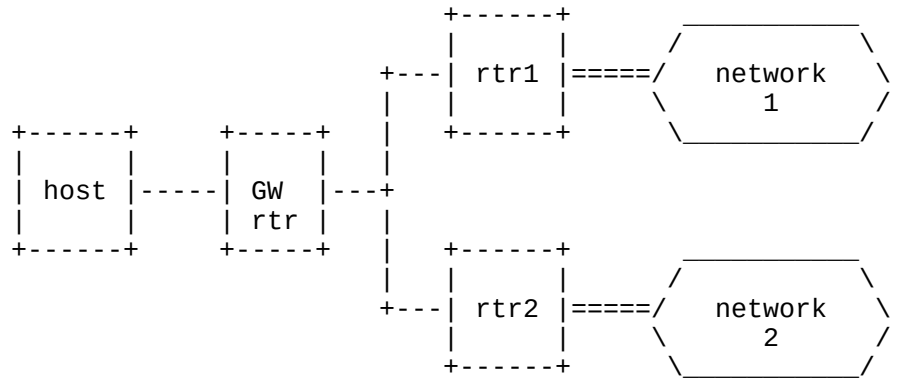


Figure 2: single uplink, single next-hop, multiple prefix (Scenario 2)

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.

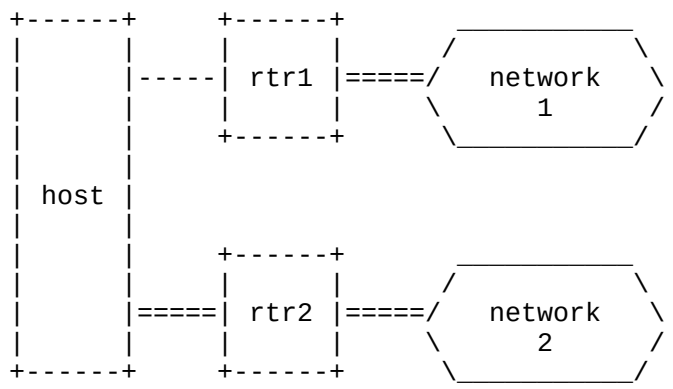


Figure 3: Multiple uplink, multiple next-hop, multiple prefix (Scenario 3)

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.

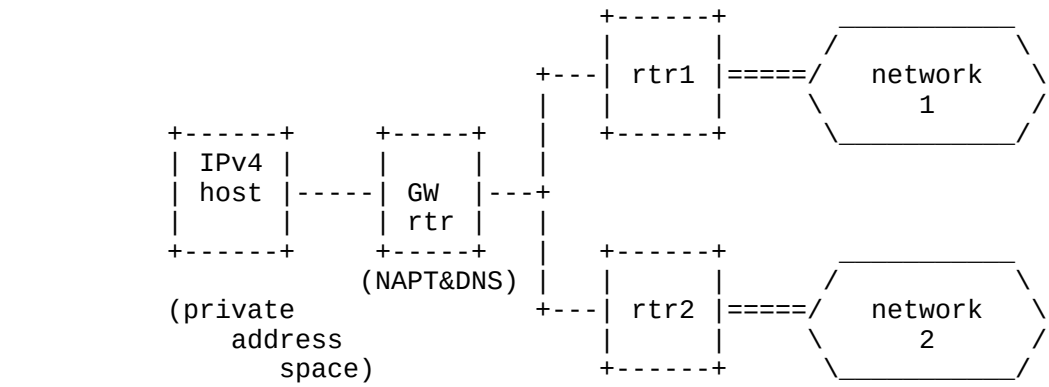


Figure 4: IPv4 Multihomed environment with Gateway Router performing NAPT

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.

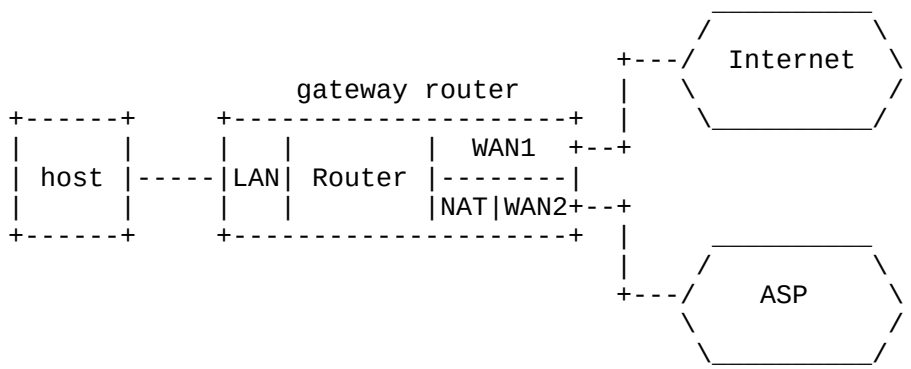


Figure 5: Legacy Host

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
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 3
1.1. Requirements Language 3
2. Terminology 3
3. Conceptual Configuration Variables 4
4. Architecture 4
5. Advanced Features and Feature Requirements 6
5.1. DNS 6
5.2. Multicast Behavior 6
5.3. ND Proxy 7
5.4. Prefix Delegation on LAN interface(s) (More details are TBD) 8
5.5. Routed network behavior(General Cases TBD) 8
5.6. Transition Technologies Support 9
5.6.1. Dual-Stack(DS)-Lite 9
5.6.2. 6rd 10
5.6.3. Transition Technologies Coexistence 10
5.7. Quality Of Service 11
5.8. Unicast Data Forwarding 11
5.9. ZeroConf 11
6. Security Considerations 11
7. Acknowledgements 11
8. Contributors 12
9. IANA Considerations 12
10. References 12
10.1. Normative References 12
10.2. Informative References 15
Authors' Addresses 15

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:

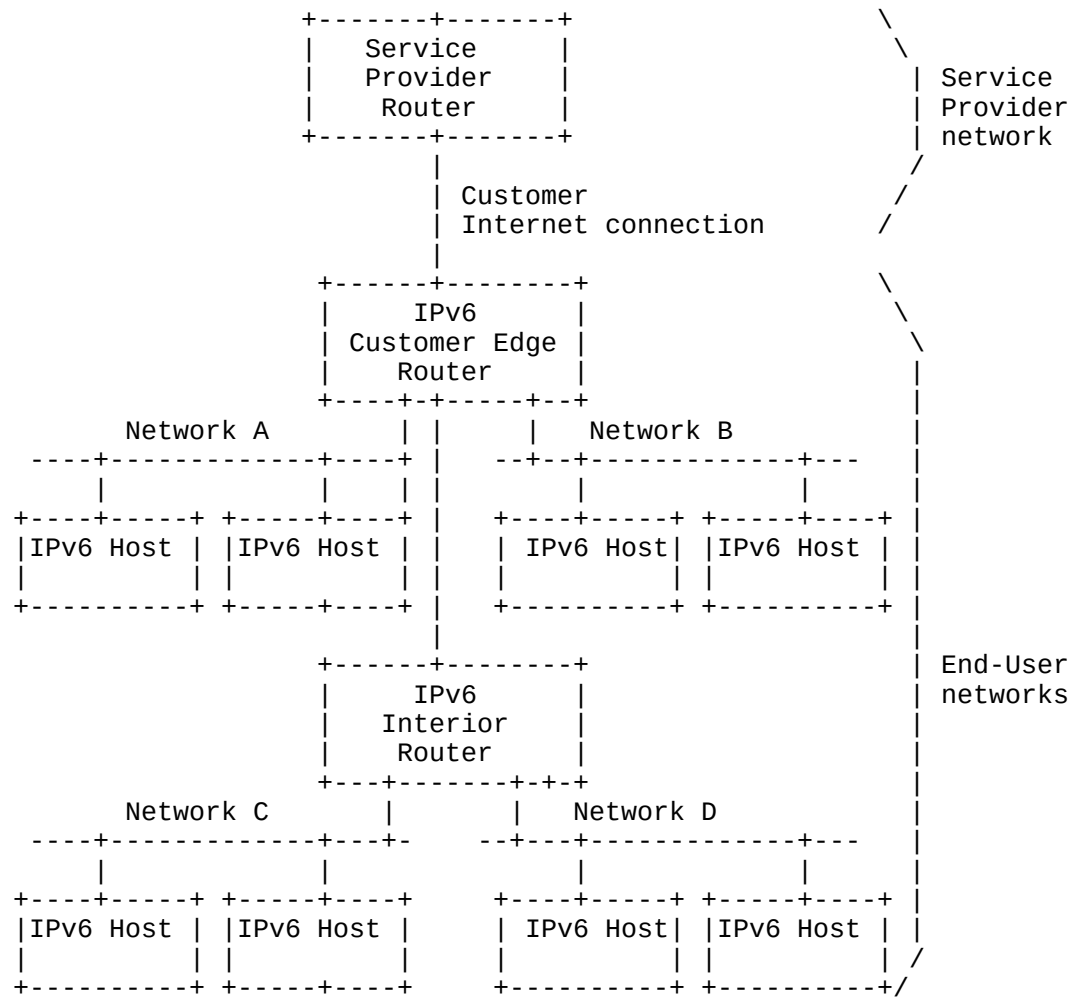


Figure 1.

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.

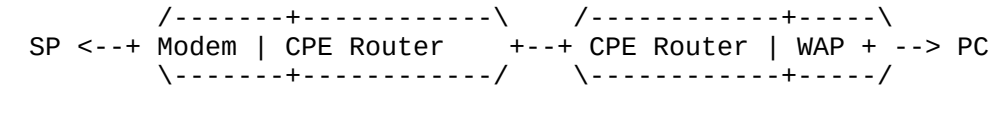
5.4. Prefix Delegation on LAN interface(s) (More details are TBD)

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