IPv6 Operations
Internet-Draft
Intended status: Informational
Expires: February 22, 2019

J. Linkova
Google
M. Stucchi
RIPE NCC

August 21, 2018

Using Conditional Router Advertisements for Enterprise Multihoming
draft-ietf-v6ops-conditional-ras-08

Abstract

This document discusses the most common scenarios of connecting an enterprise network to multiple ISPs using an address space assigned by an ISP and how the approach proposed in the "ietf-rtgwg-enterprise-pa-multihoming" draft could be applied in those scenarios. The problem of enterprise multihoming without address translation of any form has not been solved yet as it requires both the network to select the correct egress ISP based on the packet source address and hosts to select the correct source address based on the desired egress ISP for that traffic. The "ietf-rtgwg-enterprise-pa-multihoming" document proposes a solution to this problem by introducing a new routing functionality (Source Address Dependent Routing) to solve the uplink selection issue and using Router Advertisements to influence the host source address selection. While the above-mentioned document focuses on solving the general problem and on covering various complex use cases, this document adopts the approach proposed in the "ietf-rtgwg-enterprise-pa-multihoming" draft to provide a solution for a limited number of common use cases. In particular, the focus is on scenarios where an enterprise network has two Internet uplinks used either in primary/backup mode or simultaneously and hosts in that network might not yet properly support multihoming as described in RFC8028.

This Internet-Draft will expire on February 22, 2019.

Table of Contents

1. Introduction 3
1.1. Requirements Language 4
2. Common Enterprise Multihoming Scenarios 4
2.1. Two ISP Uplinks, Primary and Backup 4
2.2. Two ISP Uplinks, Used for Load Balancing 5
3. Conditional Router Advertisements 5
3.1. Solution Overview 5
3.1.1. Uplink Selection 5
3.1.2. Source Address Selection and Conditional RAs 5
3.2. Example Scenarios 8
3.2.1. Single Router, Primary/Backup Uplinks 8
3.2.2. Two Routers, Primary/Backup Uplinks 9
3.2.3. Single Router, Load Balancing Between Uplinks 12
3.2.4. Two Router, Load Balancing Between Uplinks 12
3.2.5. Topologies with Dedicated Border Routers 13
3.2.6. Intra-Site Communication during Simultaneous Uplinks Outage 15
3.2.7. Uplink Damping 15
3.2.8. Routing Packets when the Corresponding Uplink is Unavailable 16
3.3. Solution Limitations 16
3.3.1. Connections Preservation 17
4. IANA Considerations 17
5. Security Considerations 17
5.1. Privacy Considerations 18
6. Acknowledgements 18
7. References 18
7.1. Normative References 18
7.2. Informative References 20

Appendix A. Change Log 20
Authors' Addresses 20

1. Introduction

Multihoming is an obvious requirement for many enterprise networks to ensure the desired level of network reliability. However, using more than one ISP (and address space assigned by those ISPs) introduces the problem of assigning IP addresses to hosts. In IPv4 there is no choice but using [RFC1918] address space and NAT ([RFC3022]) at the network edge ([RFC4116]). Using Provider Independent (PI) address space is not always an option, since it requires running BGP between the enterprise network and the ISPs. Administrative overhead of obtaining and managing PI address space can also be a concern. As IPv6 hosts can, by design, have multiple addresses of the global scope ([RFC4291]), multihoming using provider address looks even easier for IPv6: each ISP assigns an IPv6 block (usually /48) and hosts in the enterprise network have addresses assigned from each ISP block. However using IPv6 PA blocks in multihoming scenario introduces some challenges, including but not limited to:

- Selecting the correct uplink based on the packet source address;
- Signaling to hosts that some source addresses should or should not be used (e.g. an uplink to the ISP went down or became available again).

The document [I-D.ietf-rtgwg-enterprise-pa-multihoming] discusses these and other related challenges in detail in relation to the general multihoming scenario for enterprise networks and proposes a solution which relies heavily on the rule 5.5 of the default address selection algorithm ([RFC6724]). The rule 5.5 makes hosts prefer source addresses in a prefix advertised by the next-hop and therefore is very useful in multihomed scenarios when different routers may advertise different prefixes. While [RFC6724] defines the Rule 5.5 as optional, the recent [RFC8028] recommends that multihomed hosts SHOULD support it. Unfortunately that rule has not been widely implemented when this document was written. Therefore network administrators in enterprise networks can't yet assume that all devices in their network support the rule 5.5, especially in the quite common BYOD ("Bring Your Own Device") scenario. However, while it does not seem feasible to solve all the possible multihoming scenarios without relying on rule 5.5, it is possible to provide IPv6 multihoming using provider-assigned (PA) address space for the most common use cases. This document discusses how the general approach described in [I-D.ietf-rtgwg-enterprise-pa-multihoming] can be applied to solve multihoming scenarios when:

- An enterprise network has two or more ISP uplinks;
- Those uplinks are used for Internet access in active/backup or load sharing mode w/o any sophisticated traffic engineering requirements;
- Each ISP assigns the network a subnet from its own PA address space
- Hosts in the enterprise network are not expected to support the Rule 5.5 of the default address selection algorithm ([RFC6724]).

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Common Enterprise Multihoming Scenarios

2.1. Two ISP Uplinks, Primary and Backup

This scenario has the following key characteristics:

- The enterprise network is using uplinks to two (or more) ISPs for Internet access;
- Each ISP assigns IPv6 PA address space for the network;
- Uplink(s) to one ISP is a primary (preferred) one. All other uplinks are backup and are not expected to be used while the primary one is operational;
- If the primary uplink is operational, all Internet traffic should flow via that uplink;
- When the primary uplink fails the Internet traffic needs to flow via the backup uplinks;
- Recovery of the primary uplink needs to trigger the traffic switchover from the backup uplinks back to primary one;
- Hosts in the enterprise network are not expected to support the Rule 5.5 of the default address selection algorithm ([RFC6724]).

2.2. Two ISP Uplinks, Used for Load Balancing

This scenario has the following key characteristics:

- The enterprise network is using uplinks to two (or more) ISPs for Internet access;
- Each ISP assigns an IPv6 PA address space;
- All the uplinks may be used simultaneously, with the traffic flows being randomly (not necessarily equally) distributed between them;
- Hosts in the enterprise network are not expected to support the Rule 5.5 of the default address selection algorithm ([RFC6724]).

3. Conditional Router Advertisements

3.1. Solution Overview

3.1.1. Uplink Selection

As discussed in [I-D.ietf-rtgwg-enterprise-pa-multihoming], one of the two main problems to be solved in the enterprise multihoming scenario is the problem of the next-hop (uplink) selection based on the packet source address. For example, if the enterprise network has two uplinks, to ISP_A and ISP_B, and hosts have addresses from subnet_A and subnet_B (belonging to ISP_A and ISP_B respectively) then packets sourced from subnet_A must be sent to ISP_A uplink while packets sourced from subnet_B must be sent to ISP_B uplink. Sending packets with source addresses belonging to one ISP address space to another ISP might cause those packets to be filtered out if those ISPs or their uplinks implement anti-spoofing ingress filtering ([RFC2827], [RFC3704]).

While some work is being done in the Source Address Dependent Routing (SADR) (such as [I-D.ietf-rtgwg-dst-src-routing]), the simplest way to implement the desired functionality currently is to apply a policy which selects a next-hop or an egress interface based on the packet source address. Most SMB/Enterprise grade routers have such functionality available currently.

3.1.2. Source Address Selection and Conditional RAs

Another problem to be solved in the multihoming scenario is the source address selection on hosts. In the normal situation (all uplinks are up/operational) hosts have multiple global unique addresses and can rely on the default address selection algorithm ([RFC6724]) to pick up a source address, while the network is

responsible for choosing the correct uplink based on the source address selected by a host as described in Section 3.1.1. However, some network topology changes (i.e. changing uplink status) might affect the global reachability for packets sourced from the particular prefixes and therefore such changes have to be signaled back to the hosts. For example:

- An uplink to an ISP_A went down. Hosts should not use addresses from ISP_A prefix;
- A primary uplink to ISP_A which was not operational has come back up. Hosts should start using the source addresses from ISP_A prefix.

[I-D.ietf-rtgwg-enterprise-pa-multihoming] provides a detailed explanation on why SLAAC (Stateless Address Autoconfiguration, [RFC4862]) and RAs (Router Advertisements, [RFC4861]) are the most suitable mechanism for signaling network topology changes to hosts and thereby influencing the source address selection. Sending a router advertisement to change the preferred lifetime for a given prefix provides the following functionality:

- deprecating addresses (by sending an RA with the preferred_lifetime set to 0 in the corresponding PIO (Prefix Information option, [RFC4861])) to indicate to hosts that that addresses from that prefix should not be used;
- making a previously unused (deprecated) prefix usable again (by sending an RA containing a PIO with non-zero preferred lifetime) to indicate to hosts that addresses from that prefix can be used again.

It should be notes that only preferred lifetime for the affected prefix needs to be changed. As the goal is to influence the source address selection algoorithm on hosts, not preventing them from forming addresses from a specific prefix, the valid lifetime should not be changed. Actually it would not even be possible for unauthenticated RAs (which is the most common deployment scenario) as Section 5.5.3 of [RFC4862] prevents hosts from setting valid lifetime for addresses to zero unless RAs are authenticated.

To provide the desired functionality, first-hop routers are required to

- send RA triggered by defined event policies in response to uplink status change event; and

- while sending periodic or solicted RAs, set the value in the given RA field (e.g. PIO preferred lifetime) based on the uplink status.

The exact definition of the 'uplink status' depends on the network topology and may include conditions like:

- uplink interface status change;
- presence of a particular route in the routing table;
- presence of a particular route with a particular attribute (next-hop, tag etc) in the routing table;
- protocol adjacency change.

etc.

In some scenarios, when two routers are providing first-hop redundancy via VRRP (Virtual Router Redundancy Protocol, [RFC5798]), the master-backup status can be considered as a condition for sending RAs and changing the preferred lifetime value. See Section 3.2.2 for more details.

If hosts are provided with ISP DNS servers IPv6 addresses via RDNSS (Router Advertisement Options for DNS Configuration, [RFC8106]) it might be desirable for the conditional RAs to update the Lifetime field of the RDNSS option as well.

The trigger is not only forcing the router to send an unsolicited RA to propagate the topology changes to all hosts. Obviously the RA fields values (like PIO Preferred Lifetime or DNS Server Lifetime) changed by the particular trigger need to stay the same until another event happens causing the value to be updated. E.g. if the ISP_A uplink failure causes the prefix to be deprecated, all solicited and unsolicited RAs sent by the router need to have the Preferred Lifetime for that PIO set to 0 until the uplink comes back up.

It should be noted that the proposed solution is quite similar to the existing requirement L-13 for IPv6 Customer Edge Routers ([RFC7084]) and the documented behavior of homenet devices ([RFC7788]). It is using the same mechanism of deprecating a prefix when the corresponding uplink is not operational, applying it to enterprise network scenario.

3.2. Example Scenarios

This section illustrates how the conditional RAs solution can be applied to most common enterprise multihoming scenarios, described in Section 2.

3.2.1. Single Router, Primary/Backup Uplinks

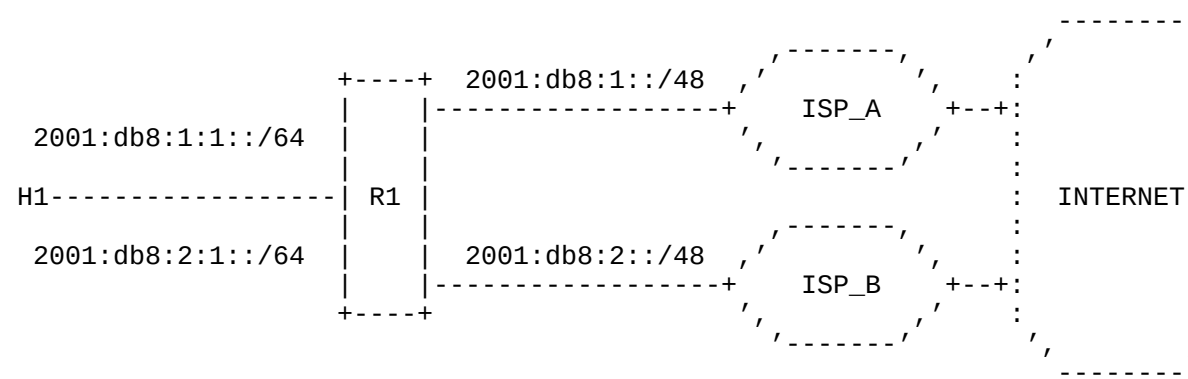

Figure 1: Single Router, Primary/Backup Uplinks

Let's look at a simple network topology where a single router acts as a border router to terminate two ISP uplinks and as a first-hop router for hosts. Each ISP assigns a /48 to the network, and the ISP_A uplink is a primary one, to be used for all Internet traffic, while the ISP_B uplink is a backup, to be used only when the primary uplink is not operational.

To ensure that packets with source addresses from ISP_A and ISP_B are only routed to ISP_A and ISP_B uplinks respectively, the network administrator needs to configure a policy on R1:

```
IF (packet_source_address is in 2001:db8:1::/48)
  and
  (packet_destination_address is not in (2001:db8:1::/48 or 2001:db8:2::/48))
 THEN
       default next-hop is ISP_A_uplink

IF (packet_source_address is in 2001:db8:2::/48)
  and
  (packet_destination_address is not in (2001:db8:1::/48 or 2001:db8:2::/48))
 THEN
      default next-hop is ISP_B_uplink
```

Under normal circumstances it is desirable that all traffic be sent via the ISP_A uplink, therefore hosts (the host H1 in the example topology figure) should be using source addresses from 2001:db8:1:1::/64. When/if ISP_A uplink fails, hosts should stop using the 2001:db8:1:1::/64 prefix and start using 2001:db8:2:1::/64 until the ISP_A uplink comes back up. To achieve this the router advertisement configuration on the R1 device for the interface facing H1 needs to have the following policy:

```
prefix 2001:db8:1:1::/64 {
    IF (ISP_A_uplink is up)
        THEN
                preferred_lifetime = 604800
        ELSE
                preferred_lifetime = 0
}

prefix 2001:db8:2:1::/64 {
    IF (ISP_A_Uplink is up)
        THEN
                preferred_lifetime = 0
        ELSE
                preferred_lifetime = 604800
}
```

A similar policy needs to be applied to the RDNSS Lifetime if ISP_A and ISP_B DNS servers are used.

3.2.2. Two Routers, Primary/Backup Uplinks

Let’s look at a more complex scenario where two border routers are terminating two ISP uplinks (one each), acting as redundant first-hop routers for hosts. The topology is shown on Fig.2

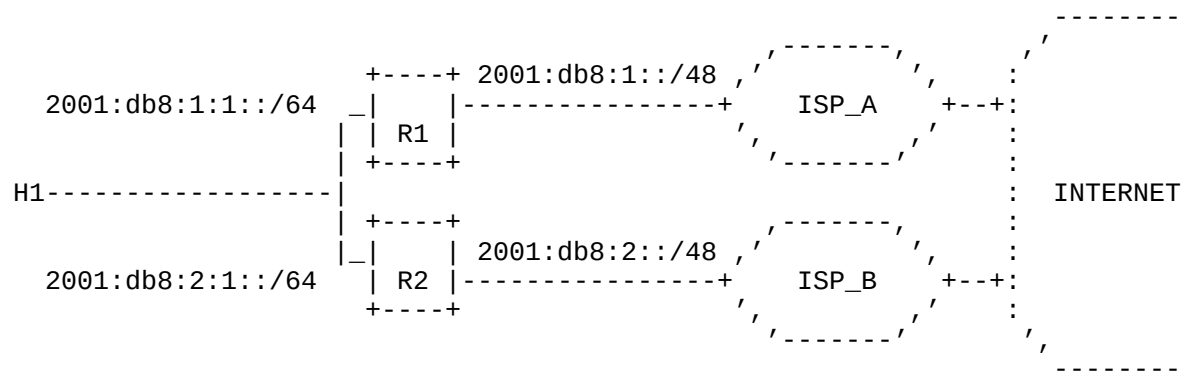

Figure 2: Two Routers, Primary/Backup Uplinks

In this scenario R1 sends RAs with PIO for 2001:db8:1:1::/64 (ISP_A address space) and R2 sends RAs with PIO for 2001:db8:2:1::/64 (ISP_B address space). Each router needs to have a forwarding policy configured for packets received on its hosts-facing interface:

```
IF (packet_source_address is in 2001:db8:1::/48)
  and
  (packet_destination_address is not in (2001:db8:1::/48 or 2001:db8:2::/48))
    THEN
        default next-hop is ISP_A_uplink

IF (packet_source_address is in 2001:db8:2::/48)
i and
  (packet_destination_address is not in (2001:db8:1::/48 or 2001:db8:2::/48))
    THEN
        default next-hop is ISP_B_uplink
```

In this case there is more than one way to ensure that hosts are selecting the correct source address based on the uplink status. If VRRP is used to provide first-hop redundancy and the master router is the one with the active uplink, then the simplest way is to use the VRRP mastership as a condition for router advertisement. So, if ISP_A is the primary uplink, the routers R1 and R2 need to be configured in the following way:

R1 is the VRRP master by default (when ISP_A uplink is up). If ISP_A uplink is down, then R1 becomes a backup (the VRRP interface status tracking is expected to be used to automatically modify the VRRP priorities and trigger the mastership switchover). Router

advertisements on R1's interface facing H1 needs to have the following policy applied:

```
prefix 2001:db8:1:1::/64 {
        IF (vrrp_master)
                THEN
                        preferred_lifetime = 604800
                ELSE
                        preferred_lifetime = 0
}
```

R2 is VRRP backup by default. Router advertsement on R2 interface facing H1 needs to have the following policy applied:

```
prefix 2001:db8:2:1::/64 {
        IF(vrrp_master)
                THEN
                        preferred_lifetime = 604800
                ELSE
                        preferred_lifetime = 0
}
```

If VRRP is not used or interface status tracking is not used for mastership switchover, then each router needs to be able to detect the uplink failure/recovery on the neighboring router, so that RAs with updated preferred lifetime values are triggered. Depending on the network setup various triggers like a route to the uplink interface subnet or a default route received from the uplink can be used. The obvious drawback of using the routing table to trigger the conditional RAs is that some additional configuration is required. For example, if a route to the prefix assigned to the ISP uplink is used as a trigger, then the conditional RA policy would have the following logic:

R1:

```
prefix 2001:db8:1:1::/64 {
        IF (ISP_A_uplink is up)
                THEN
                        preferred_lifetime = 604800
                ELSE
                        preferred_lifetime = 0
  }
```

R2:

```
prefix 2001:db8:2:1::/64 {
        IF (ISP_A_uplink_route is present)
                THEN
                        preferred_lifetime = 0
                ELSE
                        preferred_lifetime = 604800
 }
```

3.2.3. Single Router, Load Balancing Between Uplinks

Let’s look at the example topology shown in Figure 1, but with both uplinks used simultaneously. In this case R1 would send RAs containing PIOs for both prefixes, 2001:db8:1:1::/64 and 2001:db8:2:1::/64, changing the preferred lifetime based on particular uplink availability. If the interface status is used as uplink availability indicator, then the policy logic would look like the following:

```
prefix 2001:db8:1:1::/64 {
        IF (ISP_A_uplink is up)
                THEN
                        preferred_lifetime  = 604800
                ELSE
                        preferred_lifetime = 0
}
prefix 2001:db8:2:1::/64 {
        IF (ISP_B_uplink is up)
                THEN
                        preferred_lifetime  = 604800
                ELSE
                        preferred_lifetime = 0
}
```

R1 needs a forwarding policy to be applied to forward packets to the correct uplink based on the source address similar to one described in Section 3.2.1.

3.2.4. Two Router, Load Balancing Between Uplinks

In this scenario the example topology is similar to the one shown in Figure 2, but both uplinks can be used at the same time. It means that both R1 and R2 need to have the corresponding forwarding policy to forward packets based on their source addresses.

Each router would send RAs with PIO for the corresponding prefix. setting preferred_lifetime to a non-zero value when the ISP uplink is up, and deprecating the prefix by setting the preferred lifetime to 0 in case of uplink failure. The uplink recovery would trigger another

RA with non-zero preferred lifetime to make the addresses from the prefix preferred again. The example RA policy on R1 and R2 would look like:

R1:

```
prefix 2001:db8:1:1::/64 {
        IF (ISP_A_uplink is up)
                THEN
                        preferred_lifetime  = 604800
                ELSE
                        preferred_lifetime = 0
}
```

R2:

```
prefix 2001:db8:2:1::/64 {
        IF (ISP_B_uplink is up)
                THEN
                        preferred_lifetime  = 604800
                ELSE
                        preferred_lifetime = 0
}
```

3.2.5. Topologies with Dedicated Border Routers

For simplicity, all topologies above show the ISP uplinks terminated on the first-hop routers. Obviously, the proposed approach can be used in more complex topologies when dedicated devices are used for terminating ISP uplinks. In that case VRRP mastership or interface status can not be used as a trigger for conditional RAs and route presence as described above (Section 3.2.2) should be used instead.

Let’s look at the example topology shown on the Figure 3:

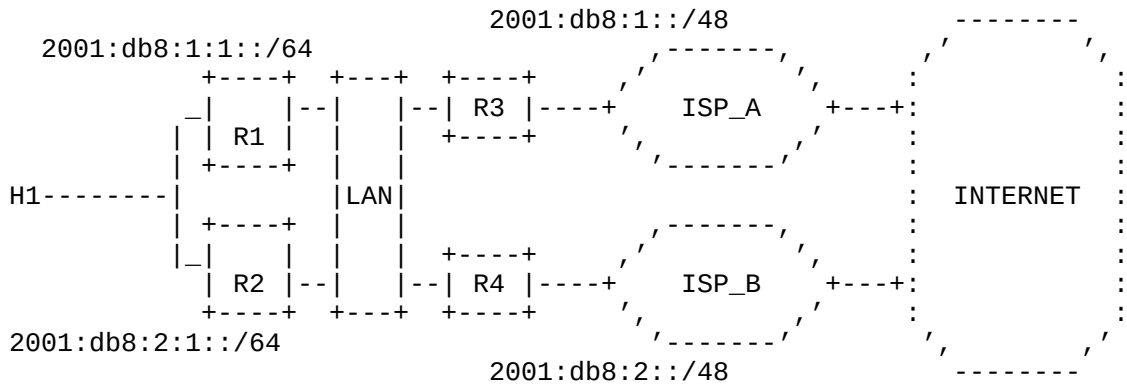

Figure 3: Dedicated Border Routers

For example, if ISP_A is a primary uplink and ISP_B is a backup one then the following policy might be used to achieve the desired behaviour (H1 is using ISP_A address space, 2001:db8:1:1::/64 while ISP_A uplink is up and only using ISP_B 2001:db8:2:1::/64 prefix if the uplink is non-operational):

R1 and R2 policy:

```
prefix 2001:db8:1:1::/64 {
        IF (ISP_A_uplink_route is present)
                THEN
                        preferred_lifetime = 604800
                ELSE
                        preferred_lifetime = 0
}

prefix 2001:db8:2:1::/64 {
        IF (ISP_A_uplink_route is present)
                THEN
                        preferred_lifetime = 0
                ELSE
                        preferred_lifetime = 604800
 }
```

For the load-balancing case the policy would look slightly different: each prefix has non-zero preferred_lifetime only if the correspoding ISP uplink route is present:

```
prefix 2001:db8:1:1::/64 {
        IF (ISP_A_uplink_route is present)
                THEN
                        preferred_lifetime = 604800
                ELSE
                        preferred_lifetime = 0
}

prefix 2001:db8:2:1::/64 {
        IF (ISP_B_uplink_route is present)
        THEN
                preferred_lifetime = 604800
        ELSE
                preferred_lifetime = 0
}
```

3.2.6. Intra-Site Communication during Simultaneous Uplinks Outage

Prefix deprecation as a result of an uplink status change might lead to a situation when all global prefixes are deprecated (all ISP uplinks are not operational for some reason). Even when there is no Internet connectivity it might be still desirable to have intra-site IPv6 connectivity (especially when the network in question is an IPv6-only one). However while an address is in a deprecated state, its use is discouraged, but not strictly forbidden ([RFC4862]). In such a scenario all IPv6 source addresses in the candidate set ([RFC6724]) are deprecated, which means that they still can be used (as there are no preferred addresses available) and the source address selection algorithm can pick up one of them, allowing the intra-site communication. However some OSes might just fall back to IPv4 if the network interface has no preferred IPv6 global addresses. Therefore if intra-site connectivity is vital during simultanious outages of multiple uplinks, administrators might consider using ULAs (Unique Local Addresses, [RFC4193]) or provisioning additional backup uplinks to protect the network from double-failure cases.

3.2.7. Uplink Damping

If an actively used uplink (primary one or one used in load balaning scenario) starts flapping, it might lead to the undesirable situation of flapping addresses on hosts (every time the uplink goes up hosts receive an RA with non-zero preferred PIO lifetime, and every time the uplink goes down all addresses in the affected prefix become deprecated). This would, undoubtedly, negatively impact the user experience, not to mention the impact of spikes of duplicate address detection traffic every time an uplink comes back up. Therefore it’s recommended that router vendors implement some form of damping policy for conditional RAs and either postpone sending an RA with non-zero

lifetime for a PIO when the uplink comes up for a number of seconds or even introduce accumulated penalties/exponential backoff algorithm for such delays. (In the case of a multiple simultaneous uplink failure scenario, when all but one uplinks are down and the last remaining is flapping it might result in all addresses being deprecated for a while after the flapping uplink recovers.)

3.2.8. Routing Packets when the Corresponding Uplink is Unavailable

Deprecating IPv6 addresses by setting the preferred lifetime to 0 discourage but not strictly forbid its usage in new communications. A deprecated address may still be used for existing connections ([RFC4862]). Therefore when an ISP uplink goes down the corresponding border router might still receive packets with source addresses belonging to that ISP address space while there is no available uplink to send those packets to.

The expected router behaviour would depend on the uplink selection mechanism. For example if some form of SADR is used then such packets will be dropped as there is no route to the destination. If policy-based routing is used to set a next-hop then the behaviour would be implementation-dependend and may vary from dropping the packets to forwarding them based on the routing table entries. It should be noted that there is no return path to the packet source (as the ISP uplink is not operational) therefore even if the outgoing packets are sent to another ISP the return traffic might not be delivered.

3.3. Solution Limitations

It should be noted that the proposed approach is not a "silver bullet" for all possible multihoming scenarios. It would work very well for networks with relatively simple topologies and straightforward routing policies. The more complex the network topology and the corresponding routing policies, the more configuration would be required to implement the solution.

Another limitation is related to the load balancing between the uplinks. In the scenario in which both uplinks are active, hosts would select the source prefix using the Default Address Selection algorithm ([RFC6724]), and therefore the load between two uplinks most likely would not be evenly distributed. (However, the proposed mechanism does allow a creative way of controlling uplinks load in software defined networks where controllers might selectively deprecate prefixes on some hosts but not others to move egress traffic between uplinks). Also the prefix selection does not take into account any other uplinks properties (such as latency etc), so egress traffic might not be sent to the nearest uplink if the

corresponding prefix is selected as a source. In general, if not all uplinks are equal and some uplinks are expected to be preferred over others, then the network administrator should ensure that prefixes from non-preferred ISP(s) are kept deprecated (so primary/backup setup is used).

3.3.1. Connections Preservation

The proposed solution is not designed to preserve connection state after an uplink failure. If all uplinks to an ISP go down, all sessions to/from addresses from that ISP address space are interrupted as there is no egress path for those packets and there is no return path from the Internet to the corresponding prefix. In this regard it is similar to IPv4 multihoming using NAT, where an uplink failure and failover to another uplink means that a public IPv4 address changes and all existing connections are interrupted.

An uplink recovery, however, does not necessarily lead to connections interruption. In the load sharing/balancing scenario an uplink recovery does not affect any existing connections at all. In the active/backup topology when the primary uplink recovers from the failure and the backup prefix is deprecated, the existing sessions (established to/from the backup ISP addresses) can be preserved if the routers are configured as described in Section 3.2.1 and send packets with the backup ISP source addresses to the backup uplink even when the primary one is operational. As a result, the primary uplink recovery makes the usage of the backup ISP addresses discouraged but still possible.

It should be noted that in IPv4 multihoming with NAT, when the egress interface is chosen without taking packet source address into account (as internal hosts usually have addresses from [RFC1918] space), sessions might not be preserved after an uplink recovery unless packet forwarding is integrated with existing NAT sessions tracking.

4. IANA Considerations

This memo asks the IANA for no new parameters.

5. Security Considerations

This memo introduces no new security considerations. It relies on Router Advertisements ([RFC4861]) and SLAAC ([RFC4862] mechanism and inherits their security properties. If an attacker is able to send a rogue RA they could deprecate IPv6 addresses on hosts or infuence source address selection processes on hosts.

The potential attack vectors are including but not limited to:

- An attacker sends a rogue RA deprecating IPv6 addresses on hosts;
- An attacker sends a rogue RA making addresses preferred while the corresponding ISP uplink is not operational;
- An attacker sends a rogue RA making addresses preferred for a backup ISP, steering traffic to undesirable (e.g. more expensive) uplink.

Therefore the network administrators SHOULD secure Router Advertisements, e.g., by deploying RA guard [RFC6105].

5.1. Privacy Considerations

This memo introduces no new privacy considerations.

6. Acknowledgements

Thanks to the following people (in alphabetical order) for their review and feedback: Mikael Abrahamsson, Lorenzo Colitti, Marcus Keane, Erik Kline, David Lamparter, Dusan Mudric, Erik Nordmark, Dave Thaler.

7. References

7.1. Normative References

[RFC1918] Rekhter, Y., Moskowitz, B., Karrenberg, D., de Groot, G., and E. Lear, "Address Allocation for Private Internets", BCP 5, RFC 1918, DOI 10.17487/RFC1918, February 1996, <https://www.rfc-editor.org/info/rfc1918>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC2827] Ferguson, P. and D. Senie, "Network Ingress Filtering: Defeating Denial of Service Attacks which employ IP Source Address Spoofing", BCP 38, RFC 2827, DOI 10.17487/RFC2827, May 2000, <https://www.rfc-editor.org/info/rfc2827>.

[RFC3022] Srisuresh, P. and K. Egevang, "Traditional IP Network Address Translator (Traditional NAT)", RFC 3022, DOI 10.17487/RFC3022, January 2001, <https://www.rfc-editor.org/info/rfc3022>.

[RFC3704] Baker, F. and P. Savola, "Ingress Filtering for Multihomed Networks", BCP 84, RFC 3704, DOI 10.17487/RFC3704, March 2004, <https://www.rfc-editor.org/info/rfc3704>.

[RFC4116] Abley, J., Lindqvist, K., Davies, E., Black, B., and V. Gill, "IPv4 Multihoming Practices and Limitations", RFC 4116, DOI 10.17487/RFC4116, July 2005, <https://www.rfc-editor.org/info/rfc4116>.

[RFC4193] Hinden, R. and B. Haberman, "Unique Local IPv6 Unicast Addresses", RFC 4193, DOI 10.17487/RFC4193, October 2005, <https://www.rfc-editor.org/info/rfc4193>.

[RFC4291] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 4291, DOI 10.17487/RFC4291, February 2006, <https://www.rfc-editor.org/info/rfc4291>.

[RFC4862] Thomson, S., Narten, T., and T. Jinmei, "IPv6 Stateless Address Autoconfiguration", RFC 4862, DOI 10.17487/RFC4862, September 2007, <https://www.rfc-editor.org/info/rfc4862>.

[RFC6105] Levy-Abegnoli, E., Van de Velde, G., Popoviciu, C., and J. Mohacsi, "IPv6 Router Advertisement Guard", RFC 6105, DOI 10.17487/RFC6105, February 2011, <https://www.rfc-editor.org/info/rfc6105>.

[RFC6724] Thaler, D., Ed., Draves, R., Matsumoto, A., and T. Chown, "Default Address Selection for Internet Protocol Version 6 (IPv6)", RFC 6724, DOI 10.17487/RFC6724, September 2012, <https://www.rfc-editor.org/info/rfc6724>.

[RFC8028] Baker, F. and B. Carpenter, "First-Hop Router Selection by Hosts in a Multi-Prefix Network", RFC 8028, DOI 10.17487/RFC8028, November 2016, <https://www.rfc-editor.org/info/rfc8028>.

[RFC8106] Jeong, J., Park, S., Beloeil, L., and S. Madanapalli, "IPv6 Router Advertisement Options for DNS Configuration", RFC 8106, DOI 10.17487/RFC8106, March 2017, <https://www.rfc-editor.org/info/rfc8106>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

7.2. Informative References

[I-D.ietf-rtgwg-dst-src-routing]
Lamparter, D. and A. Smirnov, "Destination/Source Routing", draft-ietf-rtgwg-dst-src-routing-06 (work in progress), October 2017.

[I-D.ietf-rtgwg-enterprise-pa-multihoming]
Baker, F., Bowers, C., and J. Linkova, "Enterprise Multihoming using Provider-Assigned Addresses without Network Prefix Translation: Requirements and Solution", draft-ietf-rtgwg-enterprise-pa-multihoming-07 (work in progress), June 2018.

[RFC4861] Narten, T., Nordmark, E., Simpson, W., and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, DOI 10.17487/RFC4861, September 2007, <https://www.rfc-editor.org/info/rfc4861>.

[RFC5798] Nadas, S., Ed., "Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6", RFC 5798, DOI 10.17487/RFC5798, March 2010, <https://www.rfc-editor.org/info/rfc5798>.

[RFC7084] Singh, H., Beebee, W., Donley, C., and B. Stark, "Basic Requirements for IPv6 Customer Edge Routers", RFC 7084, DOI 10.17487/RFC7084, November 2013, <https://www.rfc-editor.org/info/rfc7084>.

[RFC7788] Stenberg, M., Barth, S., and P. Pfister, "Home Networking Control Protocol", RFC 7788, DOI 10.17487/RFC7788, April 2016, <https://www.rfc-editor.org/info/rfc7788>.

Appendix A. Change Log

Initial Version: July 2017

Authors' Addresses

Jen Linkova
Google
Mountain View, California 94043
USA

Email: furry@google.com

Massimiliano Stucchi
RIPE NCC
Stationsplein, 11
Amsterdam 1012 AB
The Netherlands

Email: mstucchi@ripe.net

v6ops
Internet-Draft
Intended status: Best Current Practice
Expires: April 11, 2018

J. Palet Martinez
Consulintel, S.L.
October 8, 2017

464XLAT Deployment Guidelines in Operator Networks
draft-palet-v6ops-464xlat-deployment-00

Abstract

This document describes how 464XLAT ([RFC6877]) can be deployed in an IPv6 operator network and the issues to be considered.

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 https://datatracker.ietf.org/drafts/current/.

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

This Internet-Draft will expire on April 11, 2018.

Table of Contents

1. Introduction 2
2. DNSSEC Considerations 3
2.1. DNSSEC validator aware of DNS64 4
2.2. Stub validator 4
2.3. CLAT with DNS proxy and validator 4
2.4. ACL of clients 4
2.5. Mapping-out IPv4 addresses 4
3. Using 464XLAT with/without DNS64 5
4. DNS64 and Reverse Mapping Considerations 5
5. CLAT Translation Considerations 6
6. Summary of deployment recommendations for 464XLAT 6
7. Security Considerations 7
8. IANA Considerations 7
9. Acknowledgements 8
10. Normative References 8
Author's Address 9

1. Introduction

464XLAT ([RFC6877]) describes an architecture that provides IPv4 connectivity across a network, or part of it, when it is only natively transporting IPv6.

In order to do that, 464XLAT ([RFC6877]) relies on the combination of existing protocols:

1. The customer-side translator (CLAT) is a stateless IPv4 to IPv6 translator (NAT46) ([RFC7915]) implemented in the end-user device or CE, located at the "customer" edge of the network.

2. The provider-side translator (PLAT) is a stateful NAT64 ([RFC6146]), implemented typically at the opposite edge of the operator network, that provides access to both IPv4 and IPv6 upstreams.

3. Optionally, DNS64 ([RFC6147]), implemented as part of the PLAT allows an optimization (a single translation at the NAT64, instead of two translations - NAT46+NAT64), when the application at the end-user device supports IPv6 DNS (uses AAAA RR).

464XLAT ([RFC6877]) is a very simple approach to cope with the major NAT64+DNS64 drawback: Not working with applications or devices that use literal IPv4 addresses or non-IPv6 compliant APIs.

464XLAT ([RFC6877]) has been used initially in IPv6 cellular networks, so providing an IPv6-only access network, the end-user

device applications can access IPv4-only end-networks/applications, despite those applications or devices use literal IPv4 addresses or non-IPv6 compliant APIs.

In addition to that, in the same example of the cellular network above, if the User Equipment (UE) provides tethering, other devices behind it will be presented with a traditional NAT44, in addition to the native IPv6 support.

Furthermore, 464XLAT ([RFC6877]) can be used in non-cellular IPv6 wired (xDSL, DOCSIS, FTTH, Ethernet, ...) and wireless (WiFi) network architectures, by implementing the CLAT functionality at the CE.

The remaining sections of this document, despite of any specific examples being used, are applicable to any operator network architecture, and introduces possible issues and general deployment guidelines to be considered when deploying 464XLAT ([RFC6877]) in an IPv6 network.

2. DNSSEC Considerations

As indicated in Section 8 of [RFC6147] (DNS64, Security Considerations), because DNS64 modifies DNS answers and DNSSEC is designed to detect such modifications, DNS64 can break DNSSEC.

If a device connected to an IPv6-only WAN queries for a domain name in a signed zone, by means of a recursive name server that supports DNS64, and the result is a synthesized AAAA record, and the recursive name server is configured to perform DNSSEC validation and has a valid chain of trust to the zone in question, it will cryptographically validate the negative response from the authoritative name server. So, the recursive name server actually lie to the client device, however in most of the cases, the client will not notice it, because generally they don’t perform validation themselves as instead rely on their recursive name servers.

If the client device performs DNSSEC validation on the AAAA record, it will fail as it is a synthesized record.

Similarly, if the client querying the recursive name server is another name server configured to use it as a forwarder, and is performing DNSSEC validation, it will also fail on any synthesized AAAA record.

There are several possible solutions to avoid breaking DNSSEC:

2.1. DNSSEC validator aware of DNS64

In general, DNS servers with DNS64 function, by default, will not synthesize AAAA responses if the DNSSEC OK (DO) flag was set in the query. In this case, as only an A record is available, it means that the CLAT will take the responsibility, as in the case of literal IPv4 addresses, to keep that traffic flow end-to-end as IPv4, so DNSSEC is not broken.

2.2. Stub validator

If the DO flag is set and the client device performs DNSSEC validation, and the Checking Disabled (CD) flag is set for a query, as the DNS64 recursive server will not synthesize AAAA responses, the client could perform the DNSSEC validation with the A record and then may query the network for a NAT64 prefix ([RFC7050]) in order to synthesize the AAAA ([RFC6052]). This allows the client device to avoid using the CLAT and still use NAT64 even with DNSSEC.

Some devices/OSs may implement, instead of CLAT, a simliar function by using Bump-in-the-Host ([RFC6535]). In this case, the considerations in the above paragraphs are also applicable.

2.3. CLAT with DNS proxy and validator

If a CE includes CLAT support and also a DNS proxy, as indicated in Section 6.4 of [RFC6877], the CE could behave as a stub validator on behalf of the client devices, following the same approach described in the precedent section (Stub validator). So the DNS proxy actually lie to the client devices, which in most of the cases will not notice it unless they perform validation themselves. Again, this allow the clients devices to avoid using the CLAT and still use NAT64 with DNSSEC.

2.4. ACL of clients

In cases of dual-stack clients, stub resolvers should send the AAAA queries before the A ones. So such clients, if DNS64 is enabled, will never get A records, even for IPv4-only servers, and they may be in the path before the NAT64 and accesible by IPv4. If DNSSEC is being used for all those flows, specific addresses or prefixes can be left-out the DNS64 synthesis by means of ACLs.

2.5. Mapping-out IPv4 addresses

If there are well-known specific IPv4 addresses or prefixes using DNSSEC, they can be mapped-out of the DNS64 synthesis.

Even if this is not related to DNSSEC, this "mapping-out" feature is actually quite commonly used to ensure that [RFC1918] addresses (for example used by LAN servers) are not synthesized to AAAA.

3. Using 464XLAT with/without DNS64

In the case the client device is IPv6-only (either because the stack is IPv6-only, or because it is connected via an IPv6-only LAN) and the server is IPv4-only (either because the stack is IPv4-only, or because it is connected via an IPv4-only LAN), only NAT64 combined with DNS64 will be able to provide access among both. Because DNS64 is then required, DNSSEC validation will be only possible if the recursive name server is validating the negative response from the authoritative name server and the client is not performing validation.

However, when the client device is dual-stack and/or connected in a dual-stack LAN by means of a CLAT (or has the built-in CLAT), DNS64 is an option.

1. With DNS64: If DNS64 is used, most of the IPv4 traffic (except if using literal IPv4 addresses or non-IPv6 compliant APIs) will not use the CLAT, so will use the IPv6 path and only one translation will be done at the NAT64. This may break DNSSEC, unless measures as described in the precedent section are taken.

2. Without DNS64: If DNS64 is not used, all the IPv4 traffic will make use of the CLAT, so two translations are required (NAT46 at the CLAT and NAT64 at the PLAT), which adds some overhead in terms of the extra NAT46 translation, however avoids the AAAA synthesis and consequently will never break DNSSEC.

When clients in an operator network use DNS from other networks, for example manually configured by users, they may support or not DNS64, so the considerations in this section will apply as well.

4. DNS64 and Reverse Mapping Considerations

When a client device, using a name server configured to perform DNS64, tries to reverse-map a synthesized IPv6 address, the name server responds with a CNAME record pointing the domain name used to reverse-map the synthesized IPv6 address (the one under ip6.arpa), to the domain name corresponding to the embedded IPv4 address (under in-addr.arpa).

This is the expected behaviour, so no issues to be considered regarding DNS reverse mapping.

5. CLAT Translation Considerations

As described in Section 6.3 of [RFC6877] (IPv6 Prefix Handling), if the CLAT can be configured with a dedicated /64 prefix for the NAT64 translation, then it will be possible to do a more efficient stateless translation.

However, if this dedicated prefix is not available, the CLAT will need to do a stateful translation, for example performing stateful NAT44 for all the IPv4 LAN packets, so they appear as coming from a single IPv4 address, and then in turn, stateless translated to a single IPv6 address.

The obvious recommended setup, in order to maximize the CLAT performance, is to configure the dedicated translation prefix. This can be easily achieved automatically, if the CE or end-user device is able to obtain a shorter prefix by means of DHCPv6-PD ([RFC3633]), so the CE can use a /64 for that.

The above recommendation is often not posible for cellular networks, when connecting UEs (some broadband cellular use DHCPv6-PD ([RFC3633]), but smartphones, in general, not), as they provide a single /64 for each PDP context and use /64 prefix sharing ([RFC6877]). So in this case, the UEs typically have a build-in CLAT client, which is doing a stateful NAT44 before the stateless NAT46.

6. Summary of deployment recommendations for 464XLAT

As indicated in the introduction of this document, operators willing to deploy 464XLAT ([RFC6877]), MUST to support, at least, the provider-side translator (PLAT).

In the case it is a non-cellular network and the operator is providing the CEs to the customers, or suggesting them some specific models, they MUST support the customer-side translator (CLAT).

If the operator offers DNS services, in order to increase performance by reducing the double translation for all the IPv4 traffic, and avoid breaking DNSSEC, they MAY support DNS64. In this case, if the DNS service is offering DNSSEC validation, then it MUST be in such way that it is aware of the DNS64. This is considered de simpler and safer approach, and MAY be combined as well with the other possible solutions described in this document:

- Devices running CLAT SHOULD follow the indications in the "Stub validator" section recommendation. However, most of the time, this is out of the control of the operator.

- o CEs SHOULD include a DNS proxy and validador. This is relevant if the operator is providing the CE or suggesting it to customers.

- o ACL of clients and Mapping-out IPv4 addresses MAY be considered by each operator, depending on their own infrastructure.

The ideal configuration for CEs supporting CLAT, is that they support DHCPv6-PD ([RFC3633]) and internally reserve one /64 for the stateless NAT46 translation. The operator MUST ensure that the customers get allocated prefixes shorter than /64 in order to support this optimization. One way or the other, this is not impacting the performance of the operator network.

As indicated in Section 7 of [RFC6877] (Deployment Considerations), operators MAY follow those suggestions in order to take advantage of traffic engineering.

In the case of cellular networks, the considerations regarding DNSSEC may appear as out-of-scope, because UEs OSs, commonly don’t support DNSSEC, however applications running on them may do, or it may be an OS "built-in" support in the future. Moreover, if those devices offer tethering, other client devices may be doing the validation, hence the relevance of a proper DNSSEC support by the operator network.

Furthermore, cellular networks supporting 464XLAT ([RFC6877]) and "Discovery of the IPv6 Prefix Used for IPv6 Address Synthesis" ([RFC7050]), allow a progressive IPv6 deployment, with a single APN supporting all types of PDP context (IPv4, IPv6, IPv4v6), in such way that the network is able to automatically serve all the possible combinations of UEs.

Finally, if the operator choose to secure the NAT64 prefix, it MUST follow the advise indicated in Section 3.1.1. of [RFC7050] (Validation of Discovered Pref64::/n).

7. Security Considerations

This document does not have any new specific security considerations.

8. IANA Considerations

This document does not have any new specific IANA considerations.

9. Acknowledgements

The author would like to acknowledge the inputs of TBD ... Christian Huitema inspired working in this document by suggesting that DNS64 should never be used, during a discussion regarding the deployment of CLAT in the IETF network.

10. Normative References

[RFC1918] Rekhter, Y., Moskowitz, B., Karrenberg, D., de Groot, G., and E. Lear, "Address Allocation for Private Internets", BCP 5, RFC 1918, DOI 10.17487/RFC1918, February 1996, <https://www.rfc-editor.org/info/rfc1918>.

[RFC3633] Troan, O. and R. Droms, "IPv6 Prefix Options for Dynamic Host Configuration Protocol (DHCP) version 6", RFC 3633, DOI 10.17487/RFC3633, December 2003, <https://www.rfc-editor.org/info/rfc3633>.

[RFC6052] Bao, C., Huitema, C., Bagnulo, M., Boucadair, M., and X. Li, "IPv6 Addressing of IPv4/IPv6 Translators", RFC 6052, DOI 10.17487/RFC6052, October 2010, <https://www.rfc-editor.org/info/rfc6052>.

[RFC6146] Bagnulo, M., Matthews, P., and I. van Beijnum, "Stateful NAT64: Network Address and Protocol Translation from IPv6 Clients to IPv4 Servers", RFC 6146, DOI 10.17487/RFC6146, April 2011, <https://www.rfc-editor.org/info/rfc6146>.

[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, DOI 10.17487/RFC6147, April 2011, <https://www.rfc-editor.org/info/rfc6147>.

[RFC6535] Huang, B., Deng, H., and T. Savolainen, "Dual-Stack Hosts Using "Bump-in-the-Host" (BIH)", RFC 6535, DOI 10.17487/RFC6535, February 2012, <https://www.rfc-editor.org/info/rfc6535>.

[RFC6877] Mawatari, M., Kawashima, M., and C. Byrne, "464XLAT: Combination of Stateful and Stateless Translation", RFC 6877, DOI 10.17487/RFC6877, April 2013, <https://www.rfc-editor.org/info/rfc6877>.

[RFC7050] Savolainen, T., Korhonen, J., and D. Wing, "Discovery of the IPv6 Prefix Used for IPv6 Address Synthesis", RFC 7050, DOI 10.17487/RFC7050, November 2013, <https://www.rfc-editor.org/info/rfc7050>.

[RFC7278] Byrne, C., Drown, D., and A. Vizdal, "Extending an IPv6 /64 Prefix from a Third Generation Partnership Project (3GPP) Mobile Interface to a LAN Link", RFC 7278, DOI 10.17487/RFC7278, June 2014, <https://www.rfc-editor.org/info/rfc7278>.

[RFC7915] Bao, C., Li, X., Baker, F., Anderson, T., and F. Gont, "IP/ICMP Translation Algorithm", RFC 7915, DOI 10.17487/RFC7915, June 2016, <https://www.rfc-editor.org/info/rfc7915>.

Author’s Address

Jordi Palet Martinez
Consulintel, S.L.
Molino de la Navata, 75
La Navata - Galapagar, Madrid 28420
Spain

Email: jordi.palet@consulintel.es
URI: http://www.consulintel.es/

v6ops
Internet-Draft
Intended status: Standards Track
Expires: May 1, 2018

J. Palet Martinez
The IPv6 Company
C. Martinez
LACNIC

October 28, 2017

Reporting of Happy Eyeballs Failures
draft-palet-v6ops-he-reporting-00

Abstract

This document describes an extension to Happy Eyeballs in order to report IPv6 failures that force the fall-back to IPv4 and consequently, facilitate the troubleshooting of IPv6 networks.

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 https://datatracker.ietf.org/drafts/current/.

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

This Internet-Draft will expire on May 1, 2018.

Copyright Notice

Table of Contents

1. Introduction 2
2. Using Syslog 2
3. Discovery of the syslog collector NSP 3
4. HE behaviour on failure detection 3
5. Privacy Considerations 3
6. Security Considerations 5
7. IANA Considerations 5
8. Acknowledgements 5
9. Normative References 5
Authors' Addresses 5

1. Introduction

Happy Eyeballs ([RFC6555]) provides a way for improving user-visible delay when IPv6 connectivity is performing worse than the IPv4 one.

However, this hides the possible IPv6 connectivity issues to the operator because users don't notice anything broken, so they aren't reporting it to their providers.

The goal of this document is to specify an extension of HE, in order to use existing protocols for providing a reporting to the operator, which can be used to setup alarms and trigger further investigation so to improve network reliability, facilitating the detection of failures as soon as they appear, without the need of external monitoring.

2. Using Syslog

In order to simplify the reporting of the HE failures, syslog ([RFC5424]) over UDP ([RFC5426]), MUST be used, by means of the default port (514) with IPv6-only.

The intend is to make this reporting very simple, so no choice of alternative ports or transport protocols is offered.

Operators willing to use this reporting MUST configure at least one syslog collector at the IPv6 prefix formed as:

Network-Specific Prefix::192.88.99.1

The Network-Specific Prefix (NSP) MUST be chosen by the operator from its RIR allocated IPv6 addressing space.

Additional collectors can be made available by using anycast at the NSP + 192.88.99.0/24 prefix

3. Discovery of the syslog collector NSP

The same mechanism described by RFC7050 ([RFC7050]) should be used to define the address of the syslog collector(s).

Because the collectors will be using an IPv6 address with the 32 low order bits from the reserved range 192.88.99.0/24, this will not be in conflict with any public addresses used in Internet, so this mechanism is compatible with the expected usage of the NSP for NAT64.

4. HE behaviour on failure detection

This section will specify the exact behaviour of HE in order to initiate the reporting and the specific format/parameters of the HE failure message to be sent to the syslog collector.

A preliminary consideration is to include, in addition to the syslog required parameters, the timeouts detected, the failed destination address and the source prefix from where the destination has failed.

TBD.

5. Privacy Considerations

The goal is to provide the operator information about the failures detected by HE, without requiring specific users traffic information. Towards this, it will be sufficient to provide to the syslog collector details about the failed destination address and source prefix. So privacy issues regarding identification of a specific device or users are avoided.

Nowadays, operators already log this information in order to comply with lawful interception regulations, and in general, data protection regulations allow this logging when technically required. Data protection regulations explicitly say that the data can’t be disclosed, and there is no need to do so.

In general, vendors also collect telemetry data from devices, in order to improve OSs and in some situations, there are regulations that enforce offering the user to enable/disable that feature. So we could consider offering the same feature for this mechanism.

When the mechanism described in this document detects a failure, the operator will need to find if the problem is related to:

- A specific user (inside the customer local networks, or even at their WAN router).

- o A group of users (e.g., one or several part of the access or distribution networks).
- o The entire operator network (e.g., core network or transit router/s).
- o The destination network.
- o Somewhere else in the path to the destination (e.g., transit providers).

Those cases, in terms of privacy considerations, will fall into one of the following categories:

a. Failure cause is internal to a specific customer (LANs or router/s): The operator may decide, depending on their country regulations and services offered to that customer, to inform the customer (and decide what information is provided), or ignore the failure and include it in a "while list" (i.e., list of "don't care" failures), so the monitoring system doesn't keep providing alerts on it.

b. Failure cause is due to the operator network: The operator will need to find the cause and fix the failure, without disclosing any personal data.

c. Failure cause is due to third parties: The operator don't need to disclose any specific user source address/prefix, because in this case, the shorter prefix (typically the RIR allocated prefix or part of it, when is being announced split among different BGP peers), from which the failure has been verified.

In the most extreme case, a more restrictive usage of this procedure, not involving logging any user source address/prefix, will be to log only the failed destination address. In a big percentage of the cases, it will be enough for the operator to detect the failure, as experience shows that HE fall-back occurs mainly because path or destination misconfiguration or issues. So, the ISP could replicate the failure from any other source address in its network to the same failed destination. If we take this approach, failures internal to a specific customer, could not be reported by the operator to the customer (as there is no source data logging), and together with partial failures of the operator network will require extra work from operator's staff to research the cause of the failure (i.e., it is in my network, part of it, a specific customer or external).

So, there is no distinction between the privacy issues from this protocol compared to regular network operation, abuse reporting, etc.

6. Security Considerations

This document does not have any specific security considerations.

7. IANA Considerations

IANA is requested to reserve 192.88.99.0/24 for this RFC, which was previously released by ([RFC7526]).

8. Acknowledgements

The author would like to acknowledge the inputs of TBD ...

9. Normative References

[RFC5424] Gerhards, R., "The Syslog Protocol", RFC 5424, DOI 10.17487/RFC5424, March 2009, <https://www.rfc-editor.org/info/rfc5424>.

[RFC5426] Okmianski, A., "Transmission of Syslog Messages over UDP", RFC 5426, DOI 10.17487/RFC5426, March 2009, <https://www.rfc-editor.org/info/rfc5426>.

[RFC6555] Wing, D. and A. Yourtchenko, "Happy Eyeballs: Success with Dual-Stack Hosts", RFC 6555, DOI 10.17487/RFC6555, April 2012, <https://www.rfc-editor.org/info/rfc6555>.

[RFC7050] Savolainen, T., Korhonen, J., and D. Wing, "Discovery of the IPv6 Prefix Used for IPv6 Address Synthesis", RFC 7050, DOI 10.17487/RFC7050, November 2013, <https://www.rfc-editor.org/info/rfc7050>.

[RFC7526] Troan, O. and B. Carpenter, Ed., "Deprecating the Anycast Prefix for 6to4 Relay Routers", BCP 196, RFC 7526, DOI 10.17487/RFC7526, May 2015, <https://www.rfc-editor.org/info/rfc7526>.

Authors' Addresses

Jordi Palet Martinez
The IPv6 Company
Molino de la Navata, 75
La Navata - Galapagar, Madrid 28420
Spain

Email: jordi.palet@theipv6company.com
URI: http://www.theipv6company.com/

Carlos Martinez
LACNIC
Rambla Republica de Mexico, 6125
Montevideo 11400
Uruguay

Email: carlos@lacnic.net
URI: http://www.lacnic.net/

v6ops
Internet-Draft
Intended status: Informational
Expires: January 4, 2020

J. Palet Martinez
The IPv6 Company

July 3, 2019

IPv6-only Terminology Definition
draft-palet-v6ops-ipv6-only-04

Abstract

This document defines the terminology regarding the usage of expressions such as "IPv6-only", in order to avoid confusions when using them in IETF and other documents. The goal is that the reference to "IPv6-only" describes the actual native functionality being used, not the actual protocol support.

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 https://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 4, 2020.

Copyright Notice

Table of Contents

1. Introduction 2
2. Context 2
3. Definition of IPv6-only 3
4. Dual-stack 3
5. Native dual-stack 3
6. IPv6-only network 4
7. IPv6-only WAN/access 4
8. IPv6-only LAN 4
9. IPv6-only host/router 4
10. Transitional IPv6 host/router 4
11. Other cases 4
12. Security Considerations 5
13. IANA Considerations 5
14. Acknowledgements 5
Author's Address 5

1. Introduction

Due to the nature of the Internet and the different types of users, parts of a network, providers, flows, etc., there is not a single and easy way to categorically say something such as "IPv6-only".

The goal of this document is to depict this situation and agree in a common language to be used for IETF and other documents, in order to facilitate ourselves and future readers, the correct understanding of what we are talking about.

Note that all the references in this document are regarding the actual usage of IPv4/IPv6, not the support of those protocols by nodes. For example, a device or access network may support both IPv4 and IPv6, however actually is only "natively" forwarding IPv6, because the link used for that communication is only natively configured for IPv6. IPv4 may be used as well, but it is being encapsulated or translated by means of IPv6. So from this perspective, this device is attached to an IPv6-only link.

2. Context

The transition from IPv4 to IPv6 is not something that can be done, in the large majority of the cases, overnight and in a single step. Consequently, in general, we are unable to talk about a whole network having a "single and uniform" status regarding the IPv6 support, at least not in the early deployment stages of an operator network.

Even if possible, it is not frequent to deploy new IPv6 networks which have no IPv4 connectivity at all, because at the current phase

of the universal goal of the IPv6 deployment, almost every network still need to provide some kind of "access" to IPv4 sites. It is not feasible for most of the operators to tell their customers "I can provide you IPv6 service, but you will not be able to access all Internet contents and apps, because some of them still don't support IPv6, so you will miss every content that it is IPv4-only".

Some networks may have IPv6-only support for specific purposes. For example, a DOCSIS provider may have decided that is worth the effort to get rid of IPv4 for the management network of the cable-modems. Or a network that provides connectivity only to IoT devices, may be IPv6-only.

However, the "end-networks", in general, need to continue supporting IPv4, as there are many devices or apps, in both corporate and end-user networks (smartTV, IP cameras, etc.), which are IPv4-only and it is not always feasible to update or replace them.

In IPv6-only access networks, IPv4 support may be provided by mechanisms that allow "IPv4-as-a-service" (IPv4aaS, for example by means of encapsulation and/or translation on top of IPv6).

3. Definition of IPv6-only

Consequently, considering the context described in the section above, if we want to be precise and avoid confusing others, we can not use the terminology "IPv6-only" in a generic way, and we need to define what part of the network we are referring to.

From that perspective, we define the "IPv6-only" status in a given part(s) of a network, depending on if there is actual native forwarding of IPv4, so IPv4 is not configured neither managed.

4. Dual-stack

This can be applied to a host, router, link, network (part), etc. It means that both, IPv4 and IPv6 are reachable, without specifying how.

5. Native dual-stack

This can be applied to a host, router, link, network (part), etc. It means that both, IPv4 and IPv6 are configured/used natively (without the need of transition mechanisms).

6. IPv6-only network

IPv6-only can be used only if, a complete network, end-to-end, is actually not natively forwarding IPv4, which will mean that no-IPv4 addresses are configured, neither used for management, neither the network is providing transition/translation support, neither there is IPv4 transit/peering.

This is the end of the road of the IPv4-to-IPv6 transition, however we aren't there yet, in general at the time of writing this document, unless we are referring to special or disconnected (from IPv4) networks.

7. IPv6-only WAN/access

IPv6-only WAN or access can be used only if the WAN or access network isn't actually natively forwarding IPv4.

8. IPv6-only LAN

IPv6-only LAN(s) can be used only if the LAN(s), isn't actually natively forwarding IPv4.

9. IPv6-only host/router

IPv6-only host/router can be used only if the host/router, isn't actually using/forwarding IPv4, so IPv4 is unconfigured and/or disabled in the external facing interfaces.

Internal interfaces, such as loopback, can still be using IPv4 (internally).

10. Transitional IPv6 host/router

Transitional IPv6 host/router is a dual-stack host/router with IPv6-only WAN where IPv4 service support is provided by means of transition mechanism, IPv4aaS (IPv4-as-a-service).

11. Other cases

Similar other cases or parts of the network can be considered as IPv6-only if there is no actual native forwarding of IPv4 and in that case, after "IPv6-only" some word/short text pointing to the specific case or part of the network needs to be used. For instance, we could talk about "IPv6-only core" if a core network is only natively forwarding IPv6.

12. Security Considerations

This document does not have any specific security considerations.

13. IANA Considerations

This document does not have any IANA considerations.

14. Acknowledgements

The author would like to acknowledge the inputs from Tim Chown, Noah Maina, Lee Howard, Azael Fernandez Alcantara and Marcos Sanz Grosson.

Author's Address

Jordi Palet Martinez
The IPv6 Company
Molino de la Navata, 75
La Navata - Galapagar, Madrid 28420
Spain

Email: jordi.palet@theipv6company.com
URI: http://www.theipv6company.com/

v6ops
Internet-Draft
Intended status: Best Current Practice
Expires: May 1, 2018

J. Palet Martinez
The IPv6 Company

October 28, 2017

Using /64 from Customer Prefix for the Inter-Router Link
draft-palet-v6ops-p2p-from-customer-prefix-01

Abstract

This document describes the usage of a /64 from the customer prefix for numbering IPv6 point-to-point links in non-broadcast layer 2 media.

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 https://datatracker.ietf.org/drafts/current/.

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

This Internet-Draft will expire on May 1, 2018.

Table of Contents

1. Introduction 2
2. Rational for using /64 3
3. Numbering Interfaces 3
4. Routing Aggregation of the Point-to-Point Links 3
5. DHCPv6 Considerations 5
6. Router Considerations 5
7. Security Considerations 5
8. IANA Considerations 5
9. Acknowledgements 5
10. Normative References 6
Author's Address 6

1. Introduction

There are different alternatives for numbering IPv6 point-to-point links, and from an operational perspective, they may have different advantages or disadvantages that need to be taken in consideration under the scope of each specific network architecture design.

[RFC6164] describes using /127 prefixes for inter-router point-to-point links, using two different address pools, one for numbering the point-to-point links and another one for delegating the prefixes at the end of the point-to-point link. However this doesn't exclude other choices.

This document describes an alternative the approach, using a /64 from the customer prefix, which ensure compliance with standards, and consequently facilitate interoperability, avoids possible future issues if more addresses are needed (e.g., managed bridges) and simplifies the addressing plan.

The use of /64 also facilitates an easier way for routing the shorter aggregated prefix into the point-to-point link. Consequently it simplifies the "view" of a more unified addressing plan, providing an easier path for following up any issue when operating IPv6 networks.

The proposed approach is suitable for those point-to-point links connecting ISP to Customers and enterprise networks, but not limited to those cases, and in fact, is being used by a relevant number of networks worldwide, in several different scenarios.

This mechanism would not work in broadcast layer two media that rely on ND (as it will try ND for all the addresses within the shorter prefix being delegated thru the point-to-point link).

2. Rational for using /64

The IPv6 Addressing Architecture ([RFC4291]) specifies that all the Interface Identifiers for all the unicast addresses (except for 000/3) are required to be 64 bits long and to be constructed in Modified EUI-64 format.

The same document also mandates the usage of the predefined subnet-router anycast address, which has cleared to zero all the bits that do not form the subnet prefix.

[RFC6164] describes possible issues when using /64 for the point-to-point linkes, however, it also states that they can be mitigated by other means, and indeed, considering the publication date of that document, those issues should not be any longer considered. The fact is that many operators wordwide, today use /64 without any concerns, as vendors have taken the necessary code updates.

Consequently, we shall conclude that /64 it is a valid approach to use /64 prefixes for the point-to-point links.

3. Numbering Interfaces

Often, in point-to-point links, hardware tokens are not available, or there is the need to keep certain bits (u, g) cleared, so the links can be manually numbered sequentially with most of the bits cleared to zero. This numbering makes as well easier to remember the interfaces, which typically will become numbered as 1 (with 63 leading zero bits) for the provider side and 2 (with 63 leading zero bits) for the customer side.

Using interface identifiers as 1 and 2 is not only a very simple approach, but also a very common practice. Other different choices can as well be used as required in each case.

On the other hand, using the EUI-64, makes it more difficult to remember and handle the interfaces, but provides an additional degree of protection against port (actually address) scanning as described at [RFC7707].

4. Routing Aggregation of the Point-to-Point Links

Following this approach and assuming that a shorter prefix is typically delegated to a customer, for example a /48, it is possible to simplify the routing aggregation of the point-to-point links. Towards this, the point-to-point link may be numbered using the first /64 of the /48 delegated to the customer.

Let’s see a practical example:

- A service provider uses the prefix 2001:db8::/32 and is using 2001:db8:aaaa::/48 for a given customer.
- Instead of allocating the point-to-point link from a different addressing pool, it may use 2001:db8:aaaa::/64 (which is the first /64 subnet from the 2001:db8:aaaa::/48) to number the link.
- This means that, in the case the non-EUI-64 approach is used, the point-to-point link may be numbered as 2001:db8:aaaa::1/64 for the provider side and 2001:db8:aaaa::2/64 for the customer side.
- Note that using the first /64 and interface identifiers 1 and 2 is a very common practice. However other values may be chosen according to each case specific needs.

In this way, as the same address pool is being used for both, the prefix and the point-to-point link, one of the advantages of this approach is to make very easy the recognition of the point-to-point link that belongs to a given customer prefix, or in the other way around, the recognition of the prefix that is linked by a given point-to-point link.

For example, making a trace-route to debug any issue to a given address in the provider network, will show a straight view, and it becomes unnecessary one extra step to check a database that correlate an address pool for the point-to-point links and the customer prefixes, as all they are the same.

Moreover, it is possible to use the shorter prefix as the provider side numbering for the point-to-point link and keep the /64 for the customer side. In our example, it will become:

- Point-to-point link at provider side: 2001:db8:aaaa::1/48
- Point-to-point link at customer side: 2001:db8:aaaa::2/64

This provides one additional advantage as in some platforms the configuration may be easier saving one step for the route of the delegated prefix (no need for two routes to be configured, one for the delegated prefix, one for the point-to-point link). It is possible because the longest-prefix-match rule.

The behavior of this type of configuration has been successfully deployed in different operator and enterprise networks, using commonly available implementations with different routing protocols, including RIP, BGP, IS-IS, OSPF, along static routing, and no

failures or interoperability issues have been reported.

5. DHCPv6 Considerations

As stated in [RFC3633], "the requesting router MUST NOT assign any delegated prefixes or subnets from the delegated prefix(es) to the link through which is received the DHCP message from the delegating router", however the approach described in this document is still useful in other DHCPv6 scenarios or non-DHCPv6 scenarios.

Furthermore, [RFC3633] was updated by Prefix Exclude Option for DHCPv6-based Prefix Delegation ([RFC6603]), precisely to define a new DHCPv6 option, which covers the case described by this document.

Moreover, [RFC3769] has no explicit requirement that avoids the approach described in this document.

6. Router Considerations

This approach is being used by operators in both, residential/SOHO and enterprise networks, so the routers at the customer end for those networks MUST support [RFC6603] if DHCPv6-PD is used.

In the case of Customer Edge Routers there is a specific requirement ([RFC7084]) WPD-8 (Prefix delegation Requirements), marked as SHOULD for [RFC6603]. However, in an scenario where the approach described in this document is followed, together with DHCPv6-PD, the CE Router MUST support [RFC6603].

7. Security Considerations

This document does not have any new specific security considerations.

8. IANA Considerations

This document does not have any new specific IANA considerations.

9. Acknowledgements

The author would like to acknowledge the inputs of Mikael Abrahamsson ... Acknowledgement to co-authors, Cesar Olvera and Miguel Angel Diaz, of a previous related document (draft-palet-v6ops-point2point, 2006), as well as inputs for that version from Alain Durand, Chip Popoviciu, Daniel Roesen, Fred Baker, Gert Doering, Olaf Bonness, Ole Troan, Pekka Savola and Vincent Jardin, are also granted.

10. Normative References

[RFC3633] Troan, O. and R. Droms, "IPv6 Prefix Options for Dynamic Host Configuration Protocol (DHCP) version 6", RFC 3633, DOI 10.17487/RFC3633, December 2003, <https://www.rfc-editor.org/info/rfc3633>.

[RFC3769] Miyakawa, S. and R. Droms, "Requirements for IPv6 Prefix Delegation", RFC 3769, DOI 10.17487/RFC3769, June 2004, <https://www.rfc-editor.org/info/rfc3769>.

[RFC4291] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 4291, DOI 10.17487/RFC4291, February 2006, <https://www.rfc-editor.org/info/rfc4291>.

[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, DOI 10.17487/RFC6164, April 2011, <https://www.rfc-editor.org/info/rfc6164>.

[RFC6603] Korhonen, J., Ed., Savolainen, T., Krishnan, S., and O. Troan, "Prefix Exclude Option for DHCPv6-based Prefix Delegation", RFC 6603, DOI 10.17487/RFC6603, May 2012, <https://www.rfc-editor.org/info/rfc6603>.

[RFC7084] Singh, H., Beebee, W., Donley, C., and B. Stark, "Basic Requirements for IPv6 Customer Edge Routers", RFC 7084, DOI 10.17487/RFC7084, November 2013, <https://www.rfc-editor.org/info/rfc7084>.

[RFC7707] Gont, F. and T. Chown, "Network Reconnaissance in IPv6 Networks", RFC 7707, DOI 10.17487/RFC7707, March 2016, <https://www.rfc-editor.org/info/rfc7707>.

Author’s Address

Jordi Palet Martinez
The IPv6 Company
Molino de la Navata, 75
La Navata - Galapagar, Madrid 28420
Spain

Email: jordi.palet@theipv6company.com
URI: http://www.theipv6company.com/

IPv6 Operations (v6ops)
Internet-Draft
Intended status: Informational
Expires: April 20, 2018

J. Palet Martinez
Consulintel, S.L.

October 17, 2017

Transition Requirements for IPv6 Customer Edge Routers
draft-palet-v6ops-rfc7084-bis-transition-01

Abstract

This document specifies the transition requirements for an IPv6 Customer Edge (CE) router. Specifically, this document extends the "Basic Requirements for IPv6-only Customer Edge Routers" ([RFC7084]) in order to allow the provisioning of IPv6 transition services for the hosts attached to it. The document covers several transition technologies, either for delivering IPv6 in IPv4-only access networks and specially for delivering IPv4 "as-a-service" as required in a world where IPv4 addresses are no longer available, so hosts in the customer LANs with IPv4-only or IPv6-only applications or devices, requiring to communicate with IPv4-only services at the Internet, are able to do so.

Table of Contents

1. Introduction 2
1.1. Requirements Language 3
2. Terminology 3
3. Usage Scenarios 4
4. Architecture 5
4.1. Current IPv4 End-User Network Architecture 5
4.2. IPv6 End-User Network Architecture 6
5. Requirements 8
5.1. General Requirements 8
5.2. LAN-Side Configuration 8
5.3. Transition Technologies Support 8
5.3.1. IPv4 Service Continuity in Customer LANs 8
5.3.1.1. 464XLAT 8
5.3.1.2. Dual-Stack Lite (DS-Lite) 9
5.3.1.3. Lightweight 4over6 (lw4o6) 10
5.3.1.4. MAP-E 10
5.3.1.5. MAP-T 11
5.3.2. Support of IPv6 in IPv4-only WAN access 11
5.3.2.1. 6in4 11
5.3.2.2. 6rd 12
5.4. IPv4 Multicast Support 14
5.5. Security Considerations 14
6. Acknowledgements 14
7. ANNEX A: Code Considerations 14
8. References 15
8.1. Normative References 15
8.2. Informative References 17
Author's Address 17

1. Introduction

This document defines basic IPv6 transition features for a residential or small-office router, referred to as an "IPv6 Transition CE router", in order to establish an industry baseline for dual-stack and transition features to be implemented on such a router.

These routers are based on "Basic Requirements for IPv6-only Customer Edge Routers" ([RFC7084]), so the scope of this documents is to

include also IPv4 support, at least in the LAN side.

This document covers the IP transition technologies required when ISPs have already and IPv4-only access network that they can't turn to dual-stack or IPv6-only, as well as the situation in a world where IPv4 addresses are no longer available, so the service providers need to provision IPv6-only WAN access, while at the same time ensuring that IPv4-only or IPv6-only devices or applications in the customer LANs can still reach IPv4-only devices or applications in Internet, which still don't have IPv6 support.

This document specifies the transition mechanisms to be supported by an IPv6 transition CE router, relevant provisioning or configuration information differences from [RFC7084]. Automatic provisioning of more complex topology than a single router with multiple LAN interfaces may be handled by means of HNCP ([RFC7788]), which is out of the scope of this document.

1.1. Requirements Language

Take careful note: Unlike other IETF documents, the key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are not used as described in RFC 2119 [RFC2119]. This document uses these keywords not strictly for the purpose of interoperability, but rather for the purpose of establishing industry-common baseline functionality. As such, the document points to several other specifications (preferable in RFC or stable form) to provide additional guidance to implementers regarding any protocol implementation required to produce a successful IPv6 Transition CE router that interoperates successfully with a particular subset of currently deploying and planned common IPv6 access networks.

2. Terminology

This document uses the same terminology as in [RFC7084], with two minor clarifications.

The term "IPv6 transition Customer Edge Router" is defined as an "IPv6 Customer Edge Router" that provides transition support to allow IPv4-IPv6 coexistence either in the WAN, the LAN or both.

The "WAN Interface" term used across this document, means that can also support link technologies based in Internet-layer (or higher-layers) "tunnels", such as tunnels IPv4-in-IPv6 or IPv6-in-IPv4.

3. Usage Scenarios

The IPv6 Transition CE router described in this document is expected to be used typically, in any of the following scenarios:

1. Residential/household users. Common usage is any kind of Internet access (web, email, streaming, online gaming, etc.).

2. Residential with Small Office/Home Office (SOHO). Same usage as for the first scenario.

3. Small Office/Home Office (SOHO). Same usage as for the first scenario.

4. Small and Medium Enterprise (SME). Same usage as for the first scenario.

5. Residential/household with advanced requirements. Same basic usage as for the first scenario, however there may be requirements for exporting services to the WAN (IP cameras, web, DNS, email, VPN, etc.).

6. Small and Medium Enterprise (SME) with advanced requirements. Same basic usage as for the first scenario, however there may be requirements for exporting services to the WAN (IP cameras, web, DNS, email, VPN, etc.).

The above list is not intended to be comprehensive of all the possible usage scenarios, just the main ones. In fact, combinations of the above usages are also possible, for example a residential with SOHO and advanced requirements.

The mechanisms for exporting IPv6 services are commonly "naturally" available in any IPv6 router, as when using GUA, unless they are blocked by firewall rules, which may require some manual configuration by means of a GUI and/or CLI.

However, in the case of IPv4, because the usage of private addresses and NAT, it typically requires some degree of manual configuration such as setting up a DMZ, virtual servers, or port/protocol forwarding. In general, CE routers already provide GUI and/or CLI to manually configure them, or the possibility to setup the CE in bridge mode, so another CE behind it, takes care of that. It is out of the scope of this document the definition of any requirements for that.

The main difference for an IPv6 Transition CE router to support one or several of the above indicated scenarios, is related to the packet processing capabilities, performance, even other details such as the

number of WAN/LAN interfaces, their maximum speed, memory for keeping tables or tracking connections, etc. So, it is out of the scope of this document to classify them.

For example, an SME may have just 10 employees (micro-SME), which commonly will be considered same as a SOHO, but a small SME can have up to 50 employees, or 250 for a medium one. Depending on the IPv6 Transition CE router capabilities or even how it is being configured (for instance, using SLAAC or DHCPv6), it may support even a higher number of employees if the traffic in the LANs is low, or switched by another device(s), or the WAN bandwidth requirements are low, etc. The actual bandwidth capabilities of access with technologies such as FTTH, cable and even 3GPP/LTE, allows the support of such usages, and indeed, is a very common situation that access networks and the IPv6 Transition CE provided by the service provider are the same for SMEs and residential users.

There is also no difference in terms of who actually provides the IPv6 Transition CE router. In most of the cases is the service provider, and in fact is responsible, typically, of provisioning/ managing at least the WAN side. However, commonly the user has access to configure the LAN interfaces, firewall, DMZ, and many other aspects. In fact, in many cases, the user must supply, or at least can replace the IPv6 Transition CE router, which makes even more relevant that all the IPv6 Transition CE routers, support the same requirements defined in this document.

The IPv6 Transition CE router described in this document is not intended for usage in other scenarios such as bigger Enterprises, Data Centers, Content Providers, etc. So, even if the documented requirements meet their needs, may have additional requirements, which are out of the scope of this document.

4. Architecture

4.1. Current IPv4 End-User Network Architecture

An end-user network will likely support both IPv4 and IPv6. It is not expected that an end user will change their existing network topology with the introduction of IPv6. There are some differences in how IPv6 works and is provisioned; these differences have implications for the network architecture. A typical IPv4 end-user network consists of a "plug and play" router with NAT functionality and a single link behind it, connected to the service provider network.

A typical IPv4 NAT deployment by default blocks all incoming connections. Opening of ports is typically allowed using a Universal

Plug and Play Internet Gateway Device (UPnP IGD) [UPnP-IGD] or some other firewall control protocol.

Another consequence of using private address space in the end-user network is that it provides stable addressing; that is, it never changes even when you change service providers, and the addresses are always there even when the WAN interface is down or the customer edge router has not yet been provisioned.

Many existing routers support dynamic routing (which learns routes from other routers), and advanced end-users can build arbitrary, complex networks using manual configuration of address prefixes combined with a dynamic routing protocol.

4.2. IPv6 End-User Network Architecture

The end-user network architecture for IPv6 should provide equivalent or better capabilities and functionality than the current IPv4 architecture.

The end-user network is a stub network, in the sense that is not providing transit to other external networks. However HNCP ([RFC7788]) allows support for automatic provisioning of downstream routers. Figure 1 illustrates the model topology for the end-user network.

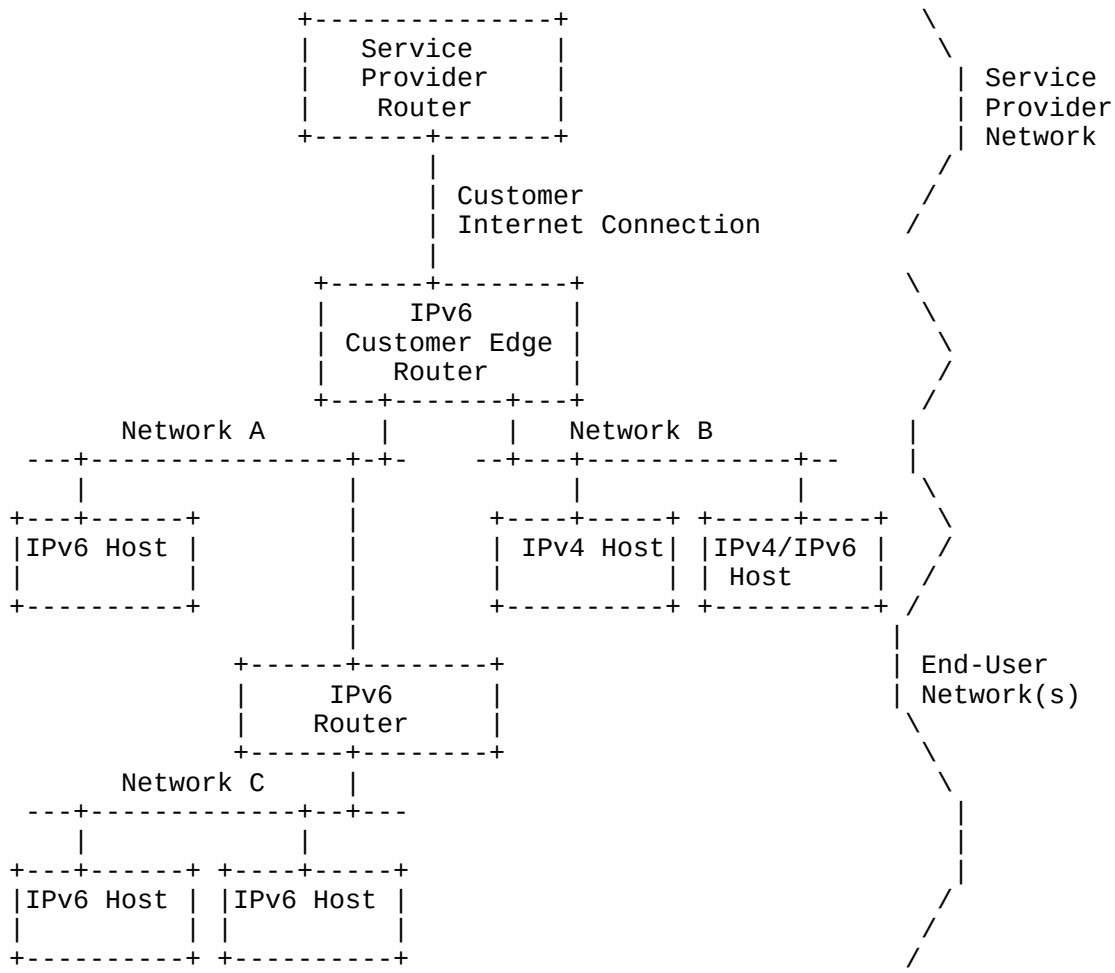

Figure 1: An Example of a Typical End-User Network

This architecture describes the:

- Basic capabilities of an IPv6 Transition CE router
- Provisioning of the WAN interface connecting to the service provider
- Provisioning of the LAN interfaces

The IPv6 Transition CE router may be manually configured in an arbitrary topology with a dynamic routing protocol or using HNCP ([RFC7788]). Automatic provisioning and configuration is described

for a single IPv6 Transition CE router only.

5. Requirements

5.1. General Requirements

The IPv6 Transition CE router must comply with the general requirements stated in [RFC7084]. Furthermore, a new general requirement is added:

G-6 The IPv6-only CE router MUST comply with [RFC7608].

5.2. LAN-Side Configuration

The IPv6 Transition CE router must comply with LAN-Side Configuration as stated in [RFC7084].

In addition, a new LAN Requirement is:

L-15 The IPv6 CE router SHOULD implement a DNS proxy as described in [RFC5625].

5.3. Transition Technologies Support

Even if the main target of this document is the support of IPv6-only WAN access, for some time, there will be a need to support IPv4-only devices and applications in the customers LANs, in one side of the picture. In the other side, some Service Providers willing to deploy IPv6, may not be able to do so in the first stage, neither as IPv6-only or dual-stack in the WAN. Consequently, transition technologies to resolve both issues should be taken in consideration.

5.3.1. IPv4 Service Continuity in Customer LANs

5.3.1.1. 464XLAT

464XLAT [RFC6877] is a technique to provide IPv4 access service to IPv6-only edge networks without encapsulation.

The IPv6 Transition CE router SHOULD support CLAT functionality. If 464XLAT is supported, it MUST be implemented according to [RFC6877]. The following CE Requirements also apply:

464XLAT requirements:

464XLAT-1: The IPv6 Transition CE router MUST perform IPv4 Network Address Translation (NAT) on IPv4 traffic translated using the CLAT, unless a dedicated /64 prefix has been

acquired using DHCPv6-PD [RFC3633].

464XLAT-2: The IPv6 Transition CE router MUST implement [RFC7050] in order to discover the PLAT-side translation IPv4 and IPv6 prefix(es)/suffix(es). In environments with PCP support, the IPv6 Transition CE SHOULD follow [RFC7225] to learn the PLAT-side translation IPv4 and IPv6 prefix(es)/suffix(es) used by an upstream PCP-controlled NAT64 device.

5.3.1.2. Dual-Stack Lite (DS-Lite)

Dual-Stack Lite [RFC6333] enables both continued support for IPv4 services and incentives for the deployment of IPv6. It also de-couples IPv6 deployment in the service provider network from the rest of the Internet, making incremental deployment easier. Dual-Stack Lite enables a broadband service provider to share IPv4 addresses among customers by combining two well-known technologies: IP in IP (IPv4-in-IPv6) and Network Address Translation (NAT). It is expected that DS-Lite traffic is forwarded over the IPv6 Transition CE router's native IPv6 WAN interface, and not encapsulated in another tunnel.

The IPv6 Transition CE router SHOULD implement DS-Lite functionality. If DS-Lite is supported, it MUST be implemented according to [RFC6333]. This document takes no position on simultaneous operation of Dual-Stack Lite and native IPv4. The following IPv6 Transition CE router requirements also apply:

DS-Lite requirements:

DSLITE-1: The IPv6 Transition CE router MUST support configuration of DS-Lite via the DS-Lite DHCPv6 option [RFC6334]. The IPv6 Transition CE router MAY use other mechanisms to configure DS-Lite parameters. Such mechanisms are outside the scope of this document.

DSLITE-2: The IPv6 Transition CE router MUST support the DHCPv6 S46 priority option described in [RFC8026].

DSLITE-3: The IPv6 Transition CE router MUST NOT perform IPv4 Network Address Translation (NAT) on IPv4 traffic encapsulated using DS-Lite.

DSLITE-4: If the IPv6 Transition CE router is configured with an IPv4 address on its WAN interface, then the IPv6 Transition CE router SHOULD disable the DS-Lite Basic Bridging BroadBand (B4) element.

5.3.1.3. Lightweight 4over6 (lw4o6)

Lw4o6 [RFC7596] specifies an extension to DS-Lite, which moves the NAPT function from the DS-Lite tunnel concentrator to the tunnel client located in the IPv6 Transition CE router, removing the requirement for a CGN function in the tunnel concentrator and reducing the amount of centralized state.

The IPv6 Transition CE router SHOULD implement lw4o6 functionality. If DS-Lite is implemented, lw4o6 MUST be supported as well. If lw4o6 is supported, it MUST be implemented according to [RFC7596]. This document takes no position on simultaneous operation of lw4o6 and native IPv4. The following IPv6 Transition CE router Requirements also apply:

Lw4o6 requirements:

LW4O6-1: The IPv6 Transition CE router MUST support configuration of lw4o6 via the lw4o6 DHCPv6 options [RFC7598]. The IPv6 Transition CE router MAY use other mechanisms to configure lw4o6 parameters. Such mechanisms are outside the scope of this document.

LW4O6-2: The IPv6 Transition CE router MUST support the DHCPv6 S46 priority option described in [RFC8026].

LW4O6-3: The IPv6 Transition CE router MUST support the DHCPv4-over-DHCPv6 (DHCP 4o6) transport described in [RFC7341].

LW4O6-4: The IPv6 Transition CE router MAY support Dynamic Allocation of Shared IPv4 Addresses as described in [RFC7618].

5.3.1.4. MAP-E

MAP-E [RFC7597] is a mechanism for transporting IPv4 packets across an IPv6 network using IP encapsulation, including a generic mechanism for mapping between IPv6 addresses and IPv4 addresses as well as transport-layer ports.

The IPv6 Transition CE router SHOULD support MAP-E functionality. If MAP-E is supported, it MUST be implemented according to [RFC7597]. The following CE Requirements also apply:

MAP-E requirements:

MAPE-1: The IPv6 Transition CE router MUST support configuration of MAP-E via the MAP-E DHCPv6 options [RFC7598]. The IPv6

Transition CE router MAY use other mechanisms to configure MAP-E parameters. Such mechanisms are outside the scope of this document.

MAPE-2: The IPv6 Transition CE router MUST support the DHCPv6 S46 priority option described in [RFC8026].

5.3.1.5. MAP-T

MAP-T [RFC7599] is a mechanism similar to MAP-E, differing from it in that MAP-T uses IPv4-IPv6 translation, rather than encapsulation, as the form of IPv6 domain transport.

The IPv6 Transition CE router SHOULD support MAP-T functionality. If MAP-T is supported, it MUST be implemented according to [RFC7599]. The following IPv6 Transition CE Requirements also apply:

MAP-T requirements:

MAPT-1: The CE router MUST support configuration of MAP-T via the MAP-E DHCPv6 options [RFC7598]. The IPv6 Transition CE router MAY use other mechanisms to configure MAP-E parameters. Such mechanisms are outside the scope of this document.

MAPT-2: The IPv6 Transition CE router MUST support the DHCPv6 S46 priority option described in [RFC8026].

5.3.2. Support of IPv6 in IPv4-only WAN access

5.3.2.1. 6in4

6in4 [RFC4213] specifies a tunneling mechanism to allow end-users to manually configure IPv6 support via a service provider’s IPv4 network infrastructure.

The IPv6 Transition CE router MAY support 6in4 functionality. 6in4 used for a manually configured tunnel requires a subset of the 6rd parameters (delegated prefix and remote IPv4 end-point). The on-wire and forwarding plane is identical for both mechanisms, however 6in4 doesn’t support mesh traffic and requires manually provisioning. Thus, if the device supports either 6rd or 6in4, it’s commonly a minor UI addition to support both. If 6in4 is supported, it MUST be implemented according to [RFC4213]. The following CE Requirements also apply:

6in4 requirements:

6IN4-1: The IPv6 Transition CE router SHOULD support 6in4 automated configuration by means of the 6rd DHCPv4 Option 212. If the IPv6 Transition CE router has obtained an IPv4 network address through some other means such as PPP, it SHOULD use the DHCPINFORM request message [RFC2131] to request the 6rd DHCPv4 Option. The IPv6 Transition CE router MAY use other mechanisms to configure 6in4 parameters. Such mechanisms are outside the scope of this document.

6IN4-2: If the IPv6 Transition CE router is capable of automated configuration of IPv4 through IPCP (i.e., over a PPP connection), it MUST support user-entered configuration of 6in4.

6IN4-3: If the IPv6 Transition CE router supports configuration mechanisms other than the 6rd DHCPv4 Option 212 (user-entered, TR-069 [TR-069], etc.), the IPv6 Transition CE router MUST support 6in4 in "hub and spoke" mode. 6in4 in "hub and spoke" requires all IPv6 traffic to go to the 6rd Border Relay, which in this case is the tunnel-end-point. In effect, this requirement removes the "direct connect to 6rd" route defined in Section 7.1.1 of [RFC5969].

6IN4-4: The IPv6 Transition CE router MUST allow 6in4 and native IPv6 WAN interfaces to be active alone as well as simultaneously in order to support coexistence of the two technologies during an incremental transition period such as a transition from 6in4 to native IPv6.

6IN4-5: Each packet sent on a 6in4 or native WAN interface MUST be directed such that its source IP address is derived from the delegated prefix associated with the particular interface from which the packet is being sent (Section 4.3 of [RFC3704]).

6IN4-6: The IPv6 Transition CE router MUST allow different as well as identical delegated prefixes to be configured via each (6in4 or native) WAN interface.

6IN4-7: In the event that forwarding rules produce a tie between 6in4 and native IPv6, by default, the IPv6 Transition CE router MUST prefer native IPv6.

5.3.2.2. 6rd

6rd [RFC5969] specifies an automatic tunneling mechanism tailored to advance deployment of IPv6 to end users via a service provider’s IPv4 network infrastructure. Key aspects include automatic IPv6 prefix

delegation to sites, stateless operation, simple provisioning, and service that is equivalent to native IPv6 at the sites that are served by the mechanism. It is expected that such traffic is forwarded over the IPv6 Transition CE router's native IPv4 WAN interface and not encapsulated in another tunnel.

The IPv6 Transition CE router MAY support 6rd functionality. If 6rd is supported, it MUST be implemented according to [RFC5969]. The following CE Requirements also apply:

6rd requirements:

6RD-1: The IPv6 Transition CE router MUST support 6rd configuration via the 6rd DHCPv4 Option 212. If the IPv6 Transition CE router has obtained an IPv4 network address through some other means such as PPP, it SHOULD use the DHCPINFORM request message [RFC2131] to request the 6rd DHCPv4 Option. The IPv6 Transition CE router MAY use other mechanisms to configure 6rd parameters. Such mechanisms are outside the scope of this document.

6RD-2: If the IPv6 Transition CE router is capable of automated configuration of IPv4 through IPCP (i.e., over a PPP connection), it MUST support user-entered configuration of 6rd.

6RD-3: If the IPv6 Transition CE router supports configuration mechanisms other than the 6rd DHCPv4 Option 212 (user-entered, TR-069 [TR-069], etc.), the IPv6 Transition CE router MUST support 6rd in "hub and spoke" mode. 6rd in "hub and spoke" requires all IPv6 traffic to go to the 6rd Border Relay. In effect, this requirement removes the "direct connect to 6rd" route defined in Section 7.1.1 of [RFC5969].

6RD-4: The IPv6 Transition CE router MUST allow 6rd and native IPv6 WAN interfaces to be active alone as well as simultaneously in order to support coexistence of the two technologies during an incremental transition period such as a transition from 6rd to native IPv6.

6RD-5: Each packet sent on a 6rd or native WAN interface MUST be directed such that its source IP address is derived from the delegated prefix associated with the particular interface from which the packet is being sent (Section 4.3 of [RFC3704]).

6RD-6: The IPv6 Transition CE router MUST allow different as well as identical delegated prefixes to be configured via each (6rd

or native) WAN interface.

6RD-7: In the event that forwarding rules produce a tie between 6rd and native IPv6, by default, the IPv6 Transition CE router MUST prefer native IPv6.

5.4. IPv4 Multicast Support

Actual deployments support IPv4 multicast for services such as IPTV. In the transition phase it is expected that multicast services will still be provided using IPv4 to the customer LANs.

In order to support the delivery of IPv4 multicast services to IPv4 clients over an IPv6 multicast network, the IPv6 Transition CE router SHOULD support [RFC8114] and [RFC8115].

5.5. Security Considerations

The IPv6 Transition CE router must comply with the Security Considerations as stated in draft-palet-v6ops-rfc7084-bis2.

6. Acknowledgements

Thanks to James Woodyatt, Mohamed Boucadair, Masanobu Kawashima, Mikael Abrahamsson, Barbara Stark, Ole Troan and Brian Carpenter for their review and comments.

7. ANNEX A: Code Considerations

One of the apparent main issues for vendors to include new functionalities, such as support for new transition mechanisms, is the lack of space in the flash (or equivalent) memory. However, it has been confirmed from existing open source implementations (OpenWRT/LEDE), that adding the support for the new transitions mechanisms, requires around 10-12 Kbytes (because most of the code is shared among several transition mechanisms), which typically means about 0,15% of the existing code size in popular CEs in the market.

It is also clear that the new requirements don’t have extra cost in terms of RAM memory, neither other hardware requirements such as more powerful CPUs.

The other issue seems to be the cost of developing the code for those new functionalities. However at the time of writing this document, it has been confirmed that there are several open source versions of the required code for supporting the new transition mechanisms, so the development cost is negligent, and only integration and testing cost may become a minor issue.

8. References

8.1. Normative References

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC2131] Droms, R., "Dynamic Host Configuration Protocol", RFC 2131, DOI 10.17487/RFC2131, March 1997, <https://www.rfc-editor.org/info/rfc2131>.

[RFC3633] Troan, O. and R. Droms, "IPv6 Prefix Options for Dynamic Host Configuration Protocol (DHCP) version 6", RFC 3633, DOI 10.17487/RFC3633, December 2003, <https://www.rfc-editor.org/info/rfc3633>.

[RFC3704] Baker, F. and P. Savola, "Ingress Filtering for Multihomed Networks", BCP 84, RFC 3704, DOI 10.17487/RFC3704, March 2004, <https://www.rfc-editor.org/info/rfc3704>.

[RFC4213] Nordmark, E. and R. Gilligan, "Basic Transition Mechanisms for IPv6 Hosts and Routers", RFC 4213, DOI 10.17487/RFC4213, October 2005, <https://www.rfc-editor.org/info/rfc4213>.

[RFC5625] Bellis, R., "DNS Proxy Implementation Guidelines", BCP 152, RFC 5625, DOI 10.17487/RFC5625, August 2009, <https://www.rfc-editor.org/info/rfc5625>.

[RFC5969] Townsley, W. and O. Troan, "IPv6 Rapid Deployment on IPv4 Infrastructures (6rd) -- Protocol Specification", RFC 5969, DOI 10.17487/RFC5969, August 2010, <https://www.rfc-editor.org/info/rfc5969>.

[RFC6333] Durand, A., Droms, R., Woodyatt, J., and Y. Lee, "Dual-Stack Lite Broadband Deployments Following IPv4 Exhaustion", RFC 6333, DOI 10.17487/RFC6333, August 2011, <https://www.rfc-editor.org/info/rfc6333>.

[RFC6334] Hankins, D. and T. Mrugalski, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6) Option for Dual-Stack Lite", RFC 6334, DOI 10.17487/RFC6334, August 2011, <https://www.rfc-editor.org/info/rfc6334>.

[RFC6877] Mawatari, M., Kawashima, M., and C. Byrne, "464XLAT: Combination of Stateful and Stateless Translation", RFC 6877, DOI 10.17487/RFC6877, April 2013, <https://www.rfc-editor.org/info/rfc6877>.

[RFC7050] Savolainen, T., Korhonen, J., and D. Wing, "Discovery of the IPv6 Prefix Used for IPv6 Address Synthesis", RFC 7050, DOI 10.17487/RFC7050, November 2013, <https://www.rfc-editor.org/info/rfc7050>.

[RFC7084] Singh, H., Beebee, W., Donley, C., and B. Stark, "Basic Requirements for IPv6 Customer Edge Routers", RFC 7084, DOI 10.17487/RFC7084, November 2013, <https://www.rfc-editor.org/info/rfc7084>.

[RFC7225] Boucadair, M., "Discovering NAT64 IPv6 Prefixes Using the Port Control Protocol (PCP)", RFC 7225, DOI 10.17487/RFC7225, May 2014, <https://www.rfc-editor.org/info/rfc7225>.

[RFC7341] Sun, Q., Cui, Y., Siodelski, M., Krishnan, S., and I. Farrer, "DHCPv4-over-DHCPv6 (DHCP 4o6) Transport", RFC 7341, DOI 10.17487/RFC7341, August 2014, <https://www.rfc-editor.org/info/rfc7341>.

[RFC7596] Cui, Y., Sun, Q., Boucadair, M., Tsou, T., Lee, Y., and I. Farrer, "Lightweight 4over6: An Extension to the Dual-Stack Lite Architecture", RFC 7596, DOI 10.17487/RFC7596, July 2015, <https://www.rfc-editor.org/info/rfc7596>.

[RFC7597] Troan, O., Ed., Dec, W., Li, X., Bao, C., Matsushima, S., Murakami, T., and T. Taylor, Ed., "Mapping of Address and Port with Encapsulation (MAP-E)", RFC 7597, DOI 10.17487/RFC7597, July 2015, <https://www.rfc-editor.org/info/rfc7597>.

[RFC7598] Mrugalski, T., Troan, O., Farrer, I., Perreault, S., Dec, W., Bao, C., Yeh, L., and X. Deng, "DHCPv6 Options for Configuration of Softwire Address and Port-Mapped Clients", RFC 7598, DOI 10.17487/RFC7598, July 2015, <https://www.rfc-editor.org/info/rfc7598>.

[RFC7599] Li, X., Bao, C., Dec, W., Ed., Troan, O., Matsushima, S., and T. Murakami, "Mapping of Address and Port using Translation (MAP-T)", RFC 7599, DOI 10.17487/RFC7599, July 2015, <https://www.rfc-editor.org/info/rfc7599>.

[RFC7608] Boucadair, M., Petrescu, A., and F. Baker, "IPv6 Prefix Length Recommendation for Forwarding", BCP 198, RFC 7608, DOI 10.17487/RFC7608, July 2015, <https://www.rfc-editor.org/info/rfc7608>.

[RFC7618] Cui, Y., Sun, Q., Farrer, I., Lee, Y., Sun, Q., and M. Boucadair, "Dynamic Allocation of Shared IPv4 Addresses", RFC 7618, DOI 10.17487/RFC7618, August 2015, <https://www.rfc-editor.org/info/rfc7618>.

[RFC8026] Boucadair, M. and I. Farrer, "Unified IPv4-in-IPv6 Softwire Customer Premises Equipment (CPE): A DHCPv6-Based Prioritization Mechanism", RFC 8026, DOI 10.17487/RFC8026, November 2016, <https://www.rfc-editor.org/info/rfc8026>.

[RFC8114] Boucadair, M., Qin, C., Jacquenet, C., Lee, Y., and Q. Wang, "Delivery of IPv4 Multicast Services to IPv4 Clients over an IPv6 Multicast Network", RFC 8114, DOI 10.17487/RFC8114, March 2017, <https://www.rfc-editor.org/info/rfc8114>.

[RFC8115] Boucadair, M., Qin, J., Tsou, T., and X. Deng, "DHCPv6 Option for IPv4-Embedded Multicast and Unicast IPv6 Prefixes", RFC 8115, DOI 10.17487/RFC8115, March 2017, <https://www.rfc-editor.org/info/rfc8115>.

8.2. Informative References

[RFC7788] Stenberg, M., Barth, S., and P. Pfister, "Home Networking Control Protocol", RFC 7788, DOI 10.17487/RFC7788, April 2016, <https://www.rfc-editor.org/info/rfc7788>.

[TR-069] Broadband Forum, "CPE WAN Management Protocol", TR-069 Amendment 4, July 2011, <http://www.broadband-forum.org/technical/trlist.php>.

[UPnP-IGD] UPnP Forum, "InternetGatewayDevice:2 Device Template Version 1.01", December 2010, <http://upnp.org/specs/gw/igd2/>.

Author’s Address

Jordi Palet Martinez
Consulintel, S.L.
Molino de la Navata, 75
La Navata - Galapagar, Madrid 28420
Spain

EMail: jordi.palet@consulintel.es
URI: http://www.consulintel.es/

Network Working Group
Internet-Draft
Intended status: Informational
Expires: December 26, 2019

F. Templin, Ed.
Boeing Research & Technology

June 24, 2019

IPv6 Prefix Delegation and Multi-Addressing Models
draft-templin-v6ops-pdhost-24.txt

Abstract

Requesting nodes typically acquire IPv6 prefixes from a prefix delegation service for the network. The requesting node can provision the prefix according to whether it acts as a router on behalf of any downstream networks and/or as a host on behalf of its local applications. In the latter case, the requesting node can use portions of the delegated prefix for its own multi-addressing purposes. This document therefore considers prefix delegation models for both the classic routing and various multi-addressing use cases.

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 https://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 December 26, 2019.

Table of Contents

1. Introduction 2
2. Terminology 6
3. Multi-Addressing Considerations 6
4. Multi-Addressing Alternatives for Delegated Prefixes 7
5. Address Autoconfiguration Considerations 8
6. MLD/DAD Implications 8
7. Dynamic Routing Protocol Implications 9
8. IPv6 Neighbor Discovery Implications 9
9. Prefix Delegation Services 9
10. IANA Considerations 10
11. Security Considerations 10
12. Acknowledgements 10
13. References 11
13.1. Normative References 11
13.2. Informative References 12
Appendix A. Change Log 13
Author's Address 14

1. Introduction

IPv6 Neighbor Discovery (ND) is the process by which nodes on the link discover each other's presence as well as advertise and receive configuration information. IPv6 Prefix Delegation (PD) entails 1) the communication of a prefix from a delegation service to a requesting node, 2) a representation of the prefix in the network's Routing Information Base (RIB) and the first-hop router's Forwarding Information Base (FIB), and 3) a control messaging service to maintain prefix lifetimes. Following delegation, the prefix is available for the requesting node's exclusive use and is not shared with any other nodes. This document considers prefix delegation models and multiaddressing considerations for requesting nodes that act as a router on behalf of any downstream networks and/or as a host on behalf of their local applications.

For nodes that connect downstream-attached networks (e.g., a cellphone that connects a "tethered" Internet of Things (IoT), a laptop computer with a complex internal network of virtual machines, etc.), the classic routing model applies as shown in Figure 1:

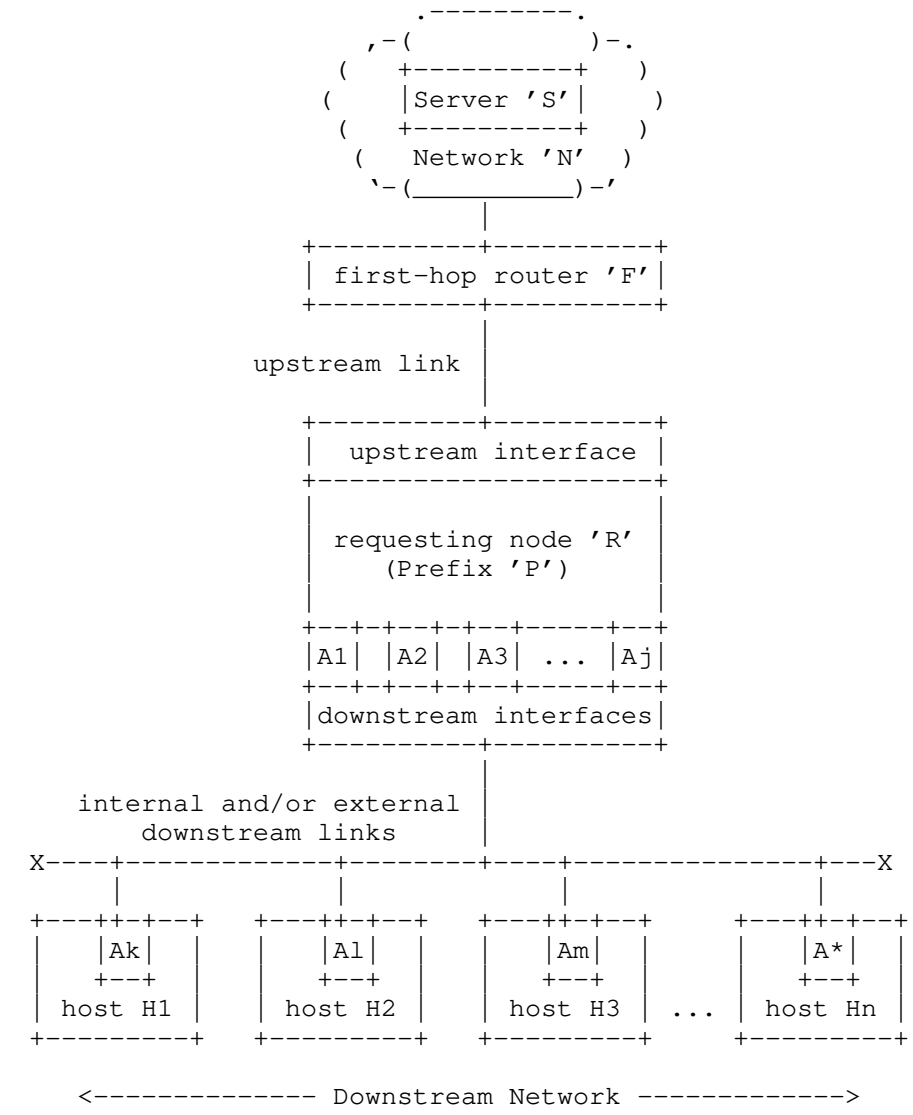

Figure 1: Classic Routing Model

In the classic routing model, requesting node 'R' has one or more upstream interfaces and connects zero or more internal and/or external downstream networks. When 'R' requests a prefix delegation, the following sequence of events transpires:

- Server 'S' located in network 'N' delegates prefix 'P' to requesting node 'R'.

- ’P’ is injected into the RIB for ’N’, and first hop router ’F’ configures a FIB entry with ’R’ as the next hop.

- R’ receives ’P’ and assigns zero or more addresses ’A(*)’ taken from ’P’ to its downstream interfaces

- ’R’ advertises zero or more sub-prefixes taken from ’P’ to hosts ’H(i)’ on downstream networks.

- ’R’ delegates zero or more sub-prefixes taken from ’P’ to requesting nodes in downstream networks.

- ’R’ acts as a router for hosts ’H(i)’ on downstream networks and as a host on behalf of its local applications.

This document also considers the case when ’R’ uses portions of ’P’ for its own internal multi-addressing purposes. [RFC7934] provides Best Current Practice (BCP) motivations for the benefits of multi-addressing, while an operational means for providing nodes with multiple addresses is given in [RFC8273]. The following multi-addressing alternatives for delegated prefixes compliment this framework.

In a first alternative, when requesting node ’R’ receives prefix ’P’, it can assign addresses taken from ’P’ to downstream virtual interfaces (e.g., a loopback) as shown in Figure 2:

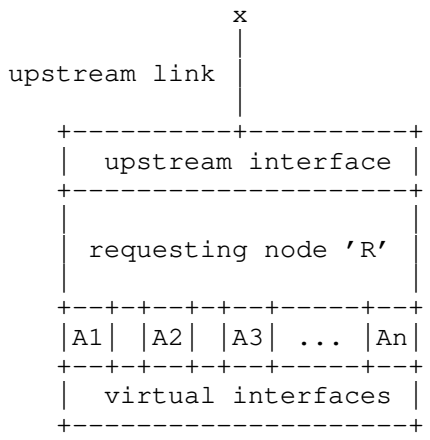

Figure 2: Address Assignment to Downstream Virtual Interfaces

In a second alternative, ’R’ could assign IPv6 addresses taken from ’P’ to the upstream interface over which the prefix was received as shown in Figure 3:

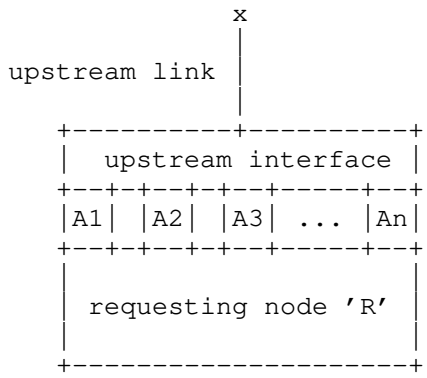

Figure 3: Upstream Interface Address Assignment

In a third alternative, 'R' could assign IPv6 addresses taken from 'P' to its local applications which appear as "psuedo" virtual interfaces as shown in Figure 4:

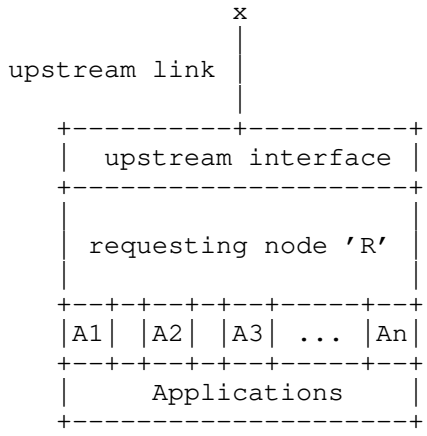

Figure 4: Application Addresssing Model

With these IPv6 PD-based multi-addressing considerations, the node can configure an unlimited supply of addresses to make them available for local applications without requiring coordination with other nodes on upstream interfaces. The following sections present considerations for nodes that employ IPv6 PD mechanisms.

2. Terminology

The terms "node", "host" and "router" are the same as defined in [RFC8200]. The terms Router Solicitation (RS), Router Advertisement (RA), Neighbor Solicitation (NS), Neighbor Advertisment (NA), Redirect and Prefix Information Option (PIO) are the same as defined in [RFC4861]. All other terminology in the normative references applies, while the following terms are defined within the context of this document:

shared prefix
: an IPv6 prefix that may be advertised to more than one node on the link. The router that advertises the prefix must consider the prefix as on-link so that the IPv6 ND address resolution function will identify the correct neighbor for each packet.

individual prefix
: an IPv6 prefix that is advertised to exactly one node on the link, where the node may be unaware that the prefix is individual and may not participate in prefix maintenance procedures. The router that advertises the prefix can consider the prefix as on-link or not on-link. In the former case, the router performs address resolution and only forwards those packets that match one of the node's configured addresses so that the node will not receive unwanted packets. In the latter case, the router can simply forward all packets matching the prefix to the node which must then drop any packets that do not match one of its configured addresses. An example individual prefix service is documented in [RFC8273].

delegated prefix
: an IPv6 prefix that is explicitly conveyed to a node for its own exclusive use, where the node is an active participant in prefix delegation and maintenance procedures. The first-hop router simply forwards all packets matching the prefix to the requesting node. The requesting node associates the prefix with downstream and/or internal virtual interfaces (i.e., and not the upstream interface).

3. Multi-Addressing Considerations

IPv6 allows nodes to assign multiple addresses to a single interface. [RFC7934] discusses options for multi-addressing as well as use cases where multi-addressing may be desirable. Address configuration options for multi-addressing include StateLess Address AutoConfiguration (SLAAC) [RFC4862], Dynamic Host Configuration Protocol for IPv6 (DHCPv6) address configuration [RFC8415], manual configuration, etc.

Nodes configure addresses from a shared or individual prefix and assign them to the upstream interface over which the prefix was received. When the node assigns the addresses, it is required to use Multicast Listener Discovery (MLD) [RFC3810] to join the appropriate solicited-node multicast group(s) and to use the Duplicate Address Detection (DAD) algorithm [RFC4862] to ensure that no other node configures a duplicate address.

In contrast, a node that configures addresses from a delegated prefix can assign them without invoking MLD/DAD on an upstream interface, since the prefix has been delegated to the node for its own exclusive use and is not shared with any other nodes.

4. Multi-Addressing Alternatives for Delegated Prefixes

When a node receives a delegated prefix, it has many alternatives for provisioning the prefix to its local interfaces and/or downstream networks. [RFC7278] discusses alternatives for provisioning a prefix obtained by a User Equipment (UE) device under the 3rd Generation Partnership Program (3GPP) service model. This document considers the more general case when the node receives a delegated prefix explicitly provided for its own exclusive use.

When the node receives the prefix, it can distribute the prefix to internal (virtual) or external (physical) downstream networks and optionally configure addresses for itself on downstream interfaces. The node then acts as a router on behalf of its downstream networks.

The node could instead (or in addition) use portions of the delegated prefix for its own multi-addressing purposes. In a first alternative, the node can assign as many addresses as it wants from the prefix to downstream virtual interfaces.

In a second alternative, the node can assign as many addresses as it wants from the prefix to the upstream interface over which the prefix was received, but in normal practice does not assign the prefix itself (or subnets from the prefix) to the upstream interface. If the node assigned the prefix to the upstream interface, any neighbors on the upstream link receiving an RA could configure addresses from the prefix and a default route with the node as the next hop. This could create a loop where upstream link neighbors send packets to the node which in turn forwards them to another upstream link neighbor. Still, there may be cases where the node provides services for dependent neighbors on the upstream link that have no other means of connecting to the network. ([RFC8415] chose to remain silent on this subject since it is operational rather than functional in nature.)

In a third alternative, the node can assign addresses taken from the delegated prefix to its local applications. The applications themselves then serve as virtual interfaces. (Note that, in the future, the practice of assigning unique non-link-local IPv6 addresses to applications could obviate the need for transport protocol port numbers.)

In these multi-addressing cases, the node normally assigns the prefix itself to a virtual interface such as a loopback so that unused portions of the prefix are correctly identified as unreachable. The node then acts as a host on behalf of its local applications even though neighbors on the upstream link consider it as a router.

5. Address Autoconfiguration Considerations

Nodes autoconfigure addresses according to Section 6 of IPv6 Node Requirements [I-D.ietf-6man-rfc6434-bis].

Nodes that connect to a network that spans more than just a single LAN configure at least one non-link-local adddress, i.e., for network management and error reporting purposes.

Nodes recognize the Subnet Router Anycast address [RFC4291] for each delegated prefix. Therefore, the node's use of the Subnet Router Anycast address must be indistinguishable from the behavior of an ordinary router when viewed from the outside world.

6. MLD/DAD Implications

When a node configures addresses for itself from a shared or individual prefix (and when the interface variable 'DupAddrDetectTransmits' is non-zero [RFC4862]), the node performs MLD/DAD by sending multicast messages over the upstream interface to test whether there is another node on the link that configures a duplicate address. When there are many such addresses and/or many such nodes, this could result in substantial multicast traffic that affects all nodes on the link.

When a node configures addresses for itself from a delegated prefix and assigns them on downstream interfaces, it can configure as many addresses as it wants without performing MLD/DAD for any of the addresses over the upstream interface.

When a node configures addresses for itself from a delegated prefix and assigns them on the upstream interface over which the prefix was received, the node honors MLD/DAD procedures according the the interface's 'DupAddrDetectTransmits' variable.

7. Dynamic Routing Protocol Implications

Nodes that receive delegated prefixes can be configured to either participate or not participate in a dynamic routing protocol over the upstream interface. When there are many nodes on the upstream link, dynamic routing protocol participation might be impractical due to scaling limitations, and may also be exacerbated by factors such as node mobility.

Unless it participates in a dynamic routing protocol, the node initially has only a default route pointing to a neighbor via an upstream interface. This means that packets sent by the node over an upstream interface will initially go through a default router even if there is a better first-hop node on the link. The node may subsequently receive Redirect messages from the default router that identify a better first-hop.

8. IPv6 Neighbor Discovery Implications

According to [RFC4861], when a node receives a shared or individual prefix with "L=1" and has a packet to send to an IPv6 destination within the prefix, it is required to use the IPv6 ND address resolution function to resolve the link-layer address of a neighbor on the link that configures the address.

Also according to [RFC4861], when a node receives a shared or individual prefix with "L=0" and has a packet to send to an IPv6 destination within the prefix, it sends the packet to a default router since "L=0" makes no statement about on-link or off-link properties of the prefix.

When a node requires a delegated prefix, it acts as a simple host by sending RS messages over the upstream interface in the manner described in Section 4.2 of [RFC7084] and invokes prefix delegation services as discussed in Section 9. The node considers the upstream interface as a non-advertising interface [RFC4861], i.e., it does not send RA messages over the upstream interface. The node further does not perform the IPv6 ND address resolution function over the upstream interface, since the delegated prefix is by definition not associated with the upstream interface.

9. Prefix Delegation Services

Selection of prefix delegation services must be considered according to specific use cases. An example service is that offered by standard DHCPv6 Prefix Delegation [RFC8415]. Alternative services based on IPv6 ND messaging have also been proposed [I-D.templin-6man-dhcpv6-ndopt][I-D.naveen-slaac-prefix-management].

Other, non-router, mechanisms may exist, such as proprietary IPAMs, [I-D.ietf-anima-prefix-management] and [I-D.li-opsawg-address-pool-management-arch]. Requirements for extending an IPv6 /64 Prefix from a Third Generation Partnership Project (3GPP) Mobile Interface to a LAN Link are discussed in [RFC7278].

10. IANA Considerations

This document introduces no IANA considerations.

11. Security Considerations

Security considerations for IPv6 Neighbor Discovery [RFC4861] and any applicable PD mechanisms apply to this document. Nodes that manage their delegated prefixes such that MLD/DAD procedures are not needed on the upstream interface can avoid introducing multicast messaging congestion on the upstream link. Also, routers that delegate prefixes keep only a single neighbor cache entry for each prefix delegation recipient, meaning that the router's neighbor cache cannot be subject to address resolution-based resource exhaustion attacks.

For shared and individual prefixes, if the advertising router considers the prefix as on-link the IPv6 ND address resolution function will prevent unwanted IPv6 packets from reaching the node. For delegated prefixes and individual prefixes that are not considered on-link, the router delivers all packets that match the prefix to the node. In that case, the node may receive unwanted IPv6 packets via an upstream interface for which it has no matching configured address. The node then drops the packets and observes the ICMPv6 "Destination Unreachable - Address/Port unreachable" procedures discussed in [RFC4443].

The node may also receive IPv6 packets via an upstream interface that do not match any of the node's delegated prefixes. In that case, the node drops the packets and observes the ICMPv6 "Destination Unreachable - No route to destination" procedures discussed in [RFC4443]. Dropping the packets is necessary to avoid a reflection attack that would cause the node to forward packets received from an upstream interface via the same or a different upstream interface.

12. Acknowledgements

This work was motivated by discussions on the v6ops list. Mark Smith, Ricardo Pelaez-Negro, Edwin Cordeiro, Fred Baker, Ron Bonica, Naveen Lakshman, Ole Troan, Bob Hinden, Brian Carpenter, Joel Halpern, Albert Manfredi, Dusan Mudric, Paul Marks, Joe Touch, Alex

Petrescu, Lorenzo Colitti, Tatuya Jinmei and Naveen Kottapalli provided useful comments that have greatly improved the document.

This work is aligned with the NASA Safe Autonomous Systems Operation (SASO) program under NASA contract number NNA16BD84C.

This work is aligned with the FAA as per the SE2025 contract number DTFAWA-15-D-00030.

This work is aligned with the Boeing Information Technology (BIT) MobileNet program and the Boeing Research & Technology (BR&T) enterprise autonomy program.

13. References

13.1. Normative References

[RFC3810] Vida, R., Ed. and L. Costa, Ed., "Multicast Listener Discovery Version 2 (MLDv2) for IPv6", RFC 3810, DOI 10.17487/RFC3810, June 2004, <https://www.rfc-editor.org/info/rfc3810>.

[RFC4443] Conta, A., Deering, S., and M. Gupta, Ed., "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", STD 89, RFC 4443, DOI 10.17487/RFC4443, March 2006, <https://www.rfc-editor.org/info/rfc4443>.

[RFC4861] Narten, T., Nordmark, E., Simpson, W., and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, DOI 10.17487/RFC4861, September 2007, <https://www.rfc-editor.org/info/rfc4861>.

[RFC4862] Thomson, S., Narten, T., and T. Jinmei, "IPv6 Stateless Address Autoconfiguration", RFC 4862, DOI 10.17487/RFC4862, September 2007, <https://www.rfc-editor.org/info/rfc4862>.

[RFC8200] Deering, S. and R. Hinden, "Internet Protocol, Version 6 (IPv6) Specification", STD 86, RFC 8200, DOI 10.17487/RFC8200, July 2017, <https://www.rfc-editor.org/info/rfc8200>.

[RFC8415] Mrugalski, T., Siodelski, M., Volz, B., Yourtchenko, A., Richardson, M., Jiang, S., Lemon, T., and T. Winters, "Dynamic Host Configuration Protocol for IPv6 (DHCPv6)", RFC 8415, DOI 10.17487/RFC8415, November 2018, <https://www.rfc-editor.org/info/rfc8415>.

13.2. Informative References

[I-D.ietf-6man-rfc6434-bis]
Chown, T., Loughney, J., and T. Winters, "IPv6 Node Requirements", draft-ietf-6man-rfc6434-bis-09 (work in progress), July 2018.

[I-D.ietf-anima-prefix-management]
Jiang, S., Du, Z., Carpenter, B., and Q. Sun, "Autonomic IPv6 Edge Prefix Management in Large-scale Networks", draft-ietf-anima-prefix-management-07 (work in progress), December 2017.

[I-D.li-opsawg-address-pool-management-arch]
Li, C., Xie, C., Kumar, R., Fioccola, G., Xu, W., LIU, W., Ma, D., and J. Bi, "Coordinated Address Space Management architecture", draft-li-opsawg-address-pool-management-arch-01 (work in progress), July 2018.

[I-D.naveen-slaac-prefix-management]
Kottapalli, N., "IPv6 Stateless Prefix Management", draft-naveen-slaac-prefix-management-00 (work in progress), November 2018.

[I-D.templin-6man-dhcpv6-ndopt]
Templin, F., "A Unified Stateful/Stateless Configuration Service for IPv6", draft-templin-6man-dhcpv6-ndopt-07 (work in progress), December 2018.

[I-D.templin-6man-rio-redirect]
Templin, F. and j. woodyatt, "Route Information Options in IPv6 Neighbor Discovery", draft-templin-6man-rio-redirect-07 (work in progress), December 2018.

[RFC4291] Hinden, R. and S. Deering, "IP Version 6 Addressing Architecture", RFC 4291, DOI 10.17487/RFC4291, February 2006, <https://www.rfc-editor.org/info/rfc4291>.

[RFC7084] Singh, H., Beebee, W., Donley, C., and B. Stark, "Basic Requirements for IPv6 Customer Edge Routers", RFC 7084, DOI 10.17487/RFC7084, November 2013, <https://www.rfc-editor.org/info/rfc7084>.

[RFC7278] Byrne, C., Drown, D., and A. Vizdal, "Extending an IPv6 /64 Prefix from a Third Generation Partnership Project (3GPP) Mobile Interface to a LAN Link", RFC 7278, DOI 10.17487/RFC7278, June 2014, <https://www.rfc-editor.org/info/rfc7278>.

[RFC7934] Colitti, L., Cerf, V., Cheshire, S., and D. Schinazi, "Host Address Availability Recommendations", BCP 204, RFC 7934, DOI 10.17487/RFC7934, July 2016, <https://www.rfc-editor.org/info/rfc7934>.

[RFC8273] Brzozowski, J. and G. Van de Velde, "Unique IPv6 Prefix per Host", RFC 8273, DOI 10.17487/RFC8273, December 2017, <https://www.rfc-editor.org/info/rfc8273>.

Appendix A. Change Log

<< RFC Editor - remove prior to publication >>

Changes from -23 to -24:

- Version and reference update

Changes from -22 to -23:

- Changed DHCPv6 references to RFC8415. Deprecate RFC3315 and RFC3633.

- New text on assignment of addresses and prefixes on the upstream interface.

Changes from -21 to -22:

- Changes to address list comments contributed by Lorenzo Colitti, Tatuya Jinmei, Brian Carpenter and Fred Baker.

- Deleted section on ICMPv6 - now defer to normative reference [RFC4443].

- Discuss 'DupAddrDetectTransmits' variable implications under MLD/ DAD considerations.

Changes from -20 to -21:

- Re-worked classic routing model section

- Included multi-addressing case where addresses may be assigned to applications

- Removed strong/weak end system discussions

Changes from -19 to -20:

- figure 1 updates to show Server as being somewhere in the network

- Introductory material to show relation to other RFCs on multi-addressing

Changes from -18 to -19:

- added new section on Prefix Delegation Services

Changes from -17 to -18:

- re-worked discussion on the prefix delegation service in Section 1
- updated figures in Section 1

Changes from -16 to -17:

- added supporting text in the introduction to discuss the Delegating Router's relationship with the Requesting Router and with supporting intrastructure in the operator's network
- updated figures in introduction to include representation of operator's network
- added new section on Address Autoconfiguration Considerations

Author's Address

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

Email: fltemplin@acm.org