INTERNET-DRAFT
July 7, 1995

Thomas Narten, IBM
Erik Nordmark, Sun Microsystems
W A Simpson, Daydreamer

Neighbor Discovery for IP Version 6 (IPv6)

<draft-ietf-ipngwg-discovery-01.txt>

Status of this Memo

Abstract

This document specifies the Neighbor Discovery protocol for the IP Version 6 protocol. IPv6 nodes on the same link use Neighbor Discovery to discover each other's presence, to determine each other's link-layer addresses, to find routers and to maintain reachability information about the paths to active neighbors.

Contents

1. INTRODUCTION

This specification defines the Neighbor Discovery (ND) protocol for the IP Version 6 protocol. Nodes (hosts and routers) use Neighbor Discovery to determine the link-layer addresses for neighbors known to reside on attached links and to quickly purge cached values that become invalid.

Hosts also use Neighbor Discovery to find neighboring routers that are willing to forward packets on their behalf. Finally, nodes use the protocol to actively keep track of which neighbors are reachable and which are not, and to detect changed link-layer addresses. Sending hosts also detect when routers fail and actively search for functioning alternates.

This document is a revision of <draft-ietf-ipngwg-discovery-00.txt> which was itself a revision of the protocol specified in the two documents:

<draft-simpson-ipv6-discov-formats-02.txt>, and
<draft-simpson-ipv6-discov-process-02.txt>

The authors would like to acknowledge the contributions the IPNGWG working group an, in particular, (in alphabetical order) Ran Atkinson, Jim Bound, Scott Bradner, Stephen Deering, Robert Hinden, Allison Mankin, Dan McDonald, and Sue Thomson.

2. TERMINOLOGY

IP - Internet Protocol Version 6. The terms IPv4 and IPv6 are used only in contexts where necessary to avoid ambiguity.

ICMP - Internet Message Control Protocol for the Internet Protocol Version 6. The terms ICMPv4 and ICMPv6 are used only in contexts where necessary to avoid ambiguity.

node - a device that implements IP.

router - a node that forwards IP packets not explicitly addressed to itself.

host - any node that is not a router.

upper layer - a protocol layer immediately above IP. Examples are transport protocols such as TCP and UDP, control protocols such as ICMP, routing protocols such as OSPF, and internet or lower-layer protocols being "tunneled" over (i.e., encapsulated in) IP such as IPX, AppleTalk,

or IP itself.

link - a communication facility or medium over which nodes can communicate at the link layer, i.e., the layer immediately below IP. Examples are Ethernets (simple or bridged); PPP links; X.25, Frame Relay, or ATM networks; and internet (or higher) layer "tunnels", such as tunnels over IPv4 or IPv6 itself.

interface - a node's attachment to a link.

neighbors - nodes attached to the same link.

address - an IP-layer identifier for an interface or a set of interfaces.

designated address
- a per-interface address selected from the interface's assigned addresses. Only routers are required to have

designated addresses, which are used as the source address in certain Neighbor Discovery messages sent by the router from the interface. Neighboring nodes use the designated address to uniquely identify a router's interface, which might have many addresses assigned to it. The designated address should change only infrequently and must be a link-local address.

anycast address
- an identifier for a set of interfaces (typically belonging to different nodes). A packet sent to an anycast address is delivered to one of the interfaces identified by that address (the "nearest" one, according to the routing protocols' measure of distance). See [ADDR-ARCH].

link-layer address
- a link-layer identifier for an interface. Examples include IEEE 802 addresses for Ethernet links and E.164 addresses for ISDN links.

on-link - an address that is assigned to a neighbor's interface

on a shared link. A host considers an address to be on-link if:
- it is covered by one of the link's prefixes, or
- a neighboring router specifies the address as the target of a Redirect message, or
- a Neighbor Advertisement message is received for the target address, or
- a Router Advertisement message is received from the address.

off-link - the opposite of "on-link"; an address that is not assigned to any interfaces attached to a shared link.

reachability
- whether or not packets sent by an IP node are properly reaching a neighboring node. For routers, reachability means that packets sent by a node's IP layer are delivered to the router's IP layer, and the router is indeed forwarding packets (i.e., it is configured as a router, not a host). For hosts, reachability means that packets sent by a node's IP layer are delivered to the neighbor host's IP layer. Note that reachability only applies to the one-way "forward" path from a node to a neighbor.

packet - an IP header plus payload.

link MTU - the maximum transmission unit, i.e., maximum packet size in octets, that can be conveyed in one piece over a link.

target - an address about which address resolution information is sought, or an address which is the new first-hop when being redirected.

proxy - a router that responds to Neighbor Discovery query messages on behalf of another node. A router acting on behalf of a mobile node that has moved off-link potentially acts as a proxy for the mobile node.

ICMP destination unreachable indication
- an error indication returned to the original sender of

a packet that cannot be delivered for the reasons outlined in [ICMPv6]. If the error occurs on a node other than the node originating the packet, an ICMP error message is generated. If the error occurs on the originating node, an implementation is not required to actually create and send an ICMP error packet to the source, as long as the sender is notified through an appropriate mechanism (e.g., return value from a procedure call). Note, however, that an implementation may find it convenient in some cases to return errors to the sender by taking the offending packet, generating an ICMP error message, and then delivering it (locally) through the generic error handling routines.

Different link layers have different properties. The ones of concern to Neighbor Discovery are:

point-to-point
- a link that connects exactly two interfaces.

multicast - a link that supports some mechanism at the link layer for sending packets to all (i.e. broadcast) or a subset of all neighbors. Multicast/broadcast can be provided by a variety of link layer mechanisms such as the physical link layer itself (for example, Ethernet), replicated unicast packets sent by the link layer software, or multicast servers (such as in ATM). Note that all point-to-point links are multicast links.

non-broadcast multi-access (NBMA)
- a link with more than two neighbors that does not

support any form of multicast or broadcast (e.g., Frame Relay).

shared media - a link that allows direct communication among a number of nodes, but attached nodes are configured in such a way that they do not complete prefix information about all on-link destinations. Examples are large (switched) public data networks

such as SMDS and B-ISDN. Also known as "large clouds". See [SH-MEDIA].

variable MTU - a link that does not have a well-defined MTU. For example, Token Ring (IEEE 802.5). Other links like Ethernet have a standard MTU defined by the link-layer protocol.

asymmetric reachability
- a link where non-reflexive and/or non-transitive reachability is part of normal operation. (Non-reflexive reachability means packets from A reach B but packets from B don't reach A. Non-transitive reachability means packets from A reach B, and packets from B reach C, but packets from A don't reach C.) Many radio links exhibit these properties.

Neighbor Discovery makes use of a number of different addresses defined in [ADDR-ARCH], including:

all-nodes multicast address
- the link scope address to reach all nodes. FF02::1

all-routers multicast address
- the link scope address to reach all routers. FF02::2

solicited-node multicast address
- a multicast address that is computed as a function of the solicited target's address. The solicited-node multicast address is formed by taking the low-order 32 bits of the IP address and appending those bits to the 96-bit prefix FF02:0:0:0:0:1, resulting in a multicast address in the range FF02::1:0:0 to FF02::1:FFFF:FFFF. For example, the solicited node multicast address corresponding to the IP address 4037::01:800:200E:8C6C is FF02::1:200E:8C6C. IP addresses that differ only in the high-order bits, e.g. due to multiple high-order prefixes associated with different providers, will map

to the same solicited-node address thereby reducing the

number of multicast addresses a node must join.

unspecified address
- the address 0:0:0:0:0:0:0:0. It indicates the absence of an address. One example of its use is in the Source Address field of Neighbor Solicitation messages sent by an initializing host verifying that an address is unique (e.g., no other node is already using the address) [ADDRCONF].

3. PROTOCOL OVERVIEW

This protocol solves a set of problems related to the interaction between nodes attached to the same link. It defines mechanisms for solving each of the following problems:

Router Discovery: How hosts locate routers that reside on an attached link.

Prefix Discovery: How hosts discover the set of address prefixes that define which destinations are on-link for an attached link. (Nodes use prefixes to distinguish destinations that reside on-link from those only reachable through a router.)

Parameter Discovery: How a node learns such link parameters as the link MTU or such Internet parameters as the maximum hop limit value to place in outgoing packets, etc.

Address Autoconfiguration: How nodes automatically configure an address for an interface.

Address Resolution: How nodes determine the link-layer address of a neighboring node given only the node's IP address.

Next-hop determination: The algorithm for mapping an IP destination address into the IP address of the neighbor to which traffic for the destination should be sent. The next-hop can be a router or the destination itself.

Neighbor Unreachability Detection: How nodes determine that a neighbor is no longer reachable. For neighbors used as routers, alternate default routers can be tried. For both routers and hosts, address resolution can be performed again.

Duplicate Address Detection: How a node detects if another node has been configured to use the same IP address.

Redirect: How a router informs a host of a better first-hop node to reach a particular destination.

Neighbor Discovery defines five different ICMP packet types: A pair of Router Solicitation and Router Advertisement messages, a pair of Neighbor Solicitation and Neighbor Advertisements messages, and a Redirect message. The messages serve the following purpose:

Router Solicitation: When an interface becomes enabled, hosts may send out Router Solicitations that request routers to generate Router Advertisements immediately rather than at their next scheduled time.

Router Advertisement: Routers advertise their presence together with various link and Internet parameters either periodically, or in response to an explicit Router Solicitation message. Router Advertisements contain prefixes that are used for on-link determination and/or address configuration, a Maximum Hop Limit value, etc.

Neighbor Solicitation: Sent by a node to determine the link-layer address of a neighbor, or to verify that a neighbor is still reachable via a cached link-layer address. Neighbor Solicitations are also used for Duplicate Address Detection.

Neighbor Advertisement: A response to a Neighbor Solicitation message. A node may also send unsolicited Neighbor Advertisements to announce a link-layer address change.

Redirect: Used by routers to inform hosts of a better first hop for a destination.

On multicast-capable links, each router periodically multicasts a Router Advertisement packet announcing its availability. A host receives Router Advertisements from all routers, building a list of default routers. Routers generate Router Advertisements frequently enough that hosts will learn of their presence within a few minutes, but not frequently enough to rely on an absence of advertisements to detect router failure; a separate Neighbor Unreachability Detection algorithm handles this condition.

Router Advertisements contain a list of prefixes that can be used for on-link determination and/or autonomous address configuration; flags in

the prefixes specify the intended uses of a particular prefix. Hosts use the advertised on-link prefixes to build and maintain a list that is used in deciding when a packet's destination is on-link or beyond a router. Note that a destination can be on-link even though it is not covered by any advertised on-link prefix. In such cases a router can send a Redirect informing the sender that the destination is a neighbor.

Router Advertisements (and per-Prefix flags) allow routers to inform hosts how to perform Address Autoconfiguration. For examples, routers can specify whether hosts should use stateful (DHCPv6) or autonomous (stateless) address configuration. Router Advertisement messages also specify lifetimes for addresses that are configured using autonomous address configuration. The exact semantics and usage of the address configuration-related information is specified in [ADDRCONF].

Router Advertisement messages also contain Internet parameters such as the maximum hop that hosts should use in outgoing packets and, optionally, link parameters such as the link MTU. This facilitates centralized administration of critical parameters that can be set on routers and automatically propagated to all attached hosts.

Nodes accomplish Address Resolution by multicasting a Neighbor Solicitation that asks the target node to return its link-layer address. Neighbor Solicitation messages are multicast to the solicited-node multicast address corresponding to the target address. The target returns its link-layer address in a unicast Neighbor Advertisement message. A single request-response pair of packets is sufficient for both the initiator and the target to resolve each other's link-layer addresses; the initiator includes its link-layer address in the Neighbor Solicitation.

Neighbor Solicitation messages can also be used to determine if more than one node has been configured to use a particular unicast address. The use of Neighbor Solicitation messages for Duplicate Address Detection is specified in [ADDRCONF].

Neighbor Unreachability Detection detects both the failure of a neighbor or the failure of the forward path to the neighbor. Doing so requires positive confirmation that packets sent to a neighbor are actually

reaching that neighbor and being processed properly by its IP layer. Neighbor Unreachability Detection uses confirmation from two sources. When possible, upper-layer protocols provide a positive confirmation that a connection is making "forward progress", that is, previously sent data is known to have been delivered correctly (e.g., new acknowledgments were received recently). When positive confirmation is not forthcoming through such "hints", a node sends explicit unicast Neighbor Solicitation messages that solicit Neighbor Advertisements as reachability confirmation from the next hop. To reduce unnecessary

network traffic, probe messages are only sent to neighbors to which the node is actively sending packets.

In addition to addressing the above general problems, Neighbor Discovery also handles the following situations:

Link-layer address change - A node that knows its link-layer address has changed can multicast a few (unsolicited) Neighbor Advertisement packets to all nodes to quickly (but unreliably) update cached link-layer addresses that have become invalid. The Neighbor Unreachability Detection algorithm ensures that all nodes will reliably discover the new address, though the delay may be somewhat longer.

Inbound load balancing - Nodes with replicated interfaces may want to load balance the reception of incoming packets across multiple network interfaces on the same link. Such nodes have multiple link-layer addresses assigned to the same interface. For example, a single network driver could represent multiple network interface cards as a single logical interface having multiple link-layer addresses. Load balancing is handled by allowing routers to omit the source link-layer address from Router Advertisement packets, thereby forcing neighbors to use Neighbor Solicitation messages to learn the link-layer addresses. Returned Neighbor Advertisement messages can then contain different link-layer addresses dependent on who issued the solicitation.

Anycast addresses - Anycast addresses identify one of a set of nodes providing an equivalent service, and multiple nodes on the same link may be configured to recognize the same Anycast address. Neighbor Discovery handles the case when a node

determines that an anycast address is on-link and sends a Neighbor Solicitation. The potentially multiple Neighbor Advertisements for the anycast address will be identified as anycast/proxy responses. When multiple such advertisements are received, rules specify precedence and how to break ties.

Proxy advertisements - A router willing to accept packets on behalf of a target address that is unable to respond to Neighbor Solicitations can issue proxy Neighbor Advertisements. There is currently no specified use of proxy, but proxy advertising could potentially be used to handle cases like mobile nodes that have moved off-link. However, it is not intended as a general mechanism to handle nodes that, e.g., do not implement this protocol. Proxy advertisements invoke the same precedence and tie-breaking rules as does Anycast.

3.1. Comparison with IPv4

The IPv6 Neighbor Discovery protocol corresponds to a combination of the IPv4 protocols ARP [ARP], ICMP Router Discovery [RDISC], and ICMP Redirect [ICMPv4]. In IPv4 there is no generally agreed upon protocol or mechanism for Neighbor Unreachability Detection, although the Hosts Requirements [HR-CL] does specify some possible algorithms for Dead Gateway Detection (which address only a subset of the problems Neighbor Unreachability Detection tackles).

The Neighbor Discovery protocol provides a multitude of improvements over the IPv4 set of protocols:

Router Discovery is part of the base protocol set; there is no need for hosts to "snoop" the routing protocols.

Router advertisements carry link-layer addresses; no additional packet exchange is needed to resolve the router's link-layer address.

Router advertisements carry prefixes for a link; there is no need to have a separate mechanism to configure the "netmask".

Router advertisements contain hooks for Address Autoconfiguration.

By default, hosts learn all on-link prefixes from Router Advertisements. However, routers may be configured to omit some or all prefixes from Router Advertisements. In such cases hosts will assume that destinations are off-link and send traffic to routers by default. A router can then issue redirects for on-link destinations as appropriate. This mechanism may be useful on shared media links where it is undesirable or not possible for nodes to know all prefixes for on-link destinations.

Routers can advertise an MTU for hosts to use on the link, ensuring that all nodes use the same MTU value on links lacking a well-defined MTU.

Address Resolution uses multicast "spread" over 4 billion (2^32) multicast addresses resulting in greatly reduced Address Resolution related interrupts for nodes other than the target and generates no interrupts on non-IPv6 nodes.

Redirects contain the link-layer address of the new first hop; separate Address Resolution is not needed upon receiving a redirect.

Nodes assume that the new next-hop target address in a Redirect is

on-link making it possible to redirect to targets that do not share a common address prefix with the sender. This is an implementation of the XRedirect idea in [SH-MEDIA], which simplifies some aspects of neighbor interaction on shared media.

Neighbor Unreachability Detection is part of the base, significantly improving the robustness of packet delivery in the presence of failing routers, partially failing or partitioned links and nodes that change their link-layer addresses. For instance, mobile nodes can move off-link without loosing any connectivity due to stale ARP caches.

Unlike ARP, Neighbor Discovery detects half-link failures and tries to avoid using neighbors with which there is not two-way connectivity.

Placing address resolution at the ICMP layer makes the protocol more media-independent than ARP and makes it possible to use

standard IP authentication and security mechanisms as appropriate [IPv6-AUTH, IPv6-ESP].

3.2. Supported Link Types

Neighbor Discovery supports links with different properties. In the presence of certain properties only a subset of the ND protocol is available:

point-to-point - Neighbor Discovery handles such links just like multicast links. (Multicast can be trivially provided on point to point links, and interfaces can be assigned link-local addresses.)

multicast - All aspects of Neighbor Discovery are available.

non-broadcast multiple access (NBMA)
- The only Neighbor Discovery mechanisms available on these links are Redirect handling and Neighbor Unreachability Detection.

  If the hosts support manual configuration of a list of default routers the hosts can dynamically acquire the link-layer addresses for their neighbors from Redirect messages.

shared media - The Redirect message is modeled after the XRedirect

message in [SH-MEDIA] in order to simplify use of the protocol on shared media links.

This specification does not address shared media issues that only relate to routers, such as:

- How routers exchange reachability information on a shared media link.

- How a router determines the link-layer address of a host, which it needs to send redirect messages

to the host.

- How a router determines that it is the first hop router for a received packet.

The protocol is extensible (through the definition of new options) so that other solutions might be possible in the future.

variable MTU - Neighbor Discovery allows the routers to specify a MTU for the link. This allows all nodes to use the same MTU. Note: It is not possible to have each node use a different MTU (or Maximum Receive Unit) due to multicast.

asymmetric reachability
- Neighbor Discovery detects the absence of symmetric reachability; a node avoids using a neighbor with which it does not have symmetric connectivity.

The protocol can presumably be extended in the future to find viable paths in environments that lack reflexive and transitive connectivity.

4. CONCEPTUAL MODEL OF A HOST

This section describes a conceptual model of one possible data structure organization that hosts (and to some extent routers) will maintain in interacting with neighboring nodes. The described organization is provided to facilitate the explanation of how the Neighbor Discovery protocol should behave. This document does not mandate that implementations adhere to this model as long as their behavior is consistent with the protocol specification.

This model is only concerned with the aspects of host behavior directly

related to Neighbor Discovery. In particular, it does not concern itself with such issues as source address selection or the selecting of an outgoing interface on a multihomed host.

4.1. Conceptual Data Structures

Hosts will need to maintain the following pieces of information about an interface:

Neighbor Cache - A set of entries about individual neighbors to which traffic has been sent recently. Entries are keyed on the neighbor's IP address and contain such information as its link-layer address, a flag indicating whether the neighbor is a router or a host (called "is_router" in this document), a pointer to any queued packets waiting for Address Resolution to complete, etc.

A Neighbor Cache entry also contains information used by the Neighbor Unreachability Detection algorithm. This includes the reachability state, the number of unanswered probes, and the time the next Neighbor Unreachability Detection event is scheduled to take place.

Destination Cache - A set of entries for each destination to which traffic has been sent recently. The Destination Cache includes both on-link and off-link destinations and provides a level of indirection into the Neighbor Cache; the Destination Cache maps a destination IP address to the IP address of the next-hop neighbor. Implementations may find it convenient to store additional information not directly related to Neighbor Discovery in destination entries, such as the Path MTU (PMTU) and round trip timers maintained by transport protocols.

Prefix List - A list of the prefixes that define a set of addresses that are on-link. Prefix list entries are created from information received in Router Advertisements. Each entry has an associated invalidation timer value (extracted from the advertisement) used to delete prefixes that routers stop advertising. The invalidation timer can have a value of infinity to make sure the prefix remains valid indefinitely until it is explicitly advertised

with a lower timer value.

Default Router List
- A list of routers to which packets may be sent. Router list entries point to entries in the Neighbor Cache so that when a router is being selected, routers known to be reachable can be favored over those whose reachability is suspect. Each entry also has an associated invalidation timer value (extracted from Router Advertisements) used to delete entries that are no longer advertised.

Note that the above conceptual data structures can be implemented using a variety of techniques. One possible implementation is to use a single longest-match routing table for all of the above data structures.

The Neighbor Cache contains information maintained by the Neighbor Unreachability Detection algorithm. A key piece of information is a neighbor's reachability state which has three possible values:

INCOMPLETE Address Resolution is in progress and the link-layer address of the neighbor has not yet been determined.

REACHABLE Roughly speaking, the neighbor is known to have been reachable recently (within tens of seconds ago).

PROBE The neighbor is probably reachable, but the last explicit reachability confirmation was received long enough ago that verification is now actively sought.

4.2. Conceptual Sending Algorithm

When sending a packet, a node uses a combination of the Destination Cache, the Prefix List, and the Default Router List to determine the IP address of the appropriate next hop, an operation known as "next-hop determination". Once the IP address of the next hop is known, the Neighbor Cache is consulted for link-level information about that neighbor.

Next-hop determination operates as follows. The sender examines the Prefix List to determine whether the packet's destination is on- or off-link. If the destination is on-link, the next-hop address is the same as the packet's destination address. If the destination is off-link, the sender selects a router from the Default Router List (following the rules described in Section 5.3.4). If there are no

routers on the Default Router List, the sender assumes that the

destination is on-link.

For efficiency reasons, next-hop determination is not performed on every packet that is sent. Instead, the results of next-hop determination computations are saved in the Destination Cache. When the sending node has a packet to send, it first examines the Destination Cache. If no entry exists for the destination, next-hop determination is invoked to create a Destination Cache entry.

Once the IP address of the next-hop node is known, the sender examines the Neighbor Cache for link-level information about that neighbor. If no entry exists, the node creates an INCOMPLETE entry, sends a Neighbor Solicitation message, and queues the packet pending completion of Address Resolution. When a Neighbor Advertisement response is received, the link-layer addresses is entered in the Neighbor Cache entry and the queued packet is transmitted. The Address Resolution mechanism is described in detail in Section 6.2.

Each time a Neighbor Cache entry is accessed while transmitting a packet, the sender checks Neighbor Unreachability Detection related information according to the Neighbor Unreachability Detection algorithm (Section 6.3). This check might result in the sender transmitting a Neighbor Solicitation to verify that the neighbor is still reachable.

Next-hop determination is done the first time traffic is sent to a destination. As long as subsequent communication to that destination proceeds successfully, the Destination Cache entry continues to be used. If at some point communication ceases to proceed, as determined by the Neighbor Unreachability Detection algorithm, next-hop determination may need to be performed again. For example, traffic through a failed router should be switched to a working router. Likewise, it may be possible to reroute traffic destined for a mobile node to a "mobility agent".

Note that when a node redoes next-hop determination there is no need to discard the complete Destination Cache entry. In fact, it is generally beneficial to retain such cached information as the PMTU and round trip timer values that may also be kept in the Destination Cache entry.

4.3. Garbage Collection and Timeout Requirements

The conceptual data structures described above use different mechanisms for discarding potentially stale, as well as unused, information.

From the perspective of correctness, there is no need to periodically purge Destination and Neighbor Cache entries. Although stale information can potentially remain in the cache indefinitely, the

Neighbor Unreachability Detection algorithm described in this document ensures that stale information is purged quickly if it is actually being used.

To limit the storage needed for the Destination and Neighbor Caches, a node may need to garbage-collect old entries. However, care must be taken to insure that sufficient space is always present to hold the working set of active entries. A small cache may result in an excessive number of Neighbor Discovery messages as discarded entries are discarded and rebuilt in quick succession. Any LRU-based policy that only reclaims entries that have not been used in some time (e.g, ten minutes or more) should be adequate for garbage-collecting unused entries.

A node should retain entries in the Default Router List and the Prefix List until their lifetimes expire. However, a node may garbage collect entries prematurely if it is low on memory. If not all routers are kept on the Default Router list, a node should retain at least two entries in the Default Router List (and preferably more) in order to maintain robust connectivity for off-link destinations.

When removing an entry from the Default Router List or the Prefix List there is no need to purge any entries from the Destination or Neighbor Caches. Neighbor Unreachability Detection will effectively purge any entries in these caches that have become stale.

5. ROUTER AND PREFIX DISCOVERY

This section describes message formats, router behavior and host behavior related to the Router Discovery portion of Neighbor Discovery. Router Discovery is used to locate neighboring routers as well as learn prefixes and configuration parameters related to address

autoconfiguration.

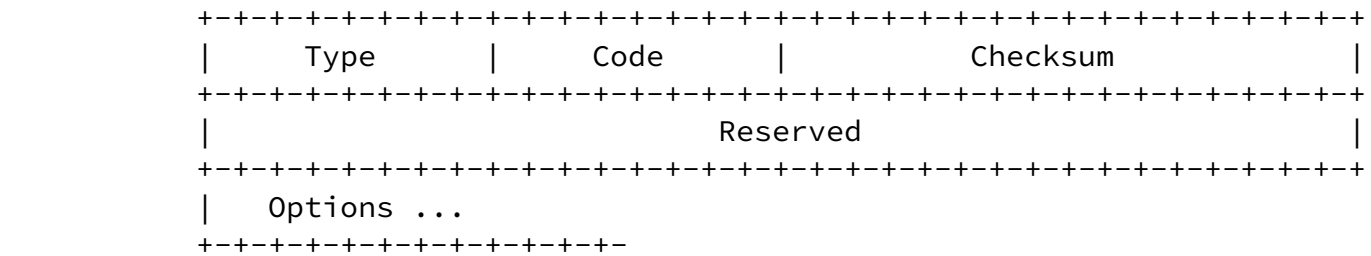
Prefix Discovery provides a mechanism through which hosts learn of ranges of IP addresses that reside on-link and thus can be reached directly without going through a router. Routers advertise a set of prefixes that cover those IP addresses that are on-link. Prefix discovery is logically separate from Router Discovery. In practice, prefix information is included in options piggybacked on Router Advertisement messages to reduce network traffic.

Address Autoconfiguration information is also logically separate from Router Discovery. To reduce network traffic, however, autoconfiguration information is piggybacked on Router Discovery messages. In fact, the same prefixes can be advertised for on-link determination and address autoconfiguration by specifying the appropriate flags in the Prefix

Information options. This document does not define how autoconfiguration information is processed. See [ADDRCONF] for details.

5.1. Message Formats

5.1.1. Router Solicitation Message Format



IP Fields:

Source Address
: MUST be the link-local address belonging to the interface from which this message is sent.

Destination Address
: The all-routers multicast address.

Hop Count 1

Authentication Header
If a security association exists between the sender and the destination the sender SHOULD include this header.

ICMP Fields:

Type 133

Code 0

Checksum The ICMP checksum. See [ICMPv6].

Reserved This field is unused. It MUST be initialized to zero by the sender and ignored by the receiver.

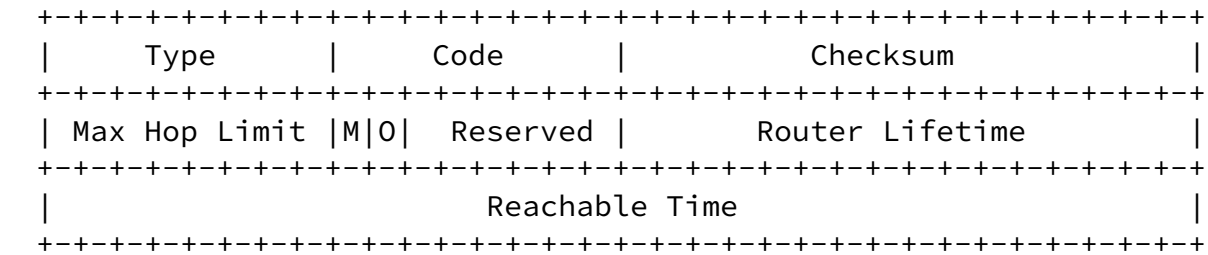
Options:

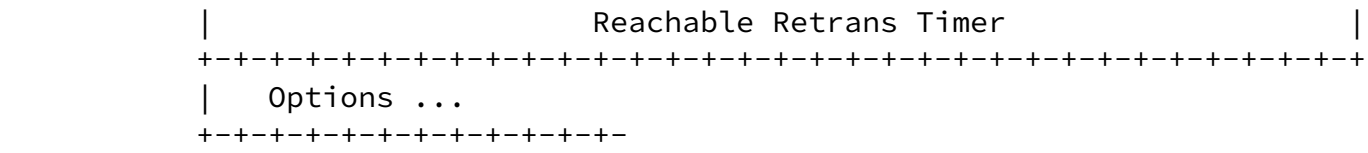
Source link-layer address
The link-layer address for the sender. SHOULD be

included on link layers that have addresses in order for the router to be able to send a Router Advertisement without having to perform address resolution on the host's address.

Future versions of this protocol may define new option types. Receivers MUST skip over and ignore any options they do not recognize and continue processing the message.

5.1.2. Router Advertisement Message Format





IP Fields:

Source Address
: The interface's designated IP address.

Destination Address
: Either the Source Address of an invoking Router Solicitation or the all-nodes multicast address.

Hop Count
: 1

Authentication Header
: If a security association exists between the sender and the destination the sender SHOULD include this header.

ICMP Fields:

Type
: 134

Code
: 0

Checksum
: The ICMP checksum. See [ICMPv6].

Max Hop Limit
: 8-bit unsigned integer. The maximum hop limit that the router suggests that hosts use when sending IP packets. A value of zero means unspecified.

M
: 1-bit "Managed address configuration" flag. Use the administered (stateful) protocol for address autoconfiguration in addition to any addresses autoconfigured using stateless address autoconfiguration. The use of this flag is described in [ADDRCONF].

O
: 1-bit "Other configuration" flag. Use the

administered (stateful) protocol for autoconfiguration of other (non-address) information. The use of this flag is described in [ADDRCONF].

Reserved
: A 6-bit unused field. It MUST be initialized to zero by the sender and ignored by the receiver.

Router Lifetime
: 16-bit unsigned integer. The lifetime associated with the default router in units of seconds. The maximum value corresponds to 18.2 hours. This lifetime does not apply to information contained in any options in the message. Options that need time limits for their information include their own lifetime fields.

Reachable Time
: 32-bit unsigned integer. The time, in milliseconds, that a node considers a neighbor being reachable after receiving some reachability confirmation. Used by the Neighbor Unreachability Detection algorithm. A value of zero means unspecified.

Reachable Retrans Timer
: 32-bit unsigned integer. The time, in milliseconds, between retransmitted Neighbor Solicitation probes to a neighbor that is not returning solicited Neighbor Advertisements. Used by the Neighbor Unreachability Detection algorithm. A value of zero means unspecified.

Options:

Source link-layer address
: The link-layer address for the router. Only used on link layers that have addresses. A router MAY omit this option in order to enable inbound load sharing across multiple link-layer addresses.

MTU
: SHOULD be sent on links that have a variable MTU. MAY be sent on other links.

Prefix Information
: A router MAY include 0 or more Prefix Information

options. These options specify the prefixes that are on-link and/or are used for address autoconfiguration. A router SHOULD include all on-link prefixes on multicast links. This enables multihomed hosts to do optimal outgoing interface selection for neighboring nodes.

Future versions of this protocol may define new option types. Receivers MUST skip over and ignore any options they do not recognize and continue processing the message.

5.2. Router Specification

5.2.1. Router Configuration Variables

A router MUST allow for the following variables to be configured by system management; default values are specified so as to make it unnecessary to configure any of these variables in many cases.

For each multicast interface:

AdvertiseDefault
: A flag indicating whether or not the router should advertise itself as a default router on the interface.

  Default: TRUE

ManagedFlag
: The true/false value to be placed in the "Managed address configuration" field in the Router Advertisement. See [ADDRCONF].

  Default: FALSE

OtherFlag
: The true/false value to be placed in the "Other configuration" field in the Router Advertisement. See [ADDRCONF].

  Default: FALSE

DesignatedAddress
: The address to be used as the source address in certain Neighbor Discovery messages sent on the interface.

  Default: The interface's link-local address.

LinkMTU
: The value to be placed in MTU options sent by the router. If the value is set to zero no MTU options are sent.

  Default: 0

ReachableTime
: The value to be placed in the Reachable Time field in the Router Advertisement messages sent by the router. The value zero means unspecified (by this router).

  Default: REACHABLE_TIME milliseconds

ReachableRetransTimer
: The value to be placed in the Reachable Retrans Timer field in the Router Advertisement messages sent by the router. The value zero means unspecified (by this router).

  Default: REACHABLE_RETRANS_TIMER milliseconds

MaximumHopLimit
: The value to be placed in the Max Hop Limit field in the Router Advertisement messages sent by the router. The value zero means unspecified (by this router).

  Default: The value specified in the most recent "Assigned Numbers" RFC [ASSIGNED].

MaxRtrAdvInterval
: The maximum time allowed between sending multicast Router Advertisements from the interface, in seconds. MUST be no less than 1 second and no greater than 1800 seconds.

  Default: 600 seconds

MinRtrAdvInterval
: The minimum time allowed between sending unsolicited

multicast Router Advertisements from the interface, in seconds. MUST be no less than 0.1 seconds and no greater than MaxRtrAdvInterval.

Default: 0.33 * MaxRtrAdvInterval

RtrAdvLifetime

The value to be placed in the Router Lifetime field of Router Advertisements sent from the interface, in seconds. MUST be no less than MaxRtrAdvInterval and no greater than 9000 seconds. Note: if AdvertiseDefault is false, the value of RtrAdvLifetime is irrelevant; a Lifetime value of 0 in Router Advertisements indicates that the router should not be used as a default router.

Default: 3 * MaxRtrAdvInterval

PrefixList

A list of prefixes to be placed in Prefix Information options in Router Advertisement messages sent from the interface.

Default: The PrefixList contains all prefixes that the router advertises via routing protocols as being on the link on which the advertisement is sent.

Each prefix is associated with:

- InvalidationLifetime
  The value to be placed in the Invalidation Lifetime in the Prefix Information option, in seconds. A designated value (like all-one bits) can be used to represent infinity.

  Default: infinity.

- OnLinkFlag
  The value to be placed in the on-link flag ("L-bit") field in the Prefix Information option.

Default: TRUE

Automatic address configuration [ADDRCONF] defines additional information associated with each the prefixes:

DeprecationLifetime
The value to be placed in the Deprecation Lifetime in the Prefix Information option, in seconds. A designated value (like all-one bits) can be used to represent infinity. See [ADDRCONF].

Default: 604800 seconds (7 days)

AutonomousFlag
The value to be placed in the Autonomous Flag field in the Prefix Information option. See [ADDRCONF].

Default: TRUE

Protocol constants are defined in Section 10.

5.2.2. Message Validation by Routers

A router MUST silently discard any received Router Solicitation messages that do not satisfy all of the following validity checks:

- IP Source Address is a link-local address.
- ICMP Checksum is valid.
- ICMP Code is 0.
- ICMP length (derived from the IP length) is 8 or more octets.
- if the message includes an Authentication Header, the message authenticates correctly.

- all included options have a length that is greater then zero.

The contents of the Reserved field, and of any unrecognized options, MUST be ignored. Future, backward-compatible changes to the protocol may specify the contents of the Reserved field or add new options; backward-incompatible changes may use different Code values.

A solicitation that passes the validity checks is called a "valid solicitation".

Routers MUST also validate Router Advertisements as described in Section

5.3.3.

Routers SHOULD receive Router Advertisements sent by other routers on the link and verify that the routers are advertising consistent information. Detected inconsistencies indicate that one or more routers might be misconfigured and SHOULD be logged to system or network management. The minimum set of information that should be checked:

- Different Max Hop Limit values (except for the unspecified value of zero).

- Different value of the M or O flags.

- Different Reachable Time values (except for the unspecified value of zero).

- Different Reachable Retrans Timer values (except for the unspecified value of zero).

- Different values in the MTU options.

- Different Invalidation Lifetime for the same prefix.

- Different Deprecation Lifetime for the same prefix.

Note that it is expected that different routers advertise different sets of prefixes. Also, some routers might leave some fields as unspecified

i.e. with the value zero.

In addition, routers can optionally examine the source address of Router Advertisements to determine which of a neighboring router's addresses is its designated address. Any other action on reception of Router Advertisement messages by a router is beyond the scope of this document.

5.2.3. Router Behavior

A router MUST join the all-routers multicast address on all multicast capable interfaces.

The term "advertising interface" refers to any functioning and enabled interface that has at least one IP address assigned to it. From each advertising interface, the router transmits periodic, multicast Router Advertisements, containing the following values consistent with the message format above:

- In the Router Lifetime field: the interface's configured

  RtrAdvLifetime. If the router's AdvertiseDefault flag is set to false, the Router Lifetime field MUST be set to 0.

- In the M and O flags: the interface's configured ManagedFlag and OtherFlag, respectively. See [ADDRCONF].

- In the Max Hop Limit field: the interface's configured MaximumHopLimit.

- In the Reachable Time field: the interface's configured ReachableTime.

- In the Reachable Retrans Timer field: the interface's configured ReachableRetransTimer.

- In the options:

  Source Link-Layer Address option: link-layer address of the sending interface. This option MAY be omitted to facilitate in-bound load balancing over replicated interfaces.

Prefix Information options: one Prefix Information option for each prefix listed in PrefixList with the option fields set from the information in the PrefixList entry as follows:

- In the "on-link" flag: the entry's OnLinkFlag.

- In the Invalidation Lifetime field: the entry's InvalidationLifetime. If the InvalidationLifetime is infinite the field is set to all one bits.

- In the "Autonomous address configuration" flag: the entry's AutonomousFlag.

- In the Deprecation Lifetime field: the entry's DeprecationLifetime. If the DeprecationLifetime is infinite the field is set to all one bits.

The advertisements are not strictly periodic: the interval between subsequent transmissions is randomized to reduce the probability of synchronization with the advertisements from other routers on the same link [SYNC]. This is done by maintaining a separate transmission interval timer for each advertising interface. Each time a multicast advertisement is sent from an interface, that interface's timer is reset to a uniformly-distributed random value between the interface's configured MinRtrAdvInterval and MaxRtrAdvInterval; expiration of the timer causes the next advertisement to be sent from the interface, and a new random value to be chosen. (It is recommended that routers include some unique value, such as one of their IP or link-layer addresses, in

the seed used to initialize their pseudo-random number generators. Although the randomization range is configured in units of seconds, the actual randomly-chosen values SHOULD not be in units of whole seconds, but rather in units of the highest available timer resolution.)

For the first few advertisements sent from an interface (up to MAX_INITIAL_RTR_ADVERTISEMENTS), if the randomly chosen interval is greater than MAX_INITIAL_RTR_ADVERT_INTERVAL, the timer SHOULD be set to MAX_INITIAL_RTR_ADVERT_INTERVAL instead. Using this smaller interval for the initial advertisements increases the likelihood of a router being discovered quickly when it first becomes available, in the presence of possible packet loss.

In addition to the periodic, unsolicited advertisements, a router sends

advertisements in response to valid solicitations received on any of its advertising interfaces. A router MAY choose to unicast the response directly to the soliciting host's address, or multicast it to the all-nodes address; in the latter case, the interface's interval timer is reset to a new random value, as with unsolicited advertisements. A unicast response MAY be delayed, and a multicast response MUST be delayed, for a small random interval not greater than MAX_RTR_RESPONSE_DELAY, in order to prevent synchronization with other responding routers, and to allow multiple, closely-spaced solicitations to be answered with a single multicast advertisement. A router that chooses to delay responses can operate as follows:

- When the router receiver a Router Solicitation it starts a timer with the above small random interval.

- If the router receives an additional Router Solicitation on the interface while the timer is running it will multicast the response. Otherwise it will unicast the response.

- The router sends the (multicast or unicast) Router Advertisement when the timer expires. If the advertisement is multicast the router resets the interface's interval timer to a new random value, as with unsolicited advertisements.

Note that a router is permitted to send multicast Router Advertisements more frequently than indicated by the MinRtrAdvInterval configuration variable if the additional advertisements are responses to explicit solicitations. In all cases, however, unsolicited multicast advertisements must not be sent more frequently than indicated by MinRtrAdvInterval.

When a router receives a Router Solicitation it records the source of the packet as being a neighbor. If the solicitation contains a Source Link-Layer Address option, and the router has a Neighbor Cache entry for

the neighbor, the link-layer address should be updated in the Neighbor Cache and the entry's "is_router" flag should be set to false. If a Neighbor Cache entry is created for the source its reachability state MUST be set to PROBE as specified in Section 6.3.2.

It should be noted that an interface may become an advertising interface at times other than system startup, as a result of recovery from an

interface failure or through actions of system management such as:

- enabling the interface, if it had been administratively disabled, and its AdvertiseDefault flag is TRUE, or

- enabling IP forwarding capability (i.e., changing the system from being a host to being a router), when the interface's AdvertiseDefault flag is TRUE, or

- changing the AdvertiseDefault flag from FALSE to TRUE.

In such cases the router MUST commence transmission of periodic advertisements on the new advertising interface, limiting the first few advertisements to intervals no greater than MAX_INITIAL_RTR_ADVERT_INTERVAL. In the case of a host becoming a router, the system MUST also join the all-routers IP multicast group on all interfaces on which the router supports IP multicast (whether or not they are advertising interfaces).

An interface may also cease to be an advertising interface, through actions of system management such as:

- administratively disabling the interface, or

- shutting down the system, or disabling the IP forwarding capability (i.e., changing the system from being a router to being a host), or

- setting the AdvertiseDefault flag of the interface to FALSE.

In such cases the router SHOULD transmit a final multicast Router Advertisement on the interface with a Router Lifetime field of zero. In the case of a router becoming a host, the system MUST also depart from the all-routers IP multicast group on all interfaces on which the router supports IP multicast (whether or not they had been advertising interfaces). In addition, the host MUST insure that subsequent Neighbor Advertisement messages sent from the interface have the Router flag set to zero.

The information advertised in Router Advertisements may change through actions of system management. For instance, the lifetime for the advertised prefixes may be changes, or the advertised MTU may change.

In such cases, the router MAY transmit a few (no more than MAX_INITIAL_RTR_AVERTISEMENTS) Router Advertisements separated by an interval of MAX_INITIAL_RTR_ADVERT_INTERVAL.

A router might want to send Router Advertisements without advertising itself as being a default router. For instance, a router might advertise prefixes for address autoconfiguration while not wishing to forward packets. Such a router MUST set the Router Lifetime field to zero in its advertisements.

A router MAY choose to not include all Prefix Information options in every Router Advertisement, if the prefix lifetimes are much longer than RtrAdvLifetime. However, when responding to a Router Solicitation the router SHOULD transmit all prefixes to allow hosts to quickly discover the prefixes during system initialization.

5.2.4. Designated Addresses

Routers should take some care in selecting their designated address and in handling any, hopefully infrequent, change of their designated address.

The designated address SHOULD be one that changes infrequently over time. Nodes receiving Neighbor Discovery messages use the source address to identify the sender. If multiple packets from the same neighbor contain different source addresses, nodes will assume they come from different nodes, leading to undesirable behavior. For example, a node will ignore Redirect messages that are believed to have been sent by a router other than the current first-hop router.

The designated address MUST be a link-local address; the link-local address does not change when a site renumbers.

If a router changes the designated address for one of its interfaces, it SHOULD inform hosts of this change. The router should multicast a few Router Advertisements with Router Lifetime field set to zero for the old designated address and also multicast a few Router Advertisements for the new designated address. The exact procedures SHOULD be the same as when an interface ceases being an advertising interface, and when an interface becomes an advertising interface, respectively.

A router MUST be able to determine the designated address for each of its neighboring routers in order to ensure that the target address in a Redirect message identifies the neighbor router by its designated address. This may require that routing protocols exchange designated addresses.

5.3. Host Specification

5.3.1. Host Configuration Variables

None.

5.3.2. Host Variables

A host maintains certain Neighbor Discovery related variables in addition to the data structures defined in Section 4.1. These variables have default values that are overridden by information received in Router Advertisement messages. The default values are used when there is no router on the link, or when all received Router Advertisements have left the value unspecified.

For each interface:

LinkMTU The MTU of the link.

Default: The valued defined in the specific document that describe how IPv6 operates over the particular link layer.

MaximumHopLimit
The maximum Hop Count to be used when sending IP packets.

Default: The value specified in the most recent "Assigned Numbers" RFC [ASSIGNED].

ReachableTime
The time a neighbor is considered reachable after receiving a reachability confirmation.

Default: REACHABLE_TIME milliseconds

ReachableRetransTimer
The time between transmissions of Neighbor Solicitation probes to a neighbor when verifying the reachability of the neighbor.

Default: REACHABLE_RETRANS_TIMER milliseconds

5.3.3. Message Validation by Nodes

A node MUST silently discard any received Router Advertisement messages that do not satisfy all of the following validity checks:

- IP Source Address is a link-local address.
- ICMP Checksum is valid.
- ICMP Code is 0.
- ICMP length (derived from the IP length) is 16 or more octets.
- if the message includes an Authentication Header, the message authenticates correctly.
- all included options have a length that is greater then zero.

The contents of the Reserved field, and of any unrecognized options, MUST be ignored. Future, backward-compatible changes to the protocol may specify the contents of the Reserved field or add new options; backward-incompatible changes may use different Code values.

An advertisement that passes the validity checks is called a "valid advertisement".

A host MUST silently discard any received Router Solicitation messages.

5.3.4. Host Behavior

The host joins the all-nodes multicast address on all multicast capable interfaces.

A host MUST NOT send a Router Advertisement message at any time.

To process a valid Router Advertisement, a host extracts the source address of the packet and does the following:

- If the address is not already present in the host's Default Router List, and the advertisement's Router Lifetime is non-zero, create a new entry in the list, and initialize its timer to value in the advertisement's Router Lifetime field.

- If the address is already present in the host's Default Router List as a result of a previously-received advertisement, reset its timer to the Router Lifetime value in the newly-received advertisement.

- If the received Router Lifetime value is zero, time-out the entry immediately and remove it from the Default Router list.

The source address in the Router Advertisement might not be covered by any of the entries in the host's Prefix List or any of the Prefix Information options in the message; a host MUST accept all valid advertisements independent of their source address.

If the received Max Hop Limit value is non-zero the host SHOULD set its MaximumHopLimit variable to the received value. Hosts use the last Max Hop Limit value they have received; routers should be configured to advertise identical values to avoid hosts switching between different values.

The host SHOULD set its ReachableTime variable to the Reachable Time field, if the received value is non-zero. Likewise it SHOULD set the ReachableRetransTimer to the Reachable Retrans Timer field, if the received value is non-zero. Hosts use the last values they have received; routers should be configured to advertise identical values to avoid hosts switching between different values as they receive advertisements from different routers.

After extracting information from the fixed part of the Router Advertisement message, the advertisement MUST be scanned for valid options. If the advertisement contains a source link-layer address option the link-layer address MUST be recorded in the Neighbor Cache entry for the router (creating an entry if necessary) and the "is_router" flag in the Neighbor Cache entry MUST be set to true. The

"is_route" flag is used by Neighbor Unreachability Detection to determine when a router changes to being a host (i.e. no longer capable of forwarding packets). If a Neighbor Cache entry is created for the router its reachability state MUST be set to PROBE as specified in Section 6.3.2.

Received MTU options are handled as specified in Section 8.4.

For each Prefix Information option that has the "on-link" (L) flag set, the host does the following:

- If the prefix is not already present in the Prefix List, create a new entry for the prefix and initialize its invalidation timer to the Invalidation Lifetime value in the Prefix Information option.

- If the prefix is already present in the host's Prefix List as the result of a previously-received advertisement, reset its invalidation timer to the Invalidation Lifetime value in the Prefix Information option.

- If the received Invalidation Lifetime value is zero, time-out the prefix immediately.

Note: Implementations can choose to process the on-link aspects of the prefixes separately from the address autoconfiguration aspects of the prefixes by e.g. passing a copy of each valid Router Advertisement message to both an "on-link" and an "addrconf" function. Each function can then operate on the prefixes that have the appropriate flag set.

Whenever the invalidation timer expires for a Prefix List entry, that entry is discarded. No existing Destination Cache entries are affected, however.

Whenever a timer expires for an entry in the Default Router List, that entry is discarded. Any entries in the Destination Cache going through that router will continue to be used. Neighbor Unreachability Detection will purge them if appropriate.

To limit the storage needed for the Default Router List, a host MAY choose not to store all of the router addresses discovered via advertisements. However, a host MUST retain at least two router

addresses and SHOULD retain more. Default router selections are made whenever communication to a destination appears to be failing. Thus, the more routers on the list, the more likely an alternative working router can be found quickly (e.g., without having to wait for the next advertisement to arrive).

The algorithm for selecting a router depends in part on whether or not a router is known to be reachable. The exact details of how a node keeps track of a neighbor's reachability state are covered in Section 6.3. The algorithm for selecting a default router is invoked only when a Destination Cache entry is incomplete or when communication through an existing router appears to be failing. Under normal conditions, a router would be selected the first time traffic is sent to a destination, with subsequent traffic for that destination using the same router as indicated in the Destination Cache. The policy for selecting routers from the Default Router List is as follows:

1) Routers known to be reachable (e.g., in the REACHABLE or PROBE state) MUST be preferred over routers whose reachability is unknown or suspect. An implementation may choose to always return the same router or cycle through the router list in a round-robin fashion as long as it always returns a reachable router when one is available.

2) When no routers on the list are known to be reachable, routers should be selected in a round-robin fashion, so that subsequent

requests for a default router do not return the same router until all other routers have been selected.

Cycling through the router list when none are known to be reachable ensures that all available routers are actively probed by the Neighbor Unreachability Detection algorithm. A request for a default router is made in conjunction with the sending of a packet to a router, and the selected router will be probed for reachability as a side effect.

3) If the Default Router List is empty, return an ICMP destination unreachable indication with code 0 (no route to destination) to the sender of the packet that triggered the search. Note that if the Default Router List contains no entries because none were ever added to the Default Router List as a result of received Router

Advertisements messages, all destinations are assumed to be on-link. Thus, ICMP errors are returned only if the Default Router List becomes empty as a result of router Lifetime expirations.

A host is permitted (but not required) to transmit up to MAX_RTR_SOLICITATIONS Router Solicitation messages from any of its multicast interfaces after any of the following events:

- The interface is initialized at system startup time.
- The interface is reinitialized after a temporary interface failure or after being temporarily disabled by system management.
- The system changes from being a router to being a host, by having its IP forwarding capability turned off by system management.
- The host is re-attached to a link after being detached for some time.

The IP destination address of the solicitations is the all-routers multicast address. The IP source address MUST be one of the interface's addresses and MUST be a link-local address. The Source Link-Layer Address option is set to the host's link-layer address.

If a host does choose to send a solicitation after one of the above events, it SHOULD delay that transmission for a random amount of time between 0 and MAX_RTR_SOLICITATION_DELAY. This serves to alleviate congestion when many hosts start up on a link at the same time, such as might happen after recovery from a power failure. (It is recommended that hosts include some unique value, such as one of their IP or link-

layer addresses, in the seed used to initialize their pseudo-random number generators. Although the randomization range is specified in units of seconds, the actual randomly-chosen value SHOULD not be in units of whole seconds, but rather in units of the highest available timer resolution.)

If a host has performed a random delay earlier during the system startup (e.g. as part of Duplicate Address Detection [ADDRCONF]) there is no

need to randomly delay the first Router Solicitation message.

A host MAY also choose to further postpone its solicitations, subsequent to one of the above events, until the first time it needs to use a default router.

Upon receiving a valid advertisement the host MUST desist from sending any solicitations on that interface (even if none have been sent yet), until the next time one of the above events occurs. The small number of retransmissions of a solicitation, which are permitted if no such advertisement is received, SHOULD be sent at intervals of RTR_SOLICITATION_INTERVAL seconds, without randomization.

6. ADDRESS RESOLUTION AND NEIGHBOR UNREACHABILITY DETECTION

This section describes the functions related to the Neighbor Solicitation and Neighbor Advertisement messages and includes descriptions of the Address Resolution and the Neighbor Unreachability Detection algorithms.

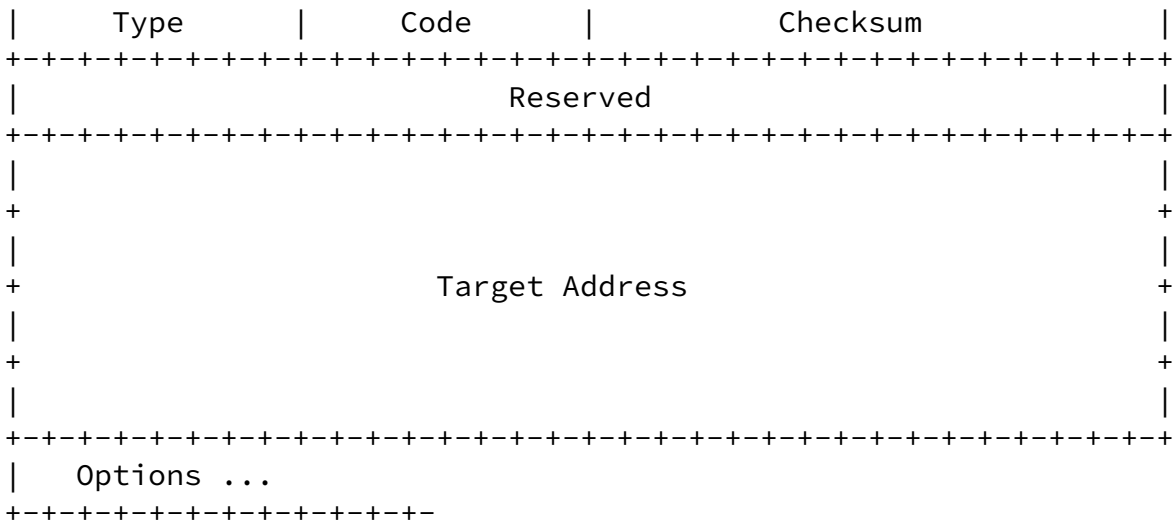
These messages are also used for Duplicate Address Detection as specified by [ADDRCONF]. In particular, Duplicate Address Detection uses the unspecified address as the Source Address in Neighbor Solicitations to prompt a node already using a particular address to multicast a Neighbor Advertisement stating that fact.

6.1. Message Formats

6.1.1. Neighbor Solicitation Message Format

Nodes send Neighbor Solicitations to request the link-layer address of a target node while providing their own link-layer address to the target. Neighbor Solicitations are multicast when the node needs to resolve an address and unicast when the node seeks to verify the reachability of a neighbor.

+-+



IP Fields:

Source Address
: Either an IP address belonging to the interface from which this message is sent, or the unspecified address. Use of the unspecified address directs the target node to multicast the resultant Neighbor Advertisement as required by duplicate address detection in [ADDRCONF].

  A node SHOULD select the same Source Address as the source address in the packet that prompts the solicitation to ensure that the receiver of the solicitation acquires the link-layer address for return traffic.

Destination Address
: Either the solicited-node multicast address corresponding to the target address, or the target address.

Hop Count
: 1

Authentication Header
: If a security association exists between the sender and the destination the sender SHOULD include this header.

ICMP Fields:

Type
: 135

Code 0

Checksum The ICMP checksum. See [ICMPv6].

Reserved This field is unused. It MUST be initialized to zero by the sender and ignored by the receiver.

Target Address The IP address of the target of the solicitation. It MUST NOT be a multicast address.

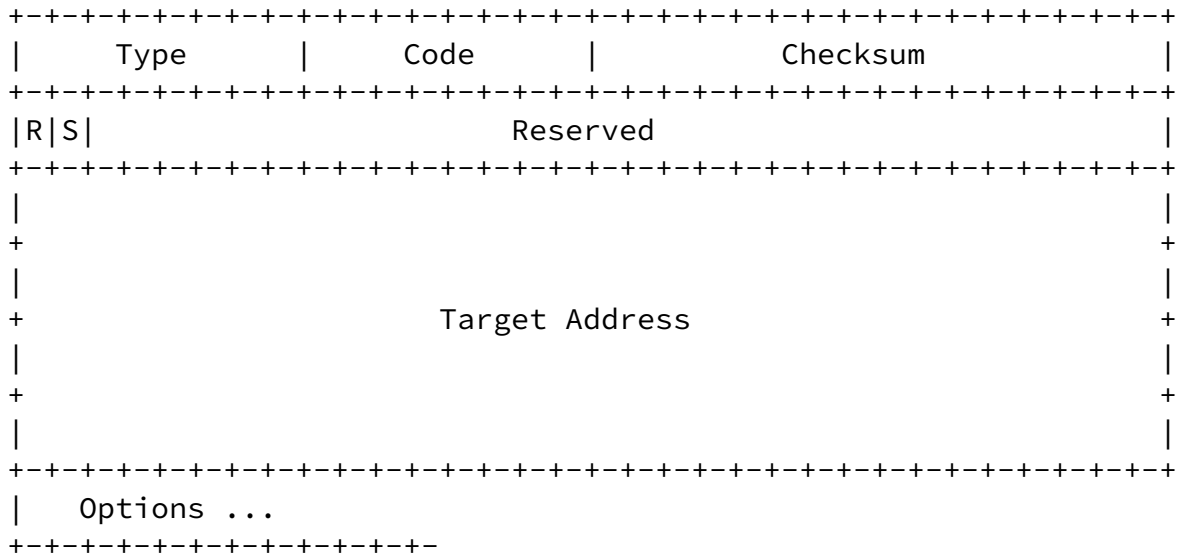
Options:

Source link-layer address
The link-layer address for the sender. MUST be included on link layers that have addresses.

Future versions of this protocol may define new option types. Receivers MUST skip over and ignore any options they do not recognize and continue processing the message.

6.1.2. Neighbor Advertisement Message Format

A node MUST send a Neighbor Advertisement in response to a Neighbor Solicitation for a target IP address that matches an assigned address on the receiving interface. A node MAY also send an unsolicited multicast Neighbor Advertisement when it knows that its link-layer address has changed.



IP Fields:

Source Address
An IP address belonging to the interface from which this message is sent. The source address MUST be the same as the target address for a "normal" (non-anycast/proxy) response.

Destination Address
Either the Source Address of an invoking Neighbor Solicitation, or the all-nodes multicast address. If the source solicitation is the unspecified address the advertisement MUST be multicast to the all-nodes address.

Hop Count 1

Authentication Header
If a security association exists between the sender and the destination the sender SHOULD include this header.

ICMP Fields:

Type 136

Code 0

Checksum The ICMP checksum. See [ICMPv6].

R Router flag. When set, the R-bit indicates that the sender is a router. The R-bit is used by Neighbor Unreachability Detection to detect a router that changes to a host.

S Solicited flag. When set, the S-bit indicates that the advertisement was sent in response to a Neighbor Solicitation from the Destination address. It MUST be zero in a multicast advertisement and in an unsolicited unicast advertisement.

Reserved 30-bit unused field. It MUST be initialized to zero by the sender and ignored by the receiver.

Target Address The address from the Target Address field in the Neighbor Solicitation message that prompted this advertisement. For an unsolicited advertisement, the address whose link-layer address has changed. The Target Address MUST NOT be a multicast address.

Options:

Target link-layer address
The link-layer address for the target. MUST be included on link layers that have addresses.

Future versions of this protocol may define new option types. Receivers MUST skip over and ignore any options they do not recognize and continue processing the message.

6.2. Address Resolution

Address Resolution provides the mechanism through which nodes determine the link-layer address of their neighbors. Address Resolution is only used for destinations that are determined to be on-link and for which the sender does not know the corresponding link-layer address. Address resolution is never used for multicast destinations.

6.2.1. Message Validation by Nodes

A node MUST silently discard any received Neighbor Solicitation or Advertisement messages that do not satisfy all of the following validity checks:

- if the message includes an Authentication Header, the message authenticates correctly.

- ICMP Checksum is valid.

- ICMP Code is 0.

- ICMP length (derived from the IP length) is 24 or more octets.

- Target Address is not a multicast address.

- for a Neighbor Advertisement; if the Destination Address is a multicast address the Solicited flag is zero.

- all included options have a length that is greater then zero.

The contents of the Reserved field, and of any unrecognized options, MUST be ignored. Future, backward-compatible changes to the protocol may specify the contents of the Reserved field or add new options; backward-incompatible changes may use different Code values.

Neighbor Solicitations and Advertisements that passes the validity checks are called "valid solicitations" and "valid advertisements", respectively.

6.2.2. Node Specification

When a multicast-capable interface is initialized the node MUST join the all-nodes multicast address on that interface, as well as the solicited-node multicast address corresponding to each of the IP addresses assigned to the interface.

The operation of automatic address configuration [ADDRCONF] may, over time, change the set of addresses assigned to an interface; new addresses might be added and old addresses might be removed. In such case the node MUST join and leave the solicited-node multicast address corresponding to the new and old addresses, respectively. Note that multiple addresses might correspond to the same solicited-node multicast address; the host should not leave the multicast address until all addresses corresponding to the multicast address have been removed.

6.2.3. Sending Node Specification

When a node has a packet to send, but does not know the next-hop's

link-layer address, it performs address resolution by creating a Neighbor Cache entry in the INCOMPLETE state and transmitting a Neighbor Solicitation message targeted at the neighbor. The packet prompting the solicitation MUST be queued in the Neighbor Cache entry and the solicitation MUST be sent to the solicited-node multicast address corresponding to the target address.

The sender MUST include its link-layer address (if it has one) in the solicitation as a Source Link-Layer Address option, so that the receiver discovers the sender's link-layer address without the need for an additional packet exchange.

While waiting for address resolution to complete, the sender MUST maintain a small queue containing packets waiting for address resolution to complete. The queue MUST hold at least one packet, and MAY contain more. However, the number of queued packets per neighbor SHOULD be limited to some small value. When a queue overflows, the new arrival SHOULD replace the oldest entry. Once address resolution completes, all queued packets SHOULD be transmitted.

While awaiting a response, the sender MUST retransmit Neighbor Solicitation messages approximately every RESOLVE_RETRANS_TIMER seconds,

even in the absence of additional traffic to the neighbor. Retransmissions MUST be rate-limited for each neighbor to at most one solicitation every RESOLVE_RETRANS_TIMER seconds.

If no advertisement is received after MAX_MULTICAST_SOLICIT solicitations, address resolution has failed. The sender MUST return ICMP destination unreachable indications with code 3 (Address Unreachable) for each packet queued awaiting address resolution.

When a valid Neighbor Advertisement is received (either solicited or unsolicited), the Neighbor Cache is searched for the target's entry. If no entry exists, the advertisement SHOULD be silently discarded. There is no need to create an entry, since the recipient has apparently not initiated any communication with the target.

If an INCOMPLETE Neighbor Cache entry exists for the target, the advertisement is the first response to a solicitation. In such cases, the receiving node records the link-layer address in the Neighbor Cache

entry and sends any packets queued for the neighbor awaiting address resolution. In addition, the receiving node MUST examine the Router flag field of the advertisement and update the "is_router" flag in the Neighbor Cache entry to reflect whether the node is a host or router. If the neighbor was previously behaving as a router, but the advertisement's Router flag is set to zero, the node MUST update the Neighbor Cache entries for all destination using that neighbor as a reouter as indicated in Section 6.3.2.

If the Solicited flag is set, the node MUST update the reachability information as described in Section 6.3.2 by setting the state to REACHABLE. In addition, the receiving node MUST examine the Router flag field of the advertisement and update the "is_router" flag in the Neighbor Cache entry to reflect whether the node is a host or router. If the neighbor was previously behaving as a router, but the advertisement's Router flag is set to zero, the node MUST update the Neighbor Cache entries for all destination using that neighbor as a reouter as indicated in Section 6.3.2.

Multiple solicited Neighbor Advertisements can be received in response to a solicitation for a anycast/proxy address. In such cases one or more of the advertisements is a anycast/proxy advertisement. Anycast/proxy advertisements are identified by having differing source and target addresses. A node MUST give preference to "normal" responses over anycast/proxy responses and, among multiple anycast/proxy responses, a node MUST prefer the first anycast/proxy response. This is accomplished by applying the following rules while processing received advertisements with the Solicited flag set to one:

1) if no link-layer address is currently recorded, install the one

contained in the advertisement.

2) if a link-layer address has already been recorded, and the advertisement is not an anycast/proxy advertisement, replace the recorded address with the one contained in the advertisement.

3) otherwise ignore the advertisement

A node MAY occasionally multicast or unicast an unsolicited Neighbor Advertisement announcing a link-layer address change. A node that

receives a Neighbor Advertisement without the Solicited flag set, does the following:

- If the node has a Neighbor Cache entry for the target, it SHOULD copy the link-layer address information contained in the advertisement's Source Link-Layer Address option into the corresponding Neighbor Cache entry.

- The node MUST not treat the receipt of a unsolicited advertisement as a confirmation that the neighbor is REACHABLE (as defined in Section 4.1). See Section 6.3.1.

6.2.4. Target Node Specification

When a node receives a valid Neighbor Solicitation, it compares the message's Target Address against the IP addresses belonging to the incoming interface as well as any anycast addresses that it has been configured to respond to. If no match is found, the node is not the target of the solicitation and it MUST silently ignore the message.

If the node is the target of the solicitation, and the solicitation's Source Address is not the unspecified address, the recipient first ensures that it has an up-to-date Neighbor Cache entry for the Source Address of the solicitation. If no entry is found one is created and its link-layer address is copied from the Source Link-Layer Address option in the message. If an entry already exists, its link-layer address is updated to match the address in the Source Link-Layer Address option. If a Neighbor Cache entry is created for the source its reachability state MUST be set to PROBE as specified in Section 6.3.2.

If the source of the solicitation is the unspecified address, the target MUST multicast an advertisement to the all-nodes address. Otherwise, the target MUST send a unicast Neighbor Advertisement to the address copied from the IP Source Address of the Neighbor Solicitation. In both cases the Target Address is copied from the solicitation message to the advertisement and the Target Link-Layer Address option is filled with

the node's link-layer address on the link. If the advertisement is multicast the Solicited flag MUST be set to zero; if it is unicast the Solicited flag MUST be set to one in order to give the source a reachability confirmation. If the node is a router it MUST set the

Router flag to one; otherwise it MUST be set to zero.

If the node is not providing anycast/proxy services for the targeted address, the IP Source Address MUST be set to the address in the Target Address field (which is one of the IP addresses belonging to the interface). Doing so guarantees that the receiver can identify the Neighbor Advertisement as being a "normal" advertisement.

If the node is providing anycast/proxy services for the target the IP Source Address MUST be set the interface's designated address (which is different than the Target Address) so that the receiver recognizes the message as an anycast/proxy advertisement.

6.2.5. Anticipated Link-Layer Address Changes

In some cases a node may be able to determine that its link-layer address has changed (e.g., hot-swap of an interface card) and may wish to inform its neighbors of the new link-layer address quickly. In such cases a node MAY send up to MAX_NEIGHBOR_ADVERTISEMENT unsolicited Neighbor Advertisement messages to the all-nodes multicast address. These advertisements MUST be separated by at least MIN_NEIGHBOR_ADVERT_INTERVAL seconds.

The Target Address field in the unsolicited advertisement is set to an IP address of the interface and the Target Link-Layer Address option is filled with the new link-layer address. The Solicited flag MUST be set to zero, in order to avoid confusing the Neighbor Unreachability Detection algorithm. If the node is a router is MUST set the Router flag to one; otherwise it MUST be set to zero.

A node that has multiple IP addresses assigned to an interface MAY multicast a separate Neighbor Advertisement for each address. In such a case the node SHOULD introduce a small delay between the sending of each advertisement to reduce the probability of the advertisements being dropped by hosts.

A proxy MAY multicast Neighbor Advertisements when its link-layer address changes or when it is configured (by system management or other mechanisms) to proxy for an address. If there are multiple nodes that are providing proxy services for the same set of addresses the proxies SHOULD provide a mechanism that prevents multiple proxies from multicasting advertisements for the same addresses.

Note that unsolicited Neighbor Advertisements do not reliably update caches in all nodes (the advertisements might not be received by all nodes) and should only be viewed as a performance optimization to quickly update the caches in most neighbors. The Neighbor Unreachability Detection algorithm will ensure that neighbors reliably update the cached link-layer address when they attempt to communicate with the node.

6.2.6. Anycast Neighbor Advertisements

An anycast address cannot be syntactically distinguished from other unicast addresses. This section shows how the rules defined above "do the right thing" for anycast addresses.

A node belonging to an anycast address MUST join the solicited-node multicast address that corresponds to the anycast address.

When a node responds to a Neighbor Solicitation for an anycast address, it by definition responds with a anycast/proxy Neighbor Advertisement. Anycast address are not permitted to appear as the source address in an IP packet, guaranteeing that the advertisement's source and target addresses differ.

A node might receive multiple Neighbor Advertisements in response to a Neighbor Solicitation for an anycast address when multiple neighbors are configured to recognize the anycast address. The precedence rules in Section 6.2.3 will make the node select the first advertisement (i.e. the fastest or most lightly loaded node) as current binding for the anycast address.

The use of Neighbor Unreachability Detection ensures that a node quickly detects when the current binding for the anycast address has gone stale e.g. due to a node no longer belonging to the anycast address.

6.2.7. Proxy Neighbor Advertisements

Under limited circumstances, a router MAY proxy for one or more other nodes, that is, through Neighbor Advertisements indicate that it is willing to accept packets not explicitly addressed to itself. For example, a router might potentially accept packets on behalf of a mobile node that has moved off-link. The mechanisms used by proxy are identical to the mechanisms needed for anycast addresses. The address being served is called a "proxee" in this section.

A proxy MUST join the solicited-node multicast address(es) that

correspond to the proxee's IP address(es).

All proxy Neighbor Advertisement messages MUST be tagged as being anycast/proxy messages; the advertisement's Source Address MUST differ from its Target Address (e.g., the proxee). In practice, this requirement poses no special burden. By definition, the advertisement's source address MUST be the designated address of the interface on which the advertisement is sent, which will be different than any proxee address.

6.3. Neighbor Unreachability Detection

Communication to or through a neighbor may fail for numerous reasons at any time, including hardware failure, hot-swap of an interface card, a mobile node moving off-link, etc. If the destination has failed, no recovery is possible and communication fails. On the other hand, if it is the path that has failed, recovery may be possible. Thus, a node actively tracks the reachability "state" for the neighbors to which it is sending packets.

Neighbor Unreachability Detection is used for all paths between hosts and neighboring nodes, including host-to-host, host-to-router, and router-to-host communication. Neighbor Unreachability Detection may also be used between routers, but is not required if an equivalent mechanism is available, for example, as part of the routing protocols.

When a path to a neighbor appears to be failing, the specific recovery attempt depends on how the neighbor is being used. For example, appropriate recovery procedures when using the neighbor as a router differ from those appropriate for the case where the neighbor is the destination.

Neighbor Unreachability Detection is performed only for neighbors to which unicast packets are sent; it is not used when sending to multicast addresses.

6.3.1. Reachability Confirmation

A neighbor is considered reachable if the node has recently received a confirmation that packets sent to the neighbor are received by its IP

layer. Positive confirmation can be gathered in two ways: hints from upper layer protocols that indicate a connection is making "forward progress", or receipt of a Neighbor Advertisement message that is a response to an explicit Neighbor Solicitation probe.

A connection makes "forward progress" if the packets received from a remote peer can only be arriving if recent packets sent to that peer are actually reaching it. For example, receipt of a (new) acknowledgement

indicates that previously sent data reached the peer. Likewise, the arrival of a new (non-duplicate) packet indicates that earlier acknowledgements are being delivered to the remote peer. If packets are reaching the peer the packets must also be reaching the sender's next-hop neighbor; thus "forward progress" is a confirmation that the next-hop neighbor is reachable. When available, this upper-layer information SHOULD be used.

In some cases (e.g, UDP-based protocols and routers forwarding packets to hosts) such reachability information may not be readily available from upper-layer protocols. When no hints are available and a node is sending packets to a neighbor, the node actively probes the neighbor using Neighbor Solicitation messages to verify that the forward path is still working.

The receipt of a solicited Neighbor Advertisement that is a response to a Neighbor Solicitation probe serves as reachability confirmation, since advertisements with the Solicited flag set to one are sent only in response to a solicitation. A received Neighbor Advertisement with the Solicited flag set to zero MUST NOT be treated as a reachability confirmation. Receipt of unsolicited advertisements only confirm the one-way path from the neighbor to the recipient node. In contrast, Neighbor Unreachability Detection requires that a path be working from the sender to the neighbor. An advertisement sent in response to an explicit solicitation confirms that a path is working in both directions; the solicitation reached the neighbor, prompting it to generate an advertisement, and the advertisement reached the querying node.

6.3.2. Node Behavior

Neighbor Unreachability Detection operates in parallel with the sending

of packets to a neighbor. While reasserting a neighbor's reachability, a node continues sending packets to that neighbor using the cached link-layer address.

A Neighbor Cache entry can be in one of three states:

INCOMPLETE Address resolution is being performed on the entry. Specifically, a Neighbor Solicitation has been sent to the solicited-node multicast address of the target, but the corresponding Neighbor Advertisement has not yet been received.

REACHABLE Positive confirmation was received within the last ReachableTime milliseconds that the forward path to the neighbor was functioning properly. While REACHABLE, no

special action takes place as packets are sent.

PROBE More than ReachableTime milliseconds have elapsed since the last positive confirmation was received that the forward path was functioning properly. Upon entering the PROBE state, the sending of an initial Neighbor Solicitation is delayed by a time of DELAY_FIRST_PROBE_TIME to give the upper layers additional time to provide reachability confirmation information. After the initial delay, Neighbor Solicitations are sent every ReachableRetransTimer milliseconds.

When an entry is created as a result of needing to perform address resolution, a Neighbor Solicitation is sent to the solicited-node multicast address of the target, a timer is started to expire RESOLVE_RETRANS_TIMER seconds later and the entry's state is set to INCOMPLETE.

As specified in Section 6.2.3, when in the INCOMPLETE state, Neighbor Solicitation messages are retransmitted every RESOLVE_RETRANS_TIMER seconds until a response is received. If no response is received within RESOLVE_RETRANS_TIMER seconds after sending MAX_MULTICAST_SOLICIT probes to the solicited-node multicast address, address resolution fails. Upon failure, ICMP destination unreachable indications with code 3 (Address unreachable) are returned for any queued packets and the entry is

deleted. Note that deleting the entry implies that all destinations using that neighbor must perform next-hop resolution again before sending a subsequent packet. Thus, if the neighbor is a router, an alternate router may be selected. Alternatively, a destination previously thought to be on-link, may now only be reachable through a router.

Unreachability detection changes a neighbor's state from REACHABLE to PROBE only on-demand, when a packet is being sent to that neighbor. When no traffic is sent to a neighbor, an entry may technically no longer be in a REACHABLE state, but the condition is not checked or acted upon until a packet is sent to the neighbor.

The first time a Neighbor Cache entry is referenced and more than ReachableTime milliseconds have passed since the last reachability confirmation was received, its state changes to PROBE. However, no Neighbor Solicitation probe is sent. Probing is deferred for an additional DELAY_FIRST_PROBE_TIME seconds, an optimization that gives the upper-layer protocol additional time to provide a reachability confirmation in those cases where ReachableTime milliseconds have passed since the last confirmation due to lack of recent traffic. Without this optimization the opening of a TCP connection after a traffic lull would

initiate probes even though the subsequent three-way handshake would provide a reachability confirmation almost immediately.

If no reachability confirmation is received within DELAY_FIRST_PROBE_TIME seconds after entering the PROBE state, a unicast Neighbor Solicitation message is sent to the neighbor using the cached link-layer address. In addition, the sender starts a timer to retransmit probe messages every ReachableRetransTimer milliseconds until the desired solicitation is received. Subsequent probes are retransmitted even if no additional packets are sent to the neighbor. If no response is received after waiting ReachableRetransTimer milliseconds after sending the MAX_UNICAST_SOLICIT solicitations, retransmissions cease and the entry SHOULD be deleted. Subsequent traffic to that neighbor recreates the entry and performs address resolution again.

Note that all Neighbor Solicitations are rate-limited on a per-neighbor basis. A node MUST NOT send Neighbor Solicitations to the same neighbor more frequently than once every ReachableRetransTimer milliseconds.

The PROBE state is also used when creating Neighbor Cache entries as a result of receiving packets other than solicited Neighbor Advertisements. This includes Router Solicitations, Router Advertisements, Redirects, and Neighbor Solicitations. These packets contain the link-layer of either the sender or, in the case of Redirect, the redirection target. However, receipt of the link-layer address does not confirm reachability of the forward-direction path to the node. Placing a newly created Neighbor Cache entry for which the link-layer address is known in the PROBE state provides assurance that path failures are detected quickly.

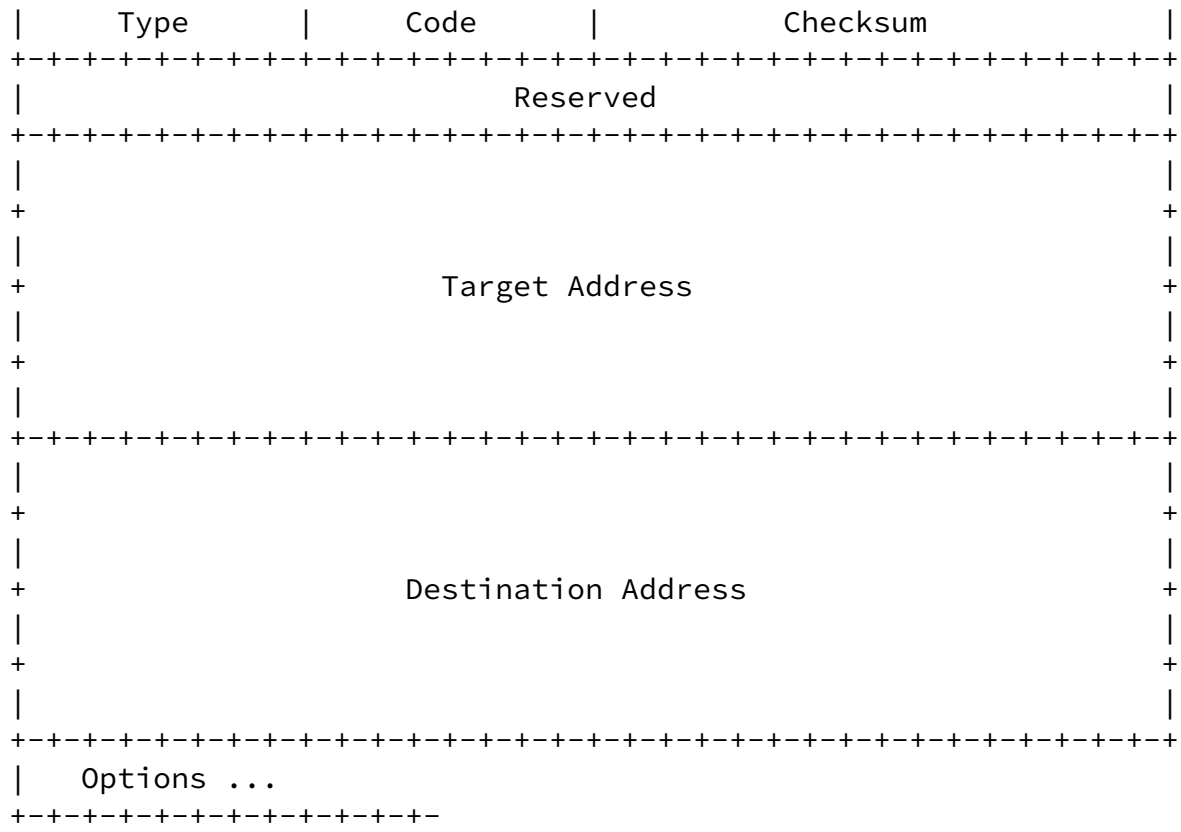
To detect a router that switches from being a router to being a host (e.g, by having its IP forwarding capability turned off by system management), a node MUST compare the Router flag field in all received Neighbor Advertisement messages with the "is_router" flag recorded in the Neighbor Cache entry. When a node detects that a neighbor has changed from being a router to being a host, the node MUST remove that router from the Default Router List and update the Destination Cache so that all entries using that neighbor as a router switch to another router. Note that a router may not be listed in the Default Router List, but still have Destination Cache entries using it, if a host was redirected to it.

7. REDIRECT FUNCTION

This section describes the set of functionality related to the sending and processing of Redirect messages.

7.1. Redirect Message Format

A Redirect packet is sent from a router to a host to inform the host of a better first-hop node on the path to a destination.



IP Fields:

Source Address
: The designated address of the interface from which the redirect is sent.

Destination Address
: The Source Address of the packet that triggered the redirect.

Hop Count
: 1

Authentication Header
: If a security association exists between the sender and the destination the sender SHOULD include this header.

ICMP Fields:

Type 5

Code 0

Checksum The ICMP checksum. See [ICMPv6].

Reserved This field is unused. It MUST be initialized to zero by the sender and ignored by the receiver.

Target Address An IP address of the node to which traffic for the Destination SHOULD be sent. When the target is a router, the Target Address MUST be the router's designated address so that hosts can uniquely identify routers. When the target is a host the target address field MUST be set to the same value as the Destination Address field.

Destination Address
The IP address of the destination which is redirected to the target.

Options:

Target link-layer address
The link-layer address for the target. It SHOULD be included on link layers that have addresses, if known.

Redirected Header
As much as possible of the IP packet that triggered the sending of the Redirect without making the redirect packet exceed 576 octets.

Future versions of this protocol may define new option types. Receivers MUST skip over and ignore any options they do not recognize and continue processing the message.

7.2. Router Specification

A router SHOULD send a redirect message, subject to rate limiting, whenever it forwards a packet in which:

- the Source Address field of the packet identifies a neighbor, and

- after consulting its routing table, the router forwards the packet to a node residing on the same link as the packet's source, and

- the Destination Address of the packet is not a multicast address, and

- the packet is not source routed through the router. A packet is source routed through the router if, when the packet is received by the router, it contains the IP route header and the router's address is in the Destination Address field.

The transmitted redirect packet contains, consistent with the above message format:

- In the Target Address field: the address to which subsequent packets for the destination SHOULD be sent. If the target is a router, that router's designated address MUST be used. If the target is a host the target address field MUST be set to the same value as the Destination Address field.

- In the Destination Address field: the destination address of the invoking IP packet.

- In the options:

  Target Link-Layer Address option: link-layer address of the target, if known.

  Redirected Header: as much of the forwarded packet as can fit without the redirect packet exceeding 576 octets in size.

A router MUST limit the rate of Redirect messages sent, in order to limit the bandwidth and processing costs incurred by the Redirect messages when the source does not correctly respond to the Redirects, or the source chooses to ignore unauthenticated Redirect messages. More details on the rate-limiting of ICMP error messages can be found in [ICMPv6].

A router MUST NOT update its routing tables upon receipt of a Redirect.

7.3. Host Specification

7.3.1. Message Validation by Hosts

A host MUST silently discard any received Redirect messages that do not satisfy all of the following validity checks:

- IP Source Address is a link-local address.
- ICMP Checksum is valid.
- ICMP Code is 0.
- ICMP length (derived from the IP length) is 40 or more octets.
- the IP source address of the Redirect is the same as the current first-hop router for the specified destination.
- the Target Address of the redirect is not a multicast address.
- the Destination Address field in the redirect message does not contain a multicast address.
- if the message includes an Authentication Header, the message authenticates correctly.
- all included options have a length that is greater then zero.

The contents of the Reserved field, and of any unrecognized options MUST be ignored. Future, backward-compatible changes to the protocol may specify the contents of the Reserved field or add new options; backward-incompatible changes may use different Code values.

A host MUST NOT consider a redirect invalid just because the Target Address of the redirect is not covered under one of the link's prefixes.

A redirect that passes the validity checks is called a "valid redirect".

7.3.2. Host Behavior

A host receiving a valid redirect SHOULD update its routing information

accordingly. When a redirect is received, the host updates the Destination Cache entry for the destination to use to the specified target as the new next-hop. If no Destination Cache entry exists for the destination such an entry is created (placing it in the PROBE state).

If the redirect contains a Target Link-Layer Address option the host either creates or updates the Neighbor Cache entry for the target. The link-layer address in the Neighbor Cache entry MUST be copied from the Target Link-Layer Address option. If a Neighbor Cache entry is created for the target its reachability state MUST be set to PROBE as specified in Section 6.3.2. In addition, if the Target Address is the same as the Destination Address the host as been redirected to the destination and it should set the "is_router" field in a created Neighbor Cache entry to false. Otherwise it should set to true.

A host MAY have a configuration switch that can be set to make it ignore a Redirect message that does not have an IP Authentication header.

A host MUST NOT send Redirect messages.

8. OPTIONS

Options provide a mechanism for encoding variable length fields, fields that may appear multiple times in the same packet, or information that is optional and may not appear in all packets. Options can also be used to add additional functionality to future versions of ND.

In order to ensure that future extensions can properly coexist with current implementations, all nodes MUST skip over any options they do not recognize in received ND packets and continue processing the packet. However, the options specified in this document MUST be implemented by all implementations.

The current set of options is defined in such a way that receivers can process multiple options in the same packet independently of each other. In order to maintain these properties future options SHOULD follow the simple rule:

The option MUST NOT depend on the presence or absence of any other options. The semantics of an option should depend only on the information in the fixed part of the ND packet and on the information contained in the option itself.

Adhering to the above rule has the following benefits:

1) Receivers can process options independently of one another. For example, an implementation can choose to process the Prefix Information option in a Router Advertisement message in a user-space process while the link-layer address in the same message is processed by routines in the kernel.

2) Should the number of options cause a packet to exceed a link's MTU, multiple packets can carry subsets of the options without any change in semantics.

3) Senders MAY send a subset of options in different packets. For instance, if the prefix Invalidation Lifetime is high it might not be necessary to include the Prefix Information option in every Router Advertisement. In addition, different routers might send different sets of options. Thus, a receiver MUST NOT associate any action with the absence of an option in a particular packet. This protocol specifies that receivers should only act on the expiration of timers and on the information that is received in the packets.

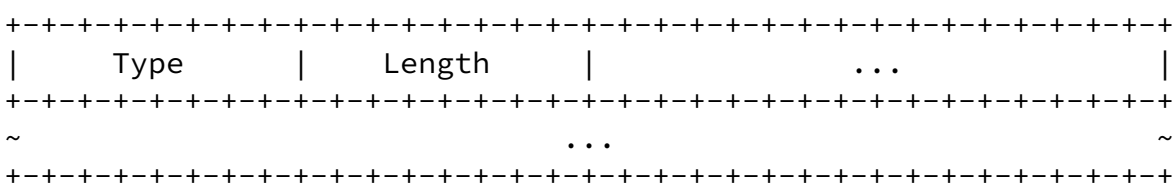
When multiple options are present in a Neighbor Discovery packet, they may appear in any order; receivers MUST be prepared to process them independently of their order. There can also be multiple instances of the same option in a message, for instance Prefix Information options.

All options have a length that is a multiple of 8 octets, ensuring appropriate alignment without any "pad" options. The fields in the options, as well as the fields in ND packets, are defined to align on their natural boundaries (e.g. a 16-bit field is aligned on a 16-bit boundary) except the 128-bit IP addresses/prefixes which are aligned on a 64-bit boundary.

The link-layer address field contains an octet string; thus it is only

aligned on an 8-bit boundary.

All options are of the form:



Fields:

Type 8-bit identifier of the type of option. The options defined in this document are:

Option Name	Type
Source Link-Layer Address	1
Target Link-Layer Address	2
Prefix Information	3
Redirected Header	4
MTU	5

Length 8-bit unsigned integer. The length of the option in units of 8 octets. The value 0 is invalid. Nodes MUST silently discard an ND packet that contains an option with length zero.

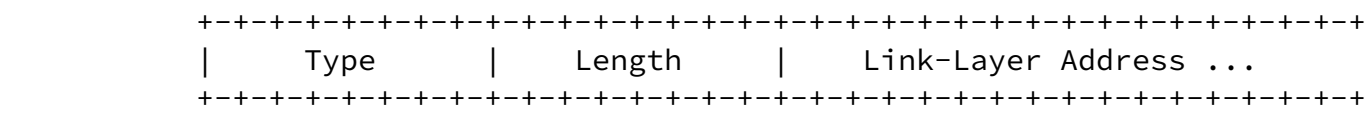
The size of an ND packet including the IP header is limited to the link MTU (which is at least 576 octets). When adding options to an ND packet a node MUST NOT exceed the link MTU. This is handled in a packet specific manner.

The only ND packets that currently can exceed the link MTU are Router Advertisements and Redirects; the former due a large number of Prefix Information options and the latter due to the Redirected Header option.

If there are more Prefix Information options than can fit in a single Router Advertisement packet the router MUST send multiple separate advertisements that each contain a subset of the set of prefixes.

In a Redirect packet the amount of data included in the Redirected Header MUST be limited so that the packet does not exceed 576 octets.

8.1. Source/Target Link-layer Address



Fields:

Type
1 for Source Link-layer Address
2 for Target Link-layer Address

Length
The length of the option in units of 8 octets. For example, the length with IEEE 802 addresses is 1.

Link-Layer Address
The variable length link-layer address.

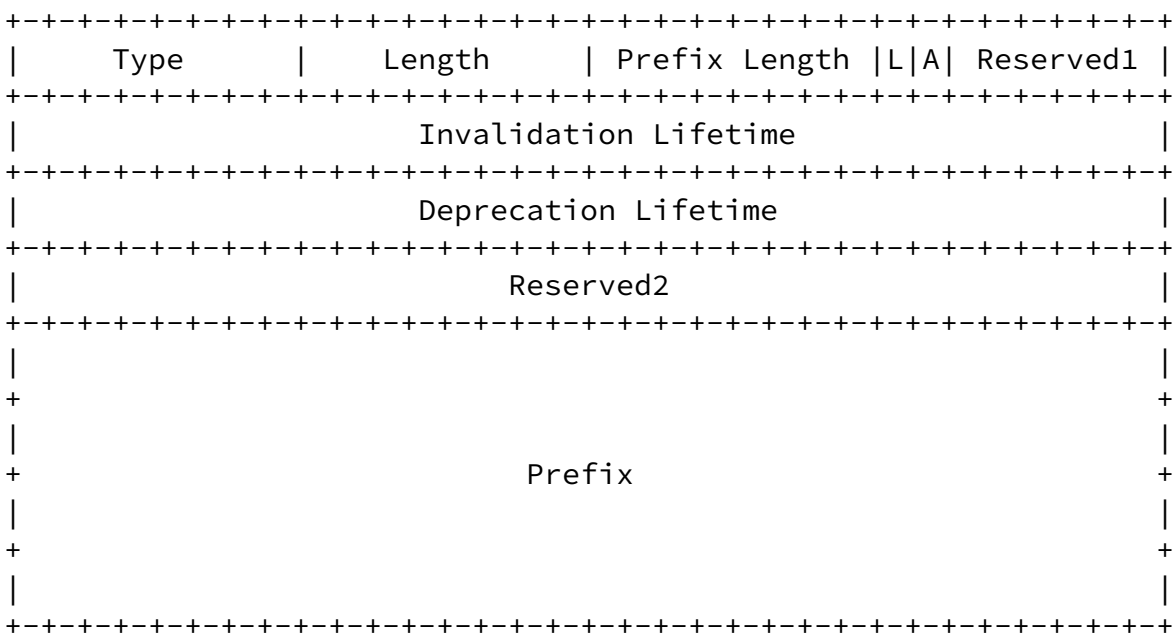
The content and format of this field is expected to be specified in specific documents that describe how IPv6 operates over different link layers. The format for IEEE 802 addresses is a 6-byte string that represents the address in Canonical bit order.

Description
The Source Link-Layer address option contains the link-layer address of the sender of the packet. It is used in the Neighbor Solicitation, Router Solicitation, and Router Advertisement packets.

The Target Link-Layer address option contains the link-layer address of the target. It is used in

Neighbor Advertisement and Redirect packets.

8.2. Prefix Information



Fields:

Type	3
Length	4
Prefix Length	8-bit unsigned integer. The number of leading bits in the Prefix that are valid. The value ranges from 0 to 128.
L	1-bit on-link flag. When set, indicates that this prefix can be used for on-link determination.
A	1-bit autonomous address-configuration flag. When set indicates that this prefix can used for autonomous address configuration as specified in [ADDRCONF].
Reserved1	6-bit unused field. It MUST be initialized to zero by

the sender and ignored by the receiver.

Invalidation Lifetime
: 32-bit unsigned integer. The lifetime of the prefix in seconds for the purpose of on-link determination. A value of all one bits (2^32-1) represents infinity. This lifetime is also used by [ADDRCONF].

Deprecation Lifetime
: 32 bits reserved for autonomous address configuration. A value of all one bits (2^32-1) represents infinity. See [ADDRCONF].

Reserved2
: This field is unused. It MUST be initialized to zero by the sender and ignored by the receiver.

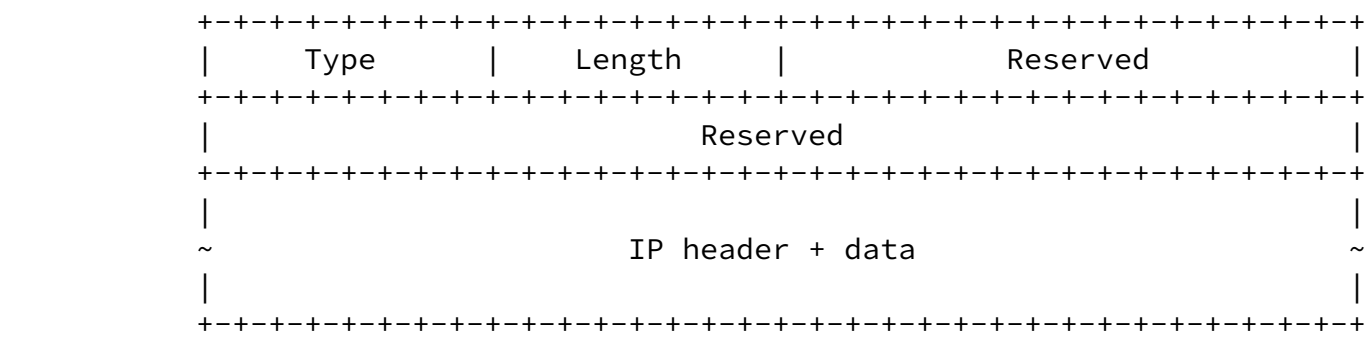
Prefix
: An IP address or a prefix of an IP address. The prefix length field contains the number of valid leading bits in the prefix.

Description

The Prefix Information option is only used in Router Advertisement packets. It provide hosts with on-link prefixes and prefixes for Address Autoconfiguration.

Implementations can choose to process the on-link aspects of the prefixes separately from the address autoconfiguration aspects of the prefixes by e.g. passing a copy of each valid Router Advertisement message to both an "on-link" and an "addrconf" function. Each function can then operate on the prefixes that have the appropriate flag set.

8.3. Redirected Header



Fields:

Type 4

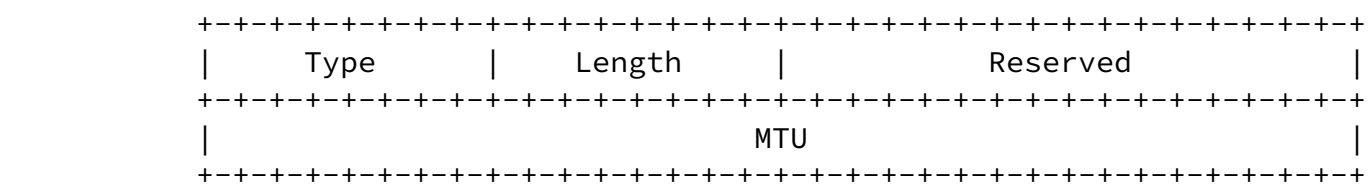
Length The length of the option in units of 8 octets.

Reserved These fields are unused. They MUST be initialized to zero by the sender and ignored by the receiver.

IP header + data
The original packet truncated to ensure that the size of the redirect message does not exceed 576 octets.

Description
The Redirected Header option MUST be included in Redirect packets.

8.4. MTU



Fields:

Type 5

Length 1

Reserved This field is unused. It MUST be initialized to zero by the sender and ignored by the receiver.

MTU 32-bit unsigned integer. The recommended MTU for the link.

Description

The MTU option SHOULD be included in Router Advertisement packets when the link has no well-known MTU and it MAY be included on links with a well-known MTU.

Hosts MUST handle this option by setting the LinkMTU variable for the interface to the received value. If the routers on the link are advertising different MTU values this will result in hosts switching between the different MTUs. Therefore, routers SHOULD verify the consistency between the MTU they advertise and that other routers advertise and log a network management event when there is a mismatch.

When a host or its interface is initialized the LinkMTU of the interface should be set to the predefined value for that type of link. If the host receives no MTU option it must continue to use that predefined value. The MTU option can be used by routers to both increase and decrease the MTU.

9. MULTIHOMED HOSTS

There are a number of complicating issues that arise when Neighbor Discovery is used by hosts that have multiple interfaces. This section does not attempt to define the proper operation of multihomed hosts with regard to Neighbor Discovery. Rather, it identifies complicating issues that require further study. Implementors are encouraged to experiment with various approaches to making Neighbor Discovery work on multihomed hosts and to report their experiences.

If a multihomed host receives Router Advertisements on all of its interfaces, it will (probably) have learned on-link prefixes for the addresses residing on each link. When a packet must be sent through a router, however, selecting the "wrong" router can result in a suboptimal or non-functioning path. There are two cases to consider:

1) If the first-hop router knows of a better first-hop on the same link as the router and sending host, it can send a redirect,

provided that the source address of the offending packet matches an address on the outgoing interface (e.g., the source address resides on the same subnet as the router). On non-multihomed hosts, this will always be the case. On multihomed hosts, however, the packet's source address may be associated with another interface.

In such cases, no redirect will be sent and suboptimal routing cannot be corrected.

2) If the selected first-hop router does not have a route at all for the destination, it will be unable to deliver the packet. However, the destination may be reachable through a router on one of the other interfaces. Neighbor Discovery does not address this scenario; it does not arise in the non-multihomed case.

3) Even if the first-hop router does have a route for a destination, there may be a better route via another interface. No mechanism exists for the multihomed host to detect this situation.

If a multihomed host fails to receive Router Advertisements on one or more of its interfaces, it will not know (in the absence of configured information) which destinations are on-link on the affected interface(s). This leads to a number of problems:

1) If no Router Advertisement is received on any interfaces, a multihomed host will have no way of knowing which interface to send packets out on, even for on-link destinations. Under similar conditions in the non-multihomed host case, a node treats all destinations as residing on-link, and communication proceeds. In the multihomed case, however, additional information is needed to select the proper outgoing interface. Alternatively, a node could attempt to perform address resolution on all interfaces. However, this entails significant complexity that is not present in the non-multihomed host case.

2) If Router Advertisements are received on some, but not all interfaces, a multihomed host could choose to only send packets out on the interfaces on which it has received Router Advertisements. A key assumption made here, however, is that routers on those other interfaces will be able to route packets to the ultimate destination, even when those destinations reside on the subnet to which the sender connects, but has no on-link prefix information.

Should the assumption be false, communication would fail. Even if the assumption holds, packets will traverse a sub-optimal path.

10. PROTOCOL CONSTANTS

Router constants:

MAX_INITIAL_RTR_ADVERT_INTERVAL	16	seconds
MAX_INITIAL_RTR_ADVERTISEMENTS	3	transmissions
MAX_RTR_RESPONSE_DELAY	2	seconds

Host constants:

MAX_RTR_SOLICITATION_DELAY	1	second
RTR_SOLICITATION_INTERVAL	3	seconds
MAX_RTR_SOLICITATIONS	3	transmissions

Node constants:

RESOLVE_RETRANS_TIMER	1	second
MAX_MULTICAST_SOLICIT	3	transmissions
MAX_UNICAST_SOLICIT	3	transmissions
MAX_NEIGHBOR_ADVERTISEMENT	3	transmissions
MIN_NEIGHBOR_ADVERT_INTERVAL	16	seconds

REACHABLE_TIME	30,000 milliseconds
REACHABLE_RETRANS_TIMER	3,000 milliseconds
DELAY_FIRST_PROBE_TIME	3 seconds

Additional protocol constants are defined with the message formats in Section 5.1, 6.1, and 7.1.

All protocol constants are subject to change in future revisions of the protocol.

11. SECURITY CONSIDERATIONS

Neighbor Discovery protocol packet exchanges can be authenticated using the IP Authentication Header [IPv6-AUTH]. A node SHOULD include an Authentication Header when sending Neighbor Discovery packets if a security association exists for the destination address. The security

associations may have been created through manual configuration or through the operation of some key management protocol.

Received Authentication Headers in Neighbor Discovery packets MUST be verified for correctness and packets with incorrect authentication MUST be ignored.

It SHOULD be possible for the system administrator to configure a node to ignore any Neighbor Discovery messages that are not authenticated using either the Authentication Header or Encapsulating Security Payload. The configuration technique for this MUST be documented. Such a switch SHOULD default to allowing unauthenticated messages.

Confidentiality issues are addressed by the IP Security Architecture and the IP Encapsulating Security Payload documents [IPv6-SA, IPv6-ESP].

The trust model for redirects is based only on trusting a redirect received from the current first-hop node as in IPv4. It is natural to trust the routers on the link. If a host has been redirected to another host (i.e. the destination is on-link) there is no way to prevent the target from issuing another redirect to some other destination. However, this exposure is no worse than it was; the target host, once

subverted, could always act as a hidden router to forward traffic elsewhere.

The protocol contains no mechanism to determine which nodes are authorized to send Router Advertisements; any node, presumably even in the presence of authentication, can send Router Advertisement messages thereby being able to cause denial of service. Furthermore, any node can send proxy Neighbor Advertisements as well as unsolicited Neighbor Advertisements as a potential denial of service attack.

REFERENCES

[ADDRCONF] S. Thomson, "IPv6 Address Autoconfiguration", Internet Draft.

[ADDR-ARCH] S. Deering, R. Hinden, Editors, "IP Version 6 Addressing Architecture", Internet Draft.

[ANYCST] C. Partridge, T. Mendez, and W. Milliken, "Host Anycasting Service", RFC 1546, November 1993.

[ARP] D. Plummer, "An Ethernet Address Resolution Protocol", STD 37, RFC 826, November 1982.

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

[ICMPv4] J. Postel, "Internet Control Message Protocol", STD 5, RFC 792, September 1981.

[ICMPv6] A. Conta, and S. Deering, "ICMP for the Internet Protocol Version 6 (IPv6)", Internet Draft.

[IPv6] S. Deering, R. Hinden, Editors, "Internet Protocol, Version 6 (IPv6) Specification", Internet Draft.

[IPv6-SA] R. Atkinson. IPv6 Security Architecture. Internet Draft, March 1995.

[IPv6-AUTH] R. Atkinson. IPv6 Authentication Header. Internet Draft, March 1995.

[IPv6-ESP] R. Atkinson. IPv6 Encapsulating Security Payload. Internet Draft, February 1995.

[RDISC] S. Deering, "ICMP Router Discovery Messages", RFC 1256, September 1991.

[SH-MEDIA] R. Braden, J. Postel, Y. Rekhter, "Internet Architecture Extensions for Shared Media", RFC 1620, May 1994.

[ASSIGNED] J. Reynolds, J. Postel, "ASSIGNED NUMBERS", RFC 1700, October 1994.

[SYNC] S. Floyd, V. Jacobsen, "The Synchronization of Periodic Routing Messages", IEEE/ACM Transactions on Networking, April 1994. ftp://ftp.ee.lbl.gov/papers/sync_94.ps.Z

AUTHORS' ADDRESSES

Erik Nordmark
Sun Microsystems, Inc.
2550 Garcia Ave
Mt. View, CA 94041
USA

Thomas Narten
IBM Corporation
P.O. Box 12195
Research Triangle Park, NC 27709-2195
USA

phone: +1 415 336 2788
fax: +1 415 336 6015
email: nordmark@sun.com

phone: +1 919 254 7798
fax: +1 919 254 4027
email: narten@vnet.ibm.com

William Allen Simpson
Daydreamer
Computer Systems Consulting Services
1384 Fontaine
Madison Heights, Michigan 48071
USA

email: Bill.Simpson@um.cc.umich.edu
       bsimpson@MorningStar.com

CHANGES SINCE PREVIOUS DOCUMENT

There are several changes since the previous version documented in:
<draft-ietf-ipngwg-discovery-00.txt>
based on feedback from the working group:

- Renamed the "Next-hop Cache" to the "Destination Cache".
- Renamed "extensions" to "options".
- Renamed "lifetime-as-default" to "Router Lifetime".
- Clarified that on-link is a property of an address whereas neighbor is a property of a node.
- Changed the solicited-node multicast address range from 256 addresses to 4 billion (2^32) addresses.
- Removed use of all-hosts multicast address. Router Advertisement messages are now sent to all-nodes. This allows routers to verify the consistency of the information different routers advertise.
- Removed the preference field from Router Advertisements. Simplified the default router selection algorithm as a result.
- Moved the information carried in the MTU and NUD timer extensions into the fixed part of the RA header. Added an additional timer for NUD. Made the NUD timers 32 bits in milliseconds.
- Made the random component for the time between subsequent Router Advertisements larger. The minimum is now 1/3 of the maximum value resulting in the range between 0.5Tp and 1.5Tp suggested in [SYNC].
- Require that host MUST maintain at least two default routers (rather than just 1)
- Made the MTU option carry a 32 bit MTU (for jumbogram capable links)
- Specified an infinity value (all one bits) for prefix Invalidation and Deprecation Lifetime.
- Made NUD use timers to retransmit probes if the first probe is not answered. This provides an hard upper bound on the

time it takes to detect an unreachable neighbor.

- Specified that Neighbor Unreachability Detection applies to router-router communication unless there is some other mechanism which ensures two-way reachability between router neighbors.

- Changed "ICMP unreachable error" to "ICMP unreachable indication" throughout, defined "indication" in the definition section explaining how errors are handled when they occur on the same node from where the packet originates.

- Require that Router Advertisement, Router Solicitation, and Redirect messages are sent with a link-local source address for improved robustness.

- Removed the use of Code 0 vs. 1 in all messages. Added a "Router flag" field to the Neighbor Advertisement message as a replacement for the Code 0/1 distinction.

- Added a Solicited flag to Neighbor Advertisements to make it possible to send unicast unsolicited advertisements without confusing the unreachability detection.

- Specified the initial neighbor reachability state when creating Neighbor Cache entries.

- Revised multihomed host section to better describe problems, without suggesting proper behavior.

- Simplified the link-layer address encoding in the options by making it link specified. Removed the address family field from the option.

- Reduced the minimum allowed time for MaxRtrAdvInterval to 1 second to make it more suitable for beaconing.

- Relaxed the constraints on sending unsolicited Neighbor Advertisements.

The changes incorporated in <draft-ietf-ipngwg-discovery-00.txt> compared to the previous version documented in:

<draft-simpson-ipv6-discov-formats-02.txt>, and
<draft-simpson-ipv6-discov-process-02.txt>

The changes agreed to at working group meetings at Xerox Parc and at Danvers IETF:

- Renamed the Media-Access extension to be the Link-Layer Address extension.
- Use of different extensions for addresses that refer to the sender of the packet and the receiver instead of using the Known-Identifier extension for both.
- Changed the processing of General/Neighbor Solicitation in order to achieve 2 packet exchange just like ARP.
- Removed the Node-Heard extension.

Other changes:

- Merged the processing and format documents into a single document with an extensive introduction to the protocol.
- Aligned the document with [ADDRCONF]. In particular this implied the removal of the Change-Identifier extension.
- Off-link prefixes are not advertized in Router Advertisements (no simple routing protocol). This removes the need for a preference in the Prefix Information extension.
- Specified a more detailed Neighbor Unreachability Detection algorithm (used to be called Dead Node Detection).
- Removed the lifetime field from Neighbor Advertisements. The protocol uses Neighbor Unreachability Detection to time out state created by Neighbor Advertisements.
- Removed the Maximum Receive Unit fields from packets since per-node MTU (or MRU) links do not work with multicast. Instead routers send an MTU extension in order to handle

links that do not have a well-defined MTU.

- Changed alignment mechanisms for extensions. All extensions are a multiple of 8 octets. Thus there is no longer a need for pad extensions.

- Added support for anycast addresses.

- Removed the ability to redirect prefixes to simplify host

processing.

- Removed lingering mobility support (Mobility-Support extension and Remote Redirect message.)

- All messages have separate ICMP types. Redirect type is now in the error range (<128) and the others in the information range (>==128)

- Moved fixed-length fields that are always present in a particular type of packet into the fixed header.

- Renamed "General" Solicitation/Advertisement to "Neighbor" Solicitation/Advertisement.

- Changed the default Router Advertisement period from 30 seconds to 600 seconds; same value as in RFC-1256. This change is possible since Neighbor Unreachability Detection will detect unreachable routers and switch a reachable router independent of the frequency of the Router Advertisements.

- Specified rules for when a node should generate ICMP address unreachable errors due to Address Resolution failures.

OPEN ISSUES

- Default timer values for NUD? Some routers might not respond in a timely manner to solicitations when they are busy processing routing updates. NUD as specified will give up after 3 transmissions spaced 3 seconds apart thereby requiring that a router respond in 9 seconds.

- Timer values and retransmissions for address resolution. Is 3 transmissions separated by 1 second sufficient or should the nodes retransmit for a longer time?

- Will all links (including point-to-point links) provide a link-local address?

- Should we remove the Redirected header option? The redirect message contains all the needed information so the only use of the included header is potentially for trouble shooting and/or if implementations want to verify the content of the included packet as being a packet that was recently sent.

- Allow for balanced load sharing between multiple default routers? This would require that hosts somehow randomly select a routers from the default router list. Do we want to require hosts to do that?

- Should it be possible to disable the Neighbor Unreachability Detection mechanism? Is it sufficient to set the Reachable Retrans timer to 2^32-1 milliseconds? (about 46 days)