IPv6 Maintenance (6man) Working Group
Internet Draft
Updates: 4861, 4862 (if approved)
Intended status: Standards Track
Expires: September 2022

E. Vasilenko
P. Volpato
Huawei Technologies
Olorunloba Olopade
Virgin Media

March 4, 2022

ND Prefix Robustness Improvements
draft-vv-6man-nd-prefix-robustness-02

Abstract

IPv6 prefixes could become invalid abruptly as a result of outages, network administrator actions, or particular product shortcomings.

That could lead to connectivity problems for the hosts attached to the subtended network.

This document has two targets: on one hand, to analyze the cases that may lead to network prefix invalidity; on the other to develop a root cause analysis for those cases and propose a solution.

This may bring to extensions of the protocols used to convey prefix information and other options.

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 September 2022.

Table of Contents

1. Terminology and pre-requisite

[ND] and [SLAAC] are pre-requisite to understand this document. The terms are inherited from these standards.

Additional terms:

Home Gateway - a small consumer-grade router that provides network access between hosts on the local area network (LAN) and the Internet behind the wide area network (WAN)

PA - Provider-Aggregatable addresses leased to the client or subscriber

MHMP - Multi-Homing Multi-Prefix. An environment with hosts connected to different PA providers (multi-homing) through different address spaces announced from different providers (multi-prefix)

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "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. Introduction

It has been reported that some number of cases could lead to loss of information (primarily prefixes) by [ND]. Current [ND] protocol's default timers may lead to many days of outage for hosts. This is not acceptable.

This document analyses all potential cases when an outage could happen and proposes solutions. Discussion is restricted to potential [ND] extensions only.

MHMP environment has been considered. It has been discovered that [ND] problems could be isolated from the overall complex [MHMP] environment, and could be fixed separately.

The document is organized to introduce, in section 3, the scenarios where the issue of prefix invalidity may happen and the cases of invalidity.

Section 4 provides a root cause analysis for the cases of invalidity and identifies the corner-cases which are subject of our discussion.

Section 5 proposes a solution for the cases identified.

Section 6 brings the discussion forward, proposing extensions to [ND].

3. Problem Scenarios

[ND] distributes prefixes as Prefix Information Options (PIOs) in Router Advertisements (RA) messages from routers.

Once a router assigns a prefix to a host, this prefix is assumed to be stable so that hosts can employ it to configure the IPv6 addresses associated with their interfaces [SLAAC] or to forward packets to the network.

Prefix changes may happen and are governed by the rules of [ND], [SLAAC].

Yet, cases exist where prefix instability may happen. An example is provided by the so-called "flash-renumbering" event: when flash-renumbering happens a network prefix in use suddenly becomes invalid because it is replaced by a new prefix.

The router causing or forced to cause the network renumbering may not be able to cope with the effects of this sudden change (for example, deprecating the previously assigned prefixes). Another possibility is that the subtended hosts do not have the means of overcoming the effects of renumbering.

This section describes problems that were found in live networks. Most of the information in this section comes from [Flash-Renumbering], [SLAAC Robustness]. Their contributions are greatly acknowledged.

3.1. Reference architectures

Home broadband networks, SOHO (Small Office Home Office) networks are the typical scenarios affected by renumbering. Some problems discussed below applicable on the more general basis.

In typical case a router (e.g. Home Gateway, Customer Premise Equipment (CPE), Customer Edge (CE), etc.) is deployed to provide connectivity to a Service Provider network for the attached devices. A second router may be deployed for redundancy, especially for business scenarios.

Two reference architecture can be considered:

Architecture #1. Hosts are directly connected to the router. For example, a Home Gateway embeds the functions of L2 device (Ethernet switch, WiFi AP) and L3 device (router).

Architecture #2. Hosts connect to an intermediate L2 device (e.g. a wired Ethernet switch or a Wi-Fi access point) that, in turn, connects to the router (or routers, if uplink redundancy is requested).

3.2. Discussion on the scenarios

The discussion provided here is introductory to both the root cause analysis provided in section 4. and the solutions proposed in section 5.

3.2.1. Non-graceful reload due to unexpected events

A router could be reloaded abruptly for many reasons: hardware or software bug, power outage, manual intervention. This last one is very probable for home broadband subscribers that tend to fix every problem with power recycle.

Usually, it does not create additional problems for [ND] and [SLAAC] because the same PIO information would be advertised by the router in RA messages after each reload. In such cases, a Home Gateway would initialize its Ethernet and WiFI connections, clearing all stale information on directly connected hosts.

It should not create problems for proper home network design where all CPEs are routers - see [HomeNet Architecture]. The delegated prefix would not be changed in the case of subtended CPE reload. Prefix change in the case of upstream CPE reload should be properly discontinued by subtended CPE. There is the need for a special protocol for prefix distribution that is out of the scope of this document - see [HNCP].

For architecture #2 implemented in home environments, there is a corner case when Home Gateway's abrupt reload would not be visible to hosts connected to subtended "bridged" CPE. If it would coincide with the situation when a different prefix would be delegated from Carrier (at 37% probability according to [Residential practices]), it would lead to the situation that hosts would receive a new prefix without deprecation of the previous one. Hosts do not have any standard mechanism to choose only the new prefix for communication. That would lead to a connectivity problem.

How long a non-preferred prefix would be kept in a stale state on the host is not important (default AdvValidLifetime is 30 days in section 6.2.1 of [ND]), because according to [Default Address] section 5 rule#3, it should have a lower priority to be chosen. [SLAAC] section 5.5.4 is another good reference highlighting that address should be avoided after it would reach the deprecated status.
How long an address would stay in the preferred state is important. [ND] instructs hosts to prefer certain prefix for 7 days - see default AdvPreferredLifetime in section 6.2.1.
It is not realistic for the subscriber to wait for 7 days.
It practically means that the subscriber in this corner case would have a few options to fix the problem: (1) reload all hosts, or (2) reconnect the physical link of every host, or (3) reload subtended bridge, or (4) manually delete the prefix on the hosts to clear stale information.

3.2.2. Graceful reload without precautions

Specifically this scenario may happen when developers don't apply precautions in case previous prefixes are not deprecated. It may happen in both architectures.

The router could be reloaded by graceful procedure (reboot or shutdown that would use "init 6" in Unix). It is still possible that software would not send RA with prefix Preferred Lifetime zero to inform hosts about prefix deprecation. This practice prevails because IPv4's centralized address assignments by DHCP does not need similar precautions.

Again, like in the previous section, it would not create a problem in the majority of the cases for directly connected hosts (architecture #1) because link layer would be reinitialized too. The same corner case (architecture #2) would lead to the same result: a connectivity problem that could be resolved only by 4 types of manual intervention mentioned in the previous section.

3.2.3. Abrupt hardware replacement without the possibility for graceful prefix deprecation

Such type of an outage is again may happen only for architecture #2. It would lead to up to 30 minutes (including time for hardware replacement) outage in all cases (to detect missing router) and up to 1 week additional outage if a different prefix would be announced after the hardware replacement.

The hardware could fail or be replaced with an abrupt power disconnect. The latter is very probable for the home environment. Graceful notification of hosts may not happen.

The new hardware may have a different link layer address and a different link local address as a result. The router would look like a new one on the link. Any communication with it could not be the reason to deprecate announcements made early by the router perceived as a different one.

[ND] section 6.2.1 has recommended the AdvDefaultLifetime as 3*MaxRtrAdvInterval. Hosts would send traffic to a non-existent router for up to 30 minutes.

According to section 4.2 of [ND] "Router Lifetime" is related only to router default status. PIO announced early may be preferred up to 7 days according to AdvPreferredLifetime in section 6.2.1 of [ND] even after the router default status is deprecated. The probability for such a situation is the same low as discussed in section 3.2.1. because a different prefix should be announced after hardware reload and a switch should be present between the host and the router. The same corner case would lead to the same result: a connectivity

problem that could be resolved only by 4 types of manual intervention mentioned in section 3.2.1. .

3.2.4. Non-graceful configuration change

This situation may happen due to abrupt prefix change on the router (in both architectures) or VLAN change on the switch (it may happen in architecture #2).

Router configuration could be changed manually, by automation tools, or by protocols (for example, prefix distribution).

Additionally for architecture #2, L2 domain could be abruptly changed by configuration (for example, VLAN change from "quarantine" to "production" without any chance for the router to send a message).

It could lead to the situation that prefix would change abruptly, without any notification to hosts about the necessity to deprecate the previous prefix. Hosts should be notified by prefix announcement with Preferred Lifetime set to zero.

It should not happen for residential CPE because [CPE Requirements] section 4.3 requirement L-13 clearly instructs: "If the delegated prefix changes, i.e., the current prefix is replaced with a new prefix without any overlapping period of time, then the IPv6 CE router MUST immediately advertise the old prefix with a Preferred Lifetime of zero".

But it is perfectly possible for other environments (except residential CPEs) because other routers are not required to do the same: [Node Requirements] does not clarify the exact router behavior in the case of abrupt prefix change. [SLAAC] does not have any recommendations either.

3.2.5. An uplink breaks connectivity without a relevant notification to the connected hosts

It may happen in both architectures #1 and #2.

A router could lose uplink. The probability for such an event is much bigger for a mobile uplink (modem). It would invalidate the possibility to use a PA prefix advertised from this carrier even in the case that another carrier uplink is available on this or redundant router (connectivity to the Internet is not lost). Some mechanism is needed to inform hosts not to use address space from

the disconnected carrier because another carrier would filter it out by anti-spoofing security protection.

The multi-homing multi-prefix PA environment has been properly explained in [MHMP]. The discussion of how traffic should be source-routed by routers in [MHMP] environment is not relevant to our [ND] discussion. Unfortunately, an improper address used as a source would cause a traffic drop as soon as traffic gets to the different carrier.

[Default Address] section 5 (source address selection) rule 5 (for different interfaces on the host) and rule 5.5 (for the same interface) partially prepare hosts for such situation: "Prefer addresses in a prefix advertised by the next-hop. If SA or SA's prefix is assigned by the selected next-hop that will be used to send to D [...] then prefer SA". This algorithm has an assumption that the source address should be chosen after the next hop.

Unfortunately, the rules mentioned above in [Default Address] section 5 would work only if the default router would cease to be default after it loses route to its carrier. It would work only in simplified topology where all hosts connect by L2 to different CPEs, each leading to its separate carrier prefix. It could be called a "common-link environment for all hosts and routers". It is not possible in practice because hosts on the most popular link layer technology (WiFi) are rooted to only one CPE (with AP inside) - they would not switch automatically to different CPE where the Internet connectivity may be still available.

[CPE Requirements] have G-3/4/5 specifically for this simplified multi-homing residential design. It recommends announcing Router Lifetime as zero on LAN if CPE does not have "default router from the uplink" - it would push the host to use another source address by the mentioned above source address selection algorithm.

It is not explained in [CPE Requirements] what should happen with PA delegated prefix after the respective uplink is disconnected. Probably, this is because it was not needed to deprecate stale prefix for the above mentioned mechanism (based on default router withdrawal) to work.

The local residential network could be left without any default router as a result of using the above mechanism - it is especially probable in the single CPE environment. Hence, [CPE Requirements] promotes [ULA] addresses for local connectivity. Default router functionality is returned specifically for [ULA] addresses by

requirement L-3: use "Route Information Option" from [Route Preferences]. It needs hosts' participation in routing through the RIO option.

Unfortunately, this long chain of fixes explained above is strictly optimized for the environment "common-link for all hosts and routers". It is not the case for single WiFi inside any CPE or other topologies.

Neither [ND] nor [SLAAC] instruct the router what to do when the PA delegated prefix is withdrawn abruptly.

[Multi-Homing] section 3 has a good discussion about the proper relationship between default routers and prefixes advertised by respective routers in a stable situation. This would be discussed in more details in section 5.1. . [Multi-Homing] does not discuss what to do in the situation when the router is available, but some uplinks (with delegated prefixes) are lost.

[MHMP] discusses the problem in deep detail with two tools proposed to regulate [ND] behavior: [Policy by DHCP] to change [Default Address] algorithm and [Route Preferences] to inform about appropriate exit points. There are more details later in section 5.1.

4. Root cause analysis

Let's further analyze to be sure that all corner cases are found.

It is assumed in all discussions below that [RA-Guard] is implemented, and all messages are from routers under legitimate administrative control. Security issues are considered as resolved by [RA-Guard], and possibly with extensions in [RA-Guard+].

DHCP is almost as vulnerable as SLAAC for cases found below. DHCP's typical lease time (hours) is shorter than SLAAC's prefix lifetime (days), but is too long for users to accept self-repairing time. Root cause analysis below applies to all possible environments: DHCP, SLAAC, and mixed.

4.1. What to protect

[ND] Router Advertisements deliver configuration information to hosts. Such information could become inaccurate in two different periods of time:

a) "Recoverable". Time is needed for some process to finish and update information (example: router reload or uplink re-connect).

b) "Non-recoverable". Time, dependent on some timer expiration (example: complete loss of prefix or default router).

A careful look at the information distributed by RA would give us the understanding that the most problematic is the information that is already protected by deprecation timers: Prefix Information Option and Default Router. Section 3 discusses that the handling of this information is still susceptible to recoverable and non-recoverable periods of inaccuracy.

For example, in the case of abrupt router reload described in sections 3.2.1. -3.2.3. , the recoverable part is the time spent by router and hosts to update their cache after the router reload. The non-recoverable part is related to the setting of the AdvPreferredLifetime timer which would probably force a user to solve the issue with manual intervention.

The next problematic case is the abrupt change of source link-layer address. This problem is not discovered yet in production because it has a low probability. Indeed, a router with a different link-layer address would be treated as a new router, the old router would just disappear from the link. It would affect primarily default router information because all other information should be immediately re-advertised from the new link layer address. Section 6.2.8 of [ND] already discusses how to properly deprecate the default router status of the old link layer address, but no recommendation is given in [ND] for prefix deprecation in this situation. A corner case is possible that software would not treat the new virtual interface as identical concerning the prefix information that should be announced. Different prefixes may be announced. Some additional precautions are needed.

Other information in RA (Hop Limit, MTU, DHCP flags, Reachable timer, and Retransmit timer) are not so sensitive because (1) it is typically static and (2) it does not affect connectivity for respective parameters change in the wide range.

Flag "A" in PIO deserves special attention. It could be cleared abruptly (signaling that hosts should not use this prefix for [SLAAC] anymore). That should not create any problem, because the prefix is still available from a respected PA provider - traffic could be routed to the global Internet. Therefore, it is not vitally important for the host to immediately deprecate the address from

this prefix.
A similar situation is with flag "M" in RA: DHCP address should be deprecated. It should not create a connectivity problem because prefixes could be routed to the global Internet.

4.2. Where to protect

[ND] is the protocol for first-hop connection between host and router. It is designed for one link only. One link could have more than one router.

It is assumed below that a more complex topology (many other routers) is shielded from this link by some other protocol that would deliver all necessary information to those routers. [HomeNet Architecture] discusses many types of information that should be distributed to every home router. Let's focus on delegated prefixes for our discussion.
The number of uplinks on every router is not important, as long as proper information about prefixes is up to date on the router.

Hence, all our topologies could be simplified into the following scenarios:

I. L2 device (switch, WiFi AP) and L3 device (router) are in the same device (sharing the fate for power, reboot) (refer to architecture #1 in section 3.1.).

II. Separate L2 device (probably a switch) and an arbitrary number of L3 devices (routers) are connected to the same IPv6 link (refer to architecture #2 in section 3.1.).

4.3. When to protect: technology scenarios

Let's reorder scenarios discussed in section 3. in the way that it would be better to map to the technology modifications and account for some corner cases found in root cause analysis:

1. Proper prefix usage for Multi-Homing Multi-Prefix environment.
   Hosts should be capable of choosing in a coordinated way
   (1) a source address (from proper PA prefix) and (2) a next hop:

   A.1. In a normal situation: all providers and prefixes are available

A.2. In a faulty situation: one provider is not reachable, but some hosts and links on the routed path to this provider may still be reachable

A.3. In the case when an administrator abruptly replaces delegated prefix

2. Proper prefix usage for the case of router outage that:

A.4. Planned for this interface (reboot, shutdown, or ceasing to be a router)

A.5. Abrupt (power outage, software or hardware bug)

A.6. Abrupt (power outage, hardware fault) with hardware replacement

3. Proper prefix usage for the case of link layer address of the router.

These cases are discussed from section 5.1. to section 5.6.

There is no big difference for [ND] between ULA and GUA at the considered link because both could be disjoined at any routed hop upstream. It would need the same invalidation mechanisms on the link. ULA could be invalidated too for the case that ULA spans many sites in a big company. The residential network would probably have a separate ULA for every household that would decrease the probability of ULA prefixes invalidation. It is the responsibility of another protocol (for example, [HNCP]) to decide when ULA should be invalidated, if ever.

5. Solutions

Let's look at the solutions for scenarios listed in section 4.3.

5.1. Multi-homing multi-prefix (MHMP) environment

Let's consider here host capability to choose a proper PA prefix and next hop router in a stable multi-homing multi-prefix (MHMP) environment.

The complex MHMP situation is properly discussed in [MHMP] section 3.1 - it is critical to read it to understand the rest of this section. Our discussion is restricted to [ND] protocol only (one link) - it would cut the number of topologies discussed in section 4.2. MHMP may need additional complex routing interactions that are out of the scope of this document.

It is possible to introduce one additional classification to clearly separate what it is possible to implement now from what needs additional standardization efforts:

1. Case "equal prefixes": Announced prefixes are fully equal by scope and value, all resources interested for hosts could be reachable through any announced PA prefix; additionally, traffic distribution between carriers could be round-robin (no any traffic engineering or policing).

2. Case "non-equal prefixes": Announced prefixes are not equal because (1) some resources could be accessed only through a particular prefix (for example walled garden of one carrier) or (2) it is desirable to have some policy for traffic distribution between PA prefixes (cost of traffic, delay, packet loss, jitter, proportional load).

There are two reminders before the discussion of the above cases:

- [ND] section 6.3.6 recommends next hop choice between default routers in a round-robin style. Traffic policy or even reachability of particular resources through a particular default router is not considered at the [ND] level.

- [Default Address] section 7 assumes that source and destination address selection should happen after the next hop (or interface) choice by [ND] or routing, source address is chosen after this.

Case "equal prefixes" does not create any requirement on what prefix should be used for the source address. It is only needed that the source address would be chosen to be compatible with the next hop that should be in the direction of the respective carrier.
No problem is possible for the topology with only one router on the link. The router itself may need source routing to choose next hop properly but it is out of the scope of ND protocol and this document.
Host on a multi-homing link would better be compliant to [Default Address] section 5 (source address selection) rule 5 (for different interfaces on the host) or rule 5.5 (for different next hops on the

same interface). It would help to properly choose a source address compliant to the next hop chosen first. Moreover, if the source address would be chosen wrongly then it is still possible to reroute the packet later by source routing.
Hence, it is possible to satisfy the "equal prefixes" case on the current level of standardization developed.

Case "non-equal prefixes" is more complicated. It would be too late to try to solve this problem on a router, because the wrong source address may be already chosen by the host - it would not be possible to contact the appropriate resource in the "walled garden". Only NAT could be left as an option, but that is not a valid choice for IPv6.

There are 2 methods to resolve the case of "non-equal prefixes":

1. The same policies could be formatted differently and fed to the host by two mechanisms: 1) "Routing Information Options" of [Route Preferences] and 2) [Policy by DHCP] to modify policies in [Default Address] selection algorithm. Then current priority of mechanisms could be preserved the same: initially [ND] or routing would choose the next hop, then [Default Address] would choose a source address (and destination if multiple answers from DNS are available). It is the method that is assumed in [MHMP].
2. Alternatively, policies could be supplied only by [Policy by DHCP] to [Default Address] selection algorithm. [Default Address] discusses potential capability in section 7 to reverse algorithm's order: source address may be chosen first, only then to choose next hop (default router).
   Source address selected from proper carrier is potentially the complete information needed for the host to choose the next hop, but not for the default round-robin distribution between available routers that specified in [ND]. [ND] extension is needed for this method for the host to prioritize default routers that have announced prefixes used for the source address of the considered flow.
   It is this method that is assumed in [Multi-Homing] section 3.2. This document is different in that the same rules are formulated not as the general advice, but as the particular extension to [ND] - see section 6.1 of this document.

The second method has the advantage that there is no need to download RIO policies by [Route Preferences]. It would simplify the implementation of the MHMP environment.
Only the second method is universal and extendable because some policies may not be translated as RIO of [Route Preferences].
For example, dynamic policies (packet loss, delay, and jitter) could

be measured on hosts. Hence, the decision about source address and next hop should be local.

5.2. A provider is not reachable in MHMP environment

Let's assume the fault situation when one provider is not reachable in the [MHMP] environment. A prefix may be very dynamic for a few reasons. It could be received from some protocols (DHCP-PD, HNCP). The prefix could become invalid (at least for the global Internet connectivity) as a result of the abrupt link loss in the upstream direction to the carrier that distributed this prefix.

Additionally, consider the more complicated case when some hosts on the upstream routed path to this provider may still be reachable using a particular prefix but Internet connectivity is broken later.

Let's consider the last problem. Because Internet connectivity is lost for this prefix, it should be announced to hosts by zero Preferred Lifetime. [Route Preferences] gives the possibility to inform hosts that particular a prefix (RIO) is still available on-site but it would be an automation challenge to dynamically calculate and announce prefix. Additionally, [Route Preferences] should be supported by hosts.
In general, it is not a good idea to involve [ND] in routing. Hence, it is better to support on-site connectivity by PI GUA or ULA that may not be invalidated. There are many reasons to promote [ULA] for internal site connectivity: (1) hosts may not have GUA address at all without initial connection to the provider, (2) PA addresses would be invalidated in 30 days of disconnect anyway, (3) it is not a good idea to use addresses from PA pool that is disconnected from global Internet - hosts may have a better option to get global reachability. ULA has better security (open transport ports that are not accessible from the Internet) which is an additional bonus.
It is effectively the request to join current [CPE Requirements] and [HomeNet Architecture] requirements in sections 2.2, 2.4, 3.4.2 that subscriber's network should have local ULA addresses.

Prefix deprecation should be done by RA with zero Lifetime for this prefix. It will put the prefix on hosts to the deprecated status that according to many standards ([ND], [SLAAC], and [Default Address]) would prioritize other addresses. Global communication would be disrupted for this prefix anyway. Local communication for deprecated addresses would continue till normal resolution because the default Valid Lifetime is 30 days. Moreover, if it would happen that this delegated prefix was the only one in the local network (no [ULA] for the same reason), then new sessions would be opened on

deprecated prefix because it is the only address available.
If connectivity would be re-established and the same prefix would be delegated to the link - it would be announced again with proper preferred lifetime. If a different prefix could be delegated by the PA provider, then the old prefix would stay in deprecated status.
It is an advantage for the host that would know about global reachability on this prefix (by deprecated status) because the host may use other means for communication at that time.

Such dynamic treatment of prefixes may have the danger of [ND] messages flood if the link on the path to PA provider would be oscillating.
[HNCP] section 1.1 states: "it is desirable for ISPs to provide large enough valid and preferred lifetimes to avoid unnecessary HNCP state churn in homes".
It makes sense to introduce dampening for the rate of prefix announcements.

Such conceptual change in the treatment of prefixes would not affect current enterprise installations where prefixes are static.

It is important to mention again that it is the responsibility of the respective protocol (that has delivered prefix to the considered router) to inform the router that prefix is not routed anymore to the respective carrier. It is easy to do it in the simplified topology when the only router could correlate uplink status with the DHCP-PD prefix delegated early. Some additional protocols like [HNCP] are needed for a more complex topology.

There is nothing in [ND] or [SLAAC] that prevents us from treating prefixes as something more dynamic than "renumbering" to reflect the dynamic path status to the PA provider. Section 6.2. proposes extensions to [CPE Requirements] and [SLAAC] that follow the logic of this section.

5.3. Administrator abruptly replaces PA prefix

This is the case when the network administrator (maybe from another domain) replaces prefix much faster than 2 hours or the remaining preferred lifetime (as per section 5.5.3 of [SLAAC] on router advertisement processing). The reason for abrupt replacement is probably not related to networking.
Abrupt prefix change may be caused by improper configuration, for example, VLAN change at the switch.
Standards recommend deprecating old prefixes but do not recommend for developers and system designers to additionally check abrupt

configuration changes to mitigate human mistakes. IPv4 cannot mitigate such type of mistake, IPv6 has an advantage here.

Section 6.3. proposes a recommendation for the additional check to make sure that prefix would be deprecated.

This problem could be exacerbated by the low reliability of multicast delivery in a wireless environment - the only packet sent (for example before VLAN change) could be lost. A long-term solution for this problem is proposed in section 6.6 that permits synchronizing host states with a new flag in router announcements.

5.4. Planned router outage

A router could be planned to be put out of service for a link (reboot, shutdown, or ceasing to be a router).

The primary Operating System for routers is LINUX. The following discussion is based on LINUX as an example - other developers can find an analogy for their operating system.

Some LINUX shutdown commands are not graceful in principle (like Halt or Poweroff). It would need extraordinary efforts to send messages discussed in this section before the system would be stopped. It is better to restrict network administrators from such tools on routers.

Other LINUX shutdown commands are safe (Reboot is safe for a long time, Shutdown and "Init 6" have been safe). It would execute shutdown scripts that would give the developer the chance to comply with requirements in this section.

It is up to the developer how reboot and shutdown should be mapped to particular OS commands in graphical user interface (GUI), command line interface (CLI), or automation interface (Netconf/YANG), and what particular actions should be taken. It SHOULD guarantee that section 6.2.5 of [ND] with updates in section 6.4 of this document properly inform hosts that the router is going out of service.

The same procedure SHOULD be automatically activated for cases when an administrator tries manually (via CLI or GUI) or automatically (via Netcong/YANG/Other) to change Link Layer Address on this router interface or disable router functionality in [ND] for this link.

5.5. Prefix information lost because of abrupt router outage

PIO could be lost because of the abrupt reload - the router may not have a chance to warn hosts, but the router could receive a different prefix after reload. Reasons could be (1) power outage, (2) software bug, or (3) hardware problem.

[HomeNet Architecture] section 3.4.3 (Delegated Prefixes) has already recommended usage of non-volatile memory:
"Provisioning such persistent prefixes may imply the need for stable storage on routing devices and also a method for a home user to 'reset' the stored prefix should a significant reconfiguration be required (though ideally the home user should not be involved at all)".
[SLAAC] section 5.7 has recommended storing acquired addresses on hosts in non-volatile memory too.
This document joins these requests and propose adding similar requirements to [CPE Requirements] and [SLAAC] - see section 6.5.

The best long-term solution is to inform the host by [ND] protocol that RA has all information in one announcement. Any missing information SHOULD be considered deprecated. It is possible to do it with the new flag in RA - see section 6.6.
"Complete" flag would become useful only when implemented on both: host and router. It is proposed to rely on storage improvements in non-volatile memory till the "Complete" flag would be supported on many hosts.

5.6. Prefix information lost after hardware replacement

Hardware fault or power outage may follow by hardware replacement.

Prefix storage in non-volatile memory and a "complete" flag would not protect in such a situation. The new router would not have the old prefix information and the "complete" flag would be sourced from a different LLA.

Initially, it would be good to speed up the detection of hardware replacement to delete the stale hardware from the default router list of hosts. It is proposed to request all routers availability by RS all-routers multicast address after new router detection on the link- see section 6.8. It would permit to detect that old hardware is not active in 13 seconds (see section 6.3.7 of [ND] for timers MAX_RTR_SOLICITATIONS * RTR_SOLICITATION_INTERVAL + MAX_RTR_SOLICITATION_DELAY). 13 seconds is considered a short enough outage compare to hardware replacement and reload.

Then it is proposed to detect stale prefixes at the event of the respective router deletion from the default router list. If the particular prefix is not announced anymore by any active router on the default router list then the prefix (and all associated addresses) should be deprecated - see section 6.9.

5.7. Link layer address of the router should be changed

Sections 6.3 and 6.4 provide an additional check also in the case of a link layer address change. Hence, additionally resolve LLA change case.

5.8. Dependency between solutions and extensions

It could be useful to map, for quick reference, the dependency between the solutions listed in this section and standard's extensions as presented in section 6.

Solution discussed in		Corresponding extension
5.1.	->	6.1.
5.2.	->	6.2. & 6.7.
5.3.	->	6.3. & 6.6.
5.4.	->	6.4.
5.5.	->	6.5. & 6.6.
5.6.	->	6.8. & 6.9.
5.7.	->	6.3. & 6.4.

6. Extensions of the existing standards

The solution requires a number of standard extensions. They are split into separate sections for better understanding. It is better to read references from section 5. before reading this section, see section 5.8. for cross-reference.

6.1. Default router choice by host

* Section 6.3.6 (Default Router Selection) of [ND], add an initial policy to default router selection:

0) For the cases when a particular implementation of ND does know the source address at the time of default router selection (it means that source address was chosen first), then default routers that advertise the prefix for respective source address SHOULD be preferred over routers that do not advertise respective prefix.

6.2. Prefixes become dynamic

* This document joins the request to [CPE Requirements] that has been proposed in section 11 (General Requirements for HNCP Nodes) of [HNCP]:

The requirement L-13 to deprecate prefixes is applied to all delegated prefixes in the network from which assignments have been made on the respective interface. Furthermore, the Prefix Information Options indicating deprecation MUST be included in Router Advertisements for the remainder of the prefixes' respective valid lifetime, but MAY be omitted after at least 2 hours have passed.

* Add section 4.2 into [SLAAC]:

4.2 Dynamic Link Renumbering

Prefix delegation (primarily by DHCP-PD) is adopted by the industry as the primary mechanism of PA address delegation in the fixed and mobile broadband environments, including cases of small business and branches of the big enterprises.
The delegated prefix is tied to dynamic link that has a considerable probability to be disconnected, especially in a mobile environment. The delegated prefix is losing the value if the remote site is disconnected from prefix provider - this fact should be propagated to all nodes on the disconnected site, including hosts. Information Options indicating deprecation (multicast RA with zero Preferred Lifetime) MUST be sent at least one time. It SHOULD be included in Router Advertisements for the remainder of the prefixes' respective valid lifetime but MAY be omitted after 2 hours of deprecation announcements.

There is a high probability that connectivity to the provider would be restored very soon then the prefix could be announced again to all nodes on the site.

There is the probability that in a small period of time the same problem would disconnect the site again (especially for mobile uplink). Such oscillation between available and not available provider could happen frequently that would flood the remote site with [ND] updates.
Dampening mechanism MAY be implemented to suppress oscillation: if the time between a particular prefix announcement and previous deprecation was less than DampeningCheck then delay the next prefix announcement for DampeningDelay and check the need for the prefix announcement after DampeningDelay seconds.
It is recommended for protocol designers to implement a dampening mechanism for protocols (like [HNCP]) that would be used to distribute prefix delegation inside the site to relieve the majority of site routers and the protocol itself from the processing of oscillating messages.

* Section 5.1 (Node Configuration Variables) of [SLAAC], add timers:

DampeningCheck - the time between prefix announcement and previous deprecation is checked against this value to decide about dampening need. The timer should use 16bit unsigned integer measured in seconds. The default value is 10 seconds.

DampeningDelay - the delay (penalty) for the next attempt to announce the same prefix again. The timer should use 16bit unsigned integer measured in seconds. The default value is 10 seconds.

These timers should be configurable like all other timers in [SLAAC] section 5.1.

6.3. Do not forget to deprecate prefixes on renumbering

* Section 4.1 (Site renumbering) of [SLAAC], add at the end:

A network administrator SHOULD avoid the situations when renumbering is done abruptly (with the time of transition that is less than the preferred time for the respective prefix). Situations could happen when it is not possible to archive the above-mentioned goal: (1) the prefix could be withdrawn by the administrator of another domain, (2) there could be the urgent need to change the prefix for reasons not related to networking, (3) prefix could be invalidated after some network event (example: loss of uplink that was used to receive this prefix), (4) L2 connection (VLAN or VPN) could be changed

abruptly by mistake or due to not a proper design.
Prefix deprecation MUST be signaled at least one time by multicast RA with Preferred Lifetime set to zero for respective PIO. It SHOULD be included in RA for the remainder of the prefixes' respective valid lifetime but MAY be omitted after 2 hours of deprecation announcements.

It is recommended for developers to check and enforce this rule in router's software: if an administrator, automated system, or other protocol would try to delete a particular prefix from the link and if that prefix has the preferred lifetime bigger than zero, then the software MUST automatically generate deprecation announcements according to the rules explained above.

System designer SHOULD make sure that in the case of abrupt change of logical connectivity at L2 (VLAN, VPN) new default router SHOULD deprecate stale prefixes inherited from the previous default router.

6.4. Do not forget to deprecate prefixes on shutdown

* Section 6.2.5 of [ND] starts from the definition of ceasing cases for the router on [ND] link. One additional reason SHOULD be added to the end of the list:

    - Link layer address of the interface should be changed.

* Section 6.2.5 (Ceasing To Be an Advertising Interface) and Section 6.2.8 (Link Local Address Change) of [ND] already discusses requirements of proper ceasing to be [ND] router advertising interface. It has requirements to announce zero for a default router lifetime. It is proposed to add at the end of both sections:

A router MUST also announce in above-mentioned announcements all previously advertised prefixes with zero Preferred LifeTime. Valid LifeTime should not be decreased from originally intended - current hosts sessions should have the possibility to be rerouted to the redundant router (if available).

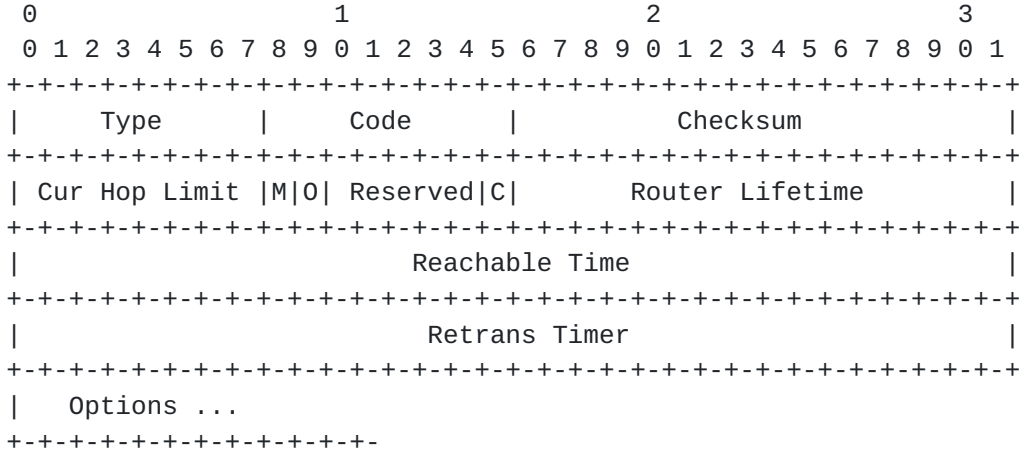
6.5. Store prefixes in non-volatile memory

Add the same text:
* [CPE Requirements], new requirement G-6 at the end of section 4.1, and
* [SLAAC], at the end of section 5.7:

The IPv6 router SHOULD keep in non-volatile memory all prefixes advertised on all links, including prefixes received by dynamic protocols with the reference to the respective protocol (DHCP-PD, HNCP, others).
A router could experience a non-graceful reload.
If another protocol would delegate any prefixes for router links then the router SHOULD immediately start announcing them in the normal way.
Additionally, the router should wait until the end of convergence for the respective prefix-delegation protocol. The way for how to decide that convergence is finished is the responsibility of the respective protocol design. It could be a simple timer after uplink would go to "up" or successful exchange by some protocol (like DHCP-PD).
If another protocol would not delegate prefix recorded in non-volatile memory after assumed convergence is achieved, then the old prefix MUST be announced on the link at least one time by multicast RA with the zero Preferred Lifetime. It SHOULD be included in RA for the remainder of the prefixes' respective valid lifetime but MAY be omitted after 2 hours of deprecation announcements.

6.6. Find lost information by "Synchronization"

* Section 4.2 (RA format) of [ND], introduce new flag:



O 1-bit "Complete configuration" flag. When set, it indicates that all configuration information has been put inside this RA. The last reserved bit has been chosen to preserve the compatibility with [Route Preferences] that already propose to use the first reserved bit.

* Section 6.2.3 (RA content) of [ND], introduce new flag:

- In the C flag: set if it was possible to put all configuration information into this RA.

* Section 6.2.3 (RA content) of [ND], add at the end:

It is recommended that all configuration information SHOULD be included in one RA (if MTU permits) for multicast and unicast distribution. If successful, then the "Complete" flag SHOULD be set to signal the possibility of synchronization with hosts.

* Section 6.3.4 (RA processing) of [ND], add at the beginning:

After: "the receipt of a Router Advertisement MUST NOT invalidate all information received in a previous advertisement or from another source".

Add: "Except for the case when RA received with "Complete" flag set, then any information from the same router (same Link Local Address) missing in this RA SHOULD be deprecated. Information protected by timers SHOULD be put into the deprecated state. Other information SHOULD be returned to the original state: in compliance to information from other routers or to default configuration if other routers do not announce respected information."

* Section 6.3.4 (RA processing) of [ND], add to the list of PIO processing options:

- If the prefix is missing in RA with the "Complete" flag set, then respective addresses should be put immediately into deprecated state up to the original valid lifetime.

[ND] section 9 mentions: "In order to ensure that future extensions properly coexist with current implementations, all nodes MUST

silently ignore any options they do not recognize in received ND packets and continue processing the packet."

There is a possibility for the gradual introduction of the "Complete" flag:

- If the host is upgraded to the new functionality first, then the router would send this bit zero (according to the basic [ND]) that would not activate new functionality on the host.

- If the router is upgraded to the new functionality first, then the host would not pay attention to the flag for Reserved bits.

6.7. Default router announcement rules

* This document joins [HNCP] section 11 (General Requirements for HNCP Nodes) request to [CPE Requirements]:

The generic requirements G-4 and G-5 are relaxed such that any known default router on any interface is sufficient for a router to announce itself as the default router; similarly, only the loss of all such default routers results in self-invalidation.

6.8. Faster detection of the stale default router

* Section 6.3.7 (sending Router Solicitations) of [ND].

The text: "When an interface becomes enabled, a host may be unwilling to wait for the next unsolicited Router Advertisement to locate default routers or learn prefixes. To obtain Router Advertisements quickly, a host SHOULD transmit up to MAX_RTR_SOLICITATIONS Router Solicitation messages, each separated by at least RTR_SOLICITATION_INTERVAL seconds. Router Solicitations may be sent after any of the following events:"

Should be replaced by the text: "

Interface enablement or new router arrival could be the signal of router replacement, a host may be unwilling to wait for the next unsolicited Router Advertisement to locate and invalidate default routers or learn prefixes. To obtain Router Advertisements quickly, a host SHOULD transmit up to MAX_RTR_SOLICITATIONS Router Solicitation messages, each separated by at least RTR_SOLICITATION_INTERVAL seconds. Router Solicitations may be sent after any of the following events:

- the new router is discovered from RA

- . . . <list of other reasons>
"

* Section 6.3.7 (sending Router Solicitations) of [ND].

After the text: "If a host sends MAX_RTR_SOLICITATIONS solicitations, and receives no Router Advertisements after having waited MAX_RTR_SOLICITATION_DELAY seconds after sending the last solicitation, the host concludes that there are no routers on the link for the purpose of [ADDRCONF]."

Add new text: "If a host sends MAX_RTR_SOLICITATIONS solicitations, and receives no Router Advertisements from the router already present on the default router list after having waited MAX_RTR_SOLICITATION_DELAY seconds, the host concludes that the router SHOULD be deprecated from the default router list."

6.9. Clean orphaned prefixes after default router list change

* Section 6.3.6 (Timing out Prefixes and Default Routers) of [ND] has:

"Whenever the Lifetime of an entry in the Default Router List expires, that entry is discarded. When removing a router from the Default Router list, the node MUST update the Destination Cache in such a way that all entries using the router perform next-hop determination again rather than continue sending traffic to the (deleted) router."

Add at the end:

"All prefixes announced by deprecated default router SHOULD be checked on the announcement from other default routers. If any prefix is not anymore announced from any router - it SHOULD be deprecated."

7. Interoperability analysis

The primary motivation for the proposed changes originated from residential broadband requirements. [ND] extensions proposed in this document should not affect other environments (enterprise WAN,

Campus). Moreover, some precautions proposed could block mistakes originated by humans in some corner cases in all environments.

This document mostly intersects with Homenet working group documents [HomeNet Architecture], [HNCP], and [MHMP]. It was shown that it is possible to isolate [ND] in the context of Homenet to solve specific [ND] problems without any potential impact to the Homenet development and directions.

[CPE Requirements] have the assumption of managing simplified topologies by manipulating routing information injection into [ND]. It has been shown in [MHMP] and in this document that it is better to signal reachability information to [ND] (reachability information to ND sounds strange) by the deprecation of delegated prefixes. This document joins [MHMP] request to change the approach.

[Route Preferences] have been avoided as the mechanism for environments with PA address space because source address is selected first. Then next hop choice can be simplified - see section 5.1 for more details.
[Route Preferences] could still be applicable for PI (Provider-Independent) address environments because only next hops need to be chosen properly.

8. Applicability analysis

Two standard extensions require changes to hosts. Hence, it would take a long time to be implemented in live networks. But workaround exists for the solution to work before it would happen:

- Absence of implementation for RA information synchronization by C flag on some hosts is not critical because router could use non-volatile memory for prefix storage.

- Not being capable of excluding a router from the default router list (for the situation when it does not advertise respective prefix) is not critical, because it is needed only for the very advanced MHMP environment with traffic distribution by the policy between different PA providers.
  It is for the far future anyway.

9. Security Considerations

This document does not introduce new vulnerabilities.

10. IANA Considerations

This document has no any request to IANA.

11. References

11.1. 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>.

[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>.

[ND] 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>.

[SLAAC] 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>.

[Route Preferences] R. Draves, D. Thaler, "Default Router Preferences and More-Specific Routes", RFC 4191, DOI 10.17487/RFC4191, November 2005, <https://www.rfc-editor.org/info/rfc4191>.

[Multi-Homing] F. Baker, 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>.

[NUD improvement] E. Nordmark, I. Gashinsky, "Neighbor Unreachability Detection Is Too Impatient", RFC 7048, DOI 10.17487/RFC7048, July 2010, <https://www.rfc-editor.org/info/rfc7048>.

[Default Address] D. Thaler, R. Draves, A. Matsumoto, 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>.

[Node Requirements] T. Chown, J. Loughney, T. Winters, "IPv6 Node Requirements", RFC 8504, DOI 10.17487/RFC8504, January 2019, <https://www.rfc-editor.org/info/rfc8504>.

[CPE Requirements] 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>.

[HomeNet Architecture] T. Chown, J. Arkko, A. Brandt, O. Troan, J. Weil, "IPv6 Home Networking Architecture Principles", RFC 7368, DOI 10.17487/RFC7368, October 2014, <https://www.rfc-editor.org/info/rfc7368>.

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

[Policy by DHCP] A. Matsumoto, T. Fujisaki, T. Chown, "Distributing Address Selection Policy Using DHCPv6", RFC 7078 DOI 10.17487/RFC7078, January 2014, <https://www.rfc-editor.org/info/rfc7078>.

[Residential practices] Palet, J., "IPv6 Deployment Survey Residential/Household Services) How IPv6 is being deployed?", UK NOF 39, January 2018, <https://indico.uknof.org.uk/event/41/contributions/542/attachments/712/866/bcop-ipv6-prefix-v9.pdf>.

[SLAAC Robustness] F. Gont, J. Zorz, R. Patterson, "Improving the Robustness of Stateless Address Autoconfiguration (SLAAC) to Flash Renumbering Events", draft-ietf-6man-slaac-renum-02 (work in progress), January 2021

11.2. Informative References

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

[Flash-Renumbering] F. Gont, J. Zorz, R. Patterson, "Reaction of Stateless Address Autoconfiguration (SLAAC) to Flash-Renumbering Events", RFC 8978, March 2021.

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

[RA-Guard+] F. Gont, "Implementation Advice for IPv6 Router Advertisement Guard (RA-Guard)", RFC 7113, DOI 10.17487/RFC7113, February 2014, <https://www.rfc-editor.org/info/rfc7113>.

[MHMP] O. Troan, D. Miles, S. Matsushima, T. Okimoto, D. Wing, "IPv6 Multihoming without Network Address Translation", RFC 7157, DOI 10.17487/RFC7157, March 2014, <https://www.rfc-editor.org/info/rfc7157>.

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

12. Acknowledgments

Thanks to 6man working group for problem discussion.

Authors' Addresses

Olorunloba Olopade
Virgin Media
270 & 280 Bartley Way, Bartley Wood Business Park, Hook,
Hampshire RG27 9UP
Email: Loba.Olopade@virginmedia.co.uk

Eduard Vasilenko
Huawei Technologies
17/4 Krylatskaya st, Moscow, Russia 121614
Email: vasilenko.eduard@huawei.com

Paolo Volpato
Huawei Technologies
Via Lorenteggio 240, 20147 Milan, Italy
Email: paolo.volpato@huawei.com