

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: January 31, 2013

G. Chen
Z. Cao
China Mobile
C. Byrne
T-Mobile USA
C. Xie
China Telecom
D. Binet
France Telecom
July 30, 2012

NAT64 Operational Experiences
draft-chen-v6ops-nat64-experience-03

Abstract

This document summarizes stateful NAT64 deployment scenarios and operational experience with NAT64-CGN(NAT64 Carrier Grade NATs) and NAT64-CE (NAT64 Customer Edge).

Status of this Memo

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

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

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

This Internet-Draft will expire on January 31, 2013.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect

to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

This document may contain material from IETF Documents or IETF Contributions published or made publicly available before November 10, 2008. The person(s) controlling the copyright in some of this material may not have granted the IETF Trust the right to allow modifications of such material outside the IETF Standards Process. Without obtaining an adequate license from the person(s) controlling the copyright in such materials, this document may not be modified outside the IETF Standards Process, and derivative works of it may not be created outside the IETF Standards Process, except to format it for publication as an RFC or to translate it into languages other than English.

Table of Contents

1. Introduction	4
2. Terminology	5
3. NAT64-CGN Deployment Experiences	6
3.1. NAT64-CGN Networking	6
3.2. High Availability Considerations	7
3.3. Traceability	7
3.4. Quality of Experience	8
3.5. Load Balancer	9
3.6. MTU Consideration	9
4. NAT64-CE Deployment Experiences	9
4.1. NAT64-CE Networking	10
4.2. Anti-DDoS/SYN Flood	11
4.3. User Behavior Analysis	11
4.4. DNS Resolving	11
4.5. Load Balancer	12
4.6. MTU Consideration	12
5. Security Considerations	12
6. IANA Considerations	12
7. Acknowledgements	12
8. Additional Author List	13
9. References	13
9.1. Normative References	13
9.2. Informative References	14
Authors' Addresses	15

1. Introduction

Continued development of global Internet demands IP address consumption. The IANA global IPv4 address pool was exhausted on February 3, 2011. IPv6 is the only sustainable solution for numbering nodes on the Internet. Network operators have to deploy IPv6 networks in order to meet the numbering needs of the expanding internet without available IPv4 addresses. IPv4 numbering resources and IPv4-only schemes to reduce the numbering utilization during the transitional will not be adequate to maintain connectivity and deliver Internet services.

As IPv6 deployment continues, IPv6 networks and hosts will need to coexist with IPv4 numbered resources. The Internet will include nodes that are Dual-stack, nodes that remain IPv4-only and IPv6-only nodes. It may be desirable in some cases for operators to deploy a single stack network, for reasons of simplicity, cost or performance relative to a dual stack network. As IPv4 utilization eventually declines, the appeal of single stack network deployments will likely increase. In a dual-stack architecture, operators have to maintain double management interfaces, provide operational support systems for two networks, track multiple addresses in different families per host, trouble shoot host behavior related to dual stack operation and engage in other activities that increase the overhead of operating the network.

Single stack IPv6 network deployment can simplify the network provisioning. Some justification has been described in [I-D.ietf-v6ops-464xlat]. IPv6-only networks confer some benefits to mobile operators employing them. In the mobile context, it enables the use of a single IPv6 PDP(Packet Data Protocol), which eliminates significant network cost caused by doubling the PDP count on a mass of legacy mobile terminals. In broadband networks overall, it can allow for the scaling of edge-network growth decoupled from IPv4 numbering limitations.

In a transition scenario, an existing network may rely on the IPv4 stack for a long time. There is also the troublesome trend of access network providers squatting on IPv4 address space that they do not own. Allowing for interconnection between IPv4-only nodes and IPv6-only nodes is a critical capability. Widespread dual-stack deployments have not materialized at the anticipated rate over the last 10 years on possible conclusion being that legacy networks will not make the jump quickly. A translation mechanism based on a NAT64[RFC6146] function might be a key element of the internet infrastructure supporting such legacy networks.

[RFC6036] reported at least 30% operators plan to run some kind of

translator (presumably NAT64/DNS64). Advice on NAT64 deployment and operation is therefore of some importance. [RFC6586] documented the implications for ipv6 only networks. This document intends to be specific to NAT64 network planning.

In regards to IPv4/IPv6 translation, [RFC6144] has described a framework of enabling networks to make interworking possible between IPv4-only and IPv6-only networks. Three scenarios are described, "An IPv6 Network to the IPv4 Internet", "The IPv6 Internet to an IPv4 Network" and "An IPv6 Network to an IPv4 Network" where a NAT64 function is relevant. The scenario of "The IPv6 Internet to the IPv4 Internet" seems to be the ideal case for inter-network translation technology. This document has focused on the three cases and further categorized different NAT64 location and use case. The principle distinction of location is if the NAT64 is located in a NAT64-CGN (Carrier Grade Nat) or NAT64-CE (Customer Edge). NAT64-CGN corresponds to the scenario "IPv6 Network to IPv4 Internet". The NAT64-CE location roughly corresponds to the "IPv6 Internet to IPv4 Network" and "IPv6 Network to IPv4 Network" scenarios. Based on different NAT64 modes, different considerations have been described for ISPs to facilitate NAT64 deployments.

2. Terminology

The terms of NAT-CGN/CE are understood to be a topological distinction indicating different features employed in a NAT64 deployment.

NAT64-CGN: A NAT64-CGN (Carrier Grade Nat) is placed in an ISP network and managed by an administrative entity, e.g. operator. From an administrator view, a NAT64-CGN usually forwards outbound traffic into an IPv4 network. IPv6 only subscribers leverage the NAT64-CGN to be served by existing IPv4 internet services. The ISP as an administrative entity takes full control on the IPv6 side, but has limited or no control on the IPv4 side. ISP's should attempt to accommodate the behavior of IPv4 networks and services.

NAT64-CE: A NAT64-CE (Customer Edge) is placed at the edge of customer network, e.g. a network operated by an Enterprise or Consumer. A NAT64-CE makes IPv4 services accessible for the IPv6 only users. An upstream entity and ISP usually operates an IPv4 and potentially IPv6 network respectively. IPv6 access is the common infrastructure behind the NAT64-CE.

3. NAT64-CGN Deployment Experiences

A NAT64-CGN deployment scenario is depicted in Figure 1

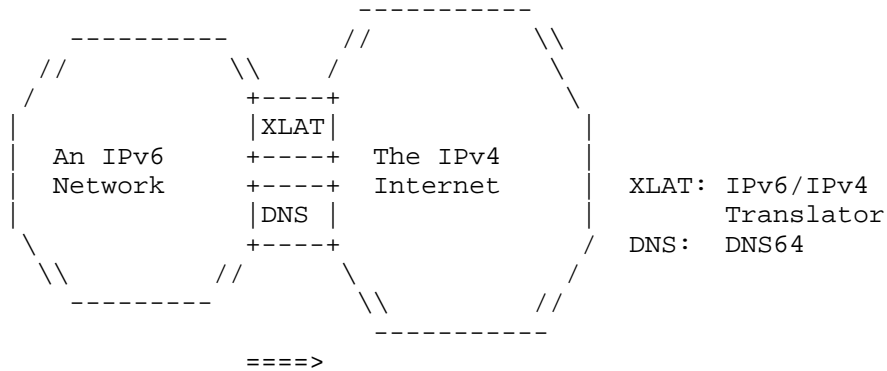


Figure 1: NAT64-CGN Scenario: IPv6 Network to IPv4 Internet

3.1. NAT64-CGN Networking

The NAT64-CGN use case is employed to connect IPv6-only users to the IPv4 Internet. The NAT64 gateway performs protocol translation from an IPv6 packet header to an IPv4 packet header and vice versa according to the Stateful NAT64 [RFC6146]. Address translation maps IPv6 addresses to IPv4 addresses and vice versa for return traffic.

All connections to the IPv4 Internet from IPv6-only clients must traverse the NAT64-CGN. It is advantageous from the vantage-point of troubleshooting and traffic engineering to carry the IPv6 traffic natively for as long as possible within an access network and translates only at or near the network egress.

In mobile networks, various possibilities can be envisaged in which to deploy the NAT64 function. Whichever option is selected, the NAT64 function will be deployed beyond the GGSN (Gateway GPRS Support Node) or PDN-GW (Public Data Network-Gateway), i.e. first IP node in currently deployed mobile architectures.

In a given implementation, NAT64 functionality can be provided by either a dedicated GW device or an multifunction gateway with integrated NAT64 functionality. In standalone NAT64, NAT64-CGN is placed to the side of a BNG or CR. An embedded NAT64 deployment would be integrated with an existing GW. Capacities of an existing GW can be potentially limited by the inserted functionality. In a mobile context, the NAT64 function can be co-located with GGSN/PDN-GW

or it can be embedded in an existing FW/NAT44 already deployed in support of IPv4 NAT or, the function can be collocated on a router. Whatever the solution retained for the co-location option, impact on existing services and legal obligations have to be assessed.

3.2. High Availability Considerations

High Availability (HA) is a major requirement for every service and network service.

Two mechanisms are typically used to achieve high availability, i.e. cold-standby and hot-standby. Cold-standby systems have synchronized configuration and mechanism to failover traffic between the hot and cold systems such as VRRP [RFC5798] . Unlike hot-standby, cold-standby does not synchronize NAT64 session state. This makes cold-standby less resource intensive and generally simpler, but it requires clients to re-establish sessions when a fail-over occurs. Hot-standby has all the features of cold-standby but must also synchronize the binding information base (BIB). Given that short lived sessions account for most of the bindings, hot-standby does not offer much benefit for those sessions. Consideration should be given to the importance (or lack thereof) of maintaining bindings for long lived sessions across failovers.

3.3. Traceability

Traceability is required in many cases to identify an attacker or a host that launches malicious attacks and/or for various other purposes, such as accounting requirements. NAT64 devices are required to log events like creation and deletion of translations and information about the occupied resources. There are two different demands for traceability, i.e. online or offline.

- o Regarding the Online requirements, XFF (X-Forwarded-For) [I-D.ietf-appsawg-http-forwarded] would be a candidate, it appends IPv6 address of subscribers to HTTP headers which is passed on to WEB servers, and the querier server can lookup radius servers for the target subscribers based on IPv6 addresses included in XFF HTTP headers. X-Forwarded-For is specific to HTTP, requires the use of an application aware gateway, cannot in general be applied to requests made over HTTPS and cannot be assumed to be preserved end-to-end as it may be overwritten by other application-aware proxies such as load balancers.
- o Some potential solutions to online traceability are explored in [I-D.ietf-intarea-nat-reveal-analysis].

- o A NAT64-CGN could also deliver NAT64 sessions (BIB and STE) to a Radius server by extension of the radius protocol. Such an extension is an alternative solution for online traceability, particularly high performance would be required on Radius servers on order to achieve this.
- o For off-line traceability, syslog might be a good choice. [RFC6269] indicates address sharing solutions generally need to record and store information for specific periods of time. A stateful NAT64 is supposed to manage one mapping per session. A large volume of logs poses a challenge for storage and processing. In order to mitigate the issue, [I-D.donley-behave-deterministic-cgn] proposed to pre-allocated a group of ports for each specific IPv6 host. A trade-off among address multiplexing efficiency, port randomization security [RFC6056] and logging storage compression should be considered during the planning. A hybrid mode combining deterministic and dynamic port assignment was recommended regarding the uncertainty of user traffic.

3.4. Quality of Experience

NAT64 is providing a translation capability between IPv6 and IPv4 end-nodes. In order to provide the reachability between two IP address families, NAT64-CGN has to implement appropriate ALGs where address translation is not itself sufficient and security mechanisms do not render it infeasible. e.g. FTP-ALG [RFC6384], RSTP-ALG, H.323-ALG, etc. It should be noted that ALGs may impact the performance on a NAT64 box to some extent. ISPs as well as content providers might choose to avoid situations where the imposition of an ALG might be required. At the same time, it is also important to remind customers that IPv6 end-to-end usage does not require ALG imposition and therefore results in a better overall user experience.

The service experience should be optimized around stateful NAT processing. Session status normally is managed by a static life-cycle. In some cases, NAT resource maybe significantly consumed by largely inactive users. The NAT translator and other customers would suffer from service degradation due to port consumption by other subscribers using the same NAT64 device. A flexible NAT session control is desirable to resolve the issues. PCP [I-D.ietf-pcp-base] could be a candidate to provide such capability. A NAT64-CGN should integrate with a PCP server, to allocate available IPv4 address/Port resources. Resources could be assigned to PCP clients through PCP MAP/PEER mode. Such an ability should also be considered to upgrade user experiences, e.g. assigning different sizes of port ranges for different subscribers. Such a mechanism is also helpful to minimize terminal battery consumption reducing the number of keepalive

messages to be sent by terminal devices.

3.5. Load Balancer

Load balancers are an essential tool to avoid the issue of single points of failure and add additional scale. It is potentially important to employ load-balancing considering that deployment of multiple NAT64 devices. Load balancers are required to achieve some service continuity and scale for customers.

[I-D.zhang-behave-nat64-load-balancing] discusses several ways of achieving NAT64 load balancing, including anycast based policy and prefix64 selection based policy, either implemented via DNS64[RFC6147] or Prefix64 assignments. Since DNS64 is normally co-located with NAT64 in some scenarios, it could be leveraged to perform the load balance. For traffic which does not require a DNS resolution, prefix64 assignment based on[I-D.ietf-behave-nat64-learn-analysis] could be adopted.

3.6. MTU Consideration

IPv6 requires that every link in the internet have an MTU of 1280 octets or greater[RFC2460]. However, in case of NAT64 translation deployments, some IPv4 MTU constrained link will be used in some communication path and originating IPv6 nodes may therefore receive an ICMP Packet Too Big message, reporting a Next-Hop MTU less than 1280. The result would be that IPv6 allows packets to contain a fragmentation header, without the packet being fragmented into multiple pieces. [I-D.ietf-6man-ipv6-atomic-fragments] discusses how this situation could be exploited by an attacker to perform fragmentation-based attacks, and also proposes an improved handling of such packets. It required enhancements on protocol level, which might imply potential upgrade/modifications on behaviors to deployed nodes. Another approach that potentially avoids this issue is to configure IPv4 MTU>=1260. It would forbid the occurrence of PTB<1280. However, such an operational consideration is hard to universally apply to the legacy "IPv4 Internet".

4. NAT64-CE Deployment Experiences

The NAT64-CE Scenario is depicted in Figure 2

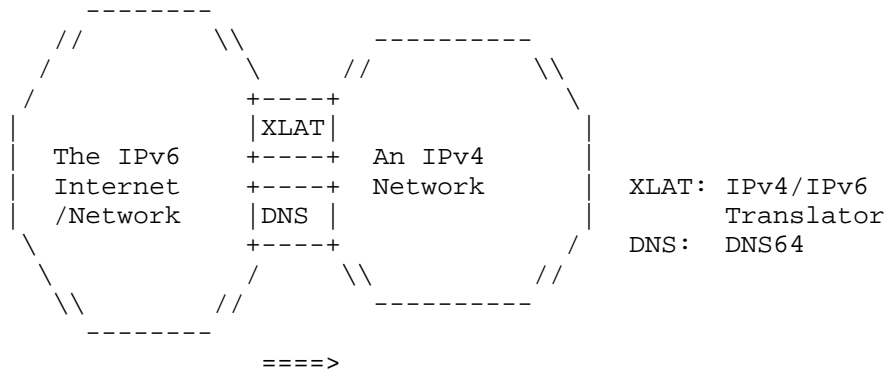


Figure 2: NAT64-CE Scenario: IPv6 Internet/Network to IPv4 Network

4.1. NAT64-CE Networking

Content providers would like to use IPv6 to serve customers since it allows for the definition of new services without having to integrate consideration of IPv4 NAT and address limitations of IPv4 networks, but they have to provide some IPv4 service continuity to their customers. In some cases, customers outside the network will have IPv6-only access provided by early adopters before the internal network. The deployment requirements could be resolved by subsidizing NAT64 to a customer edge, e.g. enterprise-GW. Those cases are sure to exist for the time being. An administrator of the IPv4 network needs to be cautious and aware of the operational issues this may cause, since the native IPv6 is always more desirable than transition solution.

One potential challenge in the scenario is NAT64-CE facing IPv6 Internet, in which a significant number of IPv6 users may initiate connections. When increasingly numerous users in IPv6 Internet access an IPv4 network, scalability concerns (e.g. additional latency, a single point of failure, IPv4 pool exhaustion, etc) are apt to be applied. For a given off-the-shelf NAT64-CE, those challenges should be seriously assessed. Potential issues should be properly identified. In order to mitigate the issues, it is suggested such usage should be restrained to a relative small-scale.

For operators who seek a clear precedent for operating reliable IPv6-only services, it should be well noted that the usage is problematic at several aspects. In some sense, it's not recommended.

4.2. Anti-DDoS/SYN Flood

For every incoming new connection from the IPv6 Internet, the NAT64-CE creates state and maps that connection to an internally-facing IPv4 address and port. An attacker can consume the resources of the NAT64-CE device by sending an excessive number of connection attempts. Without a DDOS limitation mechanism, the NAT64 is exposed to attacks from the IPv6 Internet. With service provisioning, attacks have the potential could also deteriorate service quality. One consideration in internet content providers is place a L3 load balancer with capable of line rate DDOS defense, such as the employment of SYN PROXY-COOKIE. Security domain division is necessary in this case. Load Balancers could not only serve for optimization of traffic distribution, but also serve as a DOS mitigation device

4.3. User Behavior Analysis

IP addresses are usually used as input to geo-location services. The use of address sharing will prevent these systems from resolving the location of a host based on IP address alone. Applications that assume such geographic information may not work as intended. The possible solutions listed at section 3.3 intended to bridge the gap. However, the analysis reveals those solutions can't be a optimal substitution to solve the problem of host identification, in particular it does not today mitigate problems with source identification through translation. That makes NAT64-CE usage becoming a unappealing approach, if customers require source address tracking.

For the operators, who already deployed NAT64-CE approach, the source address of the request is obscured without the source address mapping information previously obtained. It's superior to present mapping information directly to applications. Some application layer proxies e.g. XFF (X-Forwarded-For) , can convey this information in-band. Another approach is to ask application coordinating the information with NAT logging. But that is not sufficient, since the applications itself wants to know the original source address from an application message bus. The logging information may be used by administrators offline to inspect use behavior and preference analysis, and accurate advertisement delivery.

4.4. DNS Resolving

In the case of NAT64-CE, it is recommended to follow the recommendations in [RFC6144]. There is no need for the DNS to synthesize AAAA from A records, since static AAAA records can be registered in the authoritative DNS for a given domain to represent

these IPv4-only hosts. How to design the FQDN for the IPv6 service is out-of-scope of this document.

4.5. Load Balancer

Load balancing on NAT64-CE has a couple of considerations. If dictated by scale or availability requirements traffic should be balanced among multiple NAT64-CE devices. One point to be noted is that synthetic AAAA records may be added directly in authoritative DNS. load balancing based on DNS64 synthetic resource records may not work in those cases. Secondly, NAT64-CE could also serve as the load balancer for IPv4 backend servers. There are also some ways of load balance for the cases, where load balancer is placed in front of NAT64(s).

4.6. MTU Consideration

As compared to the MTU consideration in NAT64-CGN, the MTU of IPv4 network are strongly recommended to set to more than 1260. Since a CE IPv4 network is normally operated by a particular administrative entity, it should take steps to prevent the risk of fragmentation discussed in [I-D.ietf-6man-ipv6-atomic-fragments].

5. Security Considerations

This document presents the deployment experiences of NAT64 in CGN and CE scenario, some security considerations are described in detail regarding to specific NAT64 mode in section 2 and 3. In general, RFC 6146[RFC6146] provides TCP-tracking, address-dependent filtering mechanisms to protect NAT64 from DDOS. In NAT64-CGN cases, ISP also could adopt uRPF and black/white-list to enhance the security by specifying access policies. for example, NAT64-CGN should forbid establish NAT64 BIB for incoming IPv6 packets if URPF (Strict or Loose mode) check does not pass or whose source IPv6 address is associated to black-lists.

6. IANA Considerations

This memo includes no request to IANA.

7. Acknowledgements

The authors would like to thank Jari Arkko, Dan Wing, Remi Despres, Fred Baker, Lee Howard and Iljitsch van Beijnum for their helpful comments. Many thanks to Wesley George and Satoru Matsushima for

their reviews.

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

8. Additional Author List

The following are extended authors who contributed to the effort:

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

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

9. References

9.1. Normative References

- [I-D.ietf-pcp-base]
Wing, D., Cheshire, S., Boucadair, M., Penno, R., and P. Selkirk, "Port Control Protocol (PCP)",
draft-ietf-pcp-base-26 (work in progress), June 2012.
- [RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.
- [RFC5798] Nadas, S., "Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6", RFC 5798, March 2010.
- [RFC6146] Bagnulo, M., Matthews, P., and I. van Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", RFC 6146, April 2011.
- [RFC6147] Bagnulo, M., Sullivan, A., Matthews, P., and I. van

Beijnum, "DNS64: DNS Extensions for Network Address Translation from IPv6 Clients to IPv4 Servers", RFC 6147, April 2011.

[RFC6384] van Beijnum, I., "An FTP Application Layer Gateway (ALG) for IPv6-to-IPv4 Translation", RFC 6384, October 2011.

9.2. Informative References

- [I-D.donley-behave-deterministic-cgn]
Donley, C., Grundemann, C., Sarawat, V., and K. Sundaresan, "Deterministic Address Mapping to Reduce Logging in Carrier Grade NAT Deployments", draft-donley-behave-deterministic-cgn-04 (work in progress), July 2012.
- [I-D.ietf-6man-ipv6-atomic-fragments]
Gont, F., "Processing of IPv6 "atomic" fragments", draft-ietf-6man-ipv6-atomic-fragments-00 (work in progress), February 2012.
- [I-D.ietf-appsawg-http-forwarded]
Petersson, A. and M. Nilsson, "Forwarded HTTP Extension", draft-ietf-appsawg-http-forwarded-06 (work in progress), July 2012.
- [I-D.ietf-behave-nat64-learn-analysis]
Korhonen, J. and T. Savolainen, "Analysis of solution proposals for hosts to learn NAT64 prefix", draft-ietf-behave-nat64-learn-analysis-03 (work in progress), March 2012.
- [I-D.ietf-intarea-nat-reveal-analysis]
Boucadair, M., Touch, J., Levis, P., and R. Penno, "Analysis of Solution Candidates to Reveal a Host Identifier (HOST_ID) in Shared Address Deployments", draft-ietf-intarea-nat-reveal-analysis-02 (work in progress), April 2012.
- [I-D.ietf-v6ops-464xlat]
Mawatari, M., Kawashima, M., and C. Byrne, "464XLAT: Combination of Stateful and Stateless Translation", draft-ietf-v6ops-464xlat-05 (work in progress), July 2012.
- [I-D.zhang-behave-nat64-load-balancing]
Zhang, D., Xu, X., and M. Boucadair, "Considerations on NAT64 Load-Balancing", draft-zhang-behave-nat64-load-balancing-03 (work in progress), July 2012.

progress), July 2011.

- [RFC6036] Carpenter, B. and S. Jiang, "Emerging Service Provider Scenarios for IPv6 Deployment", RFC 6036, October 2010.
- [RFC6056] Larsen, M. and F. Gont, "Recommendations for Transport-Protocol Port Randomization", BCP 156, RFC 6056, January 2011.
- [RFC6144] Baker, F., Li, X., Bao, C., and K. Yin, "Framework for IPv4/IPv6 Translation", RFC 6144, April 2011.
- [RFC6269] Ford, M., Boucadair, M., Durand, A., Levis, P., and P. Roberts, "Issues with IP Address Sharing", RFC 6269, June 2011.
- [RFC6586] Arkko, J. and A. Keranen, "Experiences from an IPv6-Only Network", RFC 6586, April 2012.

Authors' Addresses

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

Email: phdgang@gmail.com

Zhen Cao
China Mobile
53A,Xibianmennei Ave.,
Xuanwu District,
Beijing 100053
China

Email: caozhen@chinamobile.com

Cameron Byrne
T-Mobile USA
Bellevue
Washington 98105
USA

Email: cameron.byrne@t-mobile.com

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

Email: xiechf@ctbri.com.cn

David Binet
France Telecom
Rennes
35000
France

Email: david.binet@orange.com

v6ops
Internet-Draft
Intended status: Informational
Expires: January 14, 2013

K. Chittimaneni
Google Inc.
T. Chown
University of Southampton
L. Howard
Time Warner Cable
V. Kuarsingh
Rogers Communications
Y. Pouffary
Hewlett Packard
E. Vyncke
Cisco Systems
July 13, 2012

Enterprise Incremental IPv6
draft-chkpvc-enterprise-incremental-ipv6-01

Abstract

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

Status of this Memo

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

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

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

This Internet-Draft will expire on January 14, 2013.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	4
1.1. Enterprise Assumptions	4
1.2. IPv4-only Considerations	5
1.3. Reasons for a Phased Approach	5
2. Requirements Language	6
3. Preparation and Assessment Phase	6
3.1. Inventory Phase	6
3.1.1. Network infrastructure readiness assessment	6
3.1.2. Applications readiness assessment	7
3.1.3. Importance of readiness validation and testing	7
3.2. Training	8
3.3. Routing	8
3.4. Security and Routing Policy	8
3.4.1. Demystifying some IPv6 Security Myths	8
3.4.2. Similarities between IPv6 and IPv4 security	9
3.4.3. Specific Security Issues for IPv6	10
3.5. Address Plan	11
3.6. Program Planning	12
3.7. Tools Assessment	13
4. External Phase	14
4.1. Connectivity	15
4.2. Security	16
4.3. Monitoring	17
4.4. Servers and Applications	17
5. Internal Phase	17
5.1. Network Infrastructure	17
5.2. End user devices	19
5.3. Corporate Systems	20
5.4. Security	20
6. Other Phases	20
6.1. Guest network	20
6.2. IPv6-only	20
7. Considerations For Specific Enterprises	22
7.1. Content Delivery Networks	22
7.2. Data Center Virtualization	22
7.3. Campus Networks	22
8. Security Considerations	22
9. Acknowledgements	22
10. IANA Considerations	23
11. References	23
11.1. Normative References	23
11.2. Informative References	23
Authors' Addresses	26

1. Introduction

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

The business reasons for spending time, effort, and money on IPv6 will be unique to each enterprise. The most common drivers are due to the fact that when Internet service providers, including mobile wireless carriers, run out of IPv4 addresses, they will provide native IPv6 and non-native IPv4. The non-native IPv4 service may be NAT64, NAT444, Dual-stack Lite, or other transition technology, but whether tunneled or translated, the native traffic will be performed better and more reliably than non-native. It is thus in the enterprise's interests to deploy native IPv6 itself.

1.1. Enterprise Assumptions

For the purposes of this document, assume:

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

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

1.2. IPv4-only Considerations

As described in [RFC6302] administrators should take certain steps even if they are not considering IPv6. Specifically, Internet-facing servers should log the source port number, timestamp (from a reliable source), and the transport protocol. This will allow investigation of malefactors behind address-sharing technologies such as NAT44 or Dual-stack Lite. Enabling this additional logging will take a few minutes on each server, and will probably require restarting the service.

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

1.3. Reasons for a Phased Approach

Given the challenges of migrating user workstations, corporate systems, and Internet-facing servers, a phased approach allows incremental deployment of IPv6, based on the administrator's own determination of priorities. The Preparation Phases is highly recommended to all administrators, as it will save errors and complexity in later phases. Each administrator must decide whether to begin with the External Phase (as recommended in [RFC5211]) or the Internal Phase. There is no "correct" answer here; the decision is one for each enterprise to make.

Some considerations:

- o In many cases, customers outside the network will have IPv6 before the internal enterprise network. For these customers, IPv6 may well perform better, especially for certain applications, than translated or tunneled IPv4, so the administrator may want to prioritize the External Phase.
- o Employees who access internal systems by VPN may find that their ISPs provide translated IPv4, which does not support the required VPN protocols. In these cases, the administrator may want to prioritize the External Phase, and any other remotely-accessible internal systems.
- o Internet-facing servers cannot be managed over IPv6 unless the management systems are IPv6-capable. These might be Network Management Systems (NMS), monitoring systems, or just remote management desktops. Thus in some cases, the Internet-facing systems are dependent on IPv6-capable internal networks. However, dual-stack Internet-facing systems can still be managed over IPv4.

- o IPv6 is enabled by default on all modern operating systems, so it may be more urgent to manage and have visibility on the internal traffic.
- o In many cases, the corporate accounting, payroll, human resource, and other internal systems may only need to be reachable from the internal network, so they may be a lower priority. But more and more internal applications support IPv6 by default and it can be expected that new applications will only support IPv6.
- o Some organizations (even when using private IPv4 addresses[RFC1918]) are facing IPv4 address exhaustion because of the internal network growth (for example the vast number of virtual machines).
- o IPv6 restores end to end reachability even for internal applications (of course security policies must still be enforced) which means that with IPv6 merging networks (after two organizations merged) is much easier and faster. Yet, another reason to move the internal network to IPv6.

These considerations are in conflict; each administrator must prioritize according to their local conditions.

2. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119] when they appear in ALL CAPS. These words may also appear in this document in lower case as plain English words, absent their normative meanings.

3. Preparation and Assessment Phase

3.1. Inventory Phase

To comprehend the inventory phase spectrum we recommended dividing the problem space in two: network infrastructure readiness and applications readiness.

3.1.1. Network infrastructure readiness assessment

The network infrastructure readiness assessment for IPv6 as its name state is focused on the network. The goal of this assessment is identify the level of readiness of network equipment. This is an important step as it will help identify the effort required to move

to an infrastructure that supports IPv6 with the same features as the existing IPv4 network (for example MPLS-VPN [RFC4364] whose equivalent in IPv6 is 6VPE [RFC4659]).

Be able to understand which network devices are already capable, which devices can be made IPv6 ready with a code/firmware upgrade, and which devices will need to be replaced. The data collection consists of a network discovery to gain an understanding of the topology and inventory network infrastructure equipment and code versions with information gathered from static files and IP address management, DNS and DHCP tools.

Remember understanding the starting point and what are the technical issues and challenges is critical as IPv6 might already be present in the environment thus creating inherent security risks.

3.1.2. Applications readiness assessment

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

There are two ways to IPv6-enable your applications. The first approach is to have separate logic for IPv4 and IPv6, thus leaving the IPv4 code path mainly untouched. This approach causes the least disruption to the existing IPv4 logic flow, but introduces more complexity, since the application now has to deal with two logic loops with complex race conditions and error recovery mechanisms between these two logic loops. The second approach is to create a combined IPv4/IPv6 logic, which ensures operation regardless of the IP version used on the network. We recommend using industry IPv6-porting tools to locate the code that need to be updated.

3.1.3. Importance of readiness validation and testing

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

testing to understand how IPv6 impacts the services and networking infrastructure will be paramount.

3.2. Training

IPv6 planning and deployment in the enterprise is not an entirely network centric affair. IPv6 adoption will be a multifaceted undertaking that will touch everyone in the organization. While technology and process transformations are taking place it is critical that people training takes place as well. Training will ensure that people and skill gaps are assessed proactively and managed accordingly. We recommend that training needs be analyzed and defined in order to successfully inform, train, and prepare staff for the impacts of the system or process changes.

3.3. Routing

When deploying IPv6, we recommend initially choosing an IGP protocol you are familiar with. That is to say if you are using OSPFv2 you should be using OSPFv3. The main advantage of this approach is that staff do not need to be trained and existing processes can be leveraged.

Enterprises could also take the opportunity the introduction of IPv6 brings to analyze your current environment and to identify which features you would like to change and what you would like to implement. Then using the validation period as a way to validate your new approach and even applying them to your IPv4 environment.

Either way IPv6 introduces the opportunity to rationalize the environment and to architect it for growth.

3.4. Security and Routing Policy

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

3.4.1. Demystifying some IPv6 Security Myths

Some people believe that IPv6 is inherently more secure than IPv4 because it is new. Nothing can be more wrong. Indeed, being a new protocol means that bugs in the implementations have yet to be

discovered and fixed and that few people have the operational security expertise needed to operate securely an IPv6 network. This lack of operational expertise is the biggest threat when deploying IPv6: the importance of training is to be stressed again.

One security myth is that thanks to its huge address space, a network cannot be scanned by enumerating all IPv6 address in a /64 LAN hence a malevolent person cannot find a victim. [RFC5157] describes some alternate techniques to find potential targets on a network, for example enumerating all DNS names in a zone.

Another security myth is that IPv6 is more secure because it mandates the use of IPsec everywhere. [RFC6434] clearly states that the IPv6 support is a SHOULD only. Moreover, if all the intra-enterprise traffic is encrypted, then this renders all the network security tools (IPS, firewall, ACL, IPFIX, etc) blind and pretty much useless. Therefore, IPsec should be used in IPv6 pretty much like in IPv4 (for example to establish a VPN overlay over a non-trusted network or reserve to some specific applications).

The last security myth is that amplification attacks (such as <http://www.cert.org/advisories/CA-1998-01.html>) do not exist in IPv6 because there is no more broadcast. Alas, this is not true as ICMP error (in some cases) or information messages can be generated by routers and hosts when forwarding or receiving a multicast message (section 2.4 of [RFC4443]). Therefore, the generation and the forwarding rate of ICMPv6 messages must be rate limited as in IPv4.

3.4.2. Similarities between IPv6 and IPv4 security

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

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

A specific case of congruence is the IPv6 ULA [RFC4193] and IPv4 private addressing [RFC1918] that do not provide any security by 'magic'. In both case, the edge router must apply strict data plane and routing policy to block those private addresses to leave and enter the network. This filtering can be done by the enterprise or

by the ISP.

IPv6 addresses can be spoofed as easily as IPv4 addresses and there are packets with bogons IPv6 addresses (see <http://www.team-cymru.org/Services/Bogons/>). The anti-bogon filtering must be done in the data and routing planes. It can be done by the enterprise or by the ISP.

3.4.3. Specific Security Issues for IPv6

Even if IPv6 is similar to IPv4, there are some differences that create some IPv6-only vulnerabilities or issues.

Privacy extension addresses [RFC4941] are usually to protect individual privacy by changing periodically the interface identifier part of the IPv6 address to avoid tracking a host by its always identical and unique EUI-64. While this presents a real advantage on the Internet, it complicates the task of audit trail when a security officer or network operator wants to trace back a log entry to a host in their network because when the tracing is done the searched IPv6 address could have disappeared from the network. A good way to prevent the use of privacy extension addresses without host configuration is to send the Router Advertisement with the M-bit set (to force the use of DHCPv6 to get an address) and with all advertised prefixes without the A-bit set (to prevent the use of stateless auto-configuration).

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

Fragmentation is different in IPv6 because it is done only by source host and never during a forwarding operation. This means that ICMPv6 packet-too-big must be allowed [RFC4890] through all filters. Fragments can also be used to evade some security mechanisms such as RA-guard [RFC6105], see also [RFC5722] which appears to be widely implemented in 2012.

But, the biggest difference is the replacement of ARP (RFC 826) by Neighbor Discovery Protocol [RFC4861]. NDP runs over ICMPv6 (this means that security policies MUST allow some ICMPv6 messages see RFC 4890) but has the same lack of security as ARP (SeND [RFC3971] and CGA [RFC3972] are not widely implemented). ARP can be made secure with the help of techniques known as DHCPv4 snooping and dynamic ARP inspection by access switches. Therefore, enterprises using those

techniques for IPv4 should use the equivalent techniques for IPv6: this is RA-guard (RFC 6105) and all work in progress from the SAVI WG ([I-D.ietf-savi-threat-scope] and others). Another DoS vulnerabilities are related to NDP cache exhaustion ([I-D.gashinsky-v6ops-v6nd-problems]) and they can be mitigated by careful tuning of the NDP cache. In 2012, there are already several vendors offering those features on their switches.

Running a dual-stack network doubles the attack exposure as a malevolent person has now two attack vectors: IPv4 and IPv6. This simply means that all routers and hosts operating in a dual-stack environment with both protocol families enabled (even if by default) must have a congruent security policy for both protocol version. For example, permit TCP ports 80 and 443 to all web servers and deny all other ports to the same servers must be implemented both for IPv4 and IPv6.

3.5. Address Plan

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

The enterprise administrator will want to evaluate whether the enterprise will request address space from its ISP (or Local Internet Registry (LIR)), or whether to request address space from the local Internet Registry (whether a Regional Internet Registry such as AfriNIC, APNIC, ARIN, LACNIC, or RIPE-NCC, or a National Internet Registry, operated in some countries). There may be a registration fee for requesting provider-independent (PI) space from and NIR/RIR, but the enterprise will avoid some complexity if renumbering is required after changing ISPs (it should be noted that renumbering caused by outgrowing the space, merger, or other internal reason might not be avoided with PI space).

Each location, no matter how small, should get at least a /48. In addition to allowing for simple planning, this can allow a site to use its prefix for local connectivity, should the need arise, and if the local ISP supports it. Generally, workstations managed by the enterprise will use stateful DHCPv6 for addressing on corporate LAN segments. DHCPv6 allows for the additional configuration options

often employed by enterprise administrators, and by using stateful DHCPv6, administrators correlating system logs know which system had which address at any given time.

In the data center or server room, assume a /64 per VLAN. This applies even if each individual system is on a separate VLAN; in a /48 assignment, typical for a site, there are 65,535 /64 blocks. Addresses are either configured manually on the server, or reserved on a DHCPv6 server, which may also synchronize forward and reverse DNS.

All user access networks MUST be a /64. Point-to-point links without MAC addresses (this is where Neighbor Discovery Protocol does not run) SHOULD be a /127 (see also [RFC6164]).

Plan to aggregate at every layer of network hierarchy. Where multiple VLANs or other layer two domains converge, allow some room for expansion. Renumbering due to outgrowing the network plan is a nuisance, so allow room within it. Generally, grow to about twice the current size can be accommodated; where rapid growth is planned, allow for twice that growth. Also, for any part of the network where DNS (or reverse DNS) zones may be delegated, it is important to delegate addresses on nibble boundaries (this is on a multiple of 4 bits), to ensure propose name delegation.

3.6. Program Planning

As with any project, an IPv6 deployment project will have its own phases. Generally, one person is identified as the project sponsor or champion, who will make sure time and talent resources are prioritized appropriately for the project. Because enabling IPv6 can be a project with many interrelated tasks, identifying a project manager is also recommended. The project manager and sponsor can initiate the project, determining the scope of work and identifying whose input is required, and who will be affected by work. The scope will generally include the Preparation Phase, and may include the Internal Phase, the External Phase, or both, and may include any or all of the Other Phases identified.

The project manager will need to spend some time planning. It is often useful for the sponsor to communicate with stakeholders at this time, to explain why IPv6 is important to the enterprise. Then, as the project manager is assessing what systems and elements will be affected, the stakeholders will understand why it is important for them to support the effort. Well-informed project participants can help significantly by explaining the relationships between components. For a large enterprise, it may take several iterations to really understand the level of effort required; some systems will

require additional development, some might require software updates, and others might need new versions or alternate vendors. Once the projects are understood, the project manager can develop a schedule and a budget, and work with the project sponsor to determine what constraints can be adjusted, if necessary.

It is tempting to roll IPv6 projects into other architectural upgrades - this can be an excellent way to improve the network and reduce costs. Project participants are advised that by increasing the scope of projects, the schedule is often affected. For instance, a major systems upgrade may take a year to complete, where just patching existing systems may take only a few months. Understanding and evaluating these trade-offs are why a project manager is important.

It is very common for assessments to continue in some areas even as execution of the project begins in other areas. This is fine, as long as recommendations in other parts of this document are considered, especially regarding security (for instance, one should not deploy IPv6 on a system before security has been evaluated). The project manager will need to continue monitoring the progress of discrete projects and tasks, to be aware of changes in schedule, budget, or scope. "Feature creep" is common, where engineers or management wish to add other features while IPv6 development or deployment is ongoing; each feature will need to be individually evaluated for its effect on the schedule and budget, and whether expanding the scope increases risk to any other part of the project.

As projects are completed, the project manager will confirm that work has been completed, often by means of seeing a completed test plan, and will report back to the project sponsor on completed parts of the project. A good project manager will remember to thank the people who executed the project.

3.7. Tools Assessment

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

The tools within the organization fall into two general categories, those which focus on managing the network, and those which are focused on managing systems and applications on the network. In either instance, the tools will run on platforms which may or may not

be capable of operating in an IPv6 network. This lack in functionality may be related to Operating system version, or based on some hardware constraint. Those systems which are found to be incapable of utilizing a IPv6 connection may need to be replaced or upgraded.

In addition to devices working on an IPv6 network natively, or via a tunnel, many tools and support systems may require additional updates to be IPv6 aware or even a hardware upgrade (mainly because of the memory utilization by IPv6 as the addresses are larger and because, for a while, IPv4 and IPv6 addresses will coexist in the tool). This awareness may include the ability to manage IPv6 elements and/or applications in addition to the ability to store and utilize IPv6 addresses.

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

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

Organizations should be aware of changes to older IPv4-Only SNMP MIB specifications have been made by the IETF related to legacy operation in [RFC2096] and [RFC2011]. Updated specifications are now available in [RFC4296] and [RFC4293] which modified the older MIB framework to be IP protocol agnostic supporting IPv4 and IPv6. Polling systems will need to be upgraded to support these updates as well as the end stations which are polled.

4. External Phase

The external phase for Enterprise IPv6 adoption covers topics which

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

How an organization connects to the outside world is very important as it is often a critical part of how a business functions, therefore must be dealt accordingly.

4.1. Connectivity

The Enterprise will need to work with one or more Service Providers to gain connectivity to the Internet or transport service infrastructure such as a BGP/MPLS IP VPN as described in [RFC4364] and [RFC4659]. One significant factor guiding how an organization may need to connect with the outside world will involve the use of PI (Provider Independent) and/or PA (Provider Aggregatable) IPv6 space.

In the case of PI, the organization will need to support BGP based connectivity for the most part since the address space is assigned direction from the Regional Registry to the local network. In this case, the local network would act as an Autonomous System on the Internet and must advertise routes accordingly. PA space is delegated from the upstream service provider and can then be assigned to the local network. If PA space is used, other forms of route exchange may be possible such as RIPng, OSPFv3 and static. PA assigned space would normally be routed to the general Internet via the upstream providers infrastructure lightening the burden on the local network administrations.

PI and PA space have additional contrasting behaviours when used such as how dual homing may work. Should an operator choose to dual home, PI space would be routed to both upstream providers and then passed on to other networks. Utilizing more than one upstream Service Provider may introduce challenges since traffic utilizing a given PA assigned block would be expected to flow through the assigning provider for entry to the Internet. Should traffic flow using source addresses which are not delegated from a given provider, reverse path forwarding rules on the operator side may reject some traffic. These considerations are quite different than those of IPv4 which relied on NAT in most cases.

When seeking IPv6 connectivity to a Service Provider, the Enterprise will want to attempt to use Native IPv6 connectivity. Native IPv6 connectivity is preferred since it provides the most robust form of connectivity. If Native IPv6 connectivity is not possible due to technical or business limitations, the Enterprise may utilize readily

available tunnelled IPv6 connectivity. There are IPv6 transit providers which provide tunnelled IPv6 connectivity which can operate over IPv4 networks. A Enterprise need not need to wait for their local Service Provider to support IPv6, as tunnelled connectivity can be used.

4.2. Security

The most important part of security for external IPv6 deployment is filtering. Filtering can be done by stateless ACL or stateful firewall. As described in section 2.4.3, the security policies must be congruent for IPv4 and IPv6 except that ICMPv6 messages must be allowed through and to the filtering device (see [RFC4890]):

- o unreachable packet-too-big
- o unreachable parameter-problem
- o neighbor solicitation
- o neighbor advertisement

**** Add some comment about setting MTU to 1280 to avoid tunnel pMTUD black holes? ****

It could also be safer to block all fragments where the transport layer header is not in the first fragment to avoid attack as described in [RFC5722]. Some filtering devices allow this filtering. To be fully compliant with [RFC5095], it can be useful to drop all packet containing the routing extension header type 0.

If Intrusion Prevention Systems (IPS) are used for IPv4 traffic, then the same IPS should also be used for IPv6 traffic. This is just a generalization of the dual-stack deployment: do for IPv6 what you do for IPv4. This also include all email content protection (anti-spam, content filtering, data leakage prevention, etc).

The peering router must also implement anti-spoofing technique based on [RFC2827].

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

The NDP cache exhaustion (see [I-D.gashinsky-v6ops-v6nd-problems]) attack can be mitigated by two techniques:

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

4.3. Monitoring

Monitoring the use of the Internet connectivity should be done for IPv6 if it is done for IPv4. This includes the use of IP flow export [RFC5102] to detect abnormal traffic pattern (such as port scanning, SYN-flooding) and SNMP MIB [RFC4293] (another way to detect abnormal bandwidth utilization).

4.4. Servers and Applications

5. Internal Phase

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

An important design paradigm to consider during this phase is "Dual Stack when you can, tunnel when you must". Dual stacking allows you to build a more robust IPv6 network that is of production quality as opposed to tunnels that are harder to troubleshoot and support. Tunnels however do provide operators with a quick and easy way to play with IPv6 and gain some operational experience with the protocol. [RFC4213] describes various transition mechanisms in more detail. [I-D.templin-v6ops-isops] suggests operational guidance when using ISATAP tunnels [RFC5214].

5.1. Network Infrastructure

The typical enterprise network infrastructure comprises of a combination of the following network elements - wired access switches, wireless access points, and routers. Although, it is fairly common to find hardware that collapses switching and routing functionality into a single device. Basic wired access switches and access points that operate only at the physical and link layer, don't

really have any special IPv6 considerations other than being able to support IPv6 addresses themselves for management purposes, if the same exists for IPv4. In many instances, these devices possess a lot more intelligence than simply switching packets. For example, some of these devices help assist with link layer security by incorporating features such as ARP inspection and DHCP Snooping.

An important design choice to be made is what IGP to use inside the network. A variety of IGPs (IS-IS, OSPFv3 and RIPng) support IPv6 today and picking one over the other is purely a design choice that will be dictated mostly by existing operational policies in an enterprise network. As mentioned earlier, it would be beneficial to maintain operational parity between IPv4 and IPv6 and therefore it might make sense to continue using the same protocol family that is being used for IPv4. For example, if you use OSPFv2 for IPv4, it might make sense to use OSPFv3 now.

Another important consideration in enterprise networks is first hop router redundancy. This directly ties into network reachability from an end host's point of view. IPv6 Neighbor Discovery (ND), [RFC4861], provides a node with the capability to maintain a list of available routers on the link, in order to be able to switch to a backup path should the primary be unreachable. By default, ND will detect a router failure in 38 seconds and cycle onto the next default router listed in its cache. While this feature does provide with a basic level of first hop router redundancy, most enterprise IPv4 networks are designed to fail over much faster. Although this delay can be improved by adjusting the default timers, care must be taken to protect against transient failures and to account for increased traffic on the link. Another option to provide robust first hop redundancy is to use the Virtual Router Redundancy Protocol for IPv6 (VRRPv3), [RFC5798]. This protocol provides a much faster switchover to an alternate default router than default ND parameters. Using VRRP, a backup router can take over for a failed default router in around three seconds (using VRRP default parameters). This is done without any interaction with the hosts and a minimum amount of VRRP traffic.

Last but not the least, one of the most important design choices to make while deploying IPv6 on the internal network is whether to use Stateless Automatic Address Configuration (SLAAC), [RFC4862], or Dynamic Host Configuration Protocol for IPv6 (DHCPv6), [RFC3315], or a combination thereof (possible when using a /64 subnet). Each option has its own unique set of pros and cons and the choice will ultimately depend on the operational policies that guide each enterprise's network design. For example, if an enterprise is looking for ease of use, rapid deployments, and less administrative overhead, then SLAAC makes more sense. However, if the operational

policies call for precise control over IP address assignment for auditing then DHCPv6 would be the way to go. DHCPv6 also allows you tie into DNS systems for host entry updates and gives you the ability to send other options information to clients. In the long term, DHCPv6 makes most sense for use in a managed environment.

5.2. End user devices

Most operating systems (OS) that are loaded on workstations and laptops in a typical enterprise support IPv6 today. However, there are various out-of-the-box nuances that one should be mindful about. For example, the default behavior of OSes vary, some may have IPv6 turned off entirely by default, some may only have certain features such as privacy addresses turned off while others have IPv6 fully enabled. It is important to note that most operating systems will, by default, prefer to use native IPv6 over IPv4 when enabled. Therefore, it is advised that enterprises investigate the default behavior of their installed OS base and account for it during the implementation of IPv6. Furthermore, some OSes may have tunneling mechanisms turned on by default and in such cases, it is recommended to administratively shut down such interfaces unless required. It is recommended that IPv6 be deployed at the network infrastructure level before it is rolled out to end user devices.

Smartphones and tablets are poised to become one of the major consumers of IP addresses and enterprises should be ready to deploy and support IPv6 on various networks that serve such devices. In general, support for IPv6 in these devices, albeit in its infancy, has been steadily rising. Most of the leading smartphone OSes have some level of support for IPv6. However, the level of configurable options are mostly at a minimum and are not consistent across all platforms. Also, it is fairly common to find IPv6 support on the wifi connection alone and not on the radio interface in these devices. This is sometimes due to the radio network not being ready or device related. An IPv6 enabled enterprise wifi network will allow the majority of these devices to connect via IPv6. Much work is still being done to bring the full IPv6 feature set across all interfaces (802.11, 3G, LTE, etc.) and platforms.

IPv6 support in peripheral equipment such as printers, IP Cameras, etc. has been steadily rising as well, although at a much slower pace than traditional OSes and Smartphones. Most newer devices are coming out with IPv6 support but there is still a large installed base of legacy peripheral devices that might need IPv4 for sometime to come. The audit phase mentioned earlier will make it easier for enterprises to plan for equipment upgrades, in line with their corporate equipment refresh cycle.

5.3. Corporate Systems

No IPv6 deployment will be successful without ensuring that all the corporate systems that enterprise uses as part of their IT infrastructure, support IPv6. Examples of such systems include, but are not limited to, Email, Video Conferencing, Telephony (VoIP), DNS, Radius, etc. All these systems must have their own detailed IPv6 rollout plan in conjunction with the network IPv6 rollout. It is important to note that DNS is one of the main anchors in an enterprise deployment, since most end hosts decide whether or not use IPv6 based on the presence of AAAA records in a reply to a DNS query. It is recommended that system administrators selectively turn on AAAA records for various systems as and when they are IPv6 enabled. Additionally, all monitoring and reporting tools across the enterprise would need to be modified to support IPv6.

5.4. Security

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

Privacy extension addresses [RFC4941] pose a real challenge for audit trail. Therefore, it is recommended not to use them within the enterprise network by using the configuration described previously.

But, the biggest problem is probably linked to all threats against Neighbor Discovery. This means that the internal network at the access layer (i.e. where hosts connect to the network over wired or wireless) must implement RA-guard [RFC6105] and the techniques being specified by SAVI WG [I-D.ietf-savi-threat-scope].

6. Other Phases

To be added.

6.1. Guest network

To be added.

6.2. IPv6-only

Although IPv4 and IPv6 networks will coexist for a long time to come, the long term enterprise network roadmap should include steps on

gradually deprecating IPv4 from the dual-stack network. In some extreme cases, deploying dual-stack networks may not even be a viable option for very large enterprises due to lack of availability of RFC 1918 addresses. In such cases, deploying IPv6-only networks might be the only choice available to sustain network growth.

If nodes in the network don't need to talk to an IPv4-only node, then deploying IPv6-only networks should be fairly trivial. However, in the current environment, given that IPv4 is the dominant protocol on the Internet, an IPv6-only node most likely needs to talk to an IPv4-only node on the Internet. It is therefore important to provide such nodes with a translation mechanism to ensure communication between nodes configured with different address families. As [RFC6144] points out, it is important to look at address translation as a transition strategy that will get you to an IPv6-only network.

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

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

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

address those issues.

7. Considerations For Specific Enterprises

7.1. Content Delivery Networks

To be added.

7.2. Data Center Virtualization

Another document ([I-D.lopez-v6ops-dc-ipv6]) describes in details the specifics about IPv6 Data Center.

7.3. Campus Networks

A number of campus networks have made some initial IPv6 deployment. There are generally three areas in which such deployments may be made, which correspond to the Internal Phase, External Phase and Other Phase (Guest Network) descrobed above.

In particular the areas commonly approached are:

- o External-facing services. Typically the campus web presence and commonly also external-facing DNS and MX services.
- o Computer science department. This is where IPv6-related research and/or teaching is most likely to occur, so enabling some or all of the campus compauter science department network is a sensible first step.
- o The eduroam wireless network. Eduroam is the defacto wireless roaming system for academic networks, and uses 802.1X based authentication, which is agnostic to the IP version used (unlike web-redirection gateway systems).

8. Security Considerations

9. Acknowledgements

The authors would like to thank Chris Grundemann, Ray Hunter, Brian Carpenter, Tina Tsou for their comments on the mailing list.

10. IANA Considerations

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

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.

11.2. 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.
- [RFC2011] McCloghrie, K., "SNMPv2 Management Information Base for the Internet Protocol using SMIV2", RFC 2011, November 1996.
- [RFC2096] Baker, F., "IP Forwarding Table MIB", RFC 2096, January 1997.
- [RFC2827] Ferguson, P. and D. Senie, "Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing", BCP 38, RFC 2827, May 2000.
- [RFC3315] Droms, R., Bound, J., Volz, B., Lemon, T., Perkins, C., and M. Carney, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", RFC 3315, July 2003.
- [RFC3971] Arkko, J., Kempf, J., Zill, B., and P. Nikander, "Secure Neighbor Discovery (SEND)", RFC 3971, March 2005.
- [RFC3972] Aura, T., "Cryptographically Generated Addresses (CGA)", RFC 3972, March 2005.
- [RFC4057] Bound, J., "IPv6 Enterprise Network Scenarios", RFC 4057, June 2005.
- [RFC4193] Hinden, R. and B. Haberman, "Unique Local IPv6 Unicast Addresses", RFC 4193, October 2005.
- [RFC4213] Nordmark, E. and R. Gilligan, "Basic Transition Mechanisms for IPv6 Hosts and Routers", RFC 4213, October 2005.

- [RFC4293] Routhier, S., "Management Information Base for the Internet Protocol (IP)", RFC 4293, April 2006.
- [RFC4296] Bailey, S. and T. Talpey, "The Architecture of Direct Data Placement (DDP) and Remote Direct Memory Access (RDMA) on Internet Protocols", RFC 4296, December 2005.
- [RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 2006.
- [RFC4443] Conta, A., Deering, S., and M. Gupta, "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", RFC 4443, March 2006.
- [RFC4659] De Clercq, J., Ooms, D., Carugi, M., and F. Le Faucheur, "BGP-MPLS IP Virtual Private Network (VPN) Extension for IPv6 VPN", RFC 4659, September 2006.
- [RFC4861] Narten, T., Nordmark, E., Simpson, W., and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, September 2007.
- [RFC4862] Thomson, S., Narten, T., and T. Jinmei, "IPv6 Stateless Address Autoconfiguration", RFC 4862, September 2007.
- [RFC4890] Davies, E. and J. Mohacsi, "Recommendations for Filtering ICMPv6 Messages in Firewalls", RFC 4890, May 2007.
- [RFC4941] Narten, T., Draves, R., and S. Krishnan, "Privacy Extensions for Stateless Address Autoconfiguration in IPv6", RFC 4941, September 2007.
- [RFC5095] Abley, J., Savola, P., and G. Neville-Neil, "Deprecation of Type 0 Routing Headers in IPv6", RFC 5095, December 2007.
- [RFC5102] Quittek, J., Bryant, S., Claise, B., Aitken, P., and J. Meyer, "Information Model for IP Flow Information Export", RFC 5102, January 2008.
- [RFC5157] Chown, T., "IPv6 Implications for Network Scanning", RFC 5157, March 2008.
- [RFC5211] Curran, J., "An Internet Transition Plan", RFC 5211, July 2008.
- [RFC5214] Templin, F., Gleeson, T., and D. Thaler, "Intra-Site Automatic Tunnel Addressing Protocol (ISATAP)", RFC 5214,

March 2008.

- [RFC5375] Van de Velde, G., Popoviciu, C., Chown, T., Bonness, O., and C. Hahn, "IPv6 Unicast Address Assignment Considerations", RFC 5375, December 2008.
- [RFC5722] Krishnan, S., "Handling of Overlapping IPv6 Fragments", RFC 5722, December 2009.
- [RFC5798] Nadas, S., "Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6", RFC 5798, March 2010.
- [RFC5952] Kawamura, S. and M. Kawashima, "A Recommendation for IPv6 Address Text Representation", RFC 5952, August 2010.
- [RFC6052] Bao, C., Huitema, C., Bagnulo, M., Boucadair, M., and X. Li, "IPv6 Addressing of IPv4/IPv6 Translators", RFC 6052, October 2010.
- [RFC6104] Chown, T. and S. Venaas, "Rogue IPv6 Router Advertisement Problem Statement", RFC 6104, February 2011.
- [RFC6105] Levy-Abegnoli, E., Van de Velde, G., Popoviciu, C., and J. Mohacsi, "IPv6 Router Advertisement Guard", RFC 6105, February 2011.
- [RFC6144] Baker, F., Li, X., Bao, C., and K. Yin, "Framework for IPv4/IPv6 Translation", RFC 6144, April 2011.
- [RFC6145] Li, X., Bao, C., and F. Baker, "IP/ICMP Translation Algorithm", RFC 6145, April 2011.
- [RFC6146] Bagnulo, M., Matthews, P., and I. van Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", RFC 6146, April 2011.
- [RFC6147] Bagnulo, M., Sullivan, A., Matthews, P., and I. van Beijnum, "DNS64: DNS Extensions for Network Address Translation from IPv6 Clients to IPv4 Servers", RFC 6147, April 2011.
- [RFC6164] Kohno, M., Nitzan, B., Bush, R., Matsuzaki, Y., Colitti, L., and T. Narten, "Using 127-Bit IPv6 Prefixes on Inter-Router Links", RFC 6164, April 2011.
- [RFC6192] Dugal, D., Pignataro, C., and R. Dunn, "Protecting the Router Control Plane", RFC 6192, March 2011.

- [RFC6302] Durand, A., Gashinsky, I., Lee, D., and S. Sheppard, "Logging Recommendations for Internet-Facing Servers", BCP 162, RFC 6302, June 2011.
- [RFC6434] Jankiewicz, E., Loughney, J., and T. Narten, "IPv6 Node Requirements", RFC 6434, December 2011.
- [I-D.xli-behave-divi]
Shang, W., Li, X., Zhai, Y., and C. Bao, "dIVI: Dual-Stateless IPv4/IPv6 Translation", draft-xli-behave-divi-04 (work in progress), October 2011.
- [I-D.gashinsky-v6ops-v6nd-problems]
Jaeggli, J., Kumari, W., and I. Gashinsky, "Operational Neighbor Discovery Problem", draft-gashinsky-v6ops-v6nd-problems-00 (work in progress), October 2011.
- [I-D.ietf-savi-threat-scope]
McPherson, D., Baker, F., and J. Halpern, "SAVI Threat Scope", draft-ietf-savi-threat-scope-05 (work in progress), April 2011.
- [I-D.lopez-v6ops-dc-ipv6]
Chen, Z., Lopez, D., Tsou, T., and C. Zhou, "A Reference Framework for DC Migration to IPv6", draft-lopez-v6ops-dc-ipv6-02 (work in progress), June 2012.
- [I-D.templin-v6ops-isops]
Templin, F., "Operational Guidance for IPv6 Deployment in IPv4 Sites using ISATAP", draft-templin-v6ops-isops-17 (work in progress), May 2012.

Authors' Addresses

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

Email: kk@google.com

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

Email: tjc@ecs.soton.ac.uk

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

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

Victor Kuarsingh
Rogers Communications
8200 Dixie Road
Brampton, Ontario
Canada

Email: victor.kuarsingh@rci.rogers.com

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

Email: Yanick.Pouffary@hp.com

Eric Vyncke
Cisco Systems
De Kleetlaan 6a
Diegem 1831
Belgium

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

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

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

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

Abstract

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

Status of This Memo

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

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

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

This Internet-Draft will expire on September 29, 2013.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	2
2. Tunnel Fragmentation and Reassembly	3
3. Jumbo Packet Accommodation	5
4. Common Tunneling Mechanisms	5
5. IANA Considerations	5
6. Security Considerations	5
7. Acknowledgments	5
8. References	6
8.1. Normative References	6
8.2. Informative References	6
Author's Address	7

1. Introduction

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

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

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

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

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

2. Tunnel Fragmentation and Reassembly

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

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

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

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

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

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

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

3. Jumbo Packet Accommodation

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

4. Common Tunneling Mechanisms

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

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

5. IANA Considerations

There are no IANA considerations for this document.

6. Security Considerations

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

7. Acknowledgments

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

8. References

8.1. Normative References

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

8.2. Informative References

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

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

Author's Address

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

Email: fltemplin@acm.org

V6OPS WG
Internet-Draft
Intended status: Informational
Expires: October 25, 2013

S. Gundavelli
M. Grayson
Cisco
P. Seite
France Telecom - Orange
Y. Lee
Comcast
April 23, 2013

Service Provider Wi-Fi Services Over Residential Architectures
draft-gundavelli-v6ops-community-wifi-svcs-06.txt

Abstract

The tremendous growth in Wi-Fi technology adoption over the last decade has met the ultimate possible goal of 100% adoption rate. All most every new mobile device is now equipped with IEEE 802.11-based wireless interface and with pre-configured policy to prefer Wi-Fi to cellular access. Matching this evolution is every service provider's desire to offer Wi-Fi based broadband services; a new business opportunity even for fixed line operators. Operators are exploring options to monetize their existing networks, most with nation-wide footprint, to build a high-speed Wi-Fi service that can be the basis for offering new wireless broadband services. This document identifies the requirements for supporting these new Wi-Fi community services and the mobility tools which have been standardized in IETF that can be used for enabling these architectures.

Status of this Memo

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

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

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

This Internet-Draft will expire on October 25, 2013.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	4
2. Conventions and Terminology	5
2.1. Conventions	5
2.2. Terminology	5
3. Deployment Models	7
4. Requirements	8
4.1. IPv6 Addressing Model for SP WiFi Architectures	9
4.2. Subscriber Authentication & Service Authorization	9
4.3. Location-based Services	9
4.4. Local Services Access & Internet Traffic Offload	10
4.5. Web-based Authentication Support	10
4.6. Transparent Auto Login (TAL)	10
4.7. Multiple WLAN SSID Support	11
4.8. Multiple Home Network Service (APN) Access	11
4.9. CPE Identity and Authorization	11
4.10. Mobility within the WLAN Access Network	11
4.11. Mobility across WLAN and Macro Access	12
4.12. Differentiated Services for Users behind RG	12
4.13. Lawful Intercept (LI)	12
4.14. Subscriber Management and Charging	13
4.15. Handling the Walk-by Users	14
4.16. Overlapping IPv4 Address Support	14
4.17. Service Provisioning & Monitoring	14
5. Solution Approaches & Considerations	15
5.1. PMIPv6 MAG on the RG: Layer-3 Encapsulation between CPE and Access Gateway	15
5.2. Ethernet-over-IP Support on the RG: Layer-2 Encapsulation between CPE and Access Gateway	15
5.3. Local Aggregation for Subscriber Control and Internet Offload	15
5.4. Mobility Chaining: Integration with Mobile Packet Core	15
6. IANA Considerations	15
7. Security Considerations	15
8. Acknowledgements	16
9. References	16
9.1. Normative References	16
9.2. Informative References	16
Authors' Addresses	17

1. Introduction

The tremendous growth in Wi-Fi technology adoption over the last decade has met the ultimate possible goal of 100% adoption rate. All most every new mobile device is now equipped with IEEE 802.11-based wireless interface and these devices are typically pre-configured with a policy to prefer Wi-Fi to cellular access. This so called, "cheap access based on unlicensed spectrum", is no longer considered an unreliable access, but with all the available protocol tools and with maturity in technology, building a reliable broadband service that can meet the committed service-level agreements is proving to be a non-issue.

Matching this evolution is every service provider's desire to offer Wi-Fi based broadband services; a new business opportunity even for both fixed and mobile operators. The demand for bandwidth is only growing with the availability of new smart devices, new technology applications and with all the content in the Internet. Furthermore, an increasing percentage of mobile consumption is happening in the home and so DSL/Cable operators are exploring options to monetize their existing networks, most with nation-wide footprint, to build a high-speed, nation-wide Wi-Fi service that can be the basis for offering new wireless broadband services and for building roaming agreements with traditional mobile operators, who are unable to meet the mobile subscriber growth due to the finite licensed spectrum available for macro-cell deployments. Every residential CPE device that the operator owns can now be enabled to provide Wi-Fi service and new community Wi-Fi hotspots can be built in any location where there is fixed line coverage. A wireless service based on unlicensed spectrum, and leveraging existing transport is a huge incentive for operators to enter this new market.

To support these business goals, operators are looking at mobility architectures for supporting various requirements. Not all requirements are well understood, and neither are the implications with the chosen solution approaches for each of those requirements. The choice of the architecture has an implication on the CPE evolution and on the core infrastructure feature requirements. Therefore, the sole purpose and the goal of this document is to present all the requirements, identify the protocol tools and any potential gaps. This analysis is important for enabling the network vendors and the mobile operators to make the right design choices and leverage the existing tools that the mobility groups in IETF have already developed and discourage them from adopting proprietary, non-standard mechanisms or developing redundant alternatives.

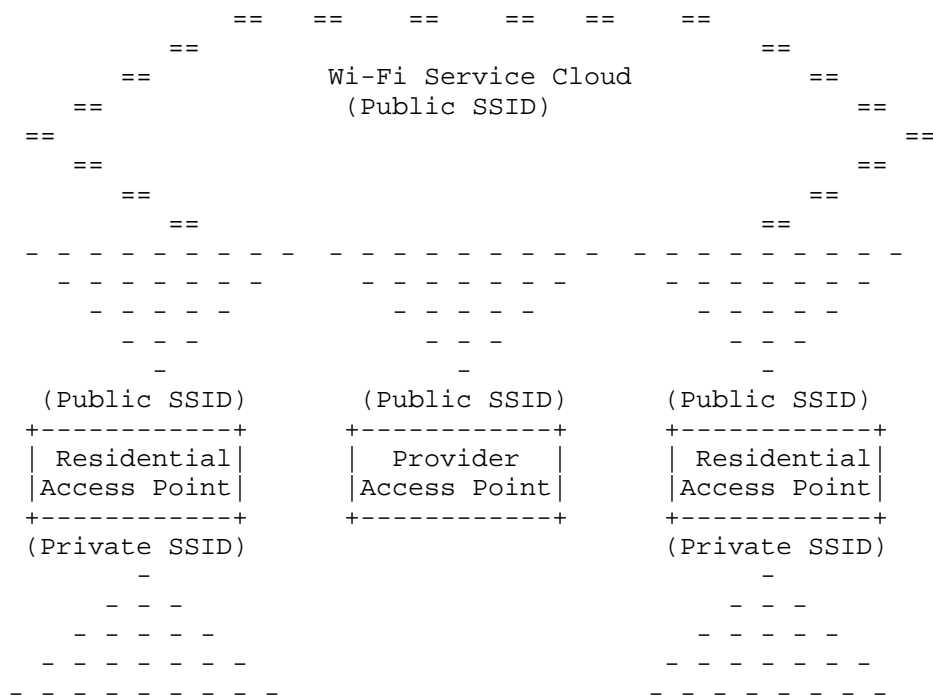


Figure 1: Wi-Fi Cloud Over Residential Gateways

2. Conventions and Terminology

2.1. Conventions

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

2.2. Terminology

This document uses the following abbreviations and definitions:

Community Wi-Fi Service

It is a Wi-Fi based broadband service offered by a service provider. The Wi-Fi Access Points that are part of this service are owned and managed by the operator, and physically located in carrier premises. These operator owned CPE's typically have a large Wi-Fi coverage area, operated on a higher signal power.

There could also be the residential Access Points that are part of this service, located in the subscriber homes, that are part of this service and allowing community access to a public SSID along with a private SSID for their personal access.

Wi-Fi Operator

A service provider that offers Community Wi-Fi services. Wi-Fi operator can be a wireline operator, mobile operator or an operator offering both wireline and mobile services.

Residential Gateway (RG)

It is a network device that is located in the Customer premises and is also referred to as Residential CPE (Customer Premises Equipment). This device is connected to service providers network and defines the demarcation point between the provider and the customer. In the context of this document this is hosting the 802.11 Access Point function.

WLAN controller (WLC)

It is an entity responsible for performing radio resource management (RRM) on the Access Points, system-wide mobility policy enforcement and centralized forwarding function for the user traffic.

Mobile Gateway

It is network entity anchoring IP traffic in the mobile core network. This entity allocates an IP address which is topologically valid in the mobile network and may act as a mobility anchor if handover between mobile and Wi-Fi is supported.

Home/Roaming User

The home user is the owner of the network where the Residential Gateway is located and is paying for the service associated with that Residential Gateway. A Roaming User is a visitor from the operator's home network, or from a partner's network and is allowed to access broadband services using that Residential Gateway and over a Public SSID.

Access Point Name (APN)

Its the name of a packet data network. This APN concept was first introduced in GPRS by 3GPP to enable legacy Intelligent Networking (IN) approaches to be applied to the newly deployed IP packet data services. In roaming deployments, the APN construct was visible to the visited network and allowed legacy IN charging solutions to be supported. Defining an application specific APN then allowed application charging to be supported.

Addressing Models

The term Per-MN-Prefix model [RFC5213] is used to refer to an addressing model where there is a unique network prefix or prefixes assigned for each mobile node. The term Shared-Prefix model [RFC5213] is used to refer to an addressing model where the prefix(es) are shared by more than one node.

3. Deployment Models

Figure 2 illustrates the most common residential and hotspots Wi-Fi deployment models.

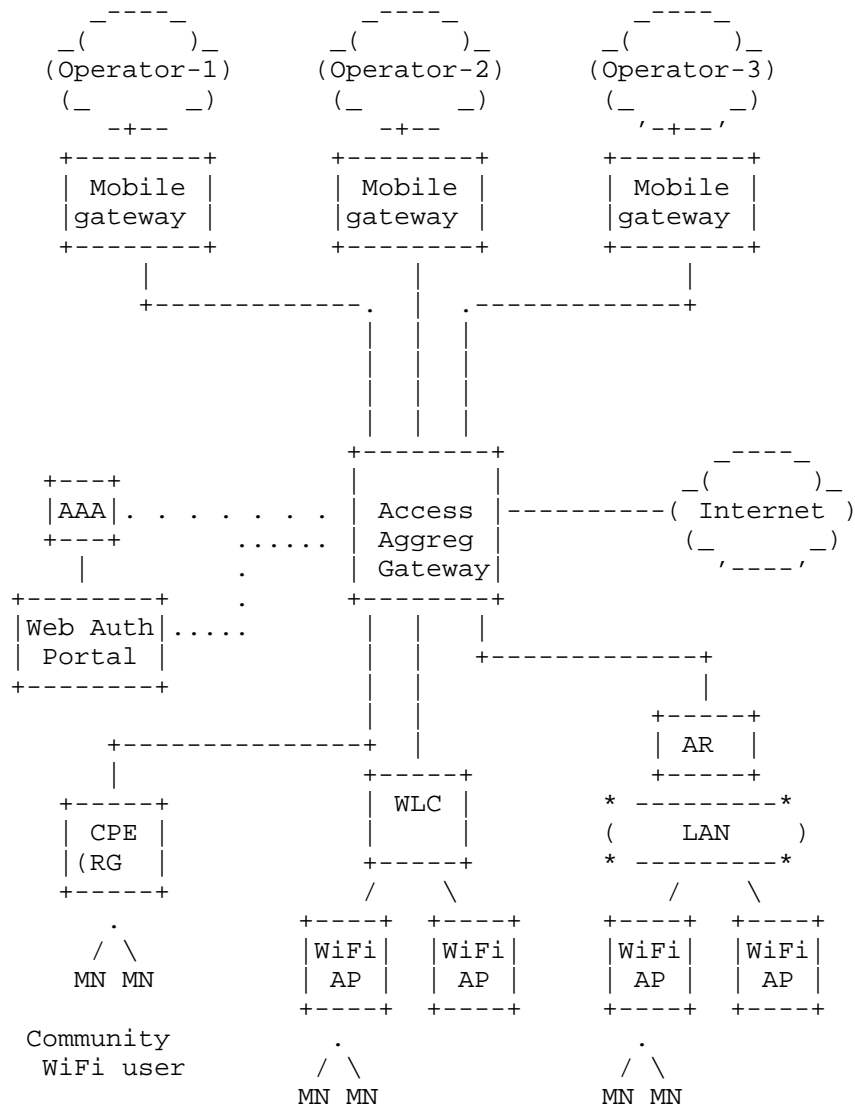


Figure 2: WLAN Service for Retail Model

4. Requirements

4.1. IPv6 Addressing Model for SP WiFi Architectures

The selection of the right IPv6 addressing model for the SP WiFi architectures is an important consideration. There are these two IPv6 addressing models:

- o Unique-Prefix Model - As per this addressing model, home network prefix(es) assigned to a mobile node are for its exclusive use and no other node shares an address from that prefix (other than the Subnet-Router anycast address [RFC4291] that is used by the IPv6 access router hosting that prefix on that link). There could be multiple unique IPv6 prefixes assigned to each mobile node.
- o Shared-Prefix model - The IPv6 prefix that is assigned to the mobile node is a shared prefix. There can be more than one mobile node that can be using IPv6 addresses from that prefix.

3GPP architecture supports Unique-Prefix model for the mobile node's PDN connections. This decision was largely influenced by the IETF recommendation to 3GPP to support this specific addressing model. In the context of SP WiFi, there are clearly scenarios where a mobile node may perform an inter-technology handover from the macro network to the WLAN access network and handoff the session and is important that the addressing model is the same in both the access architectures. Even in deployment models where such handovers are not envisioned, such as an WLAN access aggregation architecture with no mobile packet core integration, there are sufficient reasons for adopting the Unique Prefix model.

4.2. Subscriber Authentication & Service Authorization

Community Wi-Fi service is designed to be available for public access. Wi-Fi operator must authenticate users before offering services to them. Once a user is authenticated, Wi-Fi operator will authorize services based on the user identity. There are many authentication mechanisms, such as 802.1x, Web-authentication, WISPr that the operator may deploy for this purpose.

4.3. Location-based Services

In many deployments, there is a need for the mobile operator to provide differentiated services and policing to the mobile nodes based on the access network to which they are attached. Policy systems in mobility architectures such as PCC and ANDSF in 3GPP system allow configuration of policy rules with conditions based on the access network information. For example, the service treatment for the mobile node's traffic may be different when they are attached to a access network owned by the home operator than when owned by a

roaming partner. The service treatment can also be different based on the configured Service Set Identifiers (SSID) in case of IEEE 802.11 based access networks. Other examples of location services include the operator's ability to display a location specific Web Page, or apply tariff based on the location.

4.4. Local Services Access & Internet Traffic Offload

In the integrated WLAN-EPC architectures, the mobile node's IP traffic is always tunneled back from the access network to the mobile gateway in the home network. However, with the exponential growth in the mobile data traffic, mobile operators are exploring new ways to offload some of the IP traffic flows at the nearest access edge where ever there is an internet peering point, as supposed to carrying it all the way to the mobility anchor in the home network. Not all IP traffic need to be routed back to the home network, some of the non-essential traffic which does not require IP mobility support can be offloaded at the mobile access gateway in the access network. This approach provides greater leverage and efficient usage of the mobile packet core which help lowering transport cost.

4.5. Web-based Authentication Support

Most Public Wireless LAN (PWLAN) deployments today use web-based authentication for authorizing the user for network access. Web-based mode of authentication is considered a legacy mode, for its weak security properties, and there are efforts to replace it with 802.1x-based security mechanisms. However, a very high percentage of the PWLAN deployments are still using using this authentication mode and operators are not willing to move away from this mode any time soon. The reason being, lack of support for 802.1x/EAP support on the 100's of millions of handsets that are out there, and for the lack of client software in the laptops running various operating systems versions. This is forcing the operators to support web-based authentication.

4.6. Transparent Auto Login (TAL)

In many deployments, there is a need to support Transparent Auto Login capability. This is essentially an approach for maintaining Authenticated state for a user, for a duration of time. Once an authenticated user disconnects and re-attaches to the network, the network should allows instant access without forcing the user to re-authenticate.

4.7. Multiple WLAN SSID Support

A Wi-Fi Operator may broadcast multiple SSIDs. In case Residential Wi-Fi hotspots, there can be one set of private SSIDs specific to that home user and there can be another set of public SSIDs for wider community use. In case of public hotspots, the operator can advertise the public SSID for its own subscribers and also public SSID's belonging to other operators with whom the operator has roaming relationships.

4.8. Multiple Home Network Service (APN) Access

The 3GPP system architecture supports the concept of an Access Point Name (APN). An APN can identify a particular routing domain and can be used by 3GPP operators to segment user traffic. APNs are included in the session establishment signaling sent by 3GPP User Equipments (UEs), identifying which routing domain they want to be connected to. Furthermore, 3GPP has defined a system architecture which supports the ability of a single UE to have simultaneous connectivity to a plurality of APNs, and be allocated multiple IPv4 addresses and/or IPv6 prefixes from the network.

There is a need to ensure multiple APN access for a subscriber in the community Wi-Fi network.

4.9. CPE Identity and Authorization

There are two known models with respect to CPE roll out. The consumer may purchase a device off the shelf and plugin to the network, or the operator at the time of service creation may have shipped a new device with the pre-provisioned service configuration. In either case, the operator needs to be able to identify the device based on the IP address and associate that to a given location.

The Wi-Fi network performs access control of UEs, via the CPE acting as AAA supplicant. As a result, the mobile network does not authenticate directly the user but shall trust the CPE performing the authentication.

4.10. Mobility within the WLAN Access Network

The mobile node should have the ability to roam within the Wi-Fi domain. Depending on the deployment model, the mobile node may roam across different IP subnets. To survive to such handover, some applications (e.g. VPN, streaming) need the IP address to be preserved.

A WLAN network may include a large number of Wi-Fi base stations. In

some occasions, two or more Wi-Fi base stations may cover the same area. When a subscriber receives Wi-Fi service in this overlapped area, the device may bounce between different base stations. This is typical Proximity problem. In this scenario, it is important for the WLAN to offer mobility to the subscriber as such the subscriber can continue the services without changing its IP address.

4.11. Mobility across WLAN and Macro Access

A mobile node should have the ability to handover from macro network to the Wi-Fi network and be able to retain IP address configuration and be able to access the home operator services.

4.12. Differentiated Services for Users behind RG

A Wi-Fi operator enabling Hotspot Services on a residential gateway is required to ensure the service levels for the home user is not impacted as a result of opening up the service for public usage. The home user should always have preferred access over public users and the operator may be bound to meet the Service Level Agreements. This essentially requires the operator to be able to differentiate the service flows and apply differentiated service treatment. The operator should be able to enforce QoS policing and labeling of packets to enforce QoS differentiation.

A single operator has deployed both a fixed access network and a mobile access network. In this scenario, the operator may wish a harmonized QoS management on both accesses. However the fixed access network does not implement a QoS control framework. So, the operator may choose to rely on the mobile network, specifying the standard framework to provide a QoS control, to enforce the QoS policy from the mobile gateway to the Wi-Fi Access network.

4.13. Lawful Intercept (LI)

Lawful Intercept [RFC2119] stands for legally authorized interception and monitoring of communications to and from a subscriber under Surveillance by a Law Enforcement Agency. In most of the countries, there are legal obligations for Service Providers to facilitate the intercept of any subscriber's communication if requested by law enforcement agencies. Communications Assistance for Law Enforcement Act (CALEA), the United States wiretapping law passed in 1994 is an example for such legal mandates. This section talks about Lawful Intercept solution requirements that are operators are required to support when offering WLAN services.

The following are the key considerations with respect to supporting Lawful Intercept capability in Wi-Fi architectures.

- o The operator should have the ability to capture IP traffic from any of the mobile nodes for which the operator is offering Wi-Fi services.
- o The ability to identify the Geo-location of the mobile node to the nearest WLAN access point.
- o The ability to track the mobile node's roaming within the network, even when there are no active IP flows.
- o The ability to pre-provision Lawful Intercept for an inactive mobile node so that the capture of IP traffic can be initiated anytime new IP flows associated to that mobile node are detected.
- o Lawful Intercept (LI) should be undetectable by the intercept subject
- o Mechanisms should be in place to limit unauthorized personnel from performing or knowing about lawfully authorized intercepts
- o If the information being intercepted is encrypted by the service provider and the service provider has access to the keys, then the information should be decrypted before delivery to the Law Enforcement Agency (LEA) or the encryption keys should be passed to the Law Enforcement Agency to allow them to decrypt the information.

4.14. Subscriber Management and Charging

It refers to the capability to manage network resources on a per subscriber, and eventually on a per-flow, basis. Subscriber management should be able to maintain a user context associating the user identifier with specific network resource (e.g. IP address, default router, mobility/traffic anchoring point,...), QoS profile, billing context and specific network functions (e.g. legal interception). The user context includes traffic selectors if subscriber management is on a per flow basis. Subscriber management should be done according to the user subscription, the user preferences and/or operator policies.

The ability to charge the subscriber is the fundamental business requirement before an operator can deploy the Wi-Fi service. The operator should have the ability to enforce charge the subscriber by usage and enforce quota policies. This is the basis for keeping the service operational and managing inter-operator roaming agreements.

4.15. Handling the Walk-by Users

In the case of community Wi-Fi, the network is an open network with the SSID visible to any wireless LAN device. This essentially creates a situation where any walk-by user's mobile terminal automatically gets connected to the Wi-Fi network and results in a subscriber session creation. The user may not be having any intention in connecting to the Wi-Fi network and in fact may not be using the mobile device, but the device gets attached to the network and a subscriber session and other network resources get locked up for that user session. The situation is especially worse in public hotspots such as train stations, or Airports where there is high traffic. This is important that this situation is correctly handled.

4.16. Overlapping IPv4 Address Support

The transition from IPv4 to IPv6 is a long process, and during this period of transition, the Wi-Fi operators will have to continue to offer IPv4 services. However, these operators may not have sufficient public IPv4 addresses for all the Wi-Fi devices in their network. For addressing this IPv4 exhaust issue, operators may have to leverage transitioning technologies such as NAT64, Dual-Stack Lite, 6rd or other approaches. These operators may also choose to segment the network into regions and two regions may use overlapped IPv4 address space to provide IPv4 services to users.

In a different scenario, a roaming user from a partner's network, with an established mobility session with her home network, may be using a private IPv4 address and this IPv4 address may be overlapping with the address space that is being used in this access network. Furthermore, the IPv4 address space that is used for assignment to Wi-Fi subscribers should not conflict with the IPv4 addresses used on the Cable/DSL transport network.

The Wi-Fi operator should be able to handle all these scenarios related to overlapping private IPv4 address usage.

4.17. Service Provisioning & Monitoring

Deployment of any community based Wi-Fi access will require additional Wi-Fi specific configuration on a per Residential Gateway basis. In order to support scalable deployment, the Service Providers should be able to provision these configuration options remotely. This remote provisioning framework must support the following:

- o Secure provisioning of the RG with community WiFi parameters to minimize the theft of service
- o Ability to separate the private home subscriber traffic from the community WiFi traffic in the access network
- o Privacy and protection of private Residential subscriber traffic from the community WiFi users
- o Ability to remotely shut down an Residential Gateway which has been hijacked by hackers and is being used for DoS attacks.
- o Ability to temporarily disable services for the community based WiFi support while maintaining service to the Residential fixed broadband subscriber
- o Seamless integration of the WiFi provisioning aspects of the Residential Gateway into the existing RG provisioning infrastructure implemented by the Fixed Broadband Providers
- o Dynamic Service Monitoring Capability for managing the Wi-Fi Service.

5. Solution Approaches & Considerations

The following section identifies the different mobility approaches that Wi-Fi operator can leverage for deploying this Wi-Fi services.

- 5.1. PMIPv6 MAG on the RG: Layer-3 Encapsulation between CPE and Access Gateway
- 5.2. Ethernet-over-IP Support on the RG: Layer-2 Encapsulation between CPE and Access Gateway
- 5.3. Local Aggregation for Subscriber Control and Internet Offload
- 5.4. Mobility Chaining: Integration with Mobile Packet Core

6. IANA Considerations

This document does not require any IANA actions.

7. Security Considerations

This specification identifies the requirements for enabling Community

Wi-Fi Services over Residential architectures and the potential solution approaches for addressing those requirements. The security analysis for each of those requirements are covered in those respective sections.

8. Acknowledgements

The authors would like to thank Bill Choinski, John Coppola and Sangeeta Ramakrishnan for all the discussions related to Service Provider Wi-Fi Service requirements. The authors would also like to thank Byju Pularikkal for all the discussions and text contributions related to Lawful Interception and Service Provisioning.

9. References

9.1. Normative References

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

9.2. Informative References

- [I-D.gundavelli-netext-multiple-apn-pmipv6]
Gundavelli, S., Grayson, M., Lee, Y., Deng, H., and H. Yokota, "Multiple APN Support for Trusted Wireless LAN Access", draft-gundavelli-netext-multiple-apn-pmipv6-01 (work in progress), February 2012.
- [I-D.gundavelli-netext-pmipv6-wlan-applicability]
Gundavelli, S., "Applicability of Proxy Mobile IPv6 Protocol for WLAN Access Networks", draft-gundavelli-netext-pmipv6-wlan-applicability-03 (work in progress), April 2012.
- [I-D.ietf-netext-pmipv6-qos]
Liebsch, M., Seite, P., Yokota, H., Korhonen, J., and S. Gundavelli, "Quality of Service Option for Proxy Mobile IPv6", draft-ietf-netext-pmipv6-qos-00 (work in progress), June 2012.
- [I-D.ietf-netext-pmipv6-sipto-option]
Gundavelli, S., Zhou, X., Korhonen, J., and R. Koodli, "IPv4 Traffic Offload Selector Option for Proxy Mobile IPv6", draft-ietf-netext-pmipv6-sipto-option-07 (work in progress), October 2012.

- [I-D.liebsch-netext-pmip6-authiwb]
Gundavelli, S., Liebsch, M., and P. Seite, "PMIPv6 inter-working with WiFi access authentication", draft-liebsch-netext-pmip6-authiwb-05 (work in progress), September 2012.
- [RFC2663] Srisuresh, P. and M. Holdrege, "IP Network Address Translator (NAT) Terminology and Considerations", RFC 2663, August 1999.
- [RFC3924] Baker, F., Foster, B., and C. Sharp, "Cisco Architecture for Lawful Intercept in IP Networks", RFC 3924, October 2004.
- [RFC4291] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 4291, February 2006.
- [RFC5213] Gundavelli, S., Leung, K., Devarapalli, V., Chowdhury, K., and B. Patil, "Proxy Mobile IPv6", RFC 5213, August 2008.
- [RFC5844] Wakikawa, R. and S. Gundavelli, "IPv4 Support for Proxy Mobile IPv6", RFC 5844, May 2010.
- [RFC6757] Gundavelli, S., Korhonen, J., Grayson, M., Leung, K., and R. Pazhyannur, "Access Network Identifier (ANI) Option for Proxy Mobile IPv6", RFC 6757, October 2012.
- [TS23402] 3GPP, "Architecture enhancements for non-3GPP accesses", 2010.

Authors' Addresses

Sri Gundavelli
Cisco
170 West Tasman Drive
San Jose, CA 95134
USA

Email: sgundave@cisco.com

Mark Grayson
Cisco
11 New Square Park
Bedfont Lakes, FELTHAM TW14 8HA
ENGLAND

Email: mgrayson@cisco.com

Pierrick Seite
France Telecom - Orange
4, rue du clos courtel BP 91226
Cesson-Sevigne, 35512
France

Email: pierrick.seite@orange-ftgroup.com

Yiu L. Lee
Comcast
One Comcast Center
Philadelphia, PA 19103
U.S.A.

Email: yiul_lee@cable.comcast.com
URI: <http://www.comcast.com>

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: August 27, 2013

M. Mawatari
Japan Internet Exchange Co.,Ltd.
M. Kawashima
NEC AccessTechnica, Ltd.
C. Byrne
T-Mobile USA
February 23, 2013

464XLAT: Combination of Stateful and Stateless Translation
draft-ietf-v6ops-464xlat-10

Abstract

This document describes an architecture (464XLAT) for providing limited IPv4 connectivity across an IPv6-only network by combining existing and well-known stateful protocol translation RFC 6146 in the core and stateless protocol translation RFC 6145 at the edge. 464XLAT is a simple and scalable technique to quickly deploy limited IPv4 access service to IPv6-only edge networks without encapsulation.

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 27, 2013.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect

to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
2. Terminology	3
3. Motivation and Uniqueness of 464XLAT	4
4. Network Architecture	4
4.1. Wireline Network Architecture	4
4.2. Wireless 3GPP Network Architecture	5
5. Applicability	6
5.1. Wireline Network Applicability	6
5.2. Wireless 3GPP Network Applicability	7
6. Implementation Considerations	7
6.1. IPv6 Address Format	7
6.2. IPv4/IPv6 Address Translation Chart	7
6.3. IPv6 Prefix Handling	9
6.4. DNS Proxy Implementation	9
6.5. CLAT in a Gateway	9
6.6. CLAT to CLAT communications	9
7. Deployment Considerations	10
7.1. Traffic Engineering	10
7.2. Traffic Treatment Scenarios	10
8. Security Considerations	11
9. IANA Considerations	11
10. Acknowledgements	11
11. References	11
11.1. Normative References	11
11.2. Informative References	12
Appendix A. Examples of IPv4/IPv6 Address Translation	13
Authors' Addresses	14

1. Introduction

With the exhaustion of the unallocated IPv4 address pools, it will be difficult for many networks to assign IPv4 addresses to end users.

This document describes an IPv4 over IPv6 solution as one of the techniques for IPv4 service extension and encouragement of IPv6 deployment. 464XLAT is not a one-for-one replacement of full IPv4 functionality. The 464XLAT architecture only supports IPv4 in the client server model, where the server has a global IPv4 address. This means it is not fit for IPv4 peer-to-peer communication or inbound IPv4 connections. 464XLAT builds on IPv6 transport and includes full any-to-any IPv6 communication.

The 464XLAT architecture described in this document uses IPv4/IPv6 translation standardized in [RFC6145] and [RFC6146]. It does not require DNS64 [RFC6147] since an IPv4 host may simply send IPv4 packets, including packets to an IPv4 DNS server, which will be translated on the customer side translator (CLAT) to IPv6 and back to IPv4 on the provider side translator (PLAT). 464XLAT networks may use DNS64 [RFC6147] to enable single stateful translation [RFC6146] instead of 464XLAT double translation where possible. The 464XLAT architecture encourages the IPv6 transition by making IPv4 services reachable across IPv6-only networks and providing IPv6 and IPv4 connectivity to single-stack IPv4 or IPv6 servers and peers.

2. Terminology

PLAT: PLAT is Provider side translator(XLAT) that complies with [RFC6146]. It translates N:1 global IPv6 addresses to global IPv4 addresses, and vice versa.

CLAT: CLAT is Customer side translator(XLAT) that complies with [RFC6145]. It algorithmically translates 1:1 private IPv4 addresses to global IPv6 addresses, and vice versa. The CLAT function is applicable to a router or an end-node such as a mobile phone. The CLAT should perform IP routing and forwarding to facilitate packets forwarding through the stateless translation even if it is an end-node. The CLAT as a common home router or wireless Third Generation Partnership Project (3GPP) router is expected to perform gateway functions such as DHCP server and DNS proxy for local clients. The CLAT uses different IPv6 prefixes for CLAT-side and PLAT-side IPv4 addresses and therefore does not comply with the sentence "Both IPv4-translatable IPv6 addresses and IPv4-converted IPv6 addresses should use the same prefix." in Section 3.3 of [RFC6052]. The CLAT does not facilitate

communications between a local IPv4-only node and an IPv6-only node on the Internet.

3. Motivation and Uniqueness of 464XLAT

1. Minimal IPv4 resource requirements, maximum IPv4 efficiency through statistical multiplexing.
2. No new protocols required, quick deployment.
3. IPv6-only networks are simpler and therefore less expensive to operate than dual-stack networks.
4. Consistent native IP based monitoring, traffic engineering, and capacity planning techniques can be applied without the indirection or obfuscation of a tunnel.

4. Network Architecture

Examples of 464XLAT architectures are shown in the figures in the following sections.

Wireline Network Architecture can fit in the situations where there are clients behind the CLAT in the same way regardless of the type of access service, for example FTTH, DOCSIS, or WiFi.

Wireless 3GPP Network Architecture fits in the situations where a client terminates the wireless access network and may act as a router with tethered clients.

4.1. Wireline Network Architecture

The private IPv4 host on this diagram can reach global IPv4 hosts via translation on both CLAT and PLAT. On the other hand, the IPv6 host can reach other IPv6 hosts on the Internet directly without translation. This means that the CPE/CLAT can not only have the function of a CLAT but also the function of an IPv6 native router for native IPv6 traffic. The v4p host behind the CLAT on this diagram has [RFC1918] addresses.

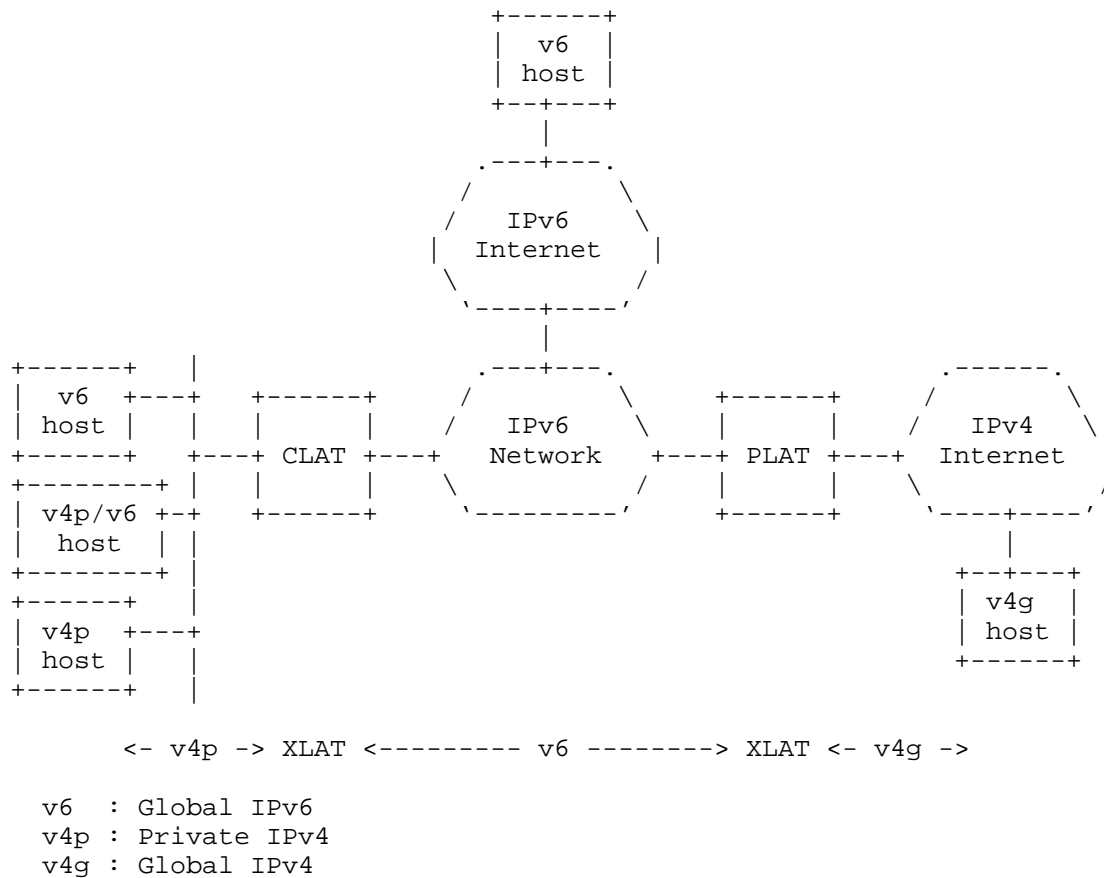


Figure 1: Wireline Network Topology

4.2. Wireless 3GPP Network Architecture

The CLAT function on the User Equipment (UE) provides an [RFC1918] address and IPv4 default route to the local node network stack. The applications on the UE can use the private IPv4 address for reaching global IPv4 hosts via translation on both the CLAT and the PLAT. On the other hand, reaching IPv6 hosts (including host presented via DNS64 [RFC6147]) does not require the CLAT function on the UE.

Presenting a private IPv4 network for tethering via NAT44 and stateless translation on the UE is also an application of the CLAT.

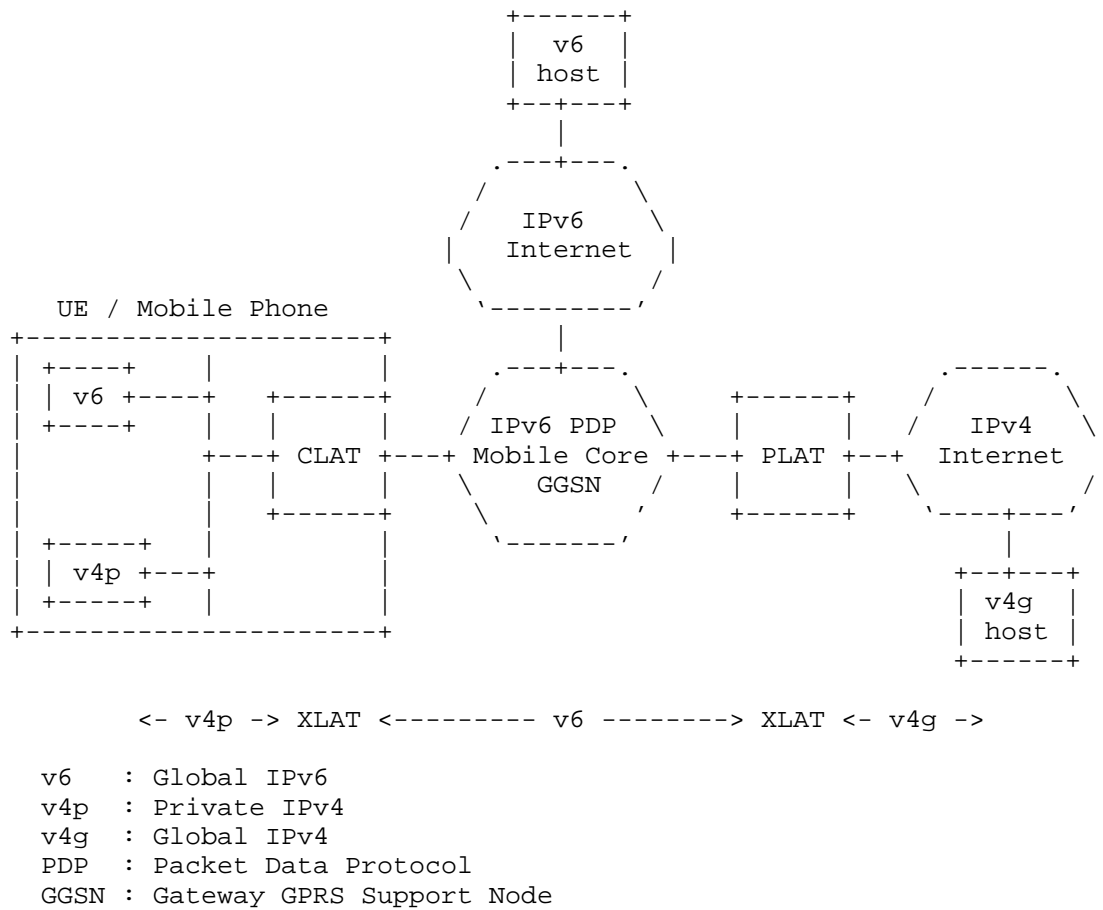


Figure 2: Wireless 3GPP Network Topology

5. Applicability

5.1. Wireline Network Applicability

When an Internet Service Provider (ISP) has IPv6 access service and provides 464XLAT, the ISP can provide outgoing IPv4 service to end users across an IPv6 access network. The result is that edge network growth is no longer tightly coupled to the availability of scarce IPv4 addresses.

If another ISP operates the PLAT, the edge ISP is only required to deploy an IPv6 access network. All ISPs do not need IPv4 access networks. They can migrate their access network to a simple and

highly scalable IPv6-only environment.

5.2. Wireless 3GPP Network Applicability

At the time of writing, in February 2013, the vast majority of mobile networks are compliant to Pre-Release 9 3GPP standards. In Pre-Release 9 3GPP networks, Global System for Mobile Communications (GSM) and Universal Mobile Telecommunications System (UMTS) networks must signal and support both IPv4 and IPv6 Packet Data Protocol (PDP) attachments to access IPv4 and IPv6 network destinations [RFC6459]. Since there are two PDPs required to support two address families, this is double the number of PDPs required to support the status quo of one address family, which is IPv4.

For the cases of connecting to an IPv4 literal or IPv4 socket that require IPv4 connectivity, the CLAT function on the UE provides a private IPv4 address and IPv4 default route on the host for the applications to reference and bind to. Connections sourced from the IPv4 interface are immediately routed to the CLAT function and passed to the IPv6-only mobile network, destined for the PLAT. In summary, the UE has the CLAT function that does a stateless translation [RFC6145], but only when required by an IPv4-only scenario such as IPv4 literals or IPv4-only sockets. The mobile network has a PLAT that does stateful translation [RFC6146].

464XLAT works with today's existing systems as much as possible. 464XLAT is compatible with existing network based deep packet inspection solutions like 3GPP standardized Policy and Charging Control (PCC) [TS.23203].

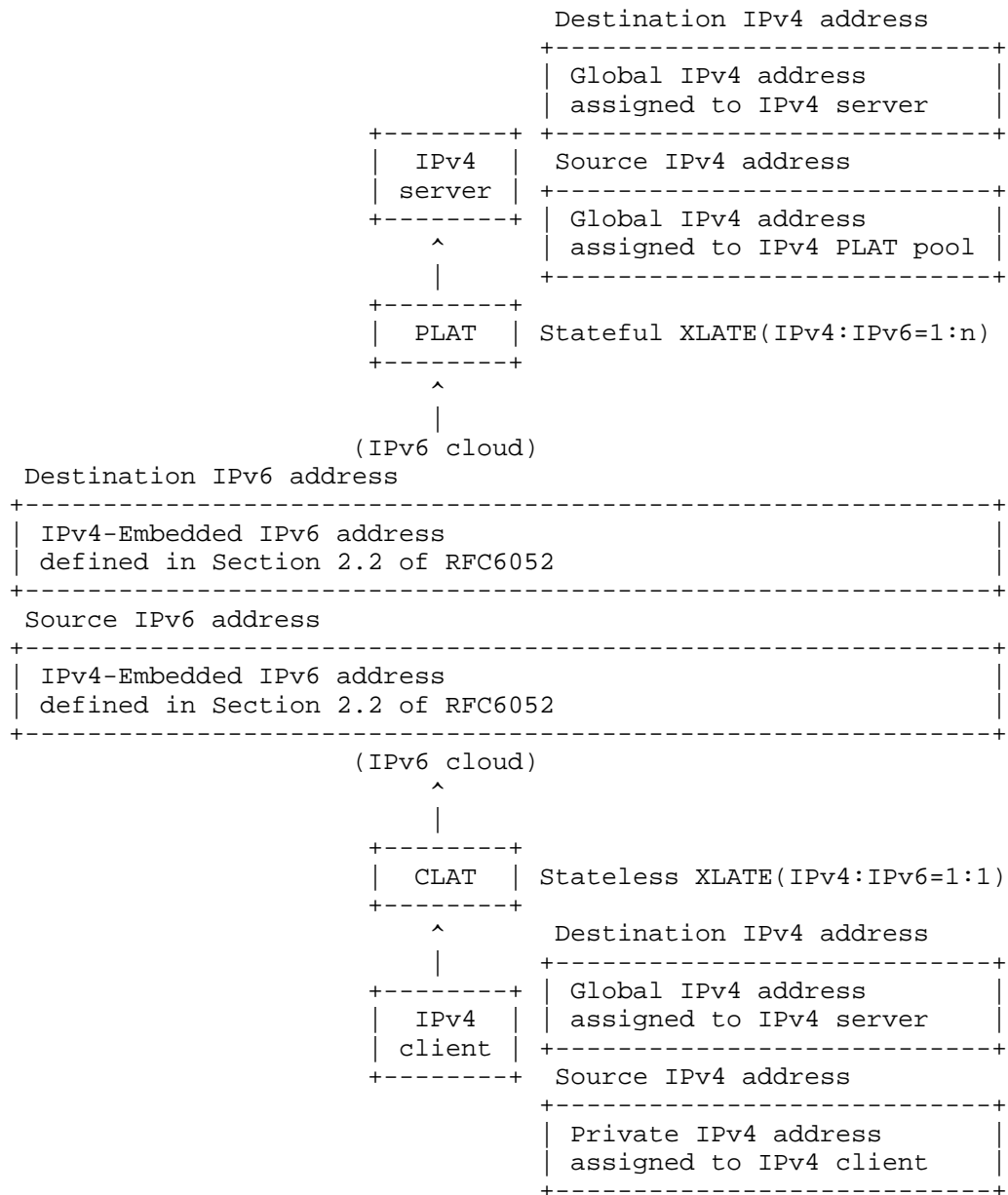
6. Implementation Considerations

6.1. IPv6 Address Format

The IPv6 address format in 464XLAT is defined in Section 2.2 of [RFC6052].

6.2. IPv4/IPv6 Address Translation Chart

This chart offers an explanation about address translation architecture using a combination of stateful translation at the PLAT and stateless translation at the CLAT. The client on this chart is delegated an IPv6 prefix from a prefix delegation mechanism such as DHCPv6-PD [RFC3633], therefore it has a dedicated IPv6 prefix for translation.



Case of enabling only stateless XLATE on CLAT

6.3. IPv6 Prefix Handling

There are two relevant IPv6 prefixes that the CLAT must be aware of.

First, CLAT must know its own IPv6 prefixes. The CLAT should acquire a /64 for the uplink interface, a /64 for all downlink interfaces, and a dedicated /64 prefix for the purpose of sending and receiving statelessly translated packets. When a dedicated /64 prefix is not available for translation from DHCPv6-PD [RFC3633], the CLAT may perform NAT44 for all IPv4 LAN packets so that all the LAN originated IPv4 packets appear from a single IPv4 address and are then statelessly translated to one interface IPv6 address that is claimed by the CLAT via NDP and defended with DAD.

Second, the CLAT must discover the PLAT-side translation IPv6 prefix used as a destination of the PLAT. The CLAT will use this prefix as the destination of all translation packets that require stateful translation to the IPv4 Internet. It may discover the PLAT-side translation prefix using [I-D.ietf-behave-nat64-discovery-heuristic]. In the future some other mechanisms, such as a new DHCPv6 option, will possibly be defined to communicate the PLAT-side translation prefix.

6.4. DNS Proxy Implementation

The CLAT should implement a DNS proxy as defined in [RFC5625]. The case of an IPv4-only node behind the CLAT querying an IPv4 DNS server is undesirable since it requires both stateful and stateless translation for each DNS lookup. The CLAT should set itself as the DNS server via DHCP or other means and proxy DNS queries for IPv4 and IPv6 LAN clients. Using the CLAT enabled home router or UE as a DNS proxy is a normal consumer gateway function and simplifies the traffic flow so that only IPv6 native queries are made across the access network. DNS queries from the client that are not sent to the DNS proxy on the CLAT must be allowed and are translated and forwarded just like any other IP traffic.

6.5. CLAT in a Gateway

The CLAT feature can be implemented in a common home router or mobile phone that has a tethering feature. Routers with a CLAT feature should also provide common router services such as DHCP of [RFC1918] addresses, DHCPv6, NDP with RA, and DNS service.

6.6. CLAT to CLAT communications

464XLAT is a hub and spoke architecture focused on enabling IPv4-only services over IPv6-only networks. ICE [RFC5245] may be used to

support peer-to-peer communication within a 464XLAT network.

7. Deployment Considerations

7.1. Traffic Engineering

Even if the ISP for end users is different from the PLAT provider (e.g. another ISP), it can implement traffic engineering independently from the PLAT provider. Detailed reasons are below:

1. The ISP for end users can figure out IPv4 destination address from translated IPv6 packet header, so it can implement traffic engineering based on IPv4 destination address (e.g. traffic monitoring for each IPv4 destination address, packet filtering for each IPv4 destination address, etc.). The tunneling methods do not have such an advantage, without any deep packet inspection for processing the inner IPv4 packet of the tunnel packet.
2. If the ISP for end users can assign an IPv6 prefix greater than /64 to each subscriber, this 464XLAT architecture can separate IPv6 prefix for native IPv6 packets and the XLAT prefixes for IPv4/IPv6 translation packets. Accordingly, it can identify the type of packets ("native IPv6 packets" and "IPv4/IPv6 translation packets"), and implement traffic engineering based on the IPv6 prefix.

7.2. Traffic Treatment Scenarios

The below table outlines how different permutations of connectivity are treated in the 464XLAT architecture.

NOTE: 464XLAT double translation treatment will be stateless when a dedicated /64 is available for translation on the CLAT. Otherwise, the CLAT will have both stateful and stateless since it requires NAT44 from the LAN to a single IPv4 address and then stateless translation to a single IPv6 address.

Server	Application and Host	Traffic Treatment	Location of Translation
IPv6	IPv6	End-to-end IPv6	None
IPv4	IPv6	Stateful Translation	PLAT
IPv4	IPv4	464XLAT	PLAT/CLAT

Traffic Treatment Scenarios

8. Security Considerations

To implement a PLAT, see security considerations presented in Section 5 of [RFC6146].

To implement a CLAT, see security considerations presented in Section 7 of [RFC6145]. The CLAT may comply with [RFC6092].

9. IANA Considerations

This document has no actions for IANA.

10. Acknowledgements

The authors would like to thank JPIX NOC members, JPIX 464XLAT trial service members, Seiichi Kawamura, Dan Drown, Brian Carpenter, Rajiv Asati, Washam Fan, Behcet Sarikaya, Jan Zorz, Tatsuya Oishi, Lorenzo Colitti, Erik Kline, Ole Troan, Maoke Chen, Gang Chen, Tom Petch, Jouni Korhonen, Bjoern A. Zeeb, Hemant Singh, Vizdal Ales, Mark ZZZ Smith, Mikael Abrahamsson, Tore Anderson, Teemu Savolainen, Alexandru Petrescu, Gert Doering, Victor Kuarsingh, Ray Hunter, James Woodyatt, Tom Taylor, and Remi Despres for their helpful comments. We also would like to thank Fred Baker and Joel Jaeggli for their support.

11. References

11.1. Normative References

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

- [RFC6144] Baker, F., Li, X., Bao, C., and K. Yin, "Framework for IPv4/IPv6 Translation", RFC 6144, April 2011.
- [RFC6145] Li, X., Bao, C., and F. Baker, "IP/ICMP Translation Algorithm", RFC 6145, April 2011.
- [RFC6146] Bagnulo, M., Matthews, P., and I. van Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", RFC 6146, April 2011.

11.2. Informative References

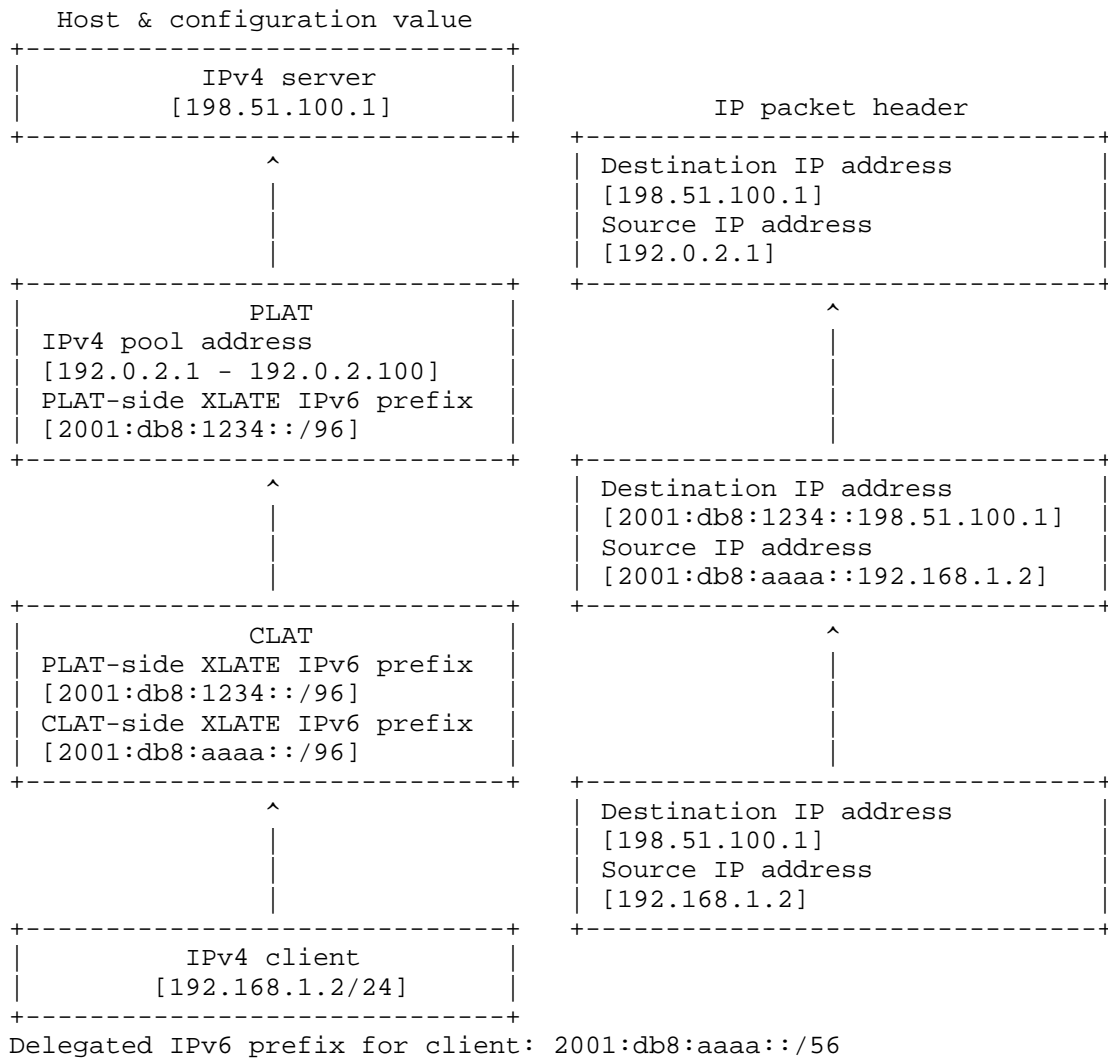
- [I-D.ietf-behave-nat64-discovery-heuristic] Savolainen, T., Korhonen, J., and D. Wing, "Discovery of the IPv6 Prefix Used for IPv6 Address Synthesis", draft-ietf-behave-nat64-discovery-heuristic-13 (work in progress), November 2012.
- [RFC1918] Rekhter, Y., Moskowitz, R., Karrenberg, D., Groot, G., and E. Lear, "Address Allocation for Private Internets", BCP 5, RFC 1918, February 1996.
- [RFC3633] Troan, O. and R. Droms, "IPv6 Prefix Options for Dynamic Host Configuration Protocol (DHCP) version 6", RFC 3633, December 2003.
- [RFC5245] Rosenberg, J., "Interactive Connectivity Establishment (ICE): A Protocol for Network Address Translator (NAT) Traversal for Offer/Answer Protocols", RFC 5245, April 2010.
- [RFC5625] Bellis, R., "DNS Proxy Implementation Guidelines", BCP 152, RFC 5625, August 2009.
- [RFC6092] Woodyatt, J., "Recommended Simple Security Capabilities in Customer Premises Equipment (CPE) for Providing Residential IPv6 Internet Service", RFC 6092, January 2011.
- [RFC6147] Bagnulo, M., Sullivan, A., Matthews, P., and I. van Beijnum, "DNS64: DNS Extensions for Network Address Translation from IPv6 Clients to IPv4 Servers", RFC 6147, April 2011.
- [RFC6459] Korhonen, J., Soininen, J., Patil, B., Savolainen, T., Bajko, G., and K. Iisakkila, "IPv6 in 3rd Generation Partnership Project (3GPP) Evolved Packet System (EPS)", RFC 6459, January 2012.

[TS.23203] 3GPP, "Policy and charging control architecture", 3GPP
TS 23.203 10.7.0, June 2012.

Appendix A. Examples of IPv4/IPv6 Address Translation

The following is a example of IPv4/IPv6 Address Translation on the
464XLAT architecture.

In the case that an IPv6 prefix greater than /64 is assigned to an
end user by such as DHCPv6-PD [RFC3633], the CLAT can use a dedicated
/64 from the assigned IPv6 prefix.



Authors' Addresses

Masataka Mawatari
Japan Internet Exchange Co.,Ltd.
KDDI Otemachi Building 19F, 1-8-1 Otemachi,
Chiyoda-ku, Tokyo 100-0004
JAPAN

Phone: +81 3 3243 9579
Email: mawatari@jpix.ad.jp

Masanobu Kawashima
NEC AccessTechnica, Ltd.
800, Shimomata
Kakegawa-shi, Shizuoka 436-8501
JAPAN

Phone: +81 537 22 8274
Email: kawashimam@vx.jp.nec.com

Cameron Byrne
T-Mobile USA
Bellevue, Washington 98006
USA

Email: cameron.byrne@t-mobile.com

V6OPS
Internet-Draft
Intended status: Informational
Expires: July 15, 2013

B. Carpenter
Univ. of Auckland
S. Jiang
Huawei Technologies Co., Ltd
January 11, 2013

IPv6 Guidance for Internet Content and Application Service Providers
draft-ietf-v6ops-icp-guidance-05

Abstract

This document provides guidance and suggestions for Internet Content Providers and Application Service Providers who wish to offer their service to both IPv6 and IPv4 customers. Many of the points will also apply to hosting providers, or to any enterprise network preparing for IPv6 users.

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 July 15, 2013.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as

described in the Simplified BSD License.

Table of Contents

1. Introduction	3
2. General Strategy	3
3. Education and Skills	5
4. Arranging IPv6 Connectivity	6
5. IPv6 Infrastructure	7
5.1. Address and subnet assignment	7
5.2. Routing	9
5.3. DNS	10
6. Load Balancers	10
7. Proxies	11
8. Servers	12
8.1. Network Stack	12
8.2. Application Layer	12
8.3. Logging	13
8.4. Geolocation	13
9. Coping with Transition Technologies	13
10. Content Delivery Networks	15
11. Business Partners	16
12. Possible Complexities	16
13. Operations and Management	17
14. Security Considerations	18
15. IANA Considerations	20
16. Acknowledgements	20
17. Change log [RFC Editor: Please remove]	20
18. References	21
18.1. Normative References	21
18.2. Informative References	23
Authors' Addresses	25

1. Introduction

The deployment of IPv6 [RFC2460] is now in progress, and users without direct IPv4 access are likely to appear in increasing numbers in the coming years. Any provider of content or application services over the Internet will need to arrange for IPv6 access or else risk losing large numbers of potential users. For users who already have dual stack connectivity, direct IPv6 access might provide more satisfactory performance than indirect access via NAT.

In this document, we often refer to the users of content or application services as "customers" to clarify the part they play, but this is not intended to limit the scope to commercial sites.

The time for action is now, while the number of IPv6-only customers is small, so that appropriate skills, software and equipment can be acquired in good time to scale up the IPv6 service as demand increases. An additional advantage of early support for IPv6 customers is that it will reduce the number of customers connecting later via IPv4 "extension" solutions such as double NAT or NAT64 [RFC6146], which will otherwise degrade the user experience.

Nevertheless, it is important that the introduction of IPv6 service should not make service for IPv4 customers worse. In some circumstances, technologies intended to assist in the transition from IPv4 to IPv6 are known to have negative effects on the user experience. A deployment strategy for IPv6 must avoid these effects as much as possible.

The purpose of this document is to provide guidance and suggestions for Internet Content Providers (ICPs) and Application Service Providers (ASPs) who wish to offer their services to both IPv6 and IPv4 customers, but who are currently supporting only IPv4. For simplicity, the term ICP is mainly used in the body of this document, but the guidance also applies to ASPs. Any hosting provider whose customers include ICPs or ASPs is also concerned. Many of the points in this document will also apply to enterprise networks that do not classify themselves as ICPs. Any enterprise or department that runs at least one externally accessible server, such as an HTTP server, may also be concerned. Although specific managerial and technical approaches are described, this is not a rule book; each operator will need to make its own plan, tailored to its own services and customers.

2. General Strategy

The most important advice here is to actually have a general

strategy. Adding support for a second network layer protocol is a new experience for most modern organisations, and it cannot be done casually on an unplanned basis. Even if it is impossible to write a precisely dated plan, the intended steps in the process need to be defined well in advance. There is no single blueprint for this. The rest of this document is meant to provide a set of topics to be taken into account in defining the strategy. Other documents about IPv6 deployment, such as [I-D.matthews-v6ops-design-guidelines], should be consulted as well.

In determining the urgency of this strategy, it should be noted that the central IPv4 registry (IANA) ran out of spare blocks of IPv4 addresses in February 2011 and the various regional registries are expected to exhaust their reserves over the next one to two years. After this, Internet Service Providers (ISPs) will run out at dates determined by their own customer base. No precise date can be given for when IPv6-only customers will appear in commercially significant numbers, but - particularly in the case of mobile users - it may be quite soon. Complacency about this is therefore not an option for any ICP that wishes to grow its customer base over the coming years.

The most common strategy for an ICP is to provide dual stack services - both IPv4 and IPv6 on an equal basis - to cover both existing and future customers. This is the recommended strategy in [RFC6180] for straightforward situations. Some ICPs who already have satisfactory operational experience with IPv6 might consider an IPv6-only strategy, with IPv4 clients being supported by translation or proxy in front of their IPv6 content servers. However, the present document is addressed to ICPs without IPv6 experience, who are likely to prefer the dual stack model to build on their existing IPv4 service.

Due to the widespread impact of supporting IPv6 everywhere within an environment, it is important to select a focussed initial approach based on clear business needs and real technical dependencies.

Within the dual stack model, two approaches could be adopted, sometimes referred to as "outside in" and "inside out":

- o Outside in: start by providing external users with an IPv6 public access to your services, for example by running a reverse proxy that handles IPv6 customers (see Section 7 for details). Progressively enable IPv6 internally.
- o Inside out: start by enabling internal networking infrastructure, hosts, and applications to support IPv6. Progressively reveal IPv6 access to external customers.

Which of these approaches to choose depends on the precise

circumstances of the ICP concerned. "Outside in" has the benefit of giving interested customers IPv6 access at an early stage, and thereby gaining precious operational experience, before meticulously updating every piece of equipment and software. For example, if some back-office system, that is never exposed to users, only supports IPv4, it will not cause delay. "Inside out" has the benefit of completing the implementation of IPv6 as a single project. Any ICP could choose this approach, but it might be most appropriate for a small ICP without complex back-end systems.

A point that must be considered in the strategy is that some customers will remain IPv4-only for many years, others will have both IPv4 and IPv6 access, and yet others will have only IPv6. Additionally, mobile customers may find themselves switching between IPv4 and IPv6 access as they travel, even within a single session. Services and applications must be able to deal with this, just as easily as they deal today with a user whose IPv4 address changes (see the discussion of cookies in Section 8.2).

Nevertheless, the end goal is to have a network that does not need major changes when at some point in the future it becomes possible to transition to IPv6-only, even if only for some parts of the network. That is, the IPv6 deployment should be designed in such a way as to more or less assume that IPv4 is already absent, so the network will function seamlessly when it is indeed no longer there.

An important step in the strategy is to determine from hardware and software suppliers details of their planned dates for providing sufficient IPv6 support, with performance equivalent to IPv4, in their products and services. Relevant specifications such as [RFC6434] [I-D.ietf-v6ops-6204bis] should be used. Even if complete information cannot be obtained, it is essential to determine which components are on the critical path during successive phases of deployment. This information will make it possible to draw up a logical sequence of events and identify any components that may cause holdups.

3. Education and Skills

Some staff may have experience of running multiprotocol networks, which were common twenty years ago before the dominance of IPv4. However, IPv6 will be new to them, and also to staff brought up only on TCP/IP. It is not enough to have one "IPv6 expert" in a team. On the contrary, everybody who knows about IPv4 needs to know about IPv6, from network architect to help desk responder. Therefore, an early and essential part of the strategy must be education, including practical training, so that all staff acquire a general understanding

of IPv6, how it affects basic features such as the DNS, and the relevant practical skills. To take a trivial example, any staff used to dotted-decimal IPv4 addresses need to become familiar with the colon-hexadecimal format used for IPv6.

There is an anecdote of one IPv6 deployment in which prefixes including the letters A to F were avoided by design, to avoid confusing system administrators unfamiliar with hexadecimal notation. This is not a desirable result. There is another anecdote of a help desk responder telling a customer to "disable one-Pv6" in order to solve a problem. It should be a goal to avoid having untrained staff who don't understand hexadecimal or who can't even spell "IPv6".

It is very useful to have a small laboratory network available for training and self-training in IPv6, where staff may experiment and make mistakes without disturbing the operational IPv4 service. This lab should run both IPv4 and IPv6, to gain experience with a dual-stack environment and new features such as having multiple addresses per interface, and addresses with lifetimes and deprecation.

Once staff are trained, they will likely need to support both IPv4, IPv6, and dual-stack customers. Rather than having separate internal escalation paths for IPv6, it generally makes sense for questions that may have an IPv6 element to follow normal escalation paths; there should not be an "IPv6 Department" once training is completed.

A final remark about training is that it should not be given too soon, or it will be forgotten. Training has a definite need to be done "just in time" in order to properly "stick." Training, lab experience, and actual deployment should therefore follow each other immediately. If possible, training should even be combined with actual operational experience.

4. Arranging IPv6 Connectivity

There are, in theory, two ways to obtain IPv6 connectivity to the Internet.

- o Native. In this case the ISP simply provides IPv6 on exactly the same basis as IPv4 - it will appear at the ICP's border router(s), which must then be configured in dual-stack mode to forward IPv6 packets in both directions. This is by far the better method. An ICP should contact all its ISPs to verify when they will provide native IPv6 support, whether this has any financial implications, and whether the same service level agreement will apply as for IPv4. Any ISP that has no definite plan to offer native IPv6 service should be avoided.

- o Managed Tunnel. It is possible to configure an IPv6-in-IPv4 tunnel to a remote ISP that offers such a service. A dual-stack router in the ICP's network will act as a tunnel end-point, or this function could be included in the ICP's border router.

A managed tunnel is a reasonable way to obtain IPv6 connectivity for initial testing and skills acquisition. However, it introduces an inevitable extra latency compared to native IPv6, giving customers a noticeably worse response time for complex web pages. A tunnel may become a performance bottleneck (especially if offered as a free service) or a target for malicious attack. It is also likely to limit the IPv6 MTU size. In normal circumstances, native IPv6 will provide an MTU size of at least 1500 bytes, but it will almost inevitably be less for a tunnel, possibly as low as 1280 bytes (the minimum MTU allowed for IPv6). Apart from the resulting loss of efficiency, there are cases in which Path MTU Discovery fails, therefore IPv6 fragmentation fails, and in this case the lower tunnel MTU will actually cause connectivity failures for customers.

For these reasons, ICPs are strongly recommended to obtain native IPv6 service before attempting to offer a production-quality service to their customers. Unfortunately it is impossible to prevent customers from using unmanaged tunnel solutions (see Section 9).

Some larger organizations may find themselves needing multiple forms of IPv6 connectivity, for their ICP data centres and for their staff working elsewhere. It is important to obtain IPv6 connectivity for both, as testing and supporting an IPv6-enabled service is challenging for staff without IPv6 connectivity. This may involve short-term alternatives to provide IPv6 connectivity to operations and support staff, such as a managed tunnel or HTTP proxy server with IPv6 connectivity. Note that unmanaged tunnels (such as 6to4 and Teredo) are generally not useful for support staff as recent client software will avoid them when accessing dual-stack sites.

5. IPv6 Infrastructure

5.1. Address and subnet assignment

An ICP must first decide whether to apply for its own Provider Independent (PI) address prefix for IPv6. This option is available either from an ISP that acts as a Local Internet Registry, or directly from the relevant Regional Internet Registry. The alternative is to obtain a Provider Aggregated (PA) prefix from an ISP. Both solutions are viable in IPv6. However, the scaling

properties of the wide area routing system (BGP4) mean that the number of PI prefixes should be limited, so only large content providers can justify obtaining a PI prefix and convincing their ISPs to route it. Millions of enterprise networks, including smaller content providers, will use PA prefixes. In this case, a change of ISP would necessitate a change of the corresponding PA prefix, using the procedure outlined in [RFC4192].

An ICP that has connections via multiple ISPs, but does not have a PI prefix, would therefore have multiple PA prefixes, one from each ISP. This would result in multiple IPv6 addresses for the ICP's servers or load balancers. If one address fails due to an ISP malfunction, sessions using that address would be lost. At the time of this writing, there is very limited operational experience with this approach [I-D.ietf-v6ops-ipv6-multihoming-without-ipv6nat].

An ICP may also choose to operate a Unique Local Address prefix [RFC4193] for internal traffic only, as described in [RFC4864].

Depending on its projected future size, an ICP might choose to obtain /48 PI or PA prefixes (allowing 16 bits of subnet address) or longer PA prefixes, e.g. /56 (allowing 8 bits of subnet address). Clearly the choice of /48 is more future-proof. Advice on the numbering of subnets may be found in [RFC5375]. An ICP with multiple locations will probably need a prefix per location

An ICP that has its service hosted by a colocation provider, cloud provider, or the like, will need to follow the addressing policy of that provider.

Since IPv6 provides for operating multiple prefixes simultaneously, it is important to check that all relevant tools, such as address management packages, can deal with this. In particular, the possible need to allow for multiple PA prefixes with IPv6, and the possible need to renumber, means that the common technique of manually assigned static addresses for servers, proxies or load balancers, with statically defined DNS entries, could be problematic [I-D.ietf-6renum-static-problem]. An ICP of reasonable size might instead choose to operate DHCPv6 [RFC3315] with standard DNS, to support stateful assignment. In either case a configuration management system is likely to be used to support stateful and/or on-demand address assignment.

Theoretically, it would also be possible to operate an ICP's IPv6 network using only Stateless Address Autoconfiguration [RFC4862], with Dynamic DNS [RFC3007] to publish server addresses for external users.

5.2. Routing

In a dual stack network, most IPv4 and IPv6 interior routing protocols operate quite independently and in parallel. The common routing protocols all support IPv6, such as OSPFv3 [RFC5340], IS-IS [RFC5308], and even RIPng [RFC2080] [RFC2081]. It is worth noting that whereas OSPF and RIP differ significantly between IPv4 and IPv6, IS-IS has the advantage of handling them both in a single instance of the protocol, with the potential for operational simplification in the long term. Some versions of OSPFv3 may also have this advantage [RFC5838]. In any case, for trained staff, there should be no particular difficulty in deploying IPv6 routing without disturbance to IPv4 services. In some cases, firmware upgrades may be needed on some network devices.

The performance impact of dual stack routing needs to be evaluated. In particular, what forwarding performance does the router vendor claim for IPv6? If the forwarding performance is significantly inferior compared to IPv4, will this be an operational problem? Is extra memory or ternary content-addressable memory (TCAM) space needed to accommodate both IPv4 and IPv6 tables? To answer these questions, the ICP will need a projected model for the amount of IPv6 traffic expected initially, and its likely rate of increase.

If a site has multiple PA prefixes as mentioned in Section 5.1, complexities will appear in routing configuration. In particular, source-based routing rules might be needed to ensure that outgoing packets are routed to the appropriate border router and ISP link. Normally, a packet sourced from an address assigned by ISP X should not be sent via ISP Y, to avoid ingress filtering by Y [RFC2827] [RFC3704]. Additional considerations may be found in [I-D.ietf-v6ops-ipv6-multihoming-without-ipv6nat]. Note that the prefix translation technique discussed in [RFC6296] does not describe a solution for enterprises that offer publicly available content servers.

Each IPv6 subnet that supports end hosts normally has a /64 prefix, leaving another 64 bits for the interface identifiers of individual hosts. In contrast, a typical IPv4 subnet will have no more than 8 bits for the host identifier, thus limiting the subnet to 256 or fewer hosts. A dual stack design will typically use the same physical or VLAN subnet topology for IPv4 and IPv6, and therefore the same router topology. In other words the IPv4 and IPv6 topologies are congruent. This means that the limited subnet size of IPv4 (such as 256 hosts) will be imposed on IPv6, even though the IPv6 prefix will allow many more hosts. It would be theoretically possible to avoid this limitation by implementing a different physical or VLAN subnet topology for IPv6. This is not advisable, as it would result

in extremely complex fault diagnosis when something went wrong.

5.3. DNS

It must be understood that as soon as an AAAA record for a well-known name is published in the DNS, the corresponding server will start to receive IPv6 traffic. Therefore, it is essential that an ICP tests thoroughly that IPv6 works on its servers, load balancers, etc., before adding their AAAA records to DNS. There have been numerous cases of ICPs breaking their sites for all IPv6 users during a roll-out by returning AAAA records for servers improperly configured for IPv6.

Once such tests have succeeded, each externally visible host (or virtual host) that has an A record for its IPv4 address needs an AAAA record [RFC3596] for its IPv6 address, and a reverse entry (in ip6.arpa) if applicable. Note that if CNAME records are in use, the AAAA record must be added alongside the A record at the end of the CNAME chain. It is not possible to have the AAAA record on the same name as a CNAME record, as per [RFC1912].

One important detail is that some clients (especially Windows XP) can only resolve DNS names via IPv4, even if they can use IPv6 for application traffic. Also, a dual stack resolver might attempt to resolve queries for A records via IPv6, or AAAA records via IPv4. It is therefore advisable for all DNS servers to respond to queries via both IPv4 and IPv6.

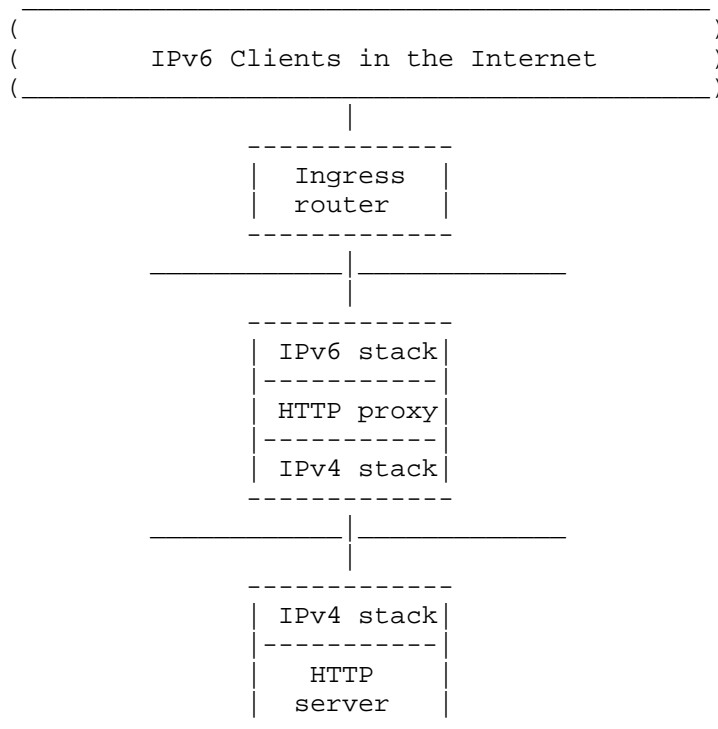
6. Load Balancers

Most available load balancers now support IPv6. However, it is important to obtain appropriate assurances from vendors about their IPv6 support, including performance aspects (as discussed for routers in Section 5.2). The update needs to be planned in anticipation of expected traffic growth. It is to be expected that IPv6 traffic will initially be low, i.e., a small but growing percentage of total traffic. For this reason, it might be acceptable to have IPv6 traffic bypass load balancing initially, by publishing an AAAA record for a specific server instead of the load balancer. However, load balancers often also provide for server fail-over, in which case it would be better to implement IPv6 load balancing immediately.

The same would apply to Transport Layer Security (TLS) or HTTP proxies used for load balancing purposes.

7. Proxies

An HTTP proxy [RFC2616] can readily be configured to handle incoming connections over IPv6 and to proxy them to a server over IPv4. Therefore, a single proxy can be used as the first step in an outside-in strategy, as shown in the following diagram:



In this case, the AAAA record for the service would provide the IPv6 address of the proxy. This approach will work for any HTTP or HTTPS applications that operate successfully via a proxy, as long as IPv6 load remains low. Additionally, many load balancer products incorporate such a proxy, in which case this approach would be possible at high load.

Note that in any proxy scenario, an ICP will need to make sure that both IPv4 and IPv6 addresses are being properly passed to application servers in any relevant HTTP headers, and that those application servers are properly handling the IPv6 addresses.

8. Servers

8.1. Network Stack

The TCP/IP network stacks in popular operating systems have supported IPv6 for many years. In most cases, it is sufficient to enable IPv6 and possibly DHCPv6; the rest will follow. Servers inside an ICP network will not need to support any transition technologies beyond a simple dual stack, with a possible exception for 6to4 mitigation noted below in Section 9.

As some operating systems have separate firewall rule sets for IPv4 and IPv6, an ICP should also evaluate those rule sets and ensure that appropriate firewall rules are configured for IPv6. More details are discussed in Section 14.

8.2. Application Layer

Basic HTTP servers have been able to handle an IPv6-enabled network stack for some years, so at the most it will be necessary to update to a more recent software version. The same is true of generic applications such as email protocols. No general statement can be made about other applications, especially proprietary ones, so each ASP will need to make its own determination. As changes to the network layer to introduce IPv6 addresses can ripple through applications, testing of both client and server applications should be performed in both IPv4-only, IPv6-only, and dual-stack environments prior to dual-stacking a production environment.

One important recommendation here is that all applications should use domain names, which are IP-version-independent, rather than IP addresses. Applications based on middleware platforms which have uniform support for IPv4 and IPv6, for example Java, may be able to support both IPv4 and IPv6 naturally without additional work. Security certificates should also contain domain names rather than addresses.

A specific issue for HTTP-based services is that IP address-based cookie authentication schemes will need to deal with dual-stack clients. Servers might create a cookie for an IPv4 connection or an IPv6 connection, depending on the setup at the client site and on the whims of the client operating system. There is no guarantee that a given client will consistently use the same address family, especially when accessing a collection of sites rather than a single site, such as when cookies are used for federated authentication. If the client is using privacy addresses [RFC4941], the IPv6 address (but usually not its /64 prefix) might change quite frequently. Any cookie mechanism based on 32-bit IPv4 addresses will need significant

remodelling.

Generic considerations on application transition are discussed in [RFC4038], but many of them will not apply to the dual-stack ICP scenario. An ICP that creates and maintains its own applications will need to review them for any dependency on IPv4.

8.3. Logging

The introduction of IPv6 clients will generally also result in IPv6 addresses appearing in the "client ip" field of server logs. It might be convenient to use the same log field to hold a client's IP address, whether it is IPv4 or IPv6. Downstream systems looking at logs and client IP addresses may also need testing to ensure that they can properly handle IPv6 addresses. This includes any of an ICP's databases recording client IP addresses, such as for recording IP addresses of online purchases and comment posters.

It is worth noting that accurate traceback from logs to individual customers requires end-to-end address transparency. This is additional motivation for an ICP to support native IPv6 connectivity, since otherwise, IPv6-only customers will inevitably connect via some form of translation mechanism, interfering with traceback.

8.4. Geolocation

Initially, ICPs may observe some weakness in geolocation for IPv6 clients. As time goes on, it is to be assumed that geolocation methods and databases will be updated to fully support IPv6 prefixes. There is no reason they will be more or less accurate in the long term than those available for IPv4. However, we can expect many more clients to be mobile as time goes on, so geolocation based on IP addresses alone may in any case become problematic. A more robust technique such as HTTP-Enabled Location Delivery (HELD) [RFC5985] could be considered.

9. Coping with Transition Technologies

As mentioned above, an ICP should obtain native IPv6 connectivity from its ISPs. In this way, the ICP can avoid most of the complexities of the numerous IPv4-to-IPv6 transition technologies that have been developed; they are all second-best solutions. However, some clients are sure to be using such technologies. An ICP needs to be aware of the operational issues this may cause and how to deal with them.

In some cases outside the ICP's control, clients might reach a

content server via a network-layer translator from IPv6 to IPv4. ICPs who are offering a dual stack service and providing both A and AAAA records, as recommended in this document, should not normally receive IPv4 traffic from NAT64 translators [RFC6146]. Exceptionally, however, such traffic could arrive via IPv4 from an IPv6-only client whose DNS resolver failed to receive the ICP's AAAA record for some reason. Such traffic would be indistinguishable from regular IPv4-via-NAT traffic.

Alternatively, ICPs who are offering a dual stack service might exceptionally receive IPv6 traffic translated from an IPv4-only client that somehow failed to receive the ICP's A record. An ICP could also receive IPv6 traffic with translated prefixes [RFC6296]. These two cases would only be an issue if the ICP was offering any service that depends on the assumption of end-to-end IPv6 address transparency.

Finally, some traffic might reach an ICP that has been translated twice en route (e.g., from IPv6 to IPv4 and back again). Again, the ICP will be unable to detect this. It is likely that real-time geolocation will be highly inaccurate for such traffic, since it will at best indicate the location of the second translator, which could be very distant from the customer.

In other cases, also outside the ICP's control, IPv6 clients may reach the IPv6 Internet via some form of IPv6-in-IPv4 tunnel. In this case a variety of problems can arise, the most acute of which affect clients connected using the Anycast 6to4 solution [RFC3068]. Advice on how ICPs may mitigate these 6to4 problems is given in Section 4.5. of [RFC6343]. For the benefit of all tunnelled clients, it is essential to verify that Path MTU Discovery works correctly (i.e., the relevant ICMPv6 packets are not blocked) and that the server-side TCP implementation correctly supports the Maximum Segment Size (MSS) negotiation mechanism [RFC2923] for IPv6 traffic.

Some ICPs have implemented an interim solution to mitigate transition problems by limiting the visibility of their AAAA records to users with validated IPv6 connectivity [RFC6589] (known as "DNS whitelisting"). At the time of this writing, this solution seems to be passing out of use, being replaced by "DNS blacklisting" of customer sites known to have problems with IPv6 connectivity. In the reverse direction, it is worth being aware that some ISPs with significant populations of clients with broken IPv6 setups have begun filtering AAAA record lookups by their clients. None of these solutions is appropriate in the long term.

Another approach taken by some ICPs is to offer IPv6-only support via a specific DNS name, e.g., ipv6.example.com, if the primary service

is `www.example.com`. In this case `ipv6.example.com` would have an AAAA record only. This has some value for testing purposes, but is otherwise only of interest to hobbyist users willing to type in special URLs.

There is little an ICP can do to deal with client-side or remote ISP deficiencies in IPv6 support, but it is hoped that the "happy eyeballs" [RFC6555] approach will improve the ability for clients to deal with such problems.

10. Content Delivery Networks

DNS-based techniques for diverting users to Content Delivery Network (CDN) points of presence (POPs) will work for IPv6, if AAAA records are provided as well as A records. In general the CDN should follow the recommendations of this document, especially by operating a full dual stack service at each POP. Additionally, each POP will need to handle IPv6 routing exactly like IPv4, for example running BGP4+ [RFC4760].

Note that if an ICP supports IPv6 but its external CDN provider does not, its clients will continue to use IPv4 and any IPv6-only clients will have to use a transition solution of some kind. This is not a desirable situation, since the ICP's work to support IPv6 will be wasted.

An ICP might face a complex situation, if its CDN provider supports IPv6 at some POPs but not at others. IPv6-only clients could only be diverted to a POP supporting IPv6. There are also scenarios where a dual-stack client would be diverted to a mixture of IPv4 and IPv6 POPs for different URLs, according to the A and AAAA records provided and the availability of optimisations such as "happy eyeballs." A related side effect is that copies of the same content viewed at the same time via IPv4 and IPv6 may be different, due to latency in the underlying data synchronisation process used by the CDN. This effect has in fact been observed in the wild for a major social network supporting dual stack. These complications do not affect the viability of relying on a dual-stack CDN, however.

The CDN itself faces related complexity: "As IPv6 rolls out, it's going to roll out in pockets, and that's going to make the routing around congestion points that much more important but also that much harder," stated John Summers of Akamai in 2010.

A converse situation that might arise is that an ICP has not yet started its deployment of IPv6, but finds that its CDN provider already supports IPv6. Then, assuming the CDN provider announces

appropriate AAAA DNS Resource Records, dual-stack and IPv6-only customers will obtain IPv6 access and the ICP's content may well be delivered to them via IPv6. In normal circumstances this should create no problems, but it is a situation that the ICP and its support staff need to be aware of. In particular, support staff should be given IPv6 connectivity in order to be able to investigate any problems that might arise (see Section 4).

11. Business Partners

As noted earlier, it is in an ICP's or ASP's best interests that their users have direct IPv6 connectivity, rather than indirect IPv4 connectivity via double NAT. If the ICP or ASP has a direct business relationship with some of their clients, or with the networks that connect them to their clients, they are advised to coordinate with those partners to ensure that they have a plan to enable IPv6. They should also verify and test that there is first-class IPv6 connectivity end-to-end between the networks concerned. This is especially true for implementations that require IPv6 support in specialized programs or systems in order for the IPv6 support on the ICP/ASP side to be useful.

12. Possible Complexities

Some additional considerations come into play for some types of complex or distributed sites and applications that an ICP may be delivering. For example, an ICP may have a site spread across many hostnames (not all under their control). Other ICPs may have their sites or applications distributed across multiple locations for availability, scale, or performance.

Many modern web sites and applications now use a collection of resources and applications, some operated by the ICP and other by third parties. While most clients support sites containing a mixture of IPv4-only and dual-stack elements, an ICP should track the IPv6 availability of embedded resources (such as images) as otherwise their site may only be partially functional or may have degraded performance for IPv6-only users.

DNS-based load balancing techniques for diverting users to servers in multiple POPs will work for IPv6, if the load balancer supports IPv6 and if AAAA records are provided. Depending on the architecture of the load balancer, an ICP may need to operate a full dual stack service at each POP. With other architectures, it may be acceptable to initially only have IPv6 at a subset of locations. Some architectures will make it preferable for IPv6 routing to mirror IPv4

routing (for example running BGP4+ [RFC4760] if appropriate) but this may not always be possible as IPv6 and IPv4 connectivity can be independent.

Some complexities may arise when a client supporting both IPv4 and IPv6 uses different POPs for each IP version (such as when IPv6 is only available in a subset of locations). There are also scenarios where a dual-stack client would be diverted to a mixture of IPv4 and IPv6 POPs for different URLs, according to the A and AAAA records provided and the availability of optimisations such as "happy eyeballs" [RFC6555]. A related side effect is that copies of the same content viewed at the same time via IPv4 and IPv6 may be different, due to latency in the underlying data synchronisation process used at the application layer. This effect has in fact been observed in the wild for a major social network supporting dual-stack.

Even with a single POP, unexpected behaviour may arise if a client switches spontaneously between IPv4 and IPv6 as a performance optimisation [RFC6555] or if its IPv6 address changes frequently for privacy reasons [RFC4941]. Such changes may affect cookies, geolocation, load balancing, and transactional integrity. Although unexpected changes of client address also occur in an IPv4-only environment, they may be more frequent with IPv6.

13. Operations and Management

There is no doubt that, initially, IPv6 deployment will have operational impact, as well as requiring education and training as mentioned in Section 3. Staff will have to update network elements such as routers, update configurations, provide information to end users, and diagnose new problems. However, for an enterprise network, there is plenty of experience, e.g. on numerous university campuses, showing that dual stack operation is no harder than IPv4-only in the steady state.

Whatever management, monitoring and logging is performed for IPv4 is also needed for IPv6. Therefore, all products and tools used for these purposes must be updated to fully support IPv6 management data. It is important to verify that tools have been fully updated to support 128 bit addresses entered and displayed in hexadecimal [RFC5952]. Since an IPv6 network may operate with more than one IPv6 prefix and therefore more than one address per host, the tools must deal with this as a normal situation. This includes any address management tool in use (see Section 5.1) as well as tools used for creating DHCP and DNS configurations. There is significant overlap here with the tools involved in site renumbering

[I-D.ietf-6renum-enterprise].

At an early stage of IPv6 deployment, it is likely that IPv6 will be mainly managed via IPv4 transport. This allows network management systems to test for dependencies between IPv4 and IPv6 management data. For example, will reports mixing IPv4 and IPv6 addresses display correctly?

In a second phase, IPv6 transport should be used to manage the network. Note that it will also be necessary for an ICP to provide IPv6 connectivity for its operations and support staff, even when working remotely. As far as possible mutual dependency between IPv4 and IPv6 should be avoided, for both the management data and the transport. Failure of one should not cause a failure of the other. One precaution to avoid this would be for network management systems to be dual-stacked. It would then be possible to use IPv4 connectivity to repair IPv6 configurations, and vice versa.

Dual stack, while necessary, does have management scaling and overhead considerations. As noted earlier, the long term goal is to move to single-stack IPv6, when the network and its customers can support this. This is an additional reason why mutual dependency between the address families should be avoided in the management system in particular; a hidden dependency on IPv4 that had been forgotten for many years would be highly inconvenient. In particular, a management tool that manages IPv6 but itself runs only over IPv4 would prove disastrous on the day that IPv4 is switched off.

An ICP should ensure that any end-to-end availability monitoring systems are updated to monitor dual-stacked servers over both IPv4 and IPv6. A particular challenge here may be monitoring systems relying on DNS names as this may result in monitoring only one of IPv4 or IPv6, resulting in a loss of visibility to failures in network connectivity over either address family.

As mentioned above, it will also be necessary for an ICP to provide IPv6 connectivity for its operations and support staff, even when working remotely.

14. Security Considerations

While many ICPs may still be in the process of experimenting with and configuring IPv6, there is mature malware in the wild that will launch attacks over IPv6. For example, if an AAAA DNS record is added for a hostname, malware using client OS libraries may automatically switch from attacking that hostname over IPv4 to

attacking that hostname over IPv6. As a result, it is crucial that firewalls and other network security appliances protecting servers support IPv6 and have rules tested and configured.

Security experience with IPv4 should be used as a guide as to the threats that may exist in IPv6, but they should not be assumed to be equally likely, and nor should they assumed to be the only threats that could exist in IPv6. However, essentially every threat that exists for IPv4 exists or will exist for IPv6, to a greater or lesser extent. It is essential to update firewalls, intrusion detection systems, denial of service precautions, and security auditing technology to fully support IPv6. Needless to say, it is also essential to turn on well-known security mechanisms such as DNS Security and DHCPv6 Authentication. Otherwise, IPv6 will become an attractive target for attackers.

When multiple PA prefixes are in use as mentioned in Section 5.1, firewall rules must allow for all valid prefixes, and must be set up to work as intended even if packets are sent via one ISP but return packets arrive via another.

Performance and memory size aspects of dual stack firewalls must be considered (as discussed for routers in Section 5.2).

In a dual stack operation, there may be a risk of cross-contamination between the two protocols. For example, a successful IPv4-based denial of service attack might also deplete resources needed by the IPv6 service, or vice versa. This risk strengthens the argument that IPv6 security must be up to the same level as IPv4. Risks can also occur with dual stack Virtual Private Network (VPN) solutions [I-D.ietf-opsec-vpn-leakages].

A general overview of techniques to protect an IPv6 network against external attack is given in [RFC4864]. Assuming an ICP has native IPv6 connectivity, it is advisable to block incoming IPv6-in-IPv4 tunnel traffic using IPv4 protocol type 41. Outgoing traffic of this kind should be blocked except for the case noted in Section 4.5 of [RFC6343]. ICMPv6 traffic should only be blocked in accordance with [RFC4890]; in particular, Packet Too Big messages, which are essential for PMTU discovery, must not be blocked.

Brute-force scanning attacks to discover the existence of hosts are much less likely to succeed for IPv6 than for IPv4 [RFC5157]. However, this should not lull an ICP into a false sense of security, as various naming or addressing conventions can result in IPv6 address space being predictable or guessable. In the extreme case, IPv6 hosts might be configured with interface identifiers that are very easy to guess; for example, hosts or subnets manually numbered

with consecutive interface identifiers starting from "1" would be much easier to guess. Such practices should be avoided, and other useful precautions are discussed in [RFC6583]. Also, attackers might find IPv6 addresses in logs, packet traces, DNS records (including reverse records), or elsewhere.

Protection against rogue Router Advertisements (RA Guard) should also be considered [RFC6105].

Transport Layer Security version 1.2 [RFC5246] and its predecessors work correctly with TCP over IPv6, meaning that HTTPS-based security solutions are immediately applicable. The same should apply to any other transport-layer or application-layer security techniques.

If an ASP uses IPsec [RFC4301] and IKE [RFC5996] in any way to secure connections with clients, these too are fully applicable to IPv6, but only if the software stack at each end has been appropriately updated.

15. IANA Considerations

This document requests no action by IANA.

16. Acknowledgements

Valuable contributions were made by Erik Kline. Useful comments were received from Tore Anderson, Cameron Byrne, Tassos Chatzithomaoglou, Wesley George, Deng Hui, Joel Jaeggli, Roger Jorgensen, Victor Kuarsingh, Bing Liu, Trent Lloyd, John Mann, Michael Newbery, Erik Nygren, Arturo Servin, Mark Smith, and other participants in the V6OPS working group.

Brian Carpenter was a visitor at the Computer Laboratory, Cambridge University during part of this work.

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

17. Change log [RFC Editor: Please remove]

draft-ietf-v6ops-icp-guidance-05: IESG comments, 2013-01-11.

draft-ietf-v6ops-icp-guidance-04: text on multiple PA prefixes tuned again, brief mention of NPTv6, 2012-09-17.

draft-ietf-v6ops-icp-guidance-03: text on multiple PA prefixes

updated, numerous other WGLC comments, 2012-08-31.

draft-ietf-v6ops-icp-guidance-02: additional WG review, 2012-07-11.

draft-ietf-v6ops-icp-guidance-01: additional WG comments, 2012-06-12.

draft-ietf-v6ops-icp-guidance-00: adopted by WG, small updates, 2012-04-17.

draft-carpenter-v6ops-icp-guidance-03: additional WG comments, 2012-02-23.

draft-carpenter-v6ops-icp-guidance-02: additional WG comments, 2012-01-07.

draft-carpenter-v6ops-icp-guidance-01: multiple clarifications after WG comments, 2011-12-06.

draft-carpenter-v6ops-icp-guidance-00: original version, 2011-10-22.

18. References

18.1. Normative References

- [RFC2080] Malkin, G. and R. Minnear, "RIPng for IPv6", RFC 2080, January 1997.
- [RFC2460] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", RFC 2460, December 1998.
- [RFC2616] Fielding, R., Gettys, J., Mogul, J., Frystyk, H., Masinter, L., Leach, P., and T. Berners-Lee, "Hypertext Transfer Protocol -- HTTP/1.1", RFC 2616, June 1999.
- [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.
- [RFC3007] Wellington, B., "Secure Domain Name System (DNS) Dynamic Update", RFC 3007, November 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.
- [RFC3596] Thomson, S., Huitema, C., Ksinant, V., and M. Souissi, "DNS Extensions to Support IP Version 6", RFC 3596,

October 2003.

- [RFC3704] Baker, F. and P. Savola, "Ingress Filtering for Multihomed Networks", BCP 84, RFC 3704, March 2004.
- [RFC4193] Hinden, R. and B. Haberman, "Unique Local IPv6 Unicast Addresses", RFC 4193, October 2005.
- [RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, December 2005.
- [RFC4760] Bates, T., Chandra, R., Katz, D., and Y. Rekhter, "Multiprotocol Extensions for BGP-4", RFC 4760, January 2007.
- [RFC4862] Thomson, S., Narten, T., and T. Jinmei, "IPv6 Stateless Address Autoconfiguration", RFC 4862, September 2007.
- [RFC4941] Narten, T., Draves, R., and S. Krishnan, "Privacy Extensions for Stateless Address Autoconfiguration in IPv6", RFC 4941, September 2007.
- [RFC5246] Dierks, T. and E. Rescorla, "The Transport Layer Security (TLS) Protocol Version 1.2", RFC 5246, August 2008.
- [RFC5308] Hopps, C., "Routing IPv6 with IS-IS", RFC 5308, October 2008.
- [RFC5340] Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF for IPv6", RFC 5340, July 2008.
- [RFC5838] Lindem, A., Mirtorabi, S., Roy, A., Barnes, M., and R. Aggarwal, "Support of Address Families in OSPFv3", RFC 5838, April 2010.
- [RFC5952] Kawamura, S. and M. Kawashima, "A Recommendation for IPv6 Address Text Representation", RFC 5952, August 2010.
- [RFC5985] Barnes, M., "HTTP-Enabled Location Delivery (HELD)", RFC 5985, September 2010.
- [RFC5996] Kaufman, C., Hoffman, P., Nir, Y., and P. Eronen, "Internet Key Exchange Protocol Version 2 (IKEv2)", RFC 5996, September 2010.
- [RFC6434] Jankiewicz, E., Loughney, J., and T. Narten, "IPv6 Node Requirements", RFC 6434, December 2011.

18.2. Informative References

- [I-D.ietf-6renum-enterprise]
Jiang, S., Liu, B., and B. Carpenter, "IPv6 Enterprise Network Renumbering Scenarios, Considerations and Methods", draft-ietf-6renum-enterprise-05 (work in progress), December 2012.
- [I-D.ietf-6renum-static-problem]
Carpenter, B. and S. Jiang, "Problem Statement for Renumbering IPv6 Hosts with Static Addresses in Enterprise Networks", draft-ietf-6renum-static-problem-03 (work in progress), December 2012.
- [I-D.ietf-opsec-vpn-leakages]
Gont, F., "Virtual Private Network (VPN) traffic leakages in dual-stack hosts/ networks", draft-ietf-opsec-vpn-leakages-00 (work in progress), December 2012.
- [I-D.ietf-v6ops-6204bis]
Singh, H., Beebee, W., Donley, C., and B. Stark, "Basic Requirements for IPv6 Customer Edge Routers", draft-ietf-v6ops-6204bis-12 (work in progress), October 2012.
- [I-D.ietf-v6ops-ipv6-multihoming-without-ipv6nat]
Matsushima, S., Okimoto, T., Troan, O., Miles, D., and D. Wing, "IPv6 Multihoming without Network Address Translation", draft-ietf-v6ops-ipv6-multihoming-without-ipv6nat-04 (work in progress), February 2012.
- [I-D.matthews-v6ops-design-guidelines]
Matthews, P., "Design Guidelines for IPv6 Networks", draft-matthews-v6ops-design-guidelines-01 (work in progress), October 2012.
- [RFC1912] Barr, D., "Common DNS Operational and Configuration Errors", RFC 1912, February 1996.
- [RFC2081] Malkin, G., "RIPng Protocol Applicability Statement", RFC 2081, January 1997.
- [RFC2629] Rose, M., "Writing I-Ds and RFCs using XML", RFC 2629, June 1999.
- [RFC2923] Lahey, K., "TCP Problems with Path MTU Discovery",

RFC 2923, September 2000.

- [RFC3068] Huitema, C., "An Anycast Prefix for 6to4 Relay Routers", RFC 3068, June 2001.
- [RFC4038] Shin, M-K., Hong, Y-G., Hagino, J., Savola, P., and E. Castro, "Application Aspects of IPv6 Transition", RFC 4038, March 2005.
- [RFC4192] Baker, F., Lear, E., and R. Droms, "Procedures for Renumbering an IPv6 Network without a Flag Day", RFC 4192, September 2005.
- [RFC4864] Van de Velde, G., Hain, T., Droms, R., Carpenter, B., and E. Klein, "Local Network Protection for IPv6", RFC 4864, May 2007.
- [RFC4890] Davies, E. and J. Mohacsi, "Recommendations for Filtering ICMPv6 Messages in Firewalls", RFC 4890, May 2007.
- [RFC5157] Chown, T., "IPv6 Implications for Network Scanning", RFC 5157, March 2008.
- [RFC5375] Van de Velde, G., Popoviciu, C., Chown, T., Bonness, O., and C. Hahn, "IPv6 Unicast Address Assignment Considerations", RFC 5375, December 2008.
- [RFC6105] Levy-Abegnoli, E., Van de Velde, G., Popoviciu, C., and J. Mohacsi, "IPv6 Router Advertisement Guard", RFC 6105, February 2011.
- [RFC6146] Bagnulo, M., Matthews, P., and I. van Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", RFC 6146, April 2011.
- [RFC6180] Arkko, J. and F. Baker, "Guidelines for Using IPv6 Transition Mechanisms during IPv6 Deployment", RFC 6180, May 2011.
- [RFC6296] Wasserman, M. and F. Baker, "IPv6-to-IPv6 Network Prefix Translation", RFC 6296, June 2011.
- [RFC6343] Carpenter, B., "Advisory Guidelines for 6to4 Deployment", RFC 6343, August 2011.
- [RFC6555] Wing, D. and A. Yourtchenko, "Happy Eyeballs: Success with Dual-Stack Hosts", RFC 6555, April 2012.

- [RFC6583] Gashinsky, I., Jaeggli, J., and W. Kumari, "Operational Neighbor Discovery Problems", RFC 6583, March 2012.
- [RFC6589] Livingood, J., "Considerations for Transitioning Content to IPv6", RFC 6589, April 2012.

Authors' Addresses

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

Email: brian.e.carpenter@gmail.com

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

Email: jiangsheng@huawei.com

Network Working Group
Internet-Draft
Intended status: Informational
Expires: January 16, 2014

S. Jiang, Ed.
Huawei Technologies Co., Ltd
Q. Sun
China Telecom
I. Farrer
Deutsche Telekom AG
Y. Bo
Huawei Technologies Co., Ltd
T. Yang
China Mobile
July 15, 2013

Analysis of Semantic Embedded IPv6 Address Schemas
draft-jiang-semantic-prefix-06

Abstract

This informational document discusses the use of embedded semantics within IPv6 address schemas. Network operators who have large IPv6 address space may choose to embed some semantics into their IPv6 addressing by assigning additional significance to specific bits within the prefix. By embedding semantics into IPv6 prefixes, the semantics of packets can be easily inspected. This can simplify the packet differentiation process. However, semantic embedded IPv6 address schemas have their own operational cost and even potential pitfalls. Some complex semantic embedded IPv6 address schemas may also require new technologies in addition to existing Internet protocols.

The document aims to understand the usage of semantic embedded IPv6 address schemas, and neutrally analyze on the associated advantages, drawbacks and technical gaps for more complex address schemas.

Status of This Memo

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

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

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

This Internet-Draft will expire on January 16, 2014.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
2. Terminology	4
3. Understanding of Semantic IPv6 Prefix Address Schema	4
3.1. Overview of Semantic IPv6 Prefix Address Schema	4
3.2. Existing Approaches to Traffic Differentiation	5
3.3. Justification for Semantics with the IPv6 Prefix	6
3.4. The Semantic Prefix Domain	7
3.5. The Embedded Semantics	8
3.6. Network Operations Based on Semantic Prefixes	8
4. Potential Benefits	9
5. Potential Drawbacks	10
6. Gaps for complex semantic prefix scenarios	11
6.1. Semantic Notification in the Network	11
6.2. Semantic Relevant Interactions between Hosts and the Network	12
6.3. Additional Technical Extensions	12
7. IANA Considerations	13
8. Change Log (removed by RFC editor)	13
9. Security Considerations	14
10. Acknowledgements	14
11. References	14
11.1. Normative References	15
11.2. Informative References	15
Appendix A. An ISP Semantic Prefix Example	15
A.1. Function Type Semantic Bits	16
A.2. Network Device Type Bits within Network Device Address Space	17
A.3. Subscriber Type Bits within Subscriber Address Space	17
A.4. Service Platform Type Bits within Service Platform	

Address Space	18
Appendix B. An Enterprise Semantic Prefix example	19
Appendix C. A Multi-Prefix Semantic example	20
Authors' Addresses	21

1. Introduction

As the global Internet expands, it is being used for an increasingly diverse range of services. These services place differentiated requirements upon packet delivery networks meaning that Internet Service Providers and enterprises need to be aware of more information about each packet in order to best meet a specific service's needs. Dividing a network into different subnets according to different semantics is already widely existing today, mostly motivated by either topological aspects, logical user/device groups, and/or trust/security domains.

In order to inspect the semantics of packets so that they can be treated differently, some network operators have chosen to embed semantics into IPv6 prefixes. Routers and other intermediary devices can easily apply relevant policies as required. User types, service types, applications, security requirements, traffic identity types, quality requirements and other criteria may be used according to how a network operator may want to differentiate its services. Packet-level differentiation can also enable flow-level and user-level differentiation. Consequently, the network operators can treat network packets differently and efficiently. It is believed this mechanism can simplify the management and maintenance of networks.

However, semantic embedded IPv6 address schemas come with their own operational cost and even pitfalls. Some complex semantic embedded IPv6 address schemas may also require technologies additional to existing Internet protocols.

While network operators, who already have large IPv6 address space allocations, are free to plan and deploy addressing in their preferred way (including semantic embedded IPv6 address schemas), it is useful to analyze the benefits and drawbacks of a semantic approach to addressing.

The document only discusses the usage of semantics within a single network, or group of interconnected networks which share a common addressing policy, referred to as a Semantic Prefix Domain.

This document does not intend to suggest the standardization of any common global semantics. It does not intend to draw any conclusions, either recommending this kind of address schemas or not. It aims to provide network operators with relevant information to use in the creation of their own addressing policy.

2. Terminology

The following terms are used throughout this document:

Semantic Prefix: A flexible-length IPv6 prefix which embeds certain semantics.

Semantic Prefix Domain: A portion of the Internet over which a consistent semantic-prefix based policy is in operation.

Semantic Prefix Policy: A policy based on the embedded semantics within IPv6 prefix.

3. Understanding of Semantic IPv6 Prefix Address Schema

Some network operators (either ISPs or enterprise network operators), who have large IPv6 address space, have chosen to embed certain pre-defined semantics into their IPv6 address schemas by assigning additional significance to specific bits within the prefix. The IPv6 addresses of each packet can then explicitly express semantics. Consequently, intermediate devices can easily apply relevant packet differentiating operations accordingly. This mechanism may divert much network complexity to the planning and management of IPv6 addressing and IP address based policies.

For illustrations of how semantic prefixes could be applied in real-world scenarios, Appendix A describes an ISP example semantic IPv6 prefix address schema; Appendix B introduces an enterprise semantic IPv6 prefix example; and Appendix C introduces an enterprise example in which a multiple-site enterprise network with several prefixes of different lengths is organized as a single, contiguous Semantic Prefix Domain.

3.1. Overview of Semantic IPv6 Prefix Address Schema

A network operator first plans their IPv6 address schema, in which useful semantics (see Section 3.5) are embedded into prefix. They then delegate prefixes with the corresponding semantics to users. The users generate their IPv6 addresses based on assigned prefixes. Then, when the IPv6 stack on the user devices forms packets, the source addresses comprise compliance semantics. For trust reasons, the filters on the edge router may drop packets which are not compliant with assigned prefixes.

The embedded semantics are only meaningful within a network domain which implements a single policy (see Section 3.4). Different service providers may make very different choices regarding the specific semantics which are relevant to their networks. Therefore, it is not possible or even desirable to attempt to standardize a general semantic prefix policy.

Forwarding policies, access control lists, policy-based routing, security isolation and other network operations (see Section 3.6) can be easily applied according to semantics, which are self-expressed by the source address of every packet. Also, the semantics of the destination address may be taken in account if the destination is in the same Semantic Prefix Domain or the peer Semantic Prefix Domain whose semantics has been notified.

3.2. Existing Approaches to Traffic Differentiation

There are several existing approaches which have been developed that can assist operators in identifying and marking traffic. These solutions were mainly developed in the IPv4 era, where the IP address is used as a host locator and little else. The limited capacity of a 32-bit IPv4 address provides very little room for encoding additional information. Correspondingly, these approaches are indirect, inefficient and expensive for operators.

3.2.1. Differentiated Services

Quality of Service (QoS) based on and Differentiated Services [RFC2474] is a widely deployed framework specifying a simple and scalable coarse-grained mechanism for classifying and managing network traffic. But in a service provider's network, DiffServ codepoint (DSCP) values cannot be trusted when they are set by the customer as these are arbitrary values.

In real-world scenarios, ISPs deploy "remarking" points at the customer edge of their network, re-classifying received packets by rewriting the DSCP field according to local policy using information such as the source/destination address, IP protocol number and transport layer source/destination ports.

The traffic classification process leads to increased packet processing overhead and complexity at the edge of the service provider's network.

DSCP mechanism abstracts all the semantics into a single-dimension service classes. This abstract processing has lost a lot of semantic information, which providers want to inspect for every packet, then process the packet accordingly.

The DSCP in the IPv6 header traffic class field allows 6-bits for encoding service provider specific information related to the contents of the packet. Whilst this is a useful part of an overall packet differentiation architecture, the relative small number of available bits (when compared to the available number of bits within the service providers prefix) means that it cannot be used in isolation.

3.2.2. Deep Packet Inspection

Deep Packet Inspection (DPI) may also be used by ISPs to learn the characteristics of users packets. This involves looking into the packet well beyond the network-layer header to identify the specific application traffic type. Once identified, the traffic type can be used as an input for setting the packet's DSCP or other actions.

But DPI is expensive both in processing costs and latency. The processing costs means that dedicated infrastructure is necessary to carry out the function. The incurred latency may be too much for use with any delay/jitter sensitive applications. As a result, DPI is difficult for large-scale deployment and it's usage is usually limited to small and specific functions in the network. In short, it is not scalable, and cannot support realtime network operations.

3.3. Justification for Semantics with the IPv6 Prefix

Although the interface identifier portion of an IPv6 address has arbitrary bits and extension headers can carry significantly more information, these fields can not be trusted by network operators. Users may easily change the setting of interface identifier or extension headers in order to obtain undeserved priorities/privileges, while servers or enterprise users may be much more self-restricted since they are charged accordingly.

With proper access control filters deployed, the prefix can be trusted by the network operators and is simple to inspect in the IP header of a packet. The packets with the noncompliance source addresses should be filtered. The prefix is delegated by the network and therefore the network is able to detect any undesired

modifications and filter the packet accordingly. This also makes it possible for the service provider to increase the level of trust in a customer-generated packet. If the packet has an source or destination address which is outside of the network operator's policy then a session will simply fail to establish.

3.4. The Semantic Prefix Domain

A Semantic Prefix Domain is a portion of the Internet over which a consistent set of semantic-prefix-based policies are administered in a coordinated fashion. It is analogous to a Differentiated Services Domain [RFC2474]. Some of the characteristics that a single Semantic Prefix Domain could represent include:

- a. Administrative domains
- b. Autonomous systems
- c. Trust regions
- d. Network technologies
- e. Hosts
- f. Routers
- g. User groups
- h. Services
- i. Traffic groups
- j. Applications

A Semantic Prefix Domain has a set of pre-defined semantic definitions, which are only meaningful locally. Without an efficient semantics notification, exchanging mechanism or service agreement, the definitions of semantics are only meaningful within local Semantic Prefix Domain. Agreements on definitions between network operators could be made. However, this may involve trust models among network operators. Sharing semantic definition among Semantic Prefix Domains enables more semantic based network operations.

An enterprise Semantic Prefix Domain may span several physical networks and traverse ISP networks. However, when an interim network is traversed (such as when an intermediary ISP is used for interconnectivity), the relevance of the semantics is limited to network domains that share a common Semantic Prefix Policy.

If an ISP has several non-contiguous address blocks, they may be organized as a single Semantic Prefix Domain if the same Semantic Prefix Policy is shared across these non-contiguous address blocks.

3.5. The Embedded Semantics

The size of the operator assigned prefix means that there is potentially much more scope for embedding semantics than has previously been possible. The following list describes some suggested semantics which may be useful to network operators besides source/destination location:

- a. User types
- b. Applications
- c. Security domain
- d. Traffic identity types
- e. Quality requirements
- f. Geo-location

The selection of semantics varies among different network operators. They may choose one or more semantics to be embedded into their IPv6 address schemas, depending on what is important for them and what may trigger packet differentiation processes in their networks. The selection criterion and the impact of each choice are out of scope of this document.

3.6. Network Operations Based on Semantic Prefixes

From the explicit semantics contained within the addresses of each packet, many network operations can be applied. Compared with traditional operations, these operations are easier to realize and stable. Although detailed operation vary depending on various embedded semantics, the network operations based on semantic prefix can be abstracted into following categories:

- a. Statistic based on certain semantic. Any embedded semantic can be set as a statistic condition. In other words, any embedded semantic can be measured independently.
- b. Differentiate packet processing. Many packet processing operations can be applied based on the semantic differentiation, such as queueing, path selection, forwarding to certain process devices, etc.

- c. Security isolation. A set of packet filters that are based on semantic can fulfil network security isolation.
- d. Access control. Resource access, authentication, service access can be directly based on semantics.
- e. Resource allocation. Resources, such as bandwidth, fast queue, caching, etc., can be allocated or reserved for certain semantic users/packets.
- f. Virtualization. Within a Semantic Prefix Domain, organizing virtual networks is simplified by assigning all the nodes the same semantic identifier so that the packets from them can be distinguished from other virtual networks.

It should also be noticed that these operations do not have to be processed on the same single device. They may be separated among network devices. In other words, if there are multiple semantics in a Semantics Prefix Domain, various semantics may be understood and treated on different network devices. It is not necessary for all network devices in such domain to capable of understanding all semantics.

4. Potential Benefits

Depending on various embedded semantics, different beneficial scenarios can be expected.

- a. Semantic prefix address schema provides a directly and explicitly mechanism for packet inspection. It improves the inspecting efficiency on IPv6 network devices.
- b. Simplified measurement and statistics gathering: the semantic prefix provides explicit identifiers which can be used for measurement and statistical information collection. This can be achieved by checking certain bits of the source and/or destination address in each packet.
- c. Simplified flow control: by applying policies according to certain bit values, packets carrying the same semantics in their source/destination addresses can.
- d. Service segregation: when service related information is encoded within the semantic prefix, this can be used to create simple access-control lists which can be applied uniformly across all network devices. Security zones are such typical services that need to be segregated.

- e. Policy aggregation: the semantic prefix allows many policies to be aggregated according to the same semantics within the policy based routing system [RFC1104].
- f. Easy dynamic reconfiguration of semantic oriented policy: network operators may want to dynamically change the policy actions that are operated on certain semantic packets. The semantic prefix allows such changes be operated easily, as only a small number of consistent policy rules need to be updated on all devices within the semantic prefix domain.
- g. Application-aware routing: embedding application information into IP addresses is the simplest way to realize application aware routing.
- h. Easy user behavior management: based on the user type reading from the addresses, any improper user behaviors can be easily detected and automatically handled by network policies.
- i. Easy network resources access rights management: the authentication of access right may already be embedded into the addresses. Simple matching policies can filter improper access requests.
- j. Easy virtualization: virtual network based on any semantics can be easily deployed using the semantic prefix mechanism.

5. Potential Drawbacks

- a. Address consumption caused by lower address utility rate.
Embedding semantics into IPv6 addresses causes the network to use more of the address space than it normally would. The wastage comes from aligning. 1) A small addressing requirement for a separate type may get the same large address space as a large addressing requirement. 2) The number of types in each semantic has to align to 2^n , for example, 5 types use to take 3 bits in the prefix.

Network operators should be aware they may not get more addresses because they have allocated their assigned address block(s) for semantic use without the addresses actually being in use - leading to a lower address utility rate. Although the current Regional Internet Registry (RIR) policies do not disallow such address usage, such usage has not been taken into account in calculating reasonable addressing quotients.

- b. Complexity that is created within the semantic prefix policy.
Encoding too many semantics into prefixes can come at the expense

of future addressing flexibility. At the same time, embedding too many semantics may induce semantic overlap. Careful consideration should be taken with semantics definition.

- c. The risk of privacy/information leakage. The semantics in the address may be guessable, or leaked to outside the organisation. Therefore, some information of either subscribers or networks may be leaked, too.
- d. Burdening the host OS. In some complex semantic prefix scenarios, the semantics prefix mechanism puts extra burden on the originator. In such scenarios, host devices are given multiple IPv6 prefixes and required to choose correctly. When forming a packet, the originator of packets (normally the host OS) has to pick the right address/prefix according to the semantics to access a service.
- e. In order to perform policies based on trusted user/prefix, tight/strict access control filter linked with prefix assignment is requested. It is the filter who makes sure the prefix right. The filter should link back to other states of the user, like user authentication, etc, in order to match the packet to its properties and check whether it is mapped to right semantics or not.

6. Gaps for complex semantic prefix scenarios

The simplest semantic prefix model is to embed only abstracted user type semantics into the prefix. Current network architectures can support this semantic prefix model, in which each subscriber is still assigned a single prefix, while they are not notified the semantic embedded in the prefix.

In order to fulfill more benefits of the semantic prefix design, additional functions are needed to allow semantic relevant operations in networks and semantic relevant interactions with hosts.

IPv6 provides a facility for multiple addresses to be configured on a single interface. This creates a precondition for the approach that user chooses addresses differently for different purposes/usages.

6.1. Semantic Notification in the Network

In order to manage semantic prefixes and their relevant network actions, the network should be able to notify semantics along with prefix delegation.

When an prefix is delegated using a DHCPv6 IA_PD [RFC3633], the associated semantics should also be propagated to the requesting router. This is particularly useful for autonomic process when a new device is connected.

6.2. Semantic Relevant Interactions between Hosts and the Network

The more that semantics are embedded into a prefix, the more complicated functions are needed for semantic relevant interactions between hosts and the network, such as prefix delegation, host notification, address selections, etc.

In practice, a single host may belong to multiple semantics. This means that several IPv6 addresses are configured on a single physical interface and should be selected for use depending on the service that a host wishes to access. A certain packet would only serve a certain semantic.

The host's IPv6 stack must have a mechanism for understanding these semantics in order to select the right source address when forming a packet. If the embedded semantic is application relevant, applications on the hosts should also be involved in the address choosing process: the host IPv6 stack reports multiple available addresses to the application through socket API (one example is "IPv6 Socket API for Source Address Selection" [RFC5014]). The application then needs to apply the semantic logic so that it can correctly select from the offered candidate addresses.

Although [RFC6724] provides an algorithm for source address selection, some semantic prefix policies may conflict with this algorithm. In this case, source address selection mechanisms may need further supporting functions to be developed.

6.3. Additional Technical Extensions

There are several areas in which the semantic prefix could be extended in order to increase the usefulness and applicability of the semantic prefix address schema. They are listed here for future study. Currently, their feasibility, usefulness and applicability are not carefully studied yet.

- Dynamic Policy Configuration

Dynamic policy configuration would simplify the distribution of policy across devices in the semantic prefix domain. New functions or protocol extension are needed to enable dynamic changes to the policy actions in operation on certain semantic packets.

- Semantics Announcements to peer networks

A network may announce all, or some of its Semantic Prefix Policy to connected peer networks. This could be used to enable more dynamic configuration and enable traffic from different semantic prefix domains to traverse different networks whilst having the same semantic prefix policy applied. To achieve this automatically by message exchanging would require new functions or protocol extensions.

- Extension of Prefix Semantics beyond the left-most 64 bits

The prefix concept refers here to the left-most bits in the IP addresses delegated by the network management plane. The prefix could be longer than 64-bits if the network operators strictly manage the address assignment by using Dynamic Host Configuration Protocol for IPv6 (DHCPv6) [RFC3315] (but in this case standard Stateless Address AutoConfiguration - SLAAC [RFC4862] cannot be used).

- Organizing consumer/home networks according to semantics

Consumers or subscribers are currently assigned /48 or /56 prefixes. They have bits, which may also count the right-most 64 bits too, to organize their networks into subnets. These subnets may be organized according to some semantics that are meaningful for the user himself. In such scenario, the user acts as the network operator for his own network. Some additional technologies/functions may be needed to make such organizing and follow-up management efficient.

7. IANA Considerations

This document has no IANA considerations.

8. Change Log (removed by RFC editor)

draft-jiang-semantic-prefix-04: add new pitfalls section; restructure to be a neutral analysis document; 2013-07-15.

draft-jiang-semantic-prefix-05: reword to emphasis this mechanism is a (not the) method that network operators use their addresses; add text to clarify the increased trust is actually from the deployment of source address filter, which is a compliance requirement by semantic prefix; restructure the document, move examples and gap analysis into appendixes, reorganize most content into a frame section; add summarized description for framework at the beginning of Section 3; add description for network operations based on semantic prefix; add a new coauthor who contributes an enterprise semantic prefix network example; combine most of draft-sun-v6ops-semantic-usecase into the draft as ISP example in appendix; 2013-5-28.

draft-jiang-semantic-prefix-04: add new coauthor, re-organize the content, and refine the English, 2013-1-31.

draft-jiang-semantic-prefix-03: add the concept of hierarchical Semantic Prefix Domain and more gap analysis, 2012-10-22.

draft-jiang-semantic-prefix-02: resubmitted to v6ops WG. Removed detailed examples and recommendations for semantics bits, 2012-10-15.

draft-jiang-semantic-prefix-01: added enterprise considerations and scenarios, emphasizing semantics only for local meaning and no intend to standardize any common global semantics, 2012-07-16.

9. Security Considerations

Embedding semantics in prefix is actually exposing more information of packets explicit. These informations may also provide convenient for malicious attackers to track or attack certain type of packets. If networks announce their local prefix semantics to their peer networks, it may also increase the vulnerable risk.

Prefix-based filters should be deployed, in order to protect against address spoofing attacks or denial of service for packets with forged source addresses.

10. Acknowledgements

Useful comments were made by Erik Nygren, Dan Wing, Nick Hilliard, Ray Hunter, David Farmer, Fred Baker, Joel Jaeggli, John Curran, Tim Chown, Ted Lemon, Owen DeLong, Lorenzo Colitti, George Michaelson, Joel Halpern, Vizdal Ales, Bless Roland, Manning Bill, Manfred Albert and other participants in the V6OPS working group.

11. References

11.1. Normative References

- [RFC1104] Braun, H., "Models of policy based routing", RFC 1104, June 1989.
- [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.
- [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.
- [RFC4291] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 4291, February 2006.
- [RFC4862] Thomson, S., Narten, T., and T. Jinmei, "IPv6 Stateless Address Autoconfiguration", RFC 4862, September 2007.
- [RFC6724] Thaler, D., Draves, R., Matsumoto, A., and T. Chown, "Default Address Selection for Internet Protocol Version 6 (IPv6)", RFC 6724, September 2012.

11.2. Informative References

- [RFC5014] Nordmark, E., Chakrabarti, S., and J. Laganier, "IPv6 Socket API for Source Address Selection", RFC 5014, September 2007.

Appendix A. An ISP Semantic Prefix Example

This ISP semantic prefix example is abstracted from a real ISP address architecture design.

Note: for now, this example only covers unicast address within IP Version 6 Addressing Architecture [RFC4291].

For ISPs, several motivations to use semantic prefixes are as follows:

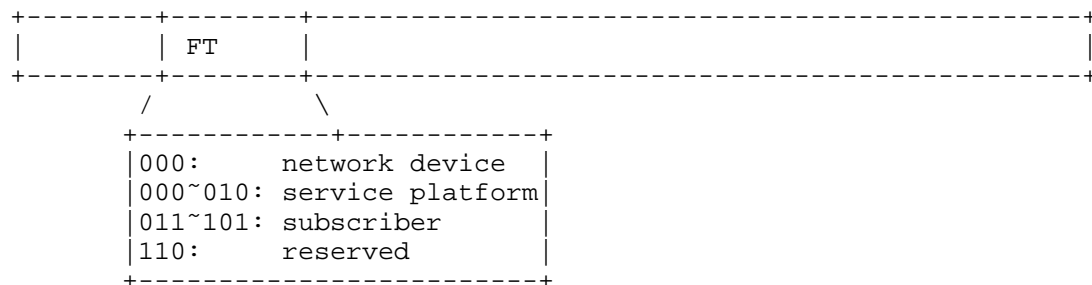
- a. Network Device management: Separated and specialized address space for network device will help to identify the network device among numerous addresses and apply policy accordingly.

- b. Differentiated user management: In ISPs' network, different kinds of customers may have different requirements for service provisioning.
- c. High-priority service guarantee: Different priorities may be divided into apply differentiated policy.
- d. Service-based Routing: ISPs may offer different routing policy for specific service platforms .e.g.video streaming, VOIP, etc.
- e. Security Control: For security requirement, operators need to take control and identify of certain devices/customers in a quick manner.
- f. Easy measurement and statistic: The semantic prefix provides explicit identifiers for measurement and statistic.

These requirements are largely falling into two categories: some is regarding to the network device features, and the others are related to services provision and subscriber identification. The functional usage of the semantics for the two categories are quite different. Therefore, an ISP semantic IPv6 prefix example is designed as a two-level hierarchical architecture, in which the first level is the function types of prefixes, and the second level is the further usage within an specific prefix type.

A.1. Function Type Semantic Bits

Function Type (FT): the value of this field is to indicate the functional usage of this prefix. The typical types for operators include network device, subscriber and service platform.



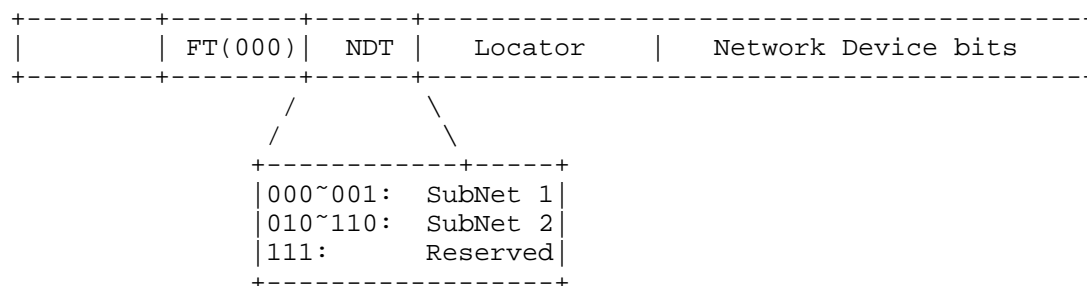
Function Type Bits Example

Figure 1

The portion of each type should be estimated according to the actual requirements for operators, in order to use the address space most efficiently. Within the above FT design, the whole ISP IPv6 address space is divided into four parts: the network device address space (1/8 of total address space), the service platform address space (2/8 of total address space), the subscriber address space (3/8 of total address space), and a reserved address space (1/8 of total address space) for future usage.

A.2. Network Device Type Bits within Network Device Address Space

Network Device Type (NDT) indicates different types of network devices. Normally, one operator may have multiple networks, e.g. backbone network, mobile network, ISP brokered service network, etc. Using NDT field to indicate specific network within an operator may help to apply some routing policies. Locating NDT bits in the left-most bits means that a single, simple access-control list implemented across all networking devices would be enough to enforce effective traffic segregation. The Locator field is followed behind NDT.



Network Device Type Bits Example

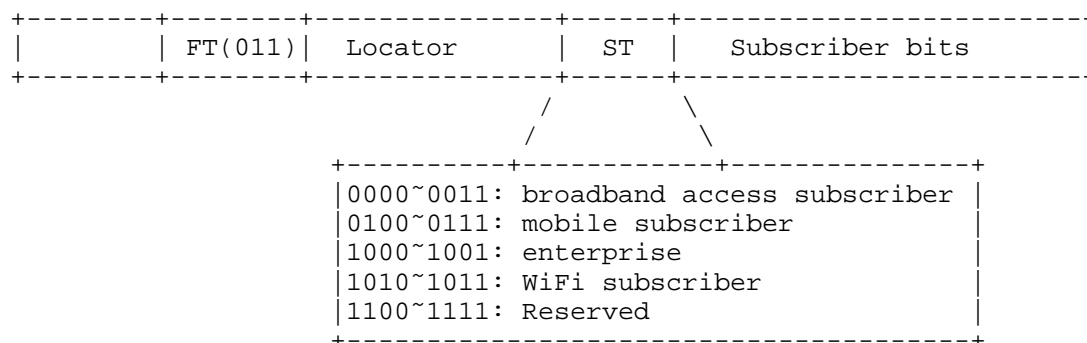
Figure 2

The portion of each subnet type should be estimated according to the actual requirements for operators, in order to use the address space most efficiently. Within the above NDT design, SubNet 1 is assigned 2/8 of the network device address space, SubNet 2 is assigned 5/8, and 1/8 is reserved.

A.3. Subscriber Type Bits within Subscriber Address Space

Subscriber Type (ST) indicates different types of subscribers, e.g. wireline broadband subscriber, mobile subscriber, enterprise, WiFi, etc. This type of prefix is allocated to end users. Further, division may be taken on subscriber's priorities within a certain subscriber type.

The Locator field within subscriber address space is put before ST for better routing aggregation.



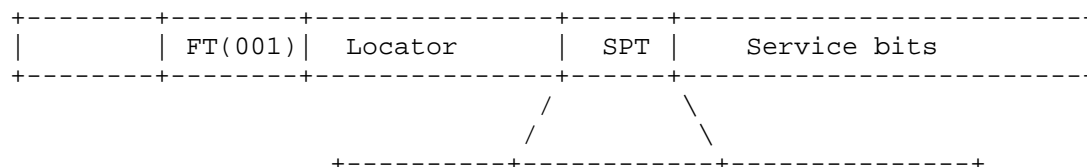
Subscriber Type Bits Example

Figure 3

The portion of each subscriber type should be estimated according to the actual requirements for operators, in order to use the address space most efficiently. Within the above ST design, the broadband access subscriber type is assigned 4/16 of the subscriber address space, the mobile subscriber is assigned 4/16, enterprise type and WiFi subscriber type are assigned 2/16 each, and 2/16 is reserved.

A.4. Service Platform Type Bits within Service Platform Address Space

Service Platform Type (SPT) indicates typical service platforms offered by operators. This field may have scalability problem since there are numerous types of services. It is recommended that only aggregated service platform types should be defined in this field. This type of prefix is usually allocated to service platforms in operator's data center.



000~001:	Self-running service platform	
001~011:	Tenant service platform	
100~101:	Independent service platform	
110~111:	Reserved	
+-----+		

Service Platform Type Bits Example

Figure 4

The portion of each subnet type should be estimated according to the actual requirements for operators, in order to use the address space most efficiently.

Appendix B. An Enterprise Semantic Prefix example

This enterprise semantic prefix example is also abstracted from an ongoing enterprise address architecture design. This example is designed for a realtime video monitor network across a city region. The semantic prefix solution is planning to be deployed along with a strict authorization system.

Note: this example only covers unicast address within IP Version 6 Addressing Architecture [RFC4291].

For this example, the below semantics are important for the network operation and require different network behaviors.

- a. Terminal type: there are two terminal types only: monitor cameras or video receivers. They are estimated to have similar number. Network devices use another different address space.
- b. Geographic location: the city has been managed in a three-level hierarchical regionalism: district, area and street. Each level has less than 28 sub-regions. This can also be considered as a replacement of topology locator within this specific network.
- c. Authorization level: the network operator is planning to administrate the authorization in three or four levels. An receiver can access the cameras that are the same or lower authorization level.
- d. Civilian or police/government.
- e. Device attribute: this indicates the attribute of a camera device. The attribute is expressed in an abstract way, such as road traffic, hospital, nursery, bank, airport, etc. The abstracted attribute type is designed to be less than 64.

- f. Receiver Attribute: this indicates the attribute of a video receiver. The attribute is based on the receiver group, such as police, firefighter, local security, etc. The attribute/receiver group type is designed to be less than 128.

This example enterprise network has obtained a /32 address block from ISP. There is another /48 dedicated for network devices.

The first bit is Terminal type, which indicates terminal type.

```

0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+
|                               ISP assigned block                               |
+-----+
|T|   Geographic   Locator   | AL|C|Device Attr|   Device Bit   |
+-----+

```

A semantic prefix design example for cameras

Figure 5

3-level hierarchical geographic locator takes 15 bits (each level 5 bits, 32 sub-regions). Authorization level takes 2 bits and 1 bit differentiates civilian or police/government. 6 bits is assigned for device attribute.

```

0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-----+
|                               ISP assigned block                               |
+-----+
|T| GeoLoc  | AL|C|Receiver Attr|   Topology Locator |ReceiverBit|
+-----+

```

An semantic prefix design example for video receivers

Figure 6

The receiver is not as much as geographically distributed as cameras. Therefore, Geographic locator is only detailed to district level. Topology locator is needed for network forwarding and aggregation within a district. It is assigned 10 bits. Authorization level bits and civilian bit are the same with camera address space. Receive attribute takes 7 bits, giving it is designed to be up to 128.

Appendix C. A Multi-Prefix Semantic example

A multiple-site enterprise may have been assigned several prefixes of different lengths by its upstream ISPs. In this situation, in order

to create a single, contiguous Semantic Prefix Domain, it is necessary to base the semantic prefix policy on the longest assigned prefix to ensure that there is enough addressing space to encode a consistent set of semantics across all of the assigned prefixes.

In this example, an enterprise has received a /38 address block for one site (A) and a /44 for a second site (B). They can be organized in the same Semantic Prefix Domain. The most-left 18 (site A) and 12 (site B) bits are allocated as locator. It provides topology based network aggregation. The 8 right-most bits (from bits 56 to 63) are assigned as the semantic field. In this design, the multiple-site enterprise that has been assigned two prefixes of different lengths can be organized as the same Semantic Prefix Domain. The semantic and the Semantic Prefix Domain can traverse the intermediate ISP networks, or even public networks.

The similar situation may happen on ISPs in the future, when an ISP used up its assigned address space, or built up multiple networks in different places.

Authors' Addresses

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

Email: jiangsheng@huawei.com

Qiong Sun
China Telecom
Room 708, No.118, Xizhimennei Street
Beijing 100084
P.R. China

Email: sunqiong@ctbri.com.cn

Ian Farrer
Deutsche Telekom AG
Bonn 53227
Germany

Email: ian.farrer@telekom.de

Yang Bo
Huawei Technologies Co., Ltd
Q21, Huawei Campus, No.156 BeiQing Road
Hai-Dian District, Beijing 100095
P.R. China

Email: boyang.bo@huawei.com

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

Email: yangtianle@chinamobile.com

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

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

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

Abstract

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

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

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

Status of this Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute

working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

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

This Internet-Draft will expire on January 15, 2014.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

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

1. Introduction

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

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

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

2. Architecture and Transition Stages

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

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

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

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

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

2.1. General Architecture

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

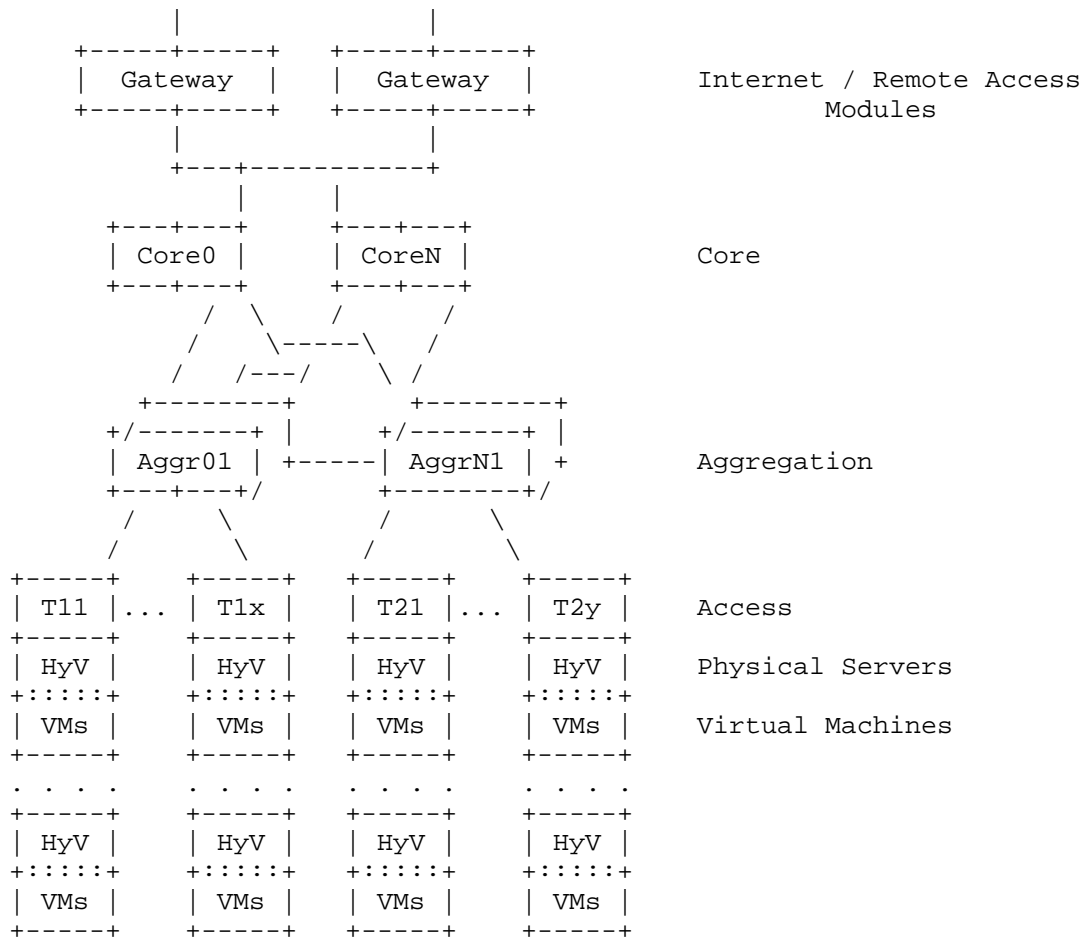


Figure 1: DC Interconnection Schema

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

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

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

2.2. Experimental Stage. Native IPv4 Infrastructure

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

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

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

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

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

2.2.1. Off-shore v6 Access

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

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

2.3. Dual Stack Stage. Internal Adaptation

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

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

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

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

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

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

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

2.3.1. Dual-stack at the Aggregation Layer

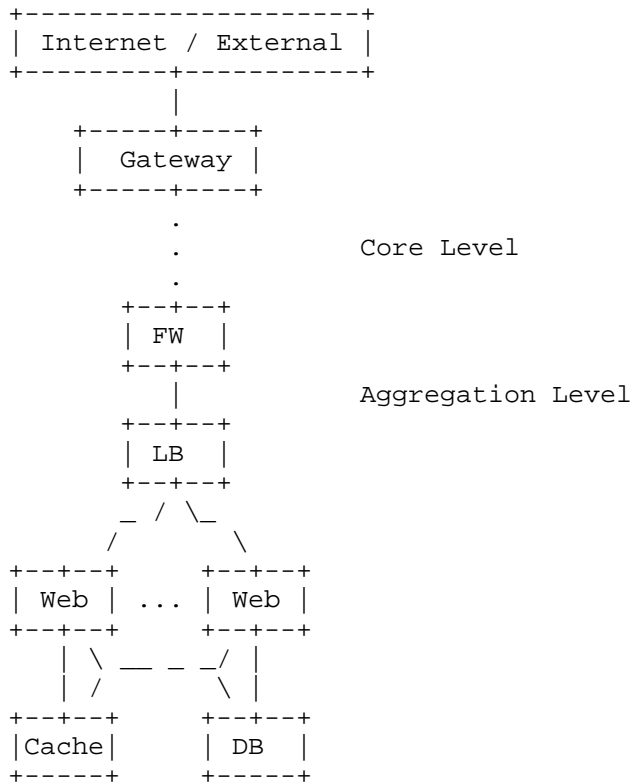


Figure 2: Data Center Application Scheme

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

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

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

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

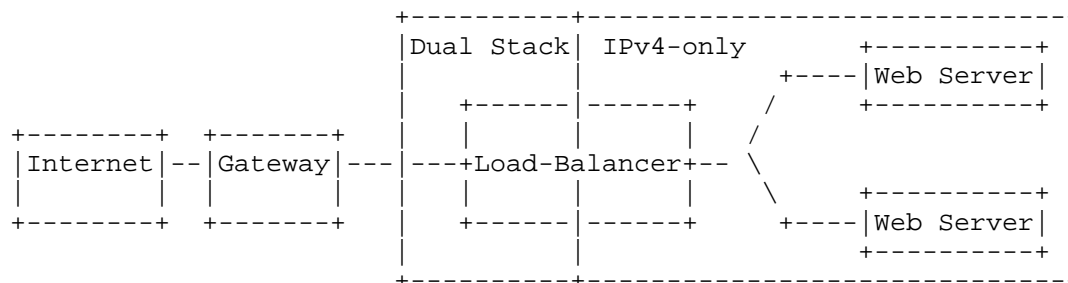


Figure 3: Dual Stack LB mechanism

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

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

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

2.3.2. Dual-stack Extended OS/Hypervisor

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

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

2.4. IPv6-Only Stage. Pervasive IPv6 Infrastructure

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

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

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

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

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

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

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

3. Other Operational Considerations

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

3.1. Addressing

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

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

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

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

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

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

3.2. Management Systems and Applications

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

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

3.3. Monitoring and Logging

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

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

3.4. Costs

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

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

4. Security Considerations

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

4.1. Neighbor Discovery Protocol attacks

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

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

4.2. Addressing

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

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

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

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

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

4.3. Edge filtering

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

4.4. Final Security Remarks

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

5. IANA Considerations

None.

6. Acknowledgements

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

7. Informative References

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

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

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

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

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

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

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

[I-D.ietf-opsec-v6]

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

[I-D.ietf-v6ops-enterprise-incremental-ipv6]

Chittimaneni, K., Chown, T., Howard, L., Kuarsingh, V., Pouffary, Y., and E. Vyncke, "Enterprise IPv6 Deployment Guidelines", draft-ietf-v6ops-enterprise-incremental-ipv6-03 (work in progress), July 2013.

[I-D.ietf-v6ops-icp-guidance]

Carpenter, B. and S. Jiang, "IPv6 Guidance for Internet Content and Application Service Providers", draft-ietf-v6ops-icp-guidance-05 (work in progress), January 2013.

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

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

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

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

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

Considerations", RFC 5375, December 2008.

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

Authors' Addresses

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

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

Zhonghua Chen
China Telecom
P.R.China

Phone:
Email: 18918588897@189.cn

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

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

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

Phone:
Email: cathy.zhou@huawei.com

Arturo Servin
LACNIC
Rambla Republica de Mexico 6125
Montevideo 11300
Uruguay

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

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

P. Matthews
Alcatel-Lucent
October 22, 2012

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

Abstract

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

Status of this Memo

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

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

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

This Internet-Draft will expire on April 25, 2013.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

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

1. Introduction

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

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

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

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

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

2. Design Choices

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

2.1. Mix IPv4 and IPv6 on the Same Link?

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

- a. Mix IPv4 and IPv6 traffic on the same layer 2 connection, OR
- b. Separate IPv4 and IPv6 by using separate physical or logical links (e.g., two physical links or two VLANs on the same link)?

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

The advantages of option (a) include:

- o Requires only half as many layer 3 interfaces as option (b), thus providing better scaling;
- o May require fewer physical ports, thus saving money;
- o Can make the QoS implementation much easier (for example, rate-limiting the combined IPv4 and IPv6 traffic to or from a customer);
- o Provides better support for the expected future of increasing IPv6 traffic and decreasing IPv4 traffic;
- o And is generally conceptually simpler.

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

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

Most networks today use option (a) wherever possible.

2.2. Links with Only Link-Local Addresses?

Should the link:

- a. Use only link-local addresses ("unnumbered"), OR
- b. Have global or unique-local addresses assigned in addition to link-locals?

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

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

- o It is not possible to ping an interface that has only a link-local address from a device that is not directly attached to the link. Thus, to troubleshoot, one must typically log into a device that is directly attached to the device in question, and execute the ping from there.
- o A traceroute passing over the unnumbered link will return the loopback or system address of the router, rather than the address of the interface itself.
- o On some devices, by default the link-layer address of the interface is derived from the MAC address assigned to interface. When this is done, swapping out the interface hardware (e.g. interface card) will cause the link-layer address to change. In some cases (peering config, ACLs, etc) this may require additional changes. However, many devices allow the link-layer address of an interface to be explicitly configured, which avoids this issue.
- o The practice of naming router interfaces using DNS names is difficult-to-impossible when using LLAs only.
- o It is not possible to identify the interface or link (in a database, email, etc) by just giving its address.

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

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

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

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

- a. Use the far-end's link-local address as the next-hop address, OR
- b. Use the far-end's GUA/ULA address as the next-hop address?

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

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

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

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

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

Today most operators use GUAs as next-hop addresses.

2.4. Separate or Combined eBGP Sessions?

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

- a. Use one BGP session to carry both IPv4 and IPv6 routes, OR
- b. Use two BGP sessions, a session over IPv4 carrying IPv4 routes and a session over IPv6 carrying IPv6 routes.

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

sessions compared with (b).

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

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

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

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

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

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

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

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

local addresses:

- o Using link-local addresses only works for single-hop eBGP sessions; it does not work for multi-hop sessions.
- o One must use "next-hop self" at both endpoints, otherwise redistributing routes learned via eBGP into iBGP will not work. (Some products enable "next-hop self" in this situation automatically).
- o Operators and their tools are used to referring to eBGP sessions by address only, something that is not possible with link-local addresses.
- o If one is configuring parallel eBGP sessions for IPv4 and IPv6 routes, then using link-local addresses for the IPv6 session introduces an extra difference between the two sessions which could otherwise be avoided.
- o On some products, an eBGP session using a link-local address is more complex to configure than a session that use a global address.
- o Finally, a strict interpretation of RFC 2545 can be seen as forbidding running eBGP between link-local addresses, as RFC 2545 requires the BGP next-hop field to contain at least a global address.

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

3. General Observations

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

3.1. Use of Link-Local Addresses

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

There are three main reasons for this current practice:

- o Network operators are concerned about the volatility of link-local addresses based on MAC addresses, despite the fact that this concern can be overcome by manually-configuring link-local addresses;
- o It is impossible to ping a link-local address from a device that is not on the same subnet. This is a troubleshooting disadvantage, though it can also be viewed as a security advantage.
- o Most operators are currently running networks that carry both IPv4 and IPv6 traffic, and wish to harmonize their IPv4 and IPv6 design and operational practices where possible.

3.2. Separation of IPv4 and IPv6

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

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

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

4. IANA Considerations

This document makes no requests of IANA.

5. Security Considerations

(TBD)

6. Acknowledgements

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

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

7. History

Version -01

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

Version -00

Initial, very preliminary, version.

8. Informative References

[I-D.ietf-idr-dynamic-cap]

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

[I-D.ietf-opsec-lla-only]

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

- [I-D.ietf-v6ops-enterprise-incremental-ipv6]
Chittimaneni, K., Chown, T., Howard, L., Kuarsingh, V.,
Pouffary, Y., and E. Vyncke, "Enterprise IPv6 Deployment
Guidelines",
draft-ietf-v6ops-enterprise-incremental-ipv6-01 (work in
progress), September 2012.
- [I-D.ietf-v6ops-icp-guidance]
Carpenter, B. and S. Jiang, "IPv6 Guidance for Internet
Content and Application Service Providers",
draft-ietf-v6ops-icp-guidance-04 (work in progress),
September 2012.
- [I-D.ietf-v6ops-wireline-incremental-ipv6]
Kuarsingh, V. and L. Howard, "Wireline Incremental IPv6",
draft-ietf-v6ops-wireline-incremental-ipv6-06 (work in
progress), September 2012.
- [RFC4852] Bound, J., Pouffary, Y., Klynsma, S., Chown, T., and D.
Green, "IPv6 Enterprise Network Analysis - IP Layer 3
Focus", RFC 4852, April 2007.
- [RFC4861] Narten, T., Nordmark, E., Simpson, W., and H. Soliman,
"Neighbor Discovery for IP version 6 (IPv6)", RFC 4861,
September 2007.
- [RFC5082] Gill, V., Heasley, J., Meyer, D., Savola, P., and C.
Pignataro, "The Generalized TTL Security Mechanism
(GTSM)", RFC 5082, October 2007.
- [RFC5308] Hopps, C., "Routing IPv6 with IS-IS", RFC 5308,
October 2008.
- [RFC5340] Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF
for IPv6", RFC 5340, July 2008.
- [RFC5375] Van de Velde, G., Popoviciu, C., Chown, T., Bonness, O.,
and C. Hahn, "IPv6 Unicast Address Assignment
Considerations", RFC 5375, December 2008.
- [RFC5925] Touch, J., Mankin, A., and R. Bonica, "The TCP
Authentication Option", RFC 5925, June 2010.
- [RFC5963] Gagliano, R., "IPv6 Deployment in Internet Exchange Points
(IXPs)", RFC 5963, August 2010.
- [RFC6180] Arkko, J. and F. Baker, "Guidelines for Using IPv6
Transition Mechanisms during IPv6 Deployment", RFC 6180,

May 2011.

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

Author's Address

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

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

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: January 17, 2013

J. Zhang
China Telecom
T. Tsou
Huawei Technologies (USA)
W. Liu
Huawei Technologies
J. Sun
China Telecom
July 16, 2012

IPv6 over ATM Interworking Function
draft-zhang-v6ops-ipv6oa-iwf-01

Abstract

This document describes an interworking function between IPv6 over ATM (Asynchronous Transfer Mode) and IPv6 over Ethernet. The interworking function enables the communication between ATM and Ethernet by maintaining the multiple states of both of them.

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 [RFC2119].

Status of this Memo

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

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

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

This Internet-Draft will expire on January 17, 2013.

Copyright Notice

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
2. Terminology	3
3. Context and Scope	4
3.1. Context and Scope	4
4. Solution Overview	4
4.1. A general topology for IPv4 and IPv6 networks	4
4.2. Address resolution for downstream	5
4.3. Address resolution for upstream	6
4.4. Link layer address change in packets	6
5. Security Considerations	7
6. Acknowledgments	7
7. IANA Considerations	7
8. References	7
8.1. Normative References	7
8.2. Informative References	7
Authors' Addresses	7

1. Introduction

The IPv4 exhaustion problem draws the attention of the world. There are a number of issues to deal with when network migrates to IPv6. The use of IPoA encapsulation on the U-interface in legacy ATM access networks is predominantly applicable to business users. IP addresses used in the network behind the RG are exchanged using routing protocols that run transparently over the ATM PVCs. Currently, IPoA encapsulation is migrated to an Ethernet aggregation network in [TR101]. This is achieved by means of an IPoA Interworking Function in IPv4 network. This model should continue to be supported in IPv6 scenario.

In this draft we define an Interworking Function for IPv6 to support IPv6 over ATM, IPv6 over Ethernet, and IPv6 over SDH. In IPv4 network, upon receiving a ARP request transmitted from a BNG, the IPoA IWF SHOULD be able to reply with the appropriate MAC address used as the source address for upstream packets. In IPv6 scenario, however, address resolution applies Neighbor Discovery Protocol [RFC4861], which is in a completely different way. As the result, IPv6oA IWF should support both the IPv6 address resolution and the functions of IPv4oA IWF in the following way.

For upstream packets, the IPoA IWF MUST use a MAC source address that is under the control of the Access Node (e.g. the MAC address of the Access Node uplink). For upstream unicast packets, the IPoA IWF MUST use a MAC unicast destination address of the BNG. For upstream multicast and broadcast packets, the IPoA IWF MUST derive the MAC destination address using the standard multicast/broadcast IP address to MAC address conversion algorithm.

For downstream, BNG need to know which downstream interface that it should forward to when receiving a packet from internet. IPv6 address resolution is necessary in this scenario. When there are multiple BNGs in the metro networks, the IPv6oA IWF need to know exactly which BNG it should send to. In this case IPv6oA IWF need do IP address resolution. If Layer 2 information (such as MAC) is included in a Neighbor Discovery packet, IPv6oA IWF need make sure the carried information is identical to the layer 2 information in packet. In a word, IPoA IWF should do some additional work when networks migrate from IPv4 to IPv6.

2. Terminology

This document makes use of the following terms:

IPv6oA IWF: IPv6 over ATM interworking function
 BNG: Broadband Network Gateway is the IP edge where user services are performed.
 ATM: Asynchronous Transfer Mode
 RG: Residential Gateway
 PVC: permanent virtual circuit

3. Context and Scope

3.1. Context and Scope

There are many kinds of access protocols between CPE and BNG device. When IPv4 network migrates to IPv6 network, the access protocol must be taken into account. This draft outlines how an IPv6 over ATM access network, or an IPv6 over SDH can be migrated to an Ethernet based IPv6 network.

This document focuses only on interworking function between IPv6 over ATM networks, IPv6 over SDH, and Ethernet based IPv6 networks. In the following, we use IPv6 over ATM to illustrate our main idea. The scenarios include CPE devices that support Inverse Neighbor Discovery [RFC3122] and BNG devices that support Neighbor Discovery based on Ethernet. The IPv6oA interworking function makes sure the two kinds of devices can communicate smoothly. The IPv6oA IWF may exist in CPEs, BNGs or Access nodes(AN).

4. Solution Overview

This section describes solutions for problems mentioned above.

4.1. A general topology for IPv4 and IPv6 networks

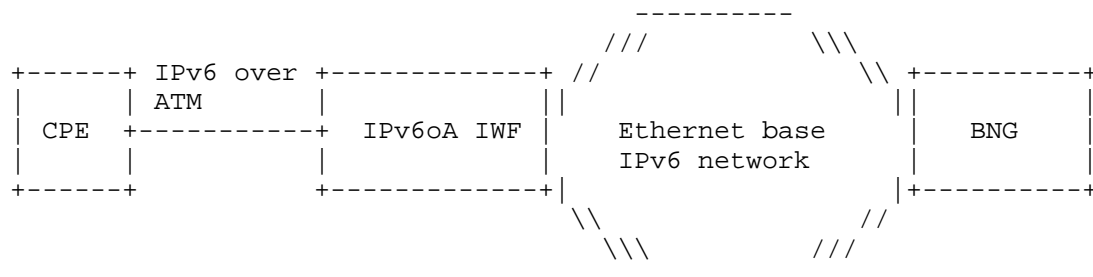


Figure 1: Topology for IPv6 over ATM IWF

IPv6oA IWF exists in a node that lies between two kinds of networks,

IPv6 over ATM and Ethernet base IPv6 network, as shown in Figure 1. [RFC3122] describes extensions to the IPv6 Neighbor Discovery that allow a node to determine and advertise an IPv6 address corresponding to a given link-layer address. These extensions are called Inverse Neighbor Discovery (IND). IPv6 Neighbor Discovery protocol based Ethernet is used to discover each other's presence, to determine each other's link-layer addresses, to find routers, and to maintain reachability information about the paths to active neighbors. The IPv6oA interworking function is proposed to make sure the two kinds of devices, CPE devices that support Inverse Neighbor Discovery and BNG devices that support Neighbor Discovery based on Ethernet, can communicate smoothly. IPv6oA IWF changes packets from IPoA to IPoE in upstream direction and ones from IPoE to IPoA in downstream direction, respectively. The encapsulation specification can be found in [TR101].

4.2. Address resolution for downstream

When a BNG receives a packet from internet sent to a CPE, if the BNG does not know the corresponding link-layer address, it checks the neighbor cache entries to get the link layer address, i.e., the MAC address in Ethernet. If the link layer address of corresponding destination IP address can not be found there, the BNG initiates an address resolution as described in chapter 7.2 of [RFC4861]. The BNG parses the destination IP address in the packet aiming to the CPE and include the parsed IP address in Target Address field of Neighbor Solicitation Message. Then BNG will send Neighbor Solicitation to its downstream interface.

When IPv6oA IWF receives the Neighbor Solicitation Message from a BNG, it has to make sure the IP address that needs to be resolved is on link IP address. IPv6oA IWF will initiate an Inverse Neighbor Discovery Solicitation to CPE. When IPv6oA IWF receives the Inverse Neighbor Discovery Advertisements from a CPE for the resolution IP address, it checks the Source/Destination Address List for the IP address to be resolved. If the IP address is in the list, then IPv6oA IWF can respond to the Neighbor Solicitation Message from BNG with Neighbor Advertisement Message. IPv6oA IWF sets the MAC of node's upstream interface to Source/Destination Link-layer Address option field in Neighbor Advertisement Message. Then IPv6oA IWF can receive packets sent to CPE and forward them according to IPoA encapsulation specification.

When IPv6oA IWF sends Inverse Neighbor Discovery Message, it can change the Neighbor Solicitation Message from BNG to Inverse Neighbor Discovery Solicitation message or change the Inverse Neighbor Discovery Advertisements message from CPE to Neighbor Advertisements message since IND protocol is just the extensions to the IPv6

Neighbor Discovery. IPv6oA IWF can work in layer 2 nodes. For upstream packets, the IPoA IWF MUST use a MAC source address that is under the control of the Access Node (e.g. the MAC address of the Access Node uplink). For upstream unicast packets, the IPoA IWF MUST use a MAC unicast destination address of the BNG. For upstream multicast and broadcast packets, the IPoA IWF MUST derive the MAC destination address using the standard multicast/broadcast IP address to MAC address conversion algorithm.

4.3. Address resolution for upstream

When there are multiple BNGs in a metro network, in other word, multiple IP edges in same networks, IPv6oA IWF needs to know which BNG it will send according to destination IP in the packets received from the CPE. IPv6oA IWF could configure the different BNG IPs and MACs as the next hop by filtering the destination IP address of packets to be sent. If we do not configure this information, IPv6oA IWF needs to finish address resolution for upstream packets. For simplicity, it is suggested to manually configure the BNG IP and MAC on IPv6oA IWF node.

When IPv6oA IWF receives Inverse Neighbor Discovery Solicitation, it responses Inverse Neighbor Discovery Advertisement with Source/Destination Address List in the message. The Source/Destination Address List includes the IPv6 address of BNG's downstream interface. Source/Destination Address List can be configured or accessed from Neighbor Cache entries.

When IPv6oA IWF receives IPv6 packets sent to a BNG, it makes address resolution. IPv6oA IWF needs to establish Neighbor Solicitation Message and support address resolution mentioned in charter 7.2 of [RFC4861]. After it receives Neighbor Advertisement message, Neighbor Cache entries will be established according to the reply from BNG. IPv6 IWF then changes IPoA packets to IPoE packets according to [TR101]. The destination MAC address and the source MAC address are set as the resolved link address and the upstream link address of IPv6oA IWF, respectively.

4.4. Link layer address change in packets

Inverse Neighbor Discovery Solicitation, Neighbor Solicitation, Router Solicitation, and Router Advertisement packets may include a Source/Destination Link-layer Address option in the packet. Inverse Neighbor Discovery Advertisement, Neighbor Advertisement and Redirect packets may include a Source/Destination Link-layer Address option in the packet. After IPv6oA IWF gets these packets from CPE or from BNG, it checks whether there is a Source/Destination Link-layer Address option in the packets. If yes, it changes the Source/

Destination Link-layer Address option value to Ethernet Link-layer Address (such as upstream interface MAC address of Access Node that IPv6oA IWF lies in) for upstream packets or ATM Link-layer Address for downstream packets.

5. Security Considerations

6. Acknowledgments

7. IANA Considerations

8. References

8.1. Normative References

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

8.2. Informative References

[RFC3122] Conta, A., "Extensions to IPv6 Neighbor Discovery for Inverse Discovery Specification", RFC 3122, June 2001.

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

[TR101] Boardband-forum, "Migration to Ethernet-Based Broadband Aggregation", Jul 2011, <http://www.broadband-forum.org/technical/download/TR-101_Issue-2.pdf>.

Authors' Addresses

Jiexin Zhang
China Telecom
NO 1835, Dongnan RD, Pudong DIST
Shanghai,
P.R. China

Email: estrellazhang2012@gmail.com

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

Phone: +1 408 330 4424
Email: tina.tsou.zouting@huawei.com

Will Liu
Huawei Technologies
Bantian, Longgang DIST
Shenzhen 518129
P.R. China

Phone: +86 755 28972315
Email: liushucheng@huawei.com

Jianping Sun
China Telecom
NO 1835, Dongnan RD, Pudong DIST
Shanghai,
China

Email: sunjp@sttri.com.cn

