Network Working Group
Request for Comments: 4191
Category: Standards Track

R. Draves
D. Thaler
Microsoft

November 2005

Default Router Preferences and More-Specific Routes

Status of This Memo

Copyright Notice

Abstract

This document describes 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.

1. Introduction

Neighbor Discovery [RFC2461] 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 uses 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. (Section 5 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.

Note that since these procedures are applicable to hosts only, the forwarding algorithm used by the routers (including hosts with enabled IP forwarding) is not affected.

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 (through multiple interfaces) 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 6to4 [RFC3056] or configured tunnel [RFC2893] 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 [RFC2080], 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 shares with Multicast Listener Discovery the property that they both define host-to-router interactions, while shielding the host from having to participate in more general router-to-router interactions. In addition, RIP is unsuitable because it does not carry route lifetimes so it requires frequent message traffic with greater processing overheads.

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 [RFC2119].

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 as a two-bit signed integer, as follows:

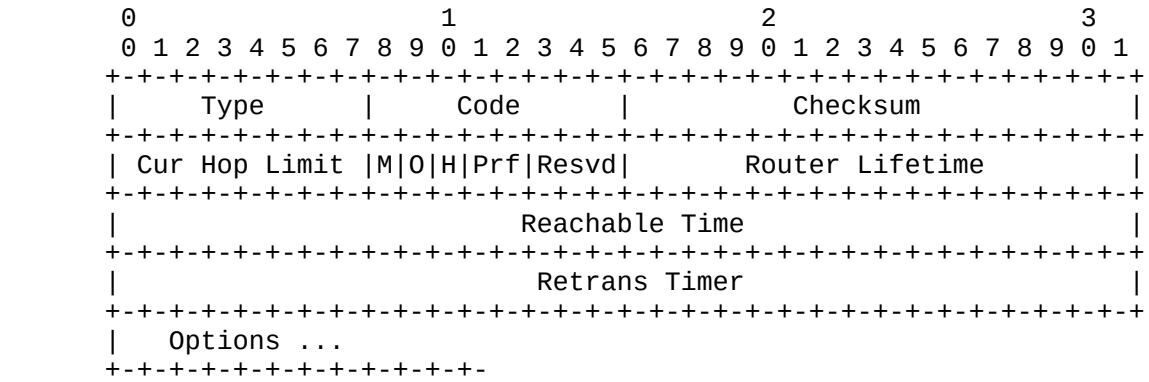
01	High
00	Medium (default)
11	Low
10	Reserved - MUST NOT be sent

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 do not appear to be necessary for reasonable scenarios.

2.2. Changes to Router Advertisement Message Format

The changes from Neighbor Discovery [RFC2461] Section 4.2 and [RFC3775] Section 7.1 are as follows:



Fields:

Prf (Default Router Preference)
: 2-bit signed integer. Indicates whether to prefer this router over other default routers. If the Router Lifetime is zero, the preference value MUST be set to (00) by the sender and MUST be ignored by the receiver. If the Reserved (10) value is received, the receiver MUST treat the value as if it were (00).

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.

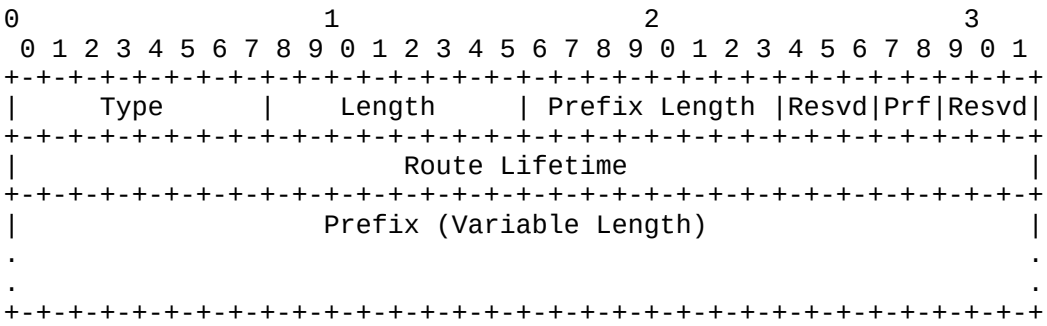
Discussion:

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 the "best router for the default route" and the "router least likely to redirect common traffic", as described below in Section 5.1.

2. When the best router for the default route is also the router least likely to redirect common traffic (which will be a common case), encoding the preference value in the message header is more efficient than sending a separate option.

2.3. Route Information Option



Fields:

Type 24

Length 8-bit unsigned integer. The length of the option (including the Type and Length fields) in units of 8 octets. The Length field is 1, 2, or 3 depending on the 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.

Prefix Length
8-bit unsigned integer. The number of leading bits in the Prefix that are valid. The value ranges from 0 to 128. The Prefix field is 0, 8, or 16 octets depending on Length.

Prf (Route Preference)
2-bit signed integer. The Route Preference indicates whether to prefer the router associated with this prefix over others, when multiple identical prefixes (for different routers) have been received. 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.

Routers MUST NOT include two Route Information Options with the same Prefix and Prefix Length in the same Router Advertisement.

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 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 a host implementation of default router preferences and more-specific routes, corresponding to different levels of support. We refer to these as type A, type B, and type C.

3.1. Conceptual Data Structures for Hosts

Type A hosts ignore default router preferences and more-specific routes. They use the conceptual data structures described in Neighbor Discovery [RFC2461].

Type B hosts use a Default Router List augmented with preference values, but ignore all Route Information Options. They use the Default Router Preference value in the Router Advertisement header. They ignore Route Information Options.

Type C hosts use 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. Type C hosts use both the Default Router Preference value in the Router Advertisement header and Route Information Options.

When a type C host receives a Router Advertisement, it modifies its Routing Table as follows. When processing a Router Advertisement, a type C host first updates a ::/0 route based on the Router Lifetime and Default Router Preference in the Router Advertisement message header. Then as the host 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. Updating each route is done as follows. A route is located in the Routing Table based on the prefix, prefix length, and next-hop router. 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.

For example, suppose hosts receive a Router Advertisement from router X with a Router Lifetime of 100 seconds and a 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, a type A host will have an entry for router X in its Default Router List with a lifetime of 100 seconds. If a type B host receives the same Router Advertisement, it will have an entry for router X in its Default Router List with a Medium preference and a lifetime of 100 seconds. A type C host will have an entry in its Routing Table for ::/0 -> router X, with a Low preference and a lifetime of 200 seconds. During processing of the Router Advertisement, a type C host MAY have a transient state, in which it has an entry in its Routing Table for ::/0 -> router X with a Medium preference and a lifetime of 100 seconds.

3.2. Conceptual Sending Algorithm for Hosts

Type A hosts use the conceptual sending algorithm described in Neighbor Discovery [RFC2461].

When a type B host 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. If the host has no information about the router's reachability, then the host assumes the router is reachable.

When a type C host does next-hop determination and consults its Routing Table for an off-link destination, it searches its routing table to find the route with the longest prefix that matches the destination, using route preference values as a tie-breaker if multiple matching routes have the same prefix length. If the best route points to a non-reachable router, this router is remembered for the algorithm described in Section 3.5 below, and the next best route is consulted. This check is repeated until a matching route is found that points to a reachable router, or no matching routes remain. Again, if the host has no information about the router's reachability, then the host assumes the router is reachable.

If there are no routes matching the destination (i.e., no default routes and no more-specific routes), then a type C host SHOULD discard the packet and report a Destination Unreachable/No Route To Destination error to the upper layer.

3.3. Destination Cache Management

When a type C host processes a Router Advertisement and updates its conceptual Routing Table, it MUST invalidate or remove Destination Cache Entries and redo next-hop determination for destinations affected by the Routing Table changes.

3.4. Client Configurability

Type B and C hosts MAY be configurable with preference values that override the values in Router Advertisements received. This is especially useful for dealing with routers that may not support preferences.

3.5. Router Reachability Probing

When a host avoids using any non-reachable router X and instead sends a data packet to another router Y, and the host would have used router X if router X were reachable, then the host SHOULD probe each such router X's reachability by sending a single Neighbor

Solicitation to that router's address. 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 time. Otherwise, the host might not notice router X's reachability and continue to use the less-desirable router Y until the next Router Advertisement is sent by X. Note that the router may have been unreachable for reasons other than being down (e.g., a switch in the middle being down), so it may be up to 30 minutes (the maximum advertisement period) before the next Router Advertisement would be sent.

For a type A host (following the algorithm in [RFC2461]), no probing is needed since all routers are equally preferable. A type B or C host, on the other hand, explicitly probes unreachable, preferable routers to notice when they become reachable again.

3.6. Example

Suppose a type C host has four entries in its Routing Table:

```
::/0 -> router W with a Medium preference
2002::/16 -> router X with a Medium preference
2001:db8::/32-> router Y with a High preference
2001:db8::/32-> router Z with a Low preference
```

and the host is sending to 2001:db8::1, an off-link destination. If all routers are reachable, then the host will choose router Y. If router Y is not reachable, then router Z will be chosen and the reachability of router Y will be probed. If routers Y and Z are not reachable, then router W will be chosen and the reachability of routers Y and Z will be probed. If routers W, Y, and Z are all not reachable, then the host should use Y 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 in which an announcement of a specific prefix is dependent on a specific condition, such as operational status of a link or presence of the same or another

prefix in the routing table installed by another source, such as a dynamic routing protocol. If done, router implementations SHOULD make sure that announcement of prefixes to hosts is decoupled from the routing table dynamics to prevent an excessive load on hosts during periods of routing instability. In particular, unstable routes SHOULD NOT be announced to hosts until their stability has improved.

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.

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 information contained in Router Advertisements may change through the actions of system management. For instance, the lifetime or preference of advertised routes may change, or new routes could be added. In such cases, the router MAY transmit up to MAX_INITIAL_RTR_ADVERTISEMENTS unsolicited advertisements, using the same rules as in [RFC2461]. When ceasing to be an advertising interface and sending Router Advertisements with a Router Lifetime of zero, the Router Advertisement SHOULD also set the Route Lifetime to zero in all Route Information Options.

4.1. Guidance to Administrators

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, should configure the router to advertise a Low Default Router Preference.

In addition, the administrator of a router should configure the router to advertise a specific route for the site prefix of the network(s) to which the router belongs. The administrator may also configure the router to advertise specific routes for directly connected subnets and any shorter 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 should advertise itself as a default router, but with a Low preference. Furthermore, the corporate router should 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.

It is worth noting that the network administrator setting up preferences and/or more specific routes in Routing Advertisements typically does not know which kind of nodes (Type A, B, and/or C) will be connected to its links. This requires that the administrator configure the settings that will work in an optimal fashion regardless of which kinds of nodes will be attached. Two examples of how to do so follow.

5. Examples

5.1. Best Router for ::/0 vs Router Least Likely to Redirect

The best router for the default route is the router with the best route toward the wider Internet. The router least likely to redirect traffic depends on the actual traffic usage. The two concepts can be different when the majority of communication actually needs to go through some other router.

For example, consider a situation in which 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. If most traffic from this site is sent to 2002:/16 destinations, then router X is the one least likely to redirect.

To make type A hosts work well, both routers should advertise themselves as default routers. In particular, if router Y goes down, type A hosts should send traffic to router X to maintain 6to4 connectivity, so router X and router Y need to be default routers.

To make type B hosts work well, router X should advertise itself with a High default router preference. This will cause type B hosts to prefer router X, minimizing the number of redirects.

To make type C hosts work well, router X should in addition advertise the ::/0 route with a Low preference and the 2002::/16 route with a Medium preference. A type C host 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. Type C hosts will not cause any redirects.

Note that when type C hosts process 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 a type C host would apply the High default router preference to its ::/0 route to router X.

5.2. Multi-Homed Host and Isolated Network

In another scenario, a multi-homed host is connected to the 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 firewalled corporate network, or it might be directly connected to an isolated test network.

In this situation, a type A multi-homed host (which has no default router preferences or more-specific routes) will have no way to intelligently choose between 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.

If the routers are configured appropriately, a multi-homed type B host in this same situation would have stable Internet connectivity, but would have no connectivity to the isolated test network.

If the routers are configured appropriately, a multi-homed type C host in this same situation can correctly choose between routers X and Y. For example, router Y 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 a Low preference. Router X should advertise itself as a default router with a Medium preference.

6. Security Considerations

A malicious node could send Router Advertisement messages, specifying a 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 a zero Router Lifetime from the other routers, causing hosts to stop using the other routes. By advertising a specific prefix, this attack could be carried out in a less noticeable way. However, this attack has no significant incremental impact on Internet infrastructure security.

A malicious node could also include an infinite lifetime in a Route Information Option causing the route to linger indefinitely. A similar attack already exists with Prefix Information Options in RFC 2461, where a malicious node causes a prefix to appear as on-link indefinitely, resulting in a lack of connectivity if it is not. In contrast, an infinite lifetime in a Route Information Option will cause router reachability probing to continue indefinitely, but will not result in a lack of connectivity.

Similarly, a malicious node could also try to overload hosts with a large number of routes in Route Information Options, or with very frequent Route Advertisements. Again, this same attack already exists with Prefix Information Options.

[RFC3756] provides more details on the trust models, and there is work in progress in the SEND WG on securing router discovery messages that will address these problems.

7. IANA Considerations

Section 2.3 defines a new Neighbor Discovery [RFC2461] option, the Route Information Option, which has been assigned the value 24 within the numbering space for IPv6 Neighbor Discovery Option Formats.

8. Acknowledgements

The authors would like to acknowledge the contributions of Balash Akbari, Steve Deering, Robert Elz, Tony Hain, Bob Hinden, Christian Huitema, JINMEI Tatuya, Erik Nordmark, Pekka Savola, Kresimir Segaric, and Brian Zill. The packet diagrams are derived from Neighbor Discovery [RFC2461].

9. Normative References

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

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

[RFC3775] Johnson, D., Perkins, C., and J. Arkko, "Mobility Support in IPv6", RFC 3775, June 2004.

10. Informative References

[RFC2080] Malkin, G. and R. Minnear, "RIPng for IPv6", RFC 2080, January 1997.

[RFC2893] Gilligan, R. and E. Nordmark, "Transition Mechanisms for IPv6 Hosts and Routers", RFC 2893, August 2000.

[RFC3056] Carpenter, B. and K. Moore, "Connection of IPv6 Domains via IPv4 Clouds", RFC 3056, February 2001.

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

Authors' Addresses

Richard Draves
Microsoft Research
One Microsoft Way
Redmond, WA 98052

Phone: +1 425 706 2268
EMail: richdr@microsoft.com

Dave Thaler
Microsoft
One Microsoft Way
Redmond, WA 98052

Phone: +1 425 703 8835
EMail: dthaler@microsoft.com

Full Copyright Statement

Intellectual Property

Acknowledgement

Funding for the RFC Editor function is currently provided by the Internet Society.