IPng Working Group
R. Draves
Internet Draft
Microsoft Research
Document: draft-ietf-ipv6-router-selection-02.txt
R. Hinden

Nokia
June 10, 2002

Default Router Preferences, More-Specific Routes, and Load Sharing

Abstract

This document describes two changes to Neighbor Discovery. The first change is an optional extension to Router Advertisement messages for communicating default router preferences and more-specific routes from routers to hosts. This improves the ability of hosts to pick an appropriate router, especially when the host is multi-homed and the routers are on different links. The preference values and specific routes advertised to hosts require administrative configuration; they are not automatically derived from routing tables. The second change is a mandatory modification of the conceptual sending algorithm to support load-sharing among equivalent routers.

1. Introduction

Neighbor Discovery [2] specifies a conceptual model for hosts that includes a Default Router List and a Prefix List. Hosts send Router Solicitation messages and receive Router Advertisement messages from routers. Hosts populate their Default Router List and Prefix List based on information in the Router Advertisement messages. A conceptual sending algorithm uses the Prefix List to determine if a destination address is on-link and the Default Router List to select a router for off-link destinations.

In some network topologies where the host has multiple routers on its Default Router List, the choice of router for an off-link destination is important. In some situations, one router may provide much better performance than another for a destination. In other situations, choosing the wrong router may result in a failure to communicate. (A later section gives specific examples of these scenarios.)

This document describes an optional extension to Neighbor Discovery Router Advertisement messages for communicating default router preferences and more-specific routes from routers to hosts. This improves the ability of hosts to pick an appropriate router for an off-link destination.

Neighbor Discovery provides a Redirect message that routers can use to correct a host's choice of router. A router can send a Redirect message to a host, telling it to use a different router for a specific destination. However, the Redirect functionality is limited to a single link. A router on one link cannot redirect a host to a router on another link. Hence, Redirect messages do not help multi-homed hosts select an appropriate router.

Multi-homed hosts are an increasingly important scenario, especially with IPv6. In addition to a wired network connection, like Ethernet, hosts may have one or more wireless connections, like 802.11 or Bluetooth. In addition to physical network connections, hosts may have virtual or tunnel network connections. For example, in addition to a direct connection to the public Internet, a host may have a tunnel into a private corporate network. Some IPv6 transition scenarios can add additional tunnels. For example, hosts may have 6-over-4 [3] or configured tunnel [4] network connections.

This document requires that the preference values and specific routes advertised to hosts require explicit administrative configuration. They are not automatically derived from routing tables. In particular, the preference values are not routing metrics and it is not recommended that routers "dump out" their entire routing tables to hosts.

We use Router Advertisement messages, instead of some other protocol like RIP [5], because Router Advertisements are an existing standard, stable protocol for router-to-host communication. Piggybacking this information on existing message traffic from routers to hosts reduces network overhead. Neighbor Discovery is to unicast routing as Multicast Listener Discovery is to multicast routing. In both cases, a single simple protocol insulates the host from the variety of router-to-router protocols. In addition, RIP is unsuitable because it does not carry route lifetimes so it requires frequent message traffic with greater processing overheads.

This document also describes a mandatory change in host behavior. Neighbor DiscoveryÆs conceptual sending algorithm is modified to require hosts to select randomly among equivalent routers. This

distributes traffic to different destinations among the routers. Traffic to a single destination continues to use a single router, because of the Destination Cache.

The mechanisms specified here are backwards-compatible, so that hosts that do not implement them continue to function as well as they did previously.

1.1. Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC-2119 [6].

2. Message Formats

2.1. Preference Values

Default router preferences and preferences for more-specific routes are encoded the same way.

Preference values are encoded in two bits, as follows:

```
     01      High
     00      Medium (default)
     11      Low
     10      Reserved - MUST NOT be sent
```

Note that implementations can treat the value as a two-bit signed integer.

Having just three values reinforces that they are not metrics and more values does not appear to be necessary for reasonable scenarios.

2.2. Changes to Router Advertisement Message Format

The changes from Neighbor Discovery [2] section 4.2 are as follows:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |     Code      |          Checksum             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Cur Hop Limit |M|O|H|Prf|Resvd|       Router Lifetime         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Reachable Time                        |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          Retrans Timer                        |
```

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Options ...
+-+-+-+-+-+-+-+-+-+-+-+-
```

Fields:

Prf (Default Router Preference)
2-bit signed integer. Indicates whether or not to prefer this router over other default routers. If Router Lifetime is zero, it MUST be initialized to zero by the sender and MUST be ignored by the receiver. If the Reserved (10) value is received, the receiver should treat the RA as having a zero Router Lifetime.

Resvd (Reserved)
A 3-bit unused field. It MUST be initialized to zero by the sender and MUST be ignored by the receiver.

Possible Options:

Route Information
These options specify prefixes that are reachable via the router.

Discussion:

Note that in addition to the preference value in the message header, a Router Advertisement can also contain a Route Information Option for ::/0, with a preference value and lifetime. Encoding a preference value in the Router Advertisement header has some advantages:

1. It allows for a distinction between "best default router" and "best router for default", as described below in section 5.1.

2. When the best default router is also the best router for default (which will be a common case), encoding the preference value in the message header is more efficient than having to send a separate option.

2.3. Route Information Option

```
0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|     Type      |    Length     | Prefix Length |Resvd|Prf|Resvd|
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                        Route Lifetime                         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
+                                                               +
```

```
|                                                               |
+                            Prefix                             +
|                                                               |
+                                                               +
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Fields:

Type TBD

Length 8-bit unsigned integer. The length of the option (including the Type and Length fields) in units of 8 octets.

Prefix Length
8-bit unsigned integer. The number of leading bits in the Prefix that are valid. The value ranges from 0 to 128.

Prf (Route Preference)
2-bit signed integer. Indicates whether or not to prefer this router for the prefix over others. If the Reserved (10) value is received, the Route Information Option MUST be ignored.

Resvd (Reserved)
Two 3-bit unused fields. They MUST be initialized to zero by the sender and MUST be ignored by the receiver.

Route Lifetime
32-bit unsigned integer. The length of time in seconds (relative to the time the packet is sent) that the prefix is valid for route determination. A value of all one bits (0xffffffff) represents infinity.

Prefix Variable-length field containing an IP address or a prefix of an IP address. The Prefix Length field contains the number of valid leading bits in the prefix. The bits in the prefix after the prefix length (if any) are reserved and MUST be initialized to zero by the sender and ignored by the receiver.

The Length field is 1, 2, or 3 depending on Prefix Length. If Prefix Length is greater than 64, then Length must be 3. If Prefix Length is greater than 0, then Length must be 2 or 3. If Prefix Length is zero, then Length must be 1, 2, or 3.

The Prefix field is 0, 8, or 16 octets depending on Length.

Routers SHOULD NOT include in a Router Advertisement two Route Information Options with the same Prefix and Prefix Length. If a host processes a Router Advertisement carrying multiple Router Information Options with the same Prefix and Prefix Length, it MUST process one of the options (unspecified which one) and it MUST effectively ignore the rest. It MUST NOT retain some information (like preference) from one option and other information (like lifetime) from another option.

Discussion:

There are several reasons for using a new Route Information Option, instead of using flag bits to overload the existing Prefix Information Option:

1. Prefixes will typically only show up in one or the other kind of option, not both, so a new option does not introduce duplication.

2. The Route Information Option is typically 16 octets while the Prefix Information Option is 32 octets.

3. Using a new option may improve backwards-compatibility with some host implementations.

3. Conceptual Model of a Host

There are three possible conceptual models for host implementation of default router preferences and more-specific routes, corresponding to different levels of support. We refer to these as host A, host B, and host C. Note that these are really classes of hosts, not individual hosts.

3.1. Conceptual Data Structures for Hosts

Host A ignores default router preferences and more-specific routes. Host A uses the conceptual data structures described in Neighbor Discovery [2].

Host B uses a Default Router List augmented with preference values. Host B does not have a routing table. Host B uses the Default Router Preference value in the Router Advertisement header. Host B ignores Route Information Options.

Host C uses a Routing Table instead of a Default Router List. (The Routing Table may also subsume the Prefix List, but that is beyond the scope of this document.) Entries in the Routing Table have a prefix, prefix length, preference value, lifetime, and next-hop router. Host C uses both the Default Router Preference value in the Router Advertisement header and Route Information Options.

When host C receives a Router Advertisement, it modifies its Routing

Table as follows. If the received route's lifetime is zero, the route is removed from the Routing Table if present. If a route's lifetime is non-zero, the route is added to the Routing Table if not present and the route's lifetime and preference is updated if the route is already present. A route is located in the Routing Table based on prefix, prefix length, and next-hop router. When processing a Router Advertisement, host C first updates a ::/0 route based on the Router Lifetime and Default Router Preference in the Router Advertisement message header. Then as host C processes Route Information Options in the Router Advertisement message body, it updates its routing table for each such option. The Router Preference and Lifetime values in a ::/0 Route Information Option

override the preference and lifetime values in the Router Advertisement header.

For example, suppose a host receives a Router Advertisement from router X with a Router Lifetime of 100 seconds and Default Router Preference of Medium. The body of the Router Advertisement contains a Route Information Option for ::/0 with a Route Lifetime of 200 seconds and a Route Preference of Low. After processing the Router Advertisement, host A will have an entry for router X in its Default Router List with lifetime 100 seconds. If host B receives the same Router Advertisement, it will have an entry in its Default Router List for router X with Medium preference and lifetime 100 seconds. Host C will have an entry in its Routing Table for ::/0 -> router X, with Low preference and lifetime 200 seconds. Host C MAY have a transient state, during processing of the Router Advertisement, in which it has an entry in its Routing Table for ::/0 -> router X with Medium preference and lifetime 100 seconds.

3.2. Conceptual Sending Algorithm for Hosts

Host A uses the conceptual sending algorithm described in Neighbor Discovery [2], modified slightly to support load sharing as described in section 3.5.

When host B does next-hop determination and consults its Default Router List, it primarily prefers reachable routers over non-reachable routers and secondarily uses the router preference values.

When host C does next-hop determination and consults its Routing Table for an off-link destination, it first prefers reachable routers over non-reachable routers, second uses longest-matching-prefix, and third uses route preference values.

If there are no routes matching the destination (i.e., no default routes and no more-specific routes), then if host C has a single interface then it SHOULD assume the destination is on-link to that interface. If host C has multiple interfaces then it SHOULD discard the packet and report a Destination Unreachable / No Route To Destination error to the upper layer.

3.3. Destination Cache Management

When host C processes a Router Advertisement and updates its conceptual Routing Table, it SHOULD invalidate or remove Destination Cache Entries and redo next-hop determination for destinations affected by the Routing Table changes. The host MAY implement this requirement by flushing its entire Destination Cache.

3.4. Router Reachability Probing

When a host avoids using a non-reachable router X and instead uses another router Y, and the host would have used router X if router X were reachable, then the host SHOULD probe router X's reachability

by sending a Neighbor Solicitation. A host MUST NOT probe a router's reachability in the absence of useful traffic that the host would have sent to the router if it were reachable. In any case, these probes MUST be rate-limited to no more than one per minute per router.

This requirement allows the host to discover when router X becomes reachable and to start using router X at that point. Otherwise, the host might not notice router XÆs reachability and continue to use the less-desirable router Y.

3.5. Host Load Sharing

Sometimes a host has a choice of multiple "equivalent" routers for a destination. We say that two routers are equivalent for a destination if they have the same reachability, the same matching prefix length (if the host supports a Routing Table), and the same preference values (if the host supports preference values).

When a host chooses from multiple equivalent routers, it MUST choose randomly.

This has the effect of distributing load for new destinations among the equivalent routers. Note that traffic to a single destination will use a single router as long as the Destination Cache Entry for the destination is not deleted. Random selection, instead of round-robin, is used to avoid synchronization issues.

3.6. Example

For example: suppose host C has five entries in its Routing Table:

```
::/0 -> router W with Medium preference
2001::/16 -> router X with Medium preference
3ffe::/16 -> router Y1 with High preference
3ffe::/16 -> router Y2 with High preference
3ffe::/16 -> router Z with Low preference
```

and host C is sending to 3ffe::1, an off-link destination. If all routers are reachable, then the host will choose randomly between routers Y1 and Y2. If routers Y1 and Y2 are not reachable, then router Z will be chosen and the reachability of routers Y1 and Y2 will be probed. If routers Y1, Y2, and Z are not reachable, then router W will be chosen and the reachability of routers Y1, Y2, and Z will be probed. If routers W, Y1, Y2, and Z are all not reachable, then host C should round-robin among Y1 and Y2 while probing the reachability of W and Z. Router X will never be chosen because its prefix does not match the destination.

4. Router Configuration

Routers should not advertise preferences or routes by default. In particular, they should not "dump out" their entire routing table to hosts. Routers MAY have a configuration mode where a filter is

applied to their routing table to obtain the routes that are advertised to hosts.

The preference values (both Default Router Preferences and Route Preferences) should not be routing metrics or automatically derived from metrics: the preference values should be configured. The High and Low (non-default) preference values should only be used when someone with knowledge of both routers and the network topology configures them explicitly. For example, it could be a common network administrator, or it could be a customer request to different administrators managing the routers.

As one exception to this general rule, the administrator of a router that does not have a connection to the internet, or is connected through a firewall that blocks general traffic, may configure the router to advertise a Low Default Router Preference.

An administrator of a router may configure the router to advertise specific routes for directly connected subnets and any shorter prefixes (eg, site, NLA, or TLA prefixes) for networks to which the router belongs.

For example, if a home user sets up a tunnel into a firewalled corporate network, the access router on the corporate network end of the tunnel can advertise itself as a default router, but with a Low preference. Furthermore the corporate router can advertise a specific route for the corporate site prefix. The net result is that destinations in the corporate network will be reached via the tunnel, and general internet destinations will be reached via the home ISP. Without these mechanisms, the home machine might choose to send internet traffic into the corporate network or corporate traffic into the internet, leading to communication failure because of the firewall.

Routers SHOULD NOT send more than 17 Route Information Options in

Router Advertisements per link. This arbitrary bound is meant to reinforce that relatively few and carefully selected routes should be advertised to hosts.

5. Examples

5.1. Best Default Router vs Best Route for Default

The best default router is not quite the same thing as the best router for default. The best default router is the router that will generate the fewest number of redirects for the host's traffic. The best router for default is the router with the best route toward the wider internet.

For example, suppose a situation where you have a link with two routers X and Y. Router X is the best for 2002::/16. (It's your 6to4 site gateway.) Router Y is the best for ::/0. (It connects to the native IPv6 internet.) Router X forwards native IPv6 traffic to

router Y; router Y forwards 6to4 traffic to router X. But most traffic from this site is sent to 2002:/16 destinations. In this scenario, router X is the best default router and router Y is the best router for default.

To make host A work well, both routers should advertise themselves as default routers. In particular, if router Y goes down host A should send traffic to router X to maintain 6to4 connectivity, so router X as well as router Y needs to be a default router.
To make host B work well, router X should in addition advertise itself with a High default router preference. This will cause host B to prefer router X, minimizing the number of redirects.

To make host C work well, router X should in addition advertise the ::/0 route with Low preference and the 2002::/16 route with Medium preference. Host C will end up with three routes in its routing table: ::/0 -> router X (Low), ::/0 -> router Y (Medium), 2002::/16 -> router X (Medium). It will send 6to4 traffic to router X and other traffic to router Y. Host C will not cause any redirects.

Note that when host C processes the Router Advertisement from router X, the Low preference for ::/0 overrides the High default router preference. If the ::/0 specific route were not present, then host C would apply the High default router preference to its ::/0 route to router X.

5.2. Multi-Homed Host and Isolated Network

Here's another scenario: a multi-homed host is connected to the 6bone/internet via router X on one link and to an isolated network via router Y on another link. The multi-homed host might have a tunnel into a fire-walled corporate network, or it might be directly connected to an isolated test network.

In this situation, a multi-homed host A (which has no default router preferences or more-specific routes) will have no way to choose between the two routers X and Y on its Default Router List. Users of the host will see unpredictable connectivity failures, depending on the destination address and the choice of router.

A multi-homed host C in this same situation can correctly choose between routers X and Y, if the routers are configured appropriately. For example, router X on the isolated network should advertise a Route Information Option for the isolated network prefix. It might not advertise itself as a default router at all (zero Router Lifetime), or it might advertise itself as a default router with Low preference. Router Y should advertise itself as a default router with Medium preference.

6. Security Considerations

A malicious node could send Router Advertisement messages, specifying High Default Router Preference or carrying specific

routes, with the effect of pulling traffic away from legitimate routers. However, a malicious node could easily achieve this same effect in other ways. For example, it could fabricate Router Advertisement messages with zero Router Lifetime from the other routers, causing hosts to stop using the other routes. Hence, this document has no new appreciable impact on Internet infrastructure security.

References

1 S. Bradner, "The Internet Standards Process -- Revision 3", BCP 9, RFC 2026, October 1996.

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

3 B. Carpenter, K. Moore. "Connection of IPv6 Domains via IPv4 Clouds", RFC 3056, February 2001.

4 R. Gilligan, E. Nordmark. "Transition Mechanisms for IPv6 Hosts and Routers", RFC 1933, April 1996.

5 G. Malkin, R. Minnear. "RIPng for IPv6", RFC 2080 , January 1997.

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

Acknowledgments

The authors would like to acknowledge the contributions of Balash Akbari, Steve Deering, Robert Elz, Tony Hain, Christian Huitema,

JINMEI Tatuya, Erik Nordmark, Pekka Savola, Dave Thaler, and Brian Zill. The packet diagrams are derived from Neighbor Discovery [2]. The description of host load sharing is derived from Bob Hinden's draft on the subject.

Author's Addresses

Richard Draves
Microsoft Research
One Microsoft Way
Redmond, WA 98052
Phone: +1 425 706 2268
Email: richdr@microsoft.com

Robert M. Hinden
Nokia
313 Fairchild Drive
Mountain View, CA 94043
Phone: +1 650 625 2004
Email: hinden@iprg.nokia.com

Revision History

Changes from draft-ietf-ipv6-router-selection-01

Added Bob Hinden as co-author.

Various clarifications and textual improvements.

Slightly simplified the specification of round-robining in next-hop determination, relying on router-reachability probing in some cases.

Clarified that router reachability probing only happens when the host is sending packets that would have gone to that router if it were reachable.

Changed load sharing to a mandatory requirement and added supporting text to the title, abstract, and introduction.

Changes from draft-ietf-ipngwg-router-selection-00

Specified reachability probing of otherwise more-preferred but currently unreachable routers.

Changed the requirement of Destination Cache invalidation, from MAY to SHOULD, but allowing flushing of the entire Destination Cache.

Added a section specifying load sharing among equivalent routers.

Changes from draft-draves-ipngwg-router-selection-01

Specified receiver processing when the Reserved preference value is seen.

Specified that routers SHOULD NOT send more than 17 Route Information Options.

Added discussion of Destination Cache invalidation, allowing but not requiring it.

Removed references to the fourth conceptual host model, host D.

Changes from draft-draves-ipngwg-router-selection-00

Made the option variable length. Must ignore prefix bits past prefix length.

Added more allowable router configuration scenarios, weakening the requirement that one administrator must coordinate the configuration of all relevant routers.

Full Copyright Statement