   IPv6 Subnet Model: The Relationship between Links and Subnet Prefixes

Abstract

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

Status of This Memo

   This is an Internet Standards Track document.

   This document is a product of the Internet Engineering Task Force
   (IETF).  It represents the consensus of the IETF community.  It has
   received public review and has been approved for publication by the
   Internet Engineering Steering Group (IESG).  Further information on
   Internet Standards is available in Section 2 of RFC 5741.

   Information about the current status of this document, any errata,
   and how to provide feedback on it may be obtained at
   http://www.rfc-editor.org/info/rfc5942.

Table of Contents

1.  Introduction

   IPv4 implementations typically associate a netmask with an address
   when an IPv4 address is assigned to an interface.  That netmask
   together with the IPv4 address designates an on-link prefix.  Nodes
   consider addresses covered by an on-link prefix to be directly
   attached to the same link as the sending node, i.e., they send
   traffic for such addresses directly rather than to a router.  See
   Section 3.3.1 of [RFC1122].  Prior to the development of subnetting
   [RFC0950] and Classless Inter-Domain Routing (CIDR) [RFC4632], an
   address's netmask could be derived directly from the address simply
   by determining whether it was a Class A, B, or C address.  Today,
   assigning an address to an interface also requires specifying a
   netmask to use.  In the absence of specifying a specific netmask when
   assigning an address, some implementations would fall back to
   deriving the netmask from the class of the address.

   The behavior of IPv6 as specified in Neighbor Discovery (ND)
   [RFC4861] is quite different.  The on-link determination is separate
   from the address assignment.  A host can have IPv6 addresses without

any related on-link prefixes or can have on-link prefixes that are
not related to any IPv6 addresses that are assigned to the host.  Any
assigned address on an interface should initially be considered as
having no internal structure as shown in [RFC4291].

In IPv6, by default, a host treats only the link-local prefix as
on-link.

The reception of a Prefix Information Option (PIO) with the L-bit set
[RFC4861] and a non-zero valid lifetime creates (or updates) an entry
in the Prefix List.  All prefixes on a host's Prefix List (i.e.,
those prefixes that have not yet timed out) are considered to be
on-link by that host.

The on-link definition in the Terminology section of [RFC4861], as
modified by this document, defines the complete list of cases in
which a host considers an address to be on-link.  Individual address
entries can be expired by the Neighbor Unreachability Detection
mechanism.

IPv6 packets sent using the Conceptual Sending Algorithm as described
in [RFC4861] only trigger address resolution for IPv6 addresses that
the sender considers to be on-link.  Packets to any other address are
sent to a default router.  If there is no default router, then the
node should send an ICMPv6 Destination Unreachable indication as
specified in [RFC4861] -- more details are provided in the "Host
Behavior" and "Host Rules" sections of this document.  (Note that
[RFC4861] changed the behavior when the Default Router List is empty.

In the old version of Neighbor Discovery [RFC2461], if the Default
Router List is empty, rather than sending the ICMPv6 Destination
Unreachable indication, the [RFC2461] node assumed that the
destination was on-link.)  Note that ND is scoped to a single link.
All Neighbor Solicitation (NS) responses are assumed to be sent out
the same interface on which the corresponding query was received
without using the Conceptual Sending Algorithm.

Failure of host implementations to correctly implement the IPv6
subnet model can result in lack of IPv6 connectivity.  See the
"Observed Incorrect Implementation Behavior" section for details.

This document deprecates the last two bullets from the definition of
"on-link" in [RFC4861] to address security concerns arising from
particular ND implementations.

Host behavior is clarified in the "Host Behavior" and "Host Rules"
sections.

2.  Requirements Language

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

3.  Host Behavior

   1.  The original Neighbor Discovery (ND) specification [RFC4861] was
       unclear in its usage of the term "on-link" in a few places.  In
       IPv6, an address is on-link (with respect to a specific link), if
       the address has been assigned to an interface attached to that
       link.  Any node attached to the link can send a datagram directly
       to an on-link address without forwarding the datagram through a
       router.  However, in order for a node to know that a destination
       is on-link, it must obtain configuration information to that
       effect.  In IPv6, there are two main ways of maintaining
       information about on-link destinations.  First, a host maintains
       a Prefix List that identifies ranges of addresses that are to be
       considered on-link.  Second, Redirects can identify individual
       destinations that are on-link; such Redirects update the
       Destination Cache.

       The Prefix List is populated via the following means:

       *  Receipt of a valid Router Advertisement (RA) that specifies a
          prefix with the L-bit set.  Such a prefix is considered
          on-link for a period specified in the Valid Lifetime and is
          added to the Prefix List.  (The link-local prefix is
          effectively considered a permanent entry on the Prefix List.)

       *  Indication of an on-link prefix (which may be a /128) via
          manual configuration, or some other yet-to-be-specified
          configuration mechanism.

       A Redirect can also signal whether an address is on-link.  If a
       host originates a packet, but the first-hop router routes the
       received packet back out onto the same link, the router also
       sends the host a Redirect.  If the Target and Destination Address
       of the Redirect are the same, the Target Address is to be treated
       as on-link as specified in Section 8 of [RFC4861].  That is, the
       host updates its Destination Cache (but not its Prefix List --
       though the impact is similar).

   2.  It should be noted that ND does not have a way to indicate a
       destination is "off-link".  Rather, a destination is assumed to
       be off-link, unless there is explicit information indicating that
       it is on-link.  Such information may later expire or be changed,

in which case a destination may revert back to being considered
off-link, but that is different than there being an explicit
mechanism for signaling that a destination is off-link.  Redirect
messages do not contain sufficient information to signal that an
address is off-link.  Instead, Redirect messages indicate a
preferred next hop that is a more appropriate choice to use than
the originator of the Redirect.

3.  IPv6 also defines the term "neighbor" to refer to nodes attached
    to the same link and that can send packets directly to each
    other.  Received ND packets that pass the required validation
    tests can only come from a neighbor attached to the link on which
    the ND packet was received.  Unfortunately, [RFC4861] is
    imprecise in its definition of "on-link" and states that a node
    considers an address to be on-link if:

    *  a Neighbor Advertisement (NA) message is received for the
       (target) address, or

    *  any Neighbor Discovery message is received from the address.

    Neither of these tests are acceptable definitions for an address
    to be considered as on-link as defined above, and this document
    deprecates and removes both of them from the formal definition of
    "on-link".  Neither of these tests should be used as
    justification for modifying the Prefix List or Destination Cache
    for an address.

    The conceptual sending algorithm of [RFC4861] defines a Prefix
    List, Destination Cache, and Default Router List.  The
    combination of Prefix List, Destination Cache, and Default Router
    List form what many implementations consider to be the IP data
    forwarding table for a host.  Note that the Neighbor Cache is a
    separate data structure referenced by the Destination Cache, but
    entries in the Neighbor Cache are not necessarily in the
    Destination Cache.  It is quite possible (and intentional) that
    entries be added to the Neighbor Cache for addresses that would
    not be considered on-link as defined above.  For example, upon
    receipt of a valid NS, Section 7.2.3 of [RFC4861] states:

       If an entry does not already exist, the node SHOULD create a
       new one and set its reachability state to STALE as specified
       in Section 7.3.3.  If an entry already exists, and the cached
       link-layer address differs from the one in the received Source
       Link-Layer option, the cached address should be replaced by
       the received address, and the entry's reachability state MUST
       be set to STALE.

The intention of the above feature is to add an address to the
Neighbor Cache, even though it might not be considered on-link
per the Prefix List.  The benefit of such a step is to have the
receiver populate the Neighbor Cache with an address it will
almost certainly be sending packets to shortly, thus avoiding the
need for an additional round of ND to perform address resolution.
But because there is no validation of the address being added to
the Neighbor Cache, an intruder could spoof the address and cause
a receiver to add an address for a remote site to its Neighbor
Cache.  This vulnerability is a specific instance of the broad
set of attacks that are possible by an on-link neighbor
[RFC3756].  This causes no problems in practice, so long as the
entry only exists in the Neighbor Cache and the address is not
considered to be on-link by the IP forwarding code (i.e., the
address is not added to the Prefix List and is not marked as
on-link in the Destination Cache).

4.  After the update to the on-link definition in [RFC4861], certain
    text from Section 7.2.3 of [RFC4861] may appear, upon a cursory
    examination, to be inconsistent with the updated definition of
    "on-link" because the text does not ensure that the source
    address is already deemed on-link through other methods:

        If the Source Address is not the unspecified address and, on
        link layers that have addresses, the solicitation includes a
        Source Link-Layer Address option, then the recipient SHOULD
        create or update the Neighbor Cache entry for the IP Source
        Address of the solicitation.

    Similarly, the following text from Section 6.2.6 of [RFC4861] may
    also seem inconsistent:

        If there is no existing Neighbor Cache entry for the
        solicitation's sender, the router creates one, installs the
        link-layer address and sets its reachability state to STALE as
        specified in Section 7.3.3.

    However, the text in the aforementioned sections of [RFC4861],
    upon closer inspection, is actually consistent with the
    deprecation of the last two bullets of the on-link definition
    because there are two different ways in which on-link
    determination can affect the state of ND: through updating the
    Prefix List or updating the Destination Cache.  Through
    deprecating the last two bullets of the on-link definition, the
    Prefix List is explicitly not to be changed when a node receives
    an NS, NA, or Router Solicitation (RS).  The Neighbor Cache can
    still be updated through receipt of an NS, NA, or RS.

   5.  [RFC4861] is written from the perspective of a host with a single
       interface on which Neighbor Discovery is run.  All ND traffic
       (whether sent or received) traverses the single interface.  On
       hosts with multiple interfaces, care must be taken to ensure that
       the scope of ND processing from one link stays local to that
       link.  That is, when responding to an NS, the NA would be sent
       out on the same link on which it was received.  Likewise, a host
       would not respond to a received NS for an address only assigned
       to an interface on a different link.  Although implementations
       may choose to implement Neighbor Discovery using a single data
       structure that merges the Neighbor Caches of all interfaces, an
       implementation's behavior must be consistent with the above
       model.

4.  Host Rules

   A correctly implemented IPv6 host MUST adhere to the following rules:

   1.  The assignment of an IPv6 address -- whether through IPv6
       stateless address autoconfiguration [RFC4862], DHCPv6 [RFC3315],
       or manual configuration -- MUST NOT implicitly cause a prefix
       derived from that address to be treated as on-link and added to
       the Prefix List.  A host considers a prefix to be on-link only
       through explicit means, such as those specified in the on-link
       definition in the Terminology section of [RFC4861] (as modified
       by this document) or via manual configuration.  Note that the
       requirement for manually configured addresses is not explicitly
       mentioned in [RFC4861].

   2.  In the absence of other sources of on-link information, including
       Redirects, if the RA advertises a prefix with the on-link (L) bit
       set and later the Valid Lifetime expires, the host MUST then
       consider addresses of the prefix to be off-link, as specified by
       the PIO paragraph of Section 6.3.4 of [RFC4861].

   3.  In the absence of other sources of on-link information, including
       Redirects, if the RA advertises a prefix with the on-link (L) bit
       set and later the Valid Lifetime expires, the host MUST then
       update its Prefix List with respect to the entry.  In most cases,
       this will result in the addresses covered by the prefix
       defaulting back to being considered off-link, as specified by the
       PIO paragraph of Section 6.3.4 of [RFC4861].  However, there are
       cases where an address could be covered by multiple entries in
       the Prefix List, where expiration of one prefix would result in
       destinations then being covered by a different entry.

4.  Implementations compliant with [RFC4861] MUST adhere to the
    following rules.  If the Default Router List is empty and there
    is no other source of on-link information about any address or
    prefix:

    a.  The host MUST NOT assume that all destinations are on-link.

    b.  The host MUST NOT perform address resolution for non-link-
        local addresses.

    c.  Since the host cannot assume the destination is on-link, and
        off-link traffic cannot be sent to a default router (since
        the Default Router List is empty), address resolution cannot
        be performed.  This case is specified in the last paragraph
        of Section 4 of [RFC4943]: when there is no route to the
        destination, the host should send an ICMPv6 Destination
        Unreachable indication (for example, a locally delivered
        error message) as specified in the Terminology section of
        [RFC4861].

    On-link information concerning particular addresses and prefixes
    can make those specific addresses and prefixes on-link, but does
    not change the default behavior mentioned above for addresses and
    prefixes not specified.  [RFC4943] provides justification for
    these rules.

5.  Observed Incorrect Implementation Behavior

    One incorrect implementation behavior illustrates the severe
    consequences when the IPv6 subnet model is not understood by the
    implementers of several popular host operating systems.  In an access
    concentrator network ([RFC4388]), a host receives a Router
    Advertisement message with no on-link prefix advertised.  An address
    could be acquired through the DHCPv6 identity association for non-
    temporary addresses (IA_NA) option from [RFC3315] (which does not
    include a prefix length), or through manual configuration (if no
    prefix length is specified).  The host incorrectly assumes an
    invented prefix is on-link.  This invented prefix typically is a /64
    that was written by the developer of the operating system network
    module API to any IPv6 application as a "default" prefix length when
    a length isn't specified.  This may cause the API to seem to work in
    the case of a network interface initiating stateless address
    autoconfiguration (SLAAC); however, it can cause connectivity
    problems in Non-Broadcast Multi-Access (NBMA) networks.  Having
    incorrectly assumed an invented prefix, the host performs address
    resolution when the host should send all non-link-local traffic to a

      default router.  Neither the router nor any other host will respond
      to the address resolution, preventing this host from sending IPv6
      traffic.

6.  Updates to RFC 4861

      This document deprecates the following two bullets from the on-link
      definition in Section 2.1 of [RFC4861]:

      o  a Neighbor Advertisement message is received for the (target)
         address, or

      o  any Neighbor Discovery message is received from the address.

7.  Conclusion

      This document clarifies and summarizes the relationship between links
      and subnet prefixes described in [RFC4861].  Configuration of an IPv6
      address does not imply the existence of corresponding on-link
      prefixes.  One should also look at API considerations for prefix
      length as described in the last paragraph of Section 4.2 of
      [RFC4903].  This document also updates the definition of "on-link"
      from [RFC4861] by deprecating the last two bullets.

8.  Security Considerations

      This document addresses a security concern present in [RFC4861].  As
      a result, the last two bullets of the on-link definition in [RFC4861]
      have been deprecated.  US-CERT Vulnerability Note VU#472363 lists the
      implementations affected.

9.  Contributors

      Thomas Narten contributed significant text and provided substantial
      guidance to the production of this document.

10.  Acknowledgements

      Thanks (in alphabetical order) to Adeel Ahmed, Jari Arkko, Ralph
      Droms, Alun Evans, Dave Forster, Prashanth Krishnamurthy, Suresh
      Krishnan, Josh Littlefield, Bert Manfredi, David Miles, Madhu Sudan,
      Jinmei Tatuya, Dave Thaler, Bernie Volz, and Vlad Yasevich for their
      consistent input, ideas, and review during the production of this
      document.  The security problem related to an NS message that
      provides one reason for invalidating a part of the on-link definition
      was found by David Miles.  Jinmei Tatuya found the security problem
      to also exist with an RS message.

## 11.  References

### 11.1.  Normative References

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

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

### 11.2.  Informative References

   [RFC0950]    Mogul, J. and J. Postel, "Internet Standard Subnetting
                Procedure", STD 5, RFC 950, August 1985.

   [RFC1122]    Braden, R., "Requirements for Internet Hosts -
                Communication Layers", STD 3, RFC 1122, October 1989.

   [RFC2461]    Narten, T., Nordmark, E., and W. Simpson, "Neighbor
                Discovery for IP Version 6 (IPv6)", RFC 2461,
                December 1998.

   [RFC3315]    Droms, R., Bound, J., Volz, B., Lemon, T., Perkins, C.,
                and M. Carney, "Dynamic Host Configuration Protocol for
                IPv6 (DHCPv6)", RFC 3315, July 2003.

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

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

   [RFC4388]    Woundy, R. and K. Kinnear, "Dynamic Host Configuration
                Protocol (DHCP) Leasequery", RFC 4388, February 2006.

   [RFC4632]    Fuller, V. and T. Li, "Classless Inter-domain Routing
                (CIDR): The Internet Address Assignment and Aggregation
                Plan", BCP 122, RFC 4632, August 2006.

   [RFC4862]    Thomson, S., Narten, T., and T. Jinmei, "IPv6 Stateless
                Address Autoconfiguration", RFC 4862, September 2007.

   [RFC4903]    Thaler, D., "Multi-Link Subnet Issues", RFC 4903,
                June 2007.

   [RFC4943]    Roy, S., Durand, A., and J. Paugh, "IPv6 Neighbor
                Discovery On-Link Assumption Considered Harmful",
                RFC 4943, September 2007.

Authors' Addresses

   Hemant Singh
   Cisco Systems, Inc.
   1414 Massachusetts Ave.
   Boxborough, MA  01719
   USA

   Phone: +1 978 936 1622
   EMail: shemant@cisco.com
   URI:   http://www.cisco.com/


   Wes Beebee
   Cisco Systems, Inc.
   1414 Massachusetts Ave.
   Boxborough, MA  01719
   USA

   Phone: +1 978 936 2030
   EMail: wbeebee@cisco.com
   URI:   http://www.cisco.com/

   Erik Nordmark
   Oracle, Inc.
   17 Network Circle
   Menlo Park, CA 94025
   USA

   Phone: +1 650 786 2921
   EMail: erik.nordmark@oracle.com