TRILL Working Group Y. Li
INTERNET-DRAFT D. Eastlake
Intended Status: Standard Track L. Dunbar
Huawei Technologies
R. Perlman
EMC
I. Gashinsky
Yahoo
Expires: October 4, 2015 April 2, 2015

TRILL: ARP/ND Optimization
draft-ietf-trill-arp-optimization-00

Abstract

This document describes mechanisms to optimize the ARP (Address Resolution Protocol) and ND (Neighbor Discovery) traffic in TRILL campus. Such optimization reduces packet flooding over a TRILL campus.

Status of this Memo

Copyright and License Notice

Table of Contents

1 Introduction

ARP [RFC826] and ND [RFC4861] are normally sent by broadcast and multicast respectively. To reduce the burden on a TRILL campus caused by these multi-destination messages, RBridges MAY implement an "optimized ARP/ND response", as specified herein, when the target's location is known by the ingress RBridge or can be obtained from a directory. This avoids ARP/ND query flooding.

1.1 Terminology

The acronyms and terminology in [RFC6325] are used herein. Some of these are listed below for convenience with the following along with some additions:

Campus: a TRILL network consisting of TRILL switches, links, and possibly bridges bounded by end stations and IP routers. For TRILL, there is no "academic" implication in the name "campus".

APPsub-TLV Application sub-Type-Length-Values

ARP Address Resolution Protocol [RFC826]

DAD Duplicate Address Detection

Data Label VLAN or FGL

ESADI End Station Address Distribution Information [RFC7357]

FGL Fine-Grained Label [RFC7172]

IA Interface Addresses, a TRILL APPsub-TLV [IA-draft]

IP Internet Protocol

MAC Media Access Control address

ND Neighbor Discoery [RFC4861]

RBridge Routing Bridge, an alternative term for a TRILL switch.

SEND secure neighbor discovery [RFC3971]

TRILL Transparent Interconnection of Lots of Links or

Tunneled Routing in the Link Layer.

TRILL switch A device implementing the TRILL protocol, an alternative term for an RBridge.

2 IP/MAC Address Mappings

Traditionally an RBridge learns the MAC and Data Label (VLAN or FGL) to nickname correspondence of a remote host, as per [RFC6325] and [RFC7172], from TRILL data frames received. No IP address information is learned directly from the TRILL data frame. Interface Addresses (IA) APPsub-TLV [IA-draft] enhances the TRILL base protocol by allowing IP and MAC address mappings to be distributed in the control plane by any RBridge. This APPsub-TLV appears inside the TRILL GENINFO TLV in ESADI [RFC7357] but the value data structure it specifies may also occur in other application contexts. Edge Directory Assist Mechanisms [DirMech] makes use of this APPsub-TLV for its push model and uses the value data structure it specifies in its pull model.

An RBridge can easily know the IP/MAC address mappings of the local hosts that it is attached to it via its access ports by receiving ARP [RFC826] or ND [RFC4861] messages. If the RBridge has extracted the sender's IP/MAC address pair from the received data packet, it may save the information and use the IA APPsub-TLV to distribute it to other RBridges through ESADI. Then the relevant remote RBridges (normally those interested in the same Data Label as the original ARP/ND messages) receive and save such mapping information also. There are others ways that RBridges save IP/MAC address mappings in advance, e.g. import from management system and distribution by directory servers [DirMech].

The examples given above shows that RBridges may have saved a host's triplet of {IP address, MAC address, ingress nickname} for a given Data Label (VLAN or FGL) before that host sends or receives any real data packet. Note such information may or may not be a complete list and may or may not exist on all RBridges. The information may possibly be from different sources. RBridges can then use the Flags Field in IA APPsub-TLV to identify if the source is a directory server or local observation by the sender. A different confidence level may also be used to indicate the reliability of the mapping information.

3 Handling ARP/ND Messages

A native frame that is an ARP [RFC826] message is detected by its Ethertype of 0x0806. A native frame that is an ND [RFC4861] is

detected by being one of five different ICMPv6 packet types. ARP/ND is commonly used on a link to (1) query for the MAC address corresponding to an IPv4 or IPv6 address, (2) test if an IPv4/IPv6 address is already in use, or (3) to announce the new or updated info on any of IPv4/IPv6 address, MAC address, and/or point of attachment.

To simplify the text, we use the following terms in this section.

1) IP address - indicated protocol address that is normally an IPv4 address in ARP or an IPv6 address in ND.

2) sender's IP/MAC address - sender protocol/hardware address in ARP, source IP address and source link-layer address in ND

3) target's IP/MAC address - target protocol/hardware address in ARP, target address and target link-layer address in ND

When an ingress RBridge receives an ARP/ND message, it can perform the steps described in the sub-sections below.

3.1 Get Sender's IP/MAC Mapping Information for Non-zero IP

If the sender's IP has not been saved by the ingress RBridge before, populate the information of sender's IP/MAC in its ARP table;

else if the sender's IP has been saved before but with a different MAC address mapped or a different ingress nickname associated with the same pair of IP/MAC, the RBridge should verify if a duplicate IP address has already been in use or a host has changed its attaching RBridge. The RBridge may use different strategies to do so, for example, ask an authoritative entity like directory servers or encapsulate and unicast the ARP/ND message to the location where it believes the address is in use. RBridge should update the saved triplet of {IP address, MAC address, ingress nickname} based on the verification.

The ingress RBridge may use the IA APPsub-TLV [IA-draft] with the Local flag set in ESADI [RFC7357] to distribute any new or updated triplet of {IP address, MAC address, ingress nickname} information obtained in this step. If a push directory server is used, such information can be distributed as per [DirMech].

3.2 Determine How to Reply to ARP/ND

a) If the message is a generic ARP/ND request and the ingress RBridge knows the target's IP address, the ingress RBridge may decide to take

one or a combination of the following actions:

a.1. Send an ARP/ND response directly to the querier, with the target's MAC address, as believed by the ingress RBridge.

a.2. Encapsulate the ARP/ND request to the target's Designated RBridge, and have the egress RBridge for the target forward the query to the target. This behavior has the advantage that a response to the request is authoritative. If the request does not reach the target, then the querier does not get a response.

a.3. Block ARP/ND requests that occur for some time after a request to the same target has been launched, and then respond to the querier when the response to the recently-launched query to that target is received.

a.4. Pull the most up-to-date records if a pull directory server is available [DirMech] and reply to the querier.

a.5. Flood the request as per [RFC6325].

b) If the message is a generic ARP request and the ingress RBridge does not know target's IP address, the ingress RBridge may take one of the following actions.

b.1. Flood the message as per [RFC6325].

b.2. Use directory server to pull the information [DirMech] and reply to the querier.

b.3. Drop the message.

c) If the message is a gratuitous ARP which can be identified by the same sender's and target's "protocol" address fields or an Unsolicited Neighbor Advertisements [RFC4861] in ND:

The RBridge may use an IA APPsub-TLV [IA-draft] with the Local flag set to distribute the sender's MAC and IP mapping information. When one or more directory servers are deployed and complete Push Directory information is used by all the TRILL switches in the Data Label, a gratuitous ARP or unsolicited NA SHOULD be discarded rather than ingressed. Otherwise, they are either ingressed and flooded as per [RFC6325] or discarded depending on local policy.

d) If the message is a Address Probe ARP Query [RFC5227] which can be identified by the sender's protocol (IPv4) address field being zero and the target's protocol address field being the IPv4 address to be

tested or a Neighbor Solicitation for DAD (Duplicate Address Detection) which has the unspecified source address [RFC4862]: it should be handled as the generic ARP message as in a) and b).

It should be noted in the case of secure neighbor discovery (SEND) [RFC3971], cryptography might prevent local reply by the ingress RBridge, since the RBridge would not be able to sign the response with the target's private key.

It is not essential that all RBridges use the same strategy for which option to select for a particular ARP/ND query. It is up to the implementation.

3.3 Determine How to Handle the ARP/ND Response

If the ingress RBridge R1 decides to unicast the ARP/ND request to the target's egress RBridge R2 as discussed in subsection 3.2 item a) or to flood the request as per [RFC6325], then R2 decapsulates the query, and initiates an ARP/ND query on the target's link. When/if the target responds, R2 encapsulates and unicasts the response to R1, which decapsulates the response and sends it to the querier. R2 should initiates a link state update to inform all the other RBridges of the target's location, layer 3 address, and layer 2 address, in addition to forwarding the reply to the querier. The update message can be carried by an IA APPsub-TLV [IA-draft] with the Local flag set in ESADI [RFC7357] or as per [DirMech] if push directory server is in use.

4 Handling RARP (Reverse Address Resolution Protocol) Messages

RARP [RFC903] uses the same packet format as ARP but a different Ethertype (0x8035) and opcode values. Its use is similar to the generic ARP Request/Response as described in 3.2 a) and b). The difference is that it is intended to query for the target "protocol" address corresponding to the target "hardware" address provided. It should be handled by doing a local cache or directory server lookup on the target "hardware" address provided to find a mapping to the desired "protocol" address. Normally, it is used to look up a MAC address to find the corresponding IP address.

5 Security Considerations

ARP and ND messages can be easily forged. Therefore the learning of MAC/IP addresses from them should not be considered as reliable. RBridge can use the confidence level in IA APPsub-TLV information received via ESADI or pull directory retrievals to determine the reliability of MAC/IP address mapping. (ESADI information can be

secured as provide in [RFC7357] and pull directory information can be secured as provide in [DirMech].) It is up to the implementation to decide if an RBridge should distribute the IP and MAC address mappings received from local native ARP/ND messages to other RBridges in the same Data Label.

The ingress RBridge should also rate limit the ARP/ND queries for the same target to be injected into the TRILL campus to prevent possible denial of service attacks.

6 IANA Considerations

No IANA action is required. RFC Editor: please delete this section before publication.

7 References

7.1 Normative References

[RFC826] Plummer, D., "An Ethernet Address Resolution Protocol", RFC 826, November 1982.

[RFC903] Finlayson, R., Mann, T., Mogul, J., and M. Theimer, "A Reverse Address Resolution Protocol", STD 38, RFC 903, June 1984

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

[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.

[RFC6165] Banerjee, A. and D. Ward, "Extensions to IS-IS for Layer-2 Systems", RFC 6165, April 2011.

[RFC6325] Perlman, R., et.al. "RBridge: Base Protocol Specification", RFC 6325, July 2011.

[RFC6439] Eastlake, D. et.al., "RBridge: Appointed Forwarder", RFC 6439, November 2011.

[RFC7172] Eastlake 3rd, D., Zhang, M., Agarwal, P., Perlman, R., and D. Dutt, "Transparent Interconnection of Lots of Links (TRILL): Fine-Grained Labeling", RFC 7172, May 2014, <http://www.rfc-editor.org/info/rfc7172>.

7.2 Informative References

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

[RFC5227] Cheshire, S., "IPv4 Address Conflict Detection", RFC 5227, July 2008.

[RFC7067] Dunbar, L., Eastlake 3rd, D., Perlman, R., and I. Gashinsky, "Directory Assistance Problem and High-Level Design Proposal", RFC 7067, November 2013.

[IA-draft] Eastlake, D., Li Y., R. Perlman, "TRILL: Interface Addresses APPsub-TLV", draft-eastlake-trill-ia-appsubtlv, work in progress.

[DirMech] Dunbar, L., Eastlake 3rd, D., Perlman, R., I. Gashinsky. and Li Y., TRILL: Edge Directory Assist Mechanisms", draft-ietf-trill-directory-assist-mechanisms, work in progress.

Authors' Addresses

Yizhou Li
Huawei Technologies
101 Software Avenue,
Nanjing 210012
China

Phone: +86-25-56625375
EMail: liyizhou@huawei.com

Donald Eastlake
Huawei R&D USA
155 Beaver Street
Milford, MA 01757 USA

Phone: +1-508-333-2270
EMail: d3e3e3@gmail.com

Linda Dunbar

Huawei Technologies
5430 Legacy Drive, Suite #175
Plano, TX 75024, USA

Phone: +1-469-277-5840
EMail: ldunbar@huawei.com

Radia Perlman
EMC
2010 256th Avenue NE, #200
Bellevue, WA 98007
USA

EMail: Radia@alum.mit.edu

Igor Gashinsky
Yahoo
45 West 18th Street 6th floor
New York, NY 10011 USA

EMail: igor@yahoo-inc.com