6MAN WG
Internet-Draft
Expires: May 17, 2012

E. Nordmark
Cisco Systems, Inc.
I. Gashinsky
Yahoo!
November 14, 2011

Neighbor Unreachability Detection is too impatient
draft-ietf-6man-impatient-nud-00.txt

Abstract

IPv6 Neighbor Discovery includes Neighbor Unreachability Detection. That function is very useful when a host has an alternative, for instance multiple default routers, since it allows the host to switch to the alternative in short time. This time is 3 seconds after the node starts probing. However, if there are no alternatives, this is far too impatient. This document proposes an approach where an implementation can choose the timeout behavior to be different based on whether or not there are alternatives.

Table of Contents

1. Introduction 3
2. Definition Of Terms 3
3. Proposed Remedy 4
4. Acknowledgements 6
5. Security Considerations 6
6. IANA Considerations 6
7. References 6
7.1. Normative References 6
7.2. Informative References 6
Authors' Addresses 6

1. Introduction

IPv6 Neighbor Discovery [RFC4861] includes Neighbor Unreachability Detection, which detects when a neighbor is no longer reachable. The timeouts specified are very short (three transmissions spaced one second apart). That can be appropriate when there are alternative paths the packet can be sent. For example, if a host has multiple default routers in its Default Router List, or if the host has a Neigbor Cache Entry (NCE) created by a Redirect message. The effect of NUD reporting a failure in those cases is that the host will try the alternative; the next router in the Default Router List, or discard the NCE which will also send using a different router.

For that reason the timeouts where chosen to be short; this ensures that if a default router fails the host can use the next router in less than 45 seconds.

However, where there is no alternative there are several benefits in making NUD try probing for a longer time. One of those benefits is to be more robust against transient failures, such as spanning tree recovergence and other layer 2 issues that can take many seconds to resolve. Marking the NCE as unreachable in that case causes additional multicast on the network. Assuming there are IP packets to send, the lack of an NCE will result in multicast Neighbor Solicitations every second instead of the unicast Neighbor Solicitations that NUD sends.

As a result IPv6 is operationally more brittle than IPv4. For IPv4 there is no mandatory time limit on the retransmission behavior for ARP [RFC0826] which allows implementors to pick more robust schemes.

The following constant values in [RFC4861] seem to have been made part of IPv6 conformance testing: MAX_MULTICAST_SOLICIT, MAX_UNICAST_SOLICIT, RETRANS_TIMER. While such strict conformance testing seems consistent with the the specificiation, it means that we need to update the standard if we want to allow IPv6 Neighbor Discovery to be as operationally robust as ARP.

Additional motivations for making IPv6 Neighbor Discovery as robust as ARP are covered in [I-D.gashinsky-v6nd-enhance].

2. Definition Of Terms

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].

3. Proposed Remedy

We can clarify that the giving up after three packets spaced one second apart is only REQUIRED when there is an alternative, such as an additional default route or a redirect.

If implementations transmit more than MAX_*CAST_SOLICIT packets they MAY use binary exponential backoff of the retransmit timer. This is so that if we end up with implementations that try for a very long time we don't end up with a steady background level of retransmissions.

However, even if there is no alternative, we still need to be able to handle the case when the link-layer address of the destination has changed. Thus at some point in time we need to switch to multicast Neighbor Solicitations.

A possible way to describe a node behavior which captures all the cases is to introduce a new, optional, UNREACHABLE state in the conceptual model described in [RFC4861]. A NCE in the UNREACHABLE state retains the link-layer address, and IPv6 packets continue to be sent to that link-layer address. But the Neighbor Soliciations are multicast, using a timeout that follows a binary exponential backoff.

In the places where RFC4861 says to to discard/delete the NCE after N probes (Section 7.3, 7.3.3 and Appendix C) we will instead transition to the UNREACHABLE state.

If the Neighbor Cache Entry was created by a redirect, a node MAY delete the NCE instead of changing its state to UNREACHABLE. In any case, the node SHOULD NOT use an NCE created by a Redirect to send packets if that NCE is in unreachable state. Packets should be sent following the next-hop selection algorithm in section XXX which disregards NCEs that are not reachable.

The default router selection in section 6.3.6 says to prefer default routers that are "known to be reachable". For the purposes of that section, if the NCE for the router is in UNREACHABLE state, it is not known to be reachable. Thus the particular text in section 6.3.6 which says "in any state other than INCOMPLETE" needs to be extended to say "in any state other than INCOMPLETE or UNREACHABLE".

Apart from the use of multicast NS instead of unicast NS, and the binary exponential backoff of the timer, the UNREACHABLE state works the same as the current PROBE state.

A node MAY garbage collect a Neighbor Cache Entry as any time as specified in RFC 4861. This does not change with the introduction of

the UNREACHABLE state in the coneptual model.

The UNREACHABLE state is conceptual and not a required part of this specification. A node merely needs to satisfy the externally observable behavior of this specificiation.

There is a non-obvious extension to the state machine description in Appendix C in RFC 4861 in the case for "NA, Solicited=1, Override=0. Different link-layer address than cached". There we need to add "UNREACHABLE" to the current list of "STALE, PROBE, Or DELAY". That is, the NCE would be unchanged. Note that there is no corresponding change necessary to the text in section 7.2.5 since it is phrased using "Otherwise" instead of explicitly listing the three states.

The other state transitions described in Appendix C handle the introduction of the UNREACHABLE state without any change, since they are described using "not INCOMPLETE".

There is also the more obvious change already described above. RFC 4861 has this:

PROBE	Retransmit timeout, N or more retransmissions.	Discard entry	-

That needs to be replaced by:

PROBE	Retransmit timeout, N or more retransmissions.	Double timeout Send multicast NS	UNREACHABLE
UNREACHABLE	Retransmit timeout	Double timeout Send multicast NS	UNREACHABLE

The binary exponential backoff SHOULD be clamped at some reasonable maximum retransmit timeout, such as 60 seconds. And if there is no IPv6 packets sent using the UNREACHABLE NCE, then it makes sense to stop the retransmits of the multicast NS until either the NCE is garbage collected, or there are IPv6 packets sent using the NCE. In essence the multicast NS and associated binary exponential backoff can be conditioned on the continued use of the NCE to send IPv6 packets to the recorded link-layer address.

A node MAY unicast the first few Neighbor Soliciation messages while in UNREACHABLE state, but it MUST switch to multicast Neighbor Soliciations. Otherwise it would not detect a link-layer address change for the target.

4. Acknowledgements

The comments from Thomas Narten and Philip Homburg have helped improve this draft.

5. Security Considerations

Relaxing the retransmission behavior for NUD has no impact on security. In particular, it doesn't impact applying Secure Neighbor Discovery [RFC3971].

6. IANA Considerations

This are no IANA considerations for this document.

7. References

7.1. Normative References

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

[RFC3971] Arkko, J., Kempf, J., Zill, B., and P. Nikander, "SEcure Neighbor Discovery (SEND)", RFC 3971, March 2005.

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

7.2. Informative References

[I-D.gashinsky-v6nd-enhance] Kumari, W., "Operational Neighbor Discovery Problems and Enhancements.", draft-gashinsky-v6nd-enhance-00 (work in progress), June 2011.

[RFC0826] Plummer, D., "Ethernet Address Resolution Protocol: Or converting network protocol addresses to 48.bit Ethernet address for transmission on Ethernet hardware", STD 37, RFC 826, November 1982.

Authors' Addresses

Erik Nordmark
Cisco Systems, Inc.
510 McCarthy Blvd.
Milpitas, CA, 95035
USA

Phone: +1 408 527 6625
Email: nordmark@cisco.com

Igor Gashinsky
Yahoo!
45 W 18th St
New York, NY
USA

Email: igor@yahoo-inc.com