

MPLS Working Group
Internet Draft
Intended status: Informational
Expires: April 2011

B. Mack-Crane
L. Dunbar
S. Hares
Huawei

October 12, 2010

IPv6 Neighbor Discovery Scalability for Large Data Centers
draft-mackcrane-armd-ipv6-nd-scaling-00.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

The list of current Internet-Drafts can be accessed at <http://www.ietf.org/ietf/lid-abstracts.txt>

The list of Internet-Draft Shadow Directories can be accessed at <http://www.ietf.org/shadow.html>

This Internet-Draft will expire on April 12, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the BSD License.

Internet-Draft

IPv6 ND Scalability

October 2010

Abstract

Server virtualization allows one physical server to support many virtual machines (VMs) so that multiple hosts (20, 30, or hundreds) can be running from one physical platform. As virtual machines are introduced into a Data Center, the number of hosts within the data center can grow dramatically, which can have tremendous impact on the network and hosts.

This document provides an analysis of the scalability of IPv6 Neighbor Discovery ([RFC 4861](#)) in data centers with a large number of virtual machines.

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](#) 0.

Table of Contents

1.	Introduction.....	3
2.	Network functions provided by IPv6 Neighbor Discovery.....	3
3.	Basic ND protocol message use.....	4
3.1.	Router Solicitation.....	4
3.2.	Router Advertisement.....	4
3.3.	Neighbor Solicitation.....	5
3.4.	Neighbor Advertisement.....	5
3.5.	Redirect.....	5
4.	Some additional protocol activities.....	5
4.1.	Duplicate Address Detection.....	5
4.2.	Anycast and Proxy address resolution.....	6
4.3.	Neighbor unreachability detection.....	6
4.4.	Host-based Load Spreading.....	6
4.5.	Router-based Load Spreading.....	6
4.6.	Holding packets while address resolution occurs.....	7
5.	Summary and conclusions.....	7
6.	Manageability Considerations.....	7
7.	Security Considerations.....	7
8.	IANA Considerations.....	8
9.	Acknowledgments.....	8

10. References.....	8
Authors' Addresses.....	8
Intellectual Property Statement.....	9
Disclaimer of Validity.....	9

[1. Introduction](#)

Server virtualization allows the sharing of the underlying physical machine (server) resources among multiple virtual machines (VMs), each running its own operating system. While Server Virtualization is a great technology for flexible management of server resources, it does impose great challenges to networks which interconnect all the servers in data center(s). Large data centers may grow to support hundreds of thousands or even millions of hosts (VMs). Even though there may be enough link bandwidth to support the traffic volume from all those VMs, other issues associated with Layer 2, like frequent ARP broadcast by hosts, broadcast unknown, etc., can create problems for the network and hosts.

This document presents an initial analysis of the scalability of IPv6 Neighbor Discovery ([RFC 4861](#)) protocols in the context of a large data center network. Two network cases are considered: 1) a single L2 VLAN connecting a very large number of hosts and a relatively small number of routers, and 2) a core VLAN connecting a large number of routers and few, if any, hosts. The analysis presented here is a rough assessment of which protocol behaviors should scale well and which may present some concern. It does not provide hard numbers and is not based on any measurements in live networks.

[2. Network functions provided by IPv6 Neighbor Discovery](#)

The protocols described in [RFC 4861](#) provide a variety of network functions used by IPv6 nodes to:

- find routers and discover link and network parameters,
- discover each other's presence,
- determine each other's link-layer addresses, and

- maintain information about the paths to active neighbors.

These functions are accomplished using five ICMP messages:

- Router Solicitation,
- Router Advertisement,
- Neighbor Solicitation,

- Neighbor Advertisement, and
- Redirect.

The first part of the analysis considers the basic ND protocol activities and how often each message is sent and to what L2 destination address to determine whether there is any concern that ND messages could take too much bandwidth or tax host processors with unnecessary work.

The second part of the analysis considers whether there may be scalability concerns related to other protocol behaviors mentioned in [RFC 4861](#) for ancillary purposes, for example duplicate address detection.

[3.](#) Basic ND protocol message use

[3.1.](#) Router Solicitation

The Router Solicitation message is sent by nodes to discover routers on the LAN, effectively requesting routers to respond to the node with a Router Advertisement message. This message is sent to the all-routers multicast address and so is not seen by other hosts on the LAN.

A Router Solicitation message is generally sent when a node is first attached to (or comes up on) the LAN. The frequency of these events should be low and so both the traffic and processing load for Router Solicitation messages is expected to be negligible.

[3.2.](#) Router Advertisement

Router Advertisement messages are sent by routers periodically to the all-nodes multicast address to announce their presence on the LAN and advertise some link parameters. As long as there are not very many routers on the LAN this should not present much traffic or processing load. In the core case where the LAN is connecting many routers the traffic and processing load will increase with the number of routers and some measures may be needed to limit the traffic, either by reducing the transmission rate or disabling the protocol (if it is not needed in an all-router environment).

Router Advertisement is also unicast to the requesting node in response to a Router Solicitation message and, as noted above, this should not present a significant load.

[3.3.](#) Neighbor Solicitation

A Neighbor Solicitation message is sent by a node when that node has no (or a stale) cache entry for the L2 address for a particular next hop IPv6 address. This message is sent to a solicited-node multicast address which is manufactured from the next hop IPv6 address. A great advantage of using a solicited-node multicast address is that only the solicited neighbor node (or perhaps a very few more) will be subscribed to this address. Therefore the processing load for this message is restricted to a small number of nodes and is not likely to present a significant burden.

In general the frequency of Neighbor Solicitation messages will be related to the number of each node's communicating peers on the LAN. Since this number is directly related to the amount of traffic the LAN must support for communications in general the fraction consumed by Neighbor Solicitation should be very small.

[3.4.](#) Neighbor Advertisement

Neighbor Advertisement messages are sent in response to Neighbor Solicitation messages. They are unicast to the originator of the Neighbor Solicitation message and so the load presented in this case should, as with Neighbor Solicitation, be a small fraction of the traffic that must be supported on the LAN.

Unsolicited Neighbor Advertisement messages may also be sent to the

all-nodes multicast address; however, as this may be done when a node's L2 address changes the frequency of these messages should be extremely low.

[3.5.](#) Redirect

Redirect messages are sent by routers to nodes to change the next hop that node is using to reach a particular destination. Although the likelihood of redirect depends on the network topology and other factors, it is not expected to present a significant load on either the network or hosts.

[4.](#) Some additional protocol activities

[4.1.](#) Duplicate Address Detection

Duplicate address detection as described in [ADDRCONF] involves sending a number of Neighbor Solicitation messages for the address to be checked (to that address's solicited-node multicast address). This is done before attempting to join the LAN using the address

being checked. Since this is an initialization procedure it is not expected to present a significant traffic or processing load during normal operation. It is also possible that address autoconfiguration will not be used in very large data centers.

[4.2.](#) Anycast and Proxy address resolution

Address resolution for Anycast addresses or addresses for which nodes are acting as a Proxy may solicit multiple Neighbor Advertisement messages in response. In this case of Anycast addresses the responses are sent with random delay so that the requesting node does not see an unmanageable burst of responses. The response traffic in this case may be greater but not likely a problem, and the additional processing load is only on the requesting node (which is in control of the rate of solicitation).

In a multi-site data center network it may be desirable to restrict the propagation of Anycast address resolution messages if it is desired that only responses local to the requesting node's site be delivered.

[4.3.](#) Neighbor unreachability detection

Neighbor unreachability detection relies on hints from higher layers to determine whether or not a given neighbor is still reachable. In some cases when connectivity is suspect and no higher layer hints are available, a Neighbor Solicitation message may be used to verify continued connectivity. This is not expected to be a common occurrence between hosts or hosts and routers (since higher layer hints are most likely available). Between routers there may not be higher layer hints available but there are likely other means to detect connectivity to router peers across the LAN making use of Neighbor Solicitation messages unnecessary.

[4.4.](#) Host-based Load Spreading

Host-based load spreading (e.g. [RFC 4311](#)) affects the selection of next hop router for particular packets. This may increase the number of routers a given host communicates with, but it is not expected to add significantly to neighbor discovery traffic or processing load.

[4.5.](#) Router-based Load Spreading

Router-based load spreading (i.e. the use of a NULL SA in a Router Advertisement message) requires hosts to solicit a next hop router address. This increases the number of solicitations for router addresses, but this should not be significant if the number of

routers on the LAN is small. This mechanism may be inappropriate (and unneeded) in a core LAN interconnecting a large number of routers and therefore not a concern in that case either.

[4.6.](#) Holding packets while address resolution occurs

In multi-site networks or virtualized networks in which the edge-to-edge delay may be increased over that in a normal (local) LAN, hold time for packets awaiting address resolution may increase significantly. This may be a concern depending on the percentage of packets that must wait for address resolution before being forwarded on the LAN.

[5.](#) Summary and conclusions

The following summarizes the analysis presented:

- IPv6 ND looks like it will scale well for the case of a large LAN with 1000s of hosts and a relatively small number of routers.
- For the case of a core LAN connecting a large number of routers there are some ND protocol behaviors that may not scale well but these are either optional or not needed between routers (i.e., there are other mechanisms available to the routers to accomplish the same end).
- Multi-site L2 networks may provide challenges for both holding time for packets while address resolution is carried out and address resolution for Anycast addresses (for example, if these are expected to select only local servers).
- The impact of network virtualization (many VLANs and virtual routers) on platforms that support many virtual networks has not been analyzed and may present additional scaling challenges.

6. Manageability Considerations

This document has no manageability considerations.

7. Security Considerations

This document adds no security considerations since it does not define any new protocol behaviors. However, it may be worthwhile to consider whether or not the size of an L2 network (as discussed here) presents any new security challenges. No analysis in this area is provided in this draft.

Mack-Crane

Expires April 12, 2011

[Page 7]

Internet-Draft

IPv6 ND Scalability

October 2010

8. IANA Considerations

This document has no IANA considerations.

9. Acknowledgments

This document was prepared using 2-Word-v2.0.template.dot.

10. References

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

September 2007.

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

Authors' Addresses

Ben Mackcrane
Huawei Technologies
1700 Alma Drive, Suite 500
Plano, TX 75075, USA
Phone: (630) 810 1132
Email: tmackcrane@huawei.com

Linda Dunbar
Huawei Technologies
1700 Alma Drive, Suite 500
Plano, TX 75075, USA
Phone: (972) 543 5849
Email: ldunbar@huawei.com

Sue Hares
Huawei Technologies
2330 Central Expressway,
Santa Clara, CA 95050, USA
Phone:
Email: shares@huawei.com

Intellectual Property Statement

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in any IETF Document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at <http://www.ietf.org/ipr>

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgment

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