

TRILL Working Group  
INTERNET-DRAFT  
Intended Status: Standard Track

Y. Li  
D. Eastlake  
L. Dunbar  
Huawei Technologies  
R. Perlman  
EMC  
M. Umair  
IPinfusion  
April 17, 2017

Expires: October 19, 2017

TRILL: ARP/ND Optimization  
draft-ietf-trill-arp-optimization-08

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

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/lid-abstracts.html>

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

Copyright and License Notice

Copyright (c) 2017 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 Simplified BSD License.

## Table of Contents

1	Introduction . . . . .	3
1.1	Terminology . . . . .	3
2	ARP/ND Optimization Requirement and Solution . . . . .	4
3	IP/MAC Address Mappings . . . . .	5
4	Handling ARP/ND/SEND Messages . . . . .	5
4.1	SEND Considerations . . . . .	6
4.2	Address Verification . . . . .	6
4.3	Get Sender's IP/MAC Mapping Information for Non-zero IP . . . . .	6
4.4	Determine How to Reply to ARP/ND . . . . .	7
4.5	Determine How to Handle the ARP/ND Response . . . . .	9
5	Handling RARP (Reverse Address Resolution Protocol) Messages . . . . .	9
6	Handling of DHCP messages . . . . .	9
7	Handling of Duplicate IP Addresses . . . . .	10
8	RBridge ARP/ND Cache Liveness and MAC Mobility . . . . .	10
9	Security Considerations . . . . .	11
10	IANA Considerations . . . . .	11
11	Acknowledgments . . . . .	11
12	References . . . . .	12
12.1	Normative References . . . . .	12
12.2	Informative References . . . . .	13
	Authors' Addresses . . . . .	13

## 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 and answer flooding.

### 1.1 Terminology

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

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

APPsub-TLV	Application sub-Type-Length-Value [RFC6823]
ARP	Address Resolution Protocol [RFC826]
Campus	A TRILL network consisting of RBridges, links, and possibly bridges bounded by end stations and IP routers [RFC6325]
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 [RFC7961]
IP	Internet Protocol, both IPv4 and IPv6
MAC	Media Access Control [RFC7042]
ND	Neighbor Discovery [RFC4861]
RBridge	A contraction of "Routing Bridge". A device implementing the TRILL protocol.
SEND	secure neighbor discovery [RFC3971]

TRILL                    Transparent Interconnection of Lots of Links or  
Tunneled Routing in the Link Layer [RFC6325] [RFC7780]

## 2 ARP/ND Optimization Requirement and Solution

IP address resolution can create significant issues in data centers due to flooded packets as discussed in [RFC6820]. Such flooding can be avoided by a proxy ARP/ND function on edge RBridges as described in this document.

To support such ARP/ND optimization, edge RBridges need to know end-station's IP to MAC mapping through manual configuration (management), through control plane mechanisms such as directories [DirMech], or through Data plane learning by snooping of messages such as ARP/ND (including DHCP or gratuitous ARP\_messages).

When all the end-stations IP/MAC address mapping is known to edge RBridges or provisioned through management or learnt via control plane on the edge RBridges, it should be possible to completely suppress flooding of ARP/ND messages in a TRILL Campus. When all end-station MAC addresses are similarly known, it should be possible to suppress unknown unicast flooding by dropping any unknown unicast received at an edge RBridge.

An ARP/ND optimization mechanism should include provisions for an edge RBridge to issue an ARP/ND request to an attached end station to confirm or update information and should allow an end station to detect duplicate IP addresses.

TRILL already provides an option to disable data-plane learning from the source MAC address of end-station frames on a per port basis (see Section 5.3 of [RFC6325]).

Most of the end station hosts either send DHCP messages requesting an IP Address or send out gratuitous ARP or RARP requests to announce themselves to the network right after they come online. Thus the local edge RBridge will immediately have the opportunity to snoop and learn their MAC and IP addresses and distribute this information to other edge RBridges through the TRILL control plane ESADI [RFC7357] protocol. Once remote RBridges received this information via the control plane they should add IP to MAC mapping information to their ARP/ND cache along with the nickname and data label of the address information. Therefore, most active IP hosts in TRILL network can be learned by the edge RBridges either through local learning or control-plane-based remote learning. As a result, ARP suppression can vastly reduce the network flooding caused by host ARP learning behavior.

### 3 IP/MAC Address Mappings

By default, an RBridge [RFC6325] [RFC7172] learns MAC Address and Data Label (VLAN or FGL) to egress nickname mapping information from TRILL data frames it receives. No IP address information is learned directly from the TRILL data frame. The Interface Addresses (IA) APPsub-TLV [RFC7961] 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 RBridge 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 end stations that it is attached to it via its access ports by receiving ARP [RFC826] or ND [RFC4861] messages. If the edge RBridge has extracted the sender's IP/MAC address pair from the received data frame (either ARP or ND), it may save the information and then use the IA APPsub-TLV to link the IP and MAC addresses and 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) also receive and save such mapping information. There are other 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 show that RBridges might have saved an end station's triplet of {IP address, MAC address, ingress nickname} for a given Data Label (VLAN or FGL) before that end station sends or receives any real data packet. Note such information might or might not be a complete list and might or might not exist on all RBridges. The information could 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.

### 4 Handling ARP/ND/SEND 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 IP/MAC address in ARP, source IP address and source link-layer address in ND
- 3) target's IP/MAC address - target IP/MAC address in ARP, target address and target link-layer address in ND

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

#### 4.1 SEND Considerations

SEND (Secure Neighbor Discovery [RFC3971] is a method of securing ND that addresses the threats discussed in [RFC3756]. Typical TRILL campuses are inside data centers, Internet exchange points, or carrier facilities. These are generally controlled and protected environments where these threats are of less concern. Nevertheless, SEND provides an additional layer of protection.

Secure SEND messages require knowledge of cryptographic keys. Methods of communicating such keys to RBridges for use in SEND are beyond the scope of this document. Thus, using the optimizations in this document, RBridges do not attempt to construct SEND messages and are generally transparent to them. RBridges only construct ARP, RARP, or insecure ND messages, as appropriate. Nevertheless, RBridges implementing ARP/ND optimization SHOULD snoop on SEND messages to extract addressing information that would be present if the message had been sent as an insecure ND message.

#### 4.2 Address Verification

RBridges may use ARP/ND to probe directly attached or remote end stations for address or liveness verification. This is typically most appropriate in less managed and/or higher mobility environments. In strongly managed environments, such as a typical data center, where a central orchestration/directory system has complete addressing knowledge [RFC7067], optimized ARP/ND responses can use that knowledge. In such cases, there is little reason for verification except for debugging operational problems or the like.

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

- o If the sender's IP is not present in the ingress RBridge's ARP/ND

cache, populate the information of sender's IP/MAC in its ARP/ND cache table. The ingress RBridge correlates its nickname and that IP/MAC mapping information. Such triplet of {IP address, MAC address, ingress nickname} information is saved locally and can be distributed to other RBridges as explain later.

o 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 an end station has changed its attaching RBridge. The RBridge may use different strategies to do so. For example, the RBridge might ask an authoritative entity like directory servers or it might 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. An RBridge might not verify an IP address if the network manager's policy is to have the network behave, for each Data Label, as if it were a single link and just believe an ARP/ND it receives.

The ingress RBridge MAY use the IA APPsub-TLV [RFC7961] 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].

#### 4.4 Determine How to Reply to ARP/ND

The options for an edge RBridge to handle a native ARP/ND are given below. For generic ARP/ND request seeking the MAC address corresponding to an IP address, if the edge RBridge knows the IP address and corresponding MAC, behavior is as in item (a), otherwise behavior is as in item (b). Behavior for gratuitous ARP and ND Unsolicited Neighbor Advertisements [RFC4861] is given in item (c). And item (d) covers handling of Address Probe ARP Query.

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.

a) If the message is a generic ARP/ND request and the ingress RBridge knows the target's IP address and associated MAC address, the ingress RBridge MUST take one or a combination of the actions below. In the case of secure neighbor discovery (SEND) [RFC3971], cryptography would prevent local reply by the ingress RBridge, since the RBridge would not be able to sign the response with the target's private key, and only action a.2 or a.5 is valid.

a.1. Send an ARP/ND response directly to the querier, using the target's MAC address present in the ingress RBridge's ARP/ND cache table. Because the edge RBridge might not have an IPv6 address, the source IP address for such an ND response MUST be that of the target end station.

a.2. Encapsulate the ARP/ND/SEND 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. Reply to the querier based on directory information [DirMech] such as information obtained from a pull directory server or directory information that the ingress RBridge has requested to be pushed to it.

a.5. Flood the /ND/SEND request as per [RFC6325].

(b) If the message is a generic ARP/ND/SEND request and the ingress RBridge does not know target's IP address, the ingress RBridge MUST take one of the following actions. In the case of secure neighbor discovery (SEND) [RFC3971], cryptography would prevent local reply by the ingress RBridge, since the RBridge would not be able to sign the response with the target's private key therefore only action b.1 is valid.

b.1. Flood the ARP/ND/SEND message as per [RFC6325].

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

b.3. Drop the message if the directory mechanism is used and you know there should be no response (query based on a non-existent IP address for example).

(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/SEND:

The RBridge MAY use an IA APPsub-TLV [RFC7961] 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 RBridges 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) or (b) above.

#### 4.5 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 initiate 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 uses an IA APPsub-TLV [IA-draft] (so the layer 3 and layer 2 addresses can be linked) with the Local flag set in ESADI [RFC7357] or as per [DirMech] if push directory server is in use.

#### 5 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" (IP) address corresponding to the target "hardware" (MAC) 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.

#### 6 Handling of DHCP messages

When a newly connected end-station exchanges messages with a DHCP [RFC2131] server an edge RBridge should snoop them (mainly the DHCPACK message) and store IP/MAC mapping information in its ARP/ND

cache and should also send the information out through the TRILL control plane using ESADI.

## 7 Handling of Duplicate IP Addresses

Duplicate IP addresses within a Data Label can occur due to an attacker sending fake ARP/ND messages or due to human/configuration errors. If complete directory information is available, then by definition the IP location information in the directory is correct. Any appearance of an IP address in a different place (different edge RBridge or port) from other sources is not correct.

Without complete directory information, the ARP/ND optimization function should support duplicate IP detection. This is critical in a Data Center to stop an attacker from using ARP/ND spoofing to divert traffic from its intended destination.

Duplicate IP addresses can be detected when an existing active IP1/MAC1 mapping gets modified. Also an edge RBridge may send a query to the former owner of IP called a DAD-query (Duplicate Address Detection query). A DAD-query is a unicast ARP/ND message with sender IP 0.0.0.0 in case of ARP (or a configurable per RBridge IP address called the DAD-Query source IP) and an IPv6 Link Local Address in case of ND with source MAC set to the DAD-querier RBridge's MAC. If the querying RBridge does not receive an answer within a given time, the new IP entry will be confirmed and activated in its ARP/ND cache.

In the case where the former owner replies, a Duplicate Address has been detected. In this case the querying RBridge SHOULD log the duplicate so that the network administrator can take appropriate action.

## 8 RBridge ARP/ND Cache Liveness and MAC Mobility

A maintenance procedure is needed for RBridge ARP/ND caching to ensure IP end-stations connected to ingress RBridges are still active.

Some links provide a physical layer indication of link liveness. A dynamic proxy-ARP/ND entry (one learned from data plane observation) MUST be removed from the table if the link over which it was learned fails.

Similarly a dynamic proxy-ARP/ND entry SHOULD be flushed out of the table if the IP/MAC mapping has not been refreshed within a given age-time. The entry is refreshed if an ARP or ND message is received for the same IP/MAC mapping entry from any location. The IP/MAC

mapping information ageing timer is configurable per RBridge and defaults to 3/4 of the MAC address learning Ageing Timer [RFC6325].

For example end-Station "A" is connected to edge-RBridge1 (RB1) and has been learnt as local entry on RB1. If end-Station "A" moves to some other location (MAC/VM Mobility) and gets connected to edge-RBridge2 (RB2), after learning on RB2's access port, RB2 advertises this entry through the TRILL control-plane and it gets learnt on RB1 as a remote entry. The old entry on RB1 SHOULD get replaced and all other edge-RBridges with end-station service enabled for that data-label should update the entry to show reachability from RB2 instead of RB1.

If an ARP/ND entry in the cache is not refreshed, then the RBridge connected to that end-station MAY send periodic refresh messages (ARP/ND "probes") to that end-station, so that the entries can be refreshed before they age out. The end-station would reply to the ARP/ND probe and the reply resets the corresponding entry age-timer.

## 9 Security Considerations

Unless Secure ND (SEND [RFC3971]) is used, ARP and ND messages can be easily forged. Therefore the learning of MAC/IP addresses by RBridges from ARP/ND should not be considered as reliable. See Section 4.1 for SEND Considerations.

An 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 provided in [RFC7357] and pull directory information can be secured as provided in [DirMech]. The implementation decides if an RBridge will distribute the IP and MAC address mappings received from local native ARP/ND messages to other RBridges in the same Data Label, if it distributes them, and with what confidence level it does so.

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.

## 10 IANA Considerations

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

## 11 Acknowledgments

The authors would like to thank Igor Gashinsky and Sue Hares for their contributions.

## 12 References

### 12.1 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.
- [RFC2131] Droms, R., "Dynamic Host Configuration Protocol", RFC 2131, 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.
- [RFC6325] Perlman, R., Eastlake 3rd, D., Dutt, D., Gai, S., and A. Ghanwani, "Routing Bridges (RBridges): Base Protocol Specification", RFC 6325, July 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.
- [RFC7356] Ginsberg, L., Previdi, S., and Y. Yang, "IS-IS Flooding Scope Link State PDUs (LSPs)", RFC 7356, September 2014.
- [RFC7357] Zhai, H., Hu, F., Perlman, R., Eastlake 3rd, D., and O. Stokes, "Transparent Interconnection of Lots of Links (TRILL): End Station Address Distribution Information (ESADI) Protocol", RFC 7357, September 2014.
- [RFC7780] Eastlake 3rd, D., Zhang, M., Perlman, R., Banerjee, A., Ghanwani, A., and S. Gupta, "Transparent Interconnection

of Lots of Links (TRILL): Clarifications, Corrections, and Updates", RFC 7780, February 2016.

- [RFC7961] Eastlake 3rd, D. and L. Yizhou, "Transparent Interconnection of Lots of Links (TRILL): Interface Addresses APPsub-TLV", RFC 7961, August 2016.
- [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.

## 12.2 Informative References

- [RFC3756] Nikander, P., Ed., Kempf, J., and E. Nordmark, "IPv6 Neighbor Discovery (ND) Trust Models and Threats", RFC 3756, May 2004.
- [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.
- [RFC6820] Narten, T., Karir, M., and I. Foo, "Address Resolution Problems in Large Data Center Networks", RFC 6820, January 2013.
- [RFC6823] Ginsberg, L., Previdi, S., and M. Shand, "Advertising Generic Information in IS-IS", RFC 6823, December 2012.
- [RFC7042] Eastlake 3rd, D. and J. Abley, "IANA Considerations and IETF Protocol and Documentation Usage for IEEE 802 Parameters", BCP 141, RFC 7042, October 2013.
- [RFC7067] Dunbar, L., Eastlake 3rd, D., Perlman, R., and I. Gashinsky, "Directory Assistance Problem and High-Level Design Proposal", RFC 7067, November 2013.

## 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

Mohammed Umair  
IPinfusion

Email: mohammed.umair2@gmail.com