TRILL Working Group
Internet-Draft
Intended Status: Standards Track

H. Zhai
JIT
T. Senevirathne
Consultant
R. Perlman
EMC
M. Zhang
Y. Li
Huawei Technologies

Expires: March 28, 2016
September 25, 2015

# TRILL: Pseudo-Nickname for Active-Active Access
**draft-ietf-trill-pseudonode-nickname-07.txt**

## Abstract

The IETF TRILL (TRansparent Interconnection of Lots of Links) protocol provides support for flow level multi-pathing for both unicast and multi-destination traffic in networks with arbitrary topology. Active-active access at the TRILL edge is the extension of these characteristics to end stations that are multiply connected to a TRILL campus as discussed in RFC 7379. In this document, the edge RBridge (TRILL switch) group providing active-active access to such an end station are represented as a Virtual RBridge. Based on the concept of Virtual RBridge along with its pseudo-nickname, this document specifies a method for TRILL active-active access by such end stations.

Table of Contents

## 1. Introduction

The IETF TRILL protocol [RFC6325] provides optimal pair-wise data frame forwarding without configuration, safe forwarding even during periods of temporary loops, and support for multi-pathing of both unicast and multicast traffic. TRILL accomplishes this by using IS-IS [IS-IS] [RFC7176] link state routing and encapsulating traffic using a header that includes a hop count. Devices that implement TRILL are called RBridges or TRILL switches.

In the base TRILL protocol, an end node can be attached to the TRILL campus via a point-to-point link or a shared link such as a bridged LAN (Local Area Network). Although there might be more than one edge RBridge on a shared link, to avoid potential forwarding loops, one and only one of the edge RBridges is permitted to provide forwarding service for end station traffic in each VLAN (Virtual LAN). That RBridge is referred to as the Appointed Forwarder (AF) for that VLAN on the link [RFC6325] [RFC6439]. However, in some practical deployments, to increase the access bandwidth and reliability, an end station might be multiply connected to several edge RBridges and all of the uplinks are handled via a Local Active-Active Link Protocol (LAALP [RFC7379]) such as Multi-Chassis Link Aggregation (MC-LAG) or Distributed Resilient Network Interconnect (DRNI [802.1AX]). In this case, it's required that traffic can be ingressed/egressed into/from the TRILL campus by any of the RBridges for each given VLAN. These RBridges constitutes an Active-Active Edge (AAE) RBridge group.

With an LAALP, traffic with the same VLAN and source MAC address but belonging to different flows will frequently be sent to different member RBridges of the AAE group and then ingressed into TRILL campus. When an egress RBridge receives such TRILL data packets ingressed by different RBridges, it learns different VLAN and MAC address to nickname correspondences continuously when decapsulating the packets if it has data plane address learning enabled. This issue is known as the "MAC flip-flopping" issue, which makes most TRILL switches behave badly and causes the returning traffic to reach the destination via different paths resulting in persistent re-ordering of the frames. In addition to this issue, other issues such as duplicate egressing and loop back of multi-destination frames may also disturb an end station multiply connected to the member RBridges of an AAE group [RFC7379].

This document addresses the AAE issues of TRILL by specifying how members of an edge RBridge group can be represented by a Virtual RBridge (RBv) and assigned a pseudo-nickname. A member RBridge of such a group uses a pseudo-nickname, instead of its own nickname, as the ingress RBridge nickname when ingressing frames received on attached LAALP links. Other methods are possible: for example the

specification in this document and the specification in [MultiAttach] could be simultaneously deployed for different AAE groups in the same campus. If the method is [MultiAttach] is used, edge TRILL switches need to support the capability indicated by the Capability Flags APPsub-TLV as specified in Section 4.2 of [MultiAttach]. If the method defined in this document is adopted, all TRILL switches need to support the Affinity sub-TLV defined in [RFC7176] and [CMT]. For a TRILL campus that deploys both these AAE methods, TRILL switches are required to support both methods. However, it is desirable to only adopt one method in a TRILL campus so that the operating expense, complexity of troubleshooting, etc, can be reduced.

The main body of this document is organized as follows: Section 2 gives an overview of the TRILL active-active access issues and the reason that a virtual RBridge (RBv) is used to resolve the issues. Section 3 gives the concept of a virtual RBridge (RBv) and its pseudo-nickname. Section 4 describes how edge RBridges can support an RBv automatically and get a pseudo-nickname for the RBv. Section 5 discusses how to protect multi-destination traffic against disruption due to Reverse Forwarding Path (RPF) check failure, duplication, forwarding loops, etc. Section 6 covers the special processing of native frames and TRILL data packets at member RBridges of an RBv (also referred to as an Active-Active Edge (AAE) RBridge group). Section 7 describes the MAC information synchronization among the member RBridges of an RBv. Section 8 discusses protection against downlink failure at a member RBridge; and Section 9 gives the necessary TRILL code points and data structures for a pseudo-nickname AAE RBridge group.

### 1.1. Terminology and Acronyms

This document uses the acronyms and terms defined in [RFC6325] and [RFC7379] and the following additional acronyms:

AAE - Active-active Edge RBridge group, a group of edge RBridges to which at least one CE is multiply attached with an LAALP. AAE is also referred to as edge group or Virtual RBridge in this document.

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".

CE - Customer Equipment (end station or bridge). The device can be either physical or virtual equipment.

Data Label - VLAN or FGL.

DF - Designated Forwarder.

DRNI: Distributed Resilient Network Interconnect. A link aggregation specified in [802.1AX] that can provide an LAALP between from 1 to 3 CEs and 2 or 3 RBridges.

E-L1FS - Extended Level 1 Flooding Scope [RFC7356].

FGL - Fine-Grained Labeling or Fine-Grained Labeled or Fine-Grained Label [RFC7172].

LAALP - Local Active-Active Link Protocol [RFC7379] such as MC-LAG or DRNI.

MC-LAG: Multi-Chassis LAG. Proprietary extensions of Link Aggregation [802.1AX] that can provide an LAALP between one CE and 2 or more RBridges.

OE flag - A flag used by the member RBridge of an LAALP to tell other edge RBridges whether it is willing to share an RBv with other LAALPs if they multiply attach to the same set of edge RBridges as it. When this flag for an LAALP is 1, it means that the LAALP needs to be served by an RBv by itself and is not willing to share, that is, it should Occupy an RBv Exclusively (OE).

RBv - virtual RBridge, an alias for active-active edge RBridge group in this document.

vDRB - The Designated RBridge in an RBv. It is responsible for deciding the pseudo-nickname for the RBv.

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 [RFC2119].

## 2. Overview

To minimize impact during failures and maximize available access bandwidth, Customer Equipment (referred to as CE in this document) may be multiply connected to TRILL campus via multiple edge RBridges.

Figure 1 shows such a typical deployment scenario, where CE1 attaches to RB1, RB2, ... RBk and treats all of the uplinks as an LAALP bundle. Then RB1, RB2, ... RBk constitute an Active-active Edge (AAE) RBridge group for CE1 in this LAALP. Even if a member RBridge or an uplink fails, CE1 will still get frame forwarding service from the

TRILL campus if there are still member RBridges and uplinks available in the AAE group. Furthermore, CE1 can make flow-based load balancing across the available member links of the LAALP bundle in the AAE group when it communicates with other CEs across the TRILL campus [RFC7379].

```
                  ----------------------
                 |                      |
                 |     TRILL Campus     |
                 |                      |
                  ----------------------
                      |       |    |
                +-----+       |    +--------+
                |             |             |
            +------+      +------+      +------+
            |(RB1) |      |(RB2) |      | (RBk)|
            +------+      +------+      +------+
              |..|          |..|          |..|
              |  +----+     |  |          |  |
              |   +---|-----|--|----------+  |
              | +-|---|-----+  +-----------+ |
              | | |   +-------------------+ | |
    LAALP1-->(| | |)                  (| | |) <--LAALPn
            +-------+    .  .  .      +-------+
            | CE1   |                 | CEn   |
            +-------+                 +-------+
```

Figure 1  Active-Active Connection to TRILL Edge RBridges

By design, an LAALP (say LAALP1) does not forward packets received on one member port to other member ports. As a result, the TRILL Hello messages sent by one member RBridge (say RB1) via a port to CE1 will not be forwarded to other member RBridges by CE1. That is to say, member RBridges will not see each other's Hellos via the LAALP. So every member RBridge of LAALP1 thinks of itself as appointed forwarder for all VLANs enabled on an LAALP1 link and can ingress/egress frames simultaneously in these VLANs [RFC6439].

The simultaneous flow-based ingressing/egressing can cause some problems. For example, simultaneous egressing of multi-destination traffic by multiple member RBridges will result in frame duplication at CE1 (see Section 3.1 of [RFC7379]); simultaneous ingressing of frames originated by CE1 for different flows in the same VLAN with the same source MAC address will result in MAC address flip-flopping at remote egress RBridges that have data plane address learning enabled (see Section 3.3 of [RFC7379]). The flip-flopping would in turn cause packet re-ordering in reverse traffic.

Edge RBridges learn Data Label and MAC address to nickname correspondences by default via decapsulating TRILL data packets (see Section 4.8.1 of [RFC6325] as updated by [RFC7172]). Assuming that the default data-plane learning is enabled at edge RBridges, MAC flip-flopping can be solved by using a Virtual RBridge together with its pseudo-nickname. This document specifies a way to do so.

## 3. Virtual RBridge and its Pseudo-nickname

A Virtual RBridge (RBv) represents a group of edge RBridges to which at least one CE is multiply attached using an LAALP. More exactly, it represents a group of ports on the edge RBridges providing end station service and the service provided to the CE(s) on these ports, through which the CE(s) are multiply attached to the TRILL campus using LAALP(s). Such end station service ports are called RBv ports; in contrast, other access ports at edge RBridges are called regular access ports in this document. RBv ports are always LAALP connecting ports, but not vice versa (see Section 4.1). For an edge RBridge, if one or more of its end station service ports are ports of an RBv, that RBridge is a member RBridge of that RBv.

For the convenience of description, a Virtual RBridge is also referred to as an Active-Active Edge (AAE) group in this document. In the TRILL campus, an RBv is identified by its pseudo-nickname, which is different from any RBridge's regular nickname(s). An RBv has one and only one pseudo-nickname. Each member RBridge (say RB1, RB2 ..., RBk) of an RBv (say RBvn) advertises RBvn's pseudo-nickname using a Nickname sub-TLV in its TRILL IS-IS LSP (Link State PDU) [RFC7176] and SHOULD do so with maximum priority of use (0xFF), along with their regular nickname(s). (Maximum priority is recommended to avoid the disruption to an AAE group that would occur if the nickname were taken away by a higher priority RBridge.) Then, from these LSPs, other RBridges outside the AAE group know that RBvn is reachable through RB1 to RBk.

A member RBridge (say RBi) loses its membership in RBvn when its last port in RBvn becomes unavailable due to failure, re-configuration, etc. Then RBi removes RBvn's pseudo-nickname from its LSP and distributes the updated LSP as usual. From those updated LSPs, other RBridges know that there is no path to RBvn through RBi now.

When member RBridges receive native frames on their RBv ports and decide to ingress the frames into the TRILL campus, they use that RBv's pseudo-nickname instead of their own regular nicknames as the ingress nickname to encapsulate them into TRILL Data packets. So when these packets arrive at an egress RBridge, even if they are originated by the same end station in the same VLAN but ingressed by

different member RBridges, no address flip-flopping is observed on the egress RBridge when decapsulating these packets. (When a member RBridge of an AAE group ingresses a frame from a non-RBv port, it still uses its own regular nickname as the ingress nickname.)

Since RBv is not a physical node and no TRILL frames are forwarded between its ports via an LAALP, pseudo-node LSP(s) MUST NOT be created for an RBv. RBv cannot act as a root when constructing distribution trees for multi-destination traffic and its pseudo-nickname is ignored when determining the distribution tree root for TRILL campus [CMT]. So the tree root priority of RBv's nickname MUST be set to 0, and this nickname MUST NOT be listed in the "s" nicknames (see Section 4.5 of [RFC6325]) by the RBridge holding the highest priority tree root nickname.

NOTE: In order to reduce the consumption of nicknames, especially in large TRILL campus with lots of RBridges and/or active-active accesses, when multiple CEs attach to the exact same set of edge RBridges via LAALPs, those edge RBridges should be considered as a single RBv with a single pseudo-nickname.

## 4. Member RBridges Auto-Discovery

Edge RBridges connected to a CE via an LAALP can automatically discover each other with minimal configuration through exchange of LAALP connection information.

From the perspective of edge RBridges, a CE that connects to edge RBridges via an LAALP can be identified by the ID of the LAALP that is unique across the TRILL campus (for example, the MC-LAG or DRNI System ID [802.1AX]), which is referred to as an LAALP ID in this document. On each of such edge RBridges, the access port to such a CE is associated with an LAALP ID for the CE. An LAALP is considered valid on an edge RBridge only if the RBridge still has an operational downlink to that LAALP. For such an edge RBridge, it advertises a list of LAALP IDs for its valid local LAALPs to other edge RBridges via its E-L1FS FS-LSP(s) [RFC7356][rfc7180bis]. Based on the LAALP IDs advertised by other RBridges, each RBridge can know which edge RBridges could constitute an AAE group (See Section 4.1 for more details). Then one RBridge is elected from the group to allocate an available nickname (the pseudo-nickname) for the group (See Section 4.2 for more details).

### 4.1. Discovering Member RBridge for an RBv

Take Figure 2 as an example, where CE1 and CE2 multiply attach to RB1, RB2 and RB3 via LAALP1 and LAALP2 respectively; CE3 and CE4

attach to RB3 and RB4 via LAALP3 and LAALP4 respectively. Assume LAALP3 is configured to occupy a Virtual RBridge by itself.

```
                      ------------------------
                    /                          \
                   |       TRILL Campus         |
                    \                          /
                      -------------------------
                       |    |             |  |
               +-------+    |             |  +----------+
               |            |             |             |
           +-------+     +-------+      +-------+     +-------+
           |  RB1  |     |  RB2  |      |  RB3  |     |  RB4  |
           +-------+     +-------+      +-------+     +-------+
             |   |        |   |          | | | |       |     |
             |   +--------|--+ | +-------|-+ | +-------|---+ |
             | +----------+  | | |       |   |         |   | |
             | | +-----------|-|-|-------+   | +-------+   | |
             | | |           | | |           | |           | |
    LAALP1->(| | |) LAALP2->(| | |) LAALP3->(| |) LAALP4->(| |)
           +-------+      +-------+     +-------+      +-------+
           |  CE1  |      |  CE2  |     |  CE3  |      |  CE4  |
           +-------+      +-------+     +-------+      +-------+
```

Figure 2  Different LAALPs to TRILL Campus

RB1 and RB2 advertise {LAALP1, LAALP2} in the PN-LAALP-Membership sub-TLV (see Section 9.1 for more details) via their TRILL E-L1FS LSPs respectively; RB3 announces {LAALP1, LAALP2, LAALP3, LAALP4}; and RB4 announces {LAALP3, LAALP4}, respectively.

An edge RBridge is called an LAALP related RBridge if it has at least one LAALP configured on an access port. On receipt of the PN-LAALP-Membership sub-TLVs, RBn ignores them if it is not an LAALP related RBridge; otherwise, RBn SHOULD use the LAALP information contained in the sub-TLVs, along with its own PN-LAALP-Membership sub-TLVs to decide which RBv(s) it should join and which edge RBridges constitute each of such RBvs. Based on the information received, each of the 4 RBridges knows the following information:

| LAALP ID | OE-flag | Set of edge RBridges |
|---|---|---|
| LAALP1 | 0 | {RB1, RB2, RB3} |
| LAALP2 | 0 | {RB1, RB2, RB3} |
| LAALP3 | 1 | {RB3, RB4} |
| LAALP4 | 0 | {RB3, RB4} |

Where the OE-flag indicates whether an LAALP is willing to share an

RBv with other LAALPs if they multiply attach to exact the same set of edge RBridges as it. For an LAALP (for example LAALP3), if its OE-flag is one, it means that LAALP3 does not want to share, so it MUST Occupy an RBv Exclusively (OE). Support of OE is optional. RBridges that do not support OE ignore the OE bit and act as if it was zero (see Section 11 on Configuration Consistency).

Otherwise, the LAALP (for example LAALP1) will share an RBv with other LAALPs if possible. By default, this flag is set to zero. For an LAALP, this flag is considered 1 if any edge RBridge advertises it as one (see Section 9.1).

In the above table, there might be some LAALPs that attach to a single RBridge due to mis-configuration or link failure, etc. Those LAALPs are considered as invalid entries. Then each of the LAALP related edge RBridges performs the following algorithm to decide which valid LAALPs can be served by an RBv.

Step 1: Take all the valid LAALPs that have their OE-flags set to 1 out of the table and create an RBv per such LAALP.

Step 2: Sort the valid LAALPs left in the table in descending order based on the number of RBridges in their associated set of multi-homed RBridges. In the case that several LAALPs have same number of RBridges, these LAALPs are then ordered in ascending order in the proper places of the table based on their LAALP IDs considered as unsigned integers. (for example, in the above table, both LAALP1 and LAALP2 have 3 member RBridges, assuming LAALP1 ID is smaller than LAALP2 ID, so LAALP1 is followed by LAALP2 in the ordered table.)

Step 3: Take the first valid LAALP (say LAALP_i) with the maximum set of RBridges, say S_i, out of the table and create a new RBv (Say RBv_i) for it.

Step 4: Walk through the remaining valid LAALPs in the table one by one, pick up all the valid LAALPs that have their sets of multi-homed RBridges contain exactly the same RBridges as that of LAALP_i and take them out of the table. Then appoint RBv_i as the servicing RBv for those LAALPs.

Step 5: Repeat Step 3-4 for any LAALPs left until all the valid entries in the table are associated with an RBv.

After performing the above steps, all the 4 RBridges know that LAALP3 is served by an RBv, say RBv1, which has RB3 and RB4 as member RBridges; LAALP1 and LAALP2 are served by another RBv, say RBv2, which has RB1, RB2 and RB3 as member RBridges; and LAALP4 is served

by RBv3, which has RB3 and RB4 as member RBridges, shown as follows:

| RBv | Serving LAALPs | Member RBridges |
| --- | --- | --- |
| RBv1 | {LAALP3} | {RB3, RB4} |
| RBv2 | {LAALP1, LAALP2} | {RB1, RB2, RB3} |
| RBv3 | {LAALP4} | {RB3, RB4} |

In each RBv, one of the member RBridges is elected as the vDRB (Designated RBridge) of the RBv. Then this RBridge picks up an available nickname as the pseudo-nickname for the RBv and announces it to all other member RBridges of the RBv via its TRILL E-L1FS LSPs (refer to Section 9.2 for the relative extended sub-TLVs).

## 4.2. Selection of Pseudo-nickname for RBv

As described in Section 3, in the TRILL campus, an RBv is identified by its pseudo-nickname. In an AAE group, one member RBridge is elected for the duty to select a pseudo-nickname for this RBv; this RBridge is called Designated RBridge of the RBv (vDRB) in this document. The winner is the RBridge with the largest IS-IS System ID considered as an unsigned integer, in the group. Then based on its TRILL IS-IS link state database and the potential pseudo-nickname(s) reported in the PN-LAALP-Membership sub-TLVs by other member RBridges of this RBv (see Section 9.1 for more details), the vDRB selects an available nickname as the pseudo-nickname for this RBv and advertises it to the other RBridges via its E-L1FS FS-LSP(s) (see Section 9.2 and [rfc7180bis]). Except as provided below, the selection of a nickname to use as the pseudo-nickname follows the usual TRILL rules given in [RFC6325] as updated by [rfc7180bis].

To reduce the traffic disruption caused by nickname changing, if possible, vDRB SHOULD attempt to reuse the pseudo-nickname recently used by the group when selecting nickname for the RBv. To help the vDRB to do so, each LAALP related RBridge advertises a re-using pseudo-nickname for each of its LAALPs in its LAALP Membership sub-TLV if it has used such a pseudo-nickname for that LAALP recently. Although it is up to the implementation of the vDRB as to how to treat the re-using pseudo-nicknames, the following is RECOMMENDED:

- o If there are multiple available re-using pseudo-nicknames that are reported by all the member RBridges of some LAALPs in this RBv, the available one that is reported by the largest number of such LAALPs is chosen as the pseudo-nickname for this RBv. If a tie exists, the re-using pseudo-nickname with the smallest value considered as an unsigned integer is chosen.

- o If only one re-using pseudo-nickname is reported, it SHOULD be

chosen if available.

If there is no available re-using pseudo-nickname reported, the vDRB selects a nickname by its usual method.

Then the selected pseudo-nickname is announced by the vDRB to other member RBridges of this RBv in the PN-RBv sub-TLV (see Section 9.2).

## 5. Distribution Trees and Designated Forwarder

In an AAE group, as each of the member RBridges thinks it is the appointed forwarder for VLAN x, without changes made for active-active connection support, they would all ingress/egress frames into/from TRILL campus for all VLANs. For multi-destination frames, more than one member RBridges ingressing them may cause some of the resulting TRILL Data packets to be discarded due to failure of Reverse Path Forwarding (RPF) Check on other RBridges; for a multi-destination traffic, more than one RBridges egressing it may cause local CE(s) receiving duplication frame. Furthermore, in an AAE group, a multi-destination frame sent by a CE (say CEi) may be ingressed into TRILL campus by one member RBridge, then another member RBridge will receive it from TRILL campus and egress it to CEi, which will result in loop back of frame for CEi. These problems are all described in [RFC7379].

In the following sub-sections, the first two issues are discussed in Section 5.1 and Section 5.2, respectively; the third one is discussed in Section 5.3.

### 5.1. Different Trees for Different Member RBridges

In TRILL, RBridges normally use distribution trees to forward multi-destination frames. (Under some circumstances they can be unicast as specified in [RFC7172].) An RPF Check along with other checking is used to avoid temporary multicast loops during topology changes (Section 4.5.2 of [RFC6325]). The RPF Check mechanism only accepts a multi-destination frame ingressed by an RBridge RBi and forwarded on a distribution tree if it arrives at another RBridge RBn on the expected port. If arriving on any other port, the frame MUST be dropped.

To avoid address flip-flopping on remote RBridges, member RBridges use RBv's pseudo-nickname instead of their regular nicknames as ingress nickname to ingress native frames, including multi-destination frames. From the view of other RBridges, these frames appear as if they were ingressed by the RBv. When multi-destination frames of different flows are ingressed by different member RBridges of an RBv and forwarded along the same distribution tree, they may

arrive at RBn on different ports. Some of them will violate the RPF Check principle at RBn and be dropped, which will result in lost traffic.

In an RBv, if different member RBridge uses different distribution trees to ingress multi-destination frames, the RPF Check violation issue can be fixed. Coordinated Multicast Trees (CMT) [CMT] proposes such an approach, and makes use of the Affinity sub-TLV defined in [RFC7176] to tell other RBridges which trees a member RBridge (say RBi) may choose when ingressing multi-destination frames; then all RBridges in the TRILL campus can calculate RPF Check information for RBi on those trees taking the tree affinity information into account [CMT].

This document uses the approach proposed in [CMT] to fix the RPF Check violation issue. Please refer to [CMT] for more details of the approach.

## 5.2. Designated Forwarder for Member RBridges

Take Figure 3 as an example, where CE1 and CE2 are served by an RBv that has RB1 and RB2 as member RBridges. In VLAN x, the three CEs can communicate with each other.

```
                    ---------------------
                  /                       \    +-----+
                 |       TRILL Campus      |---| RBn |
                  \                       /    +-----+
                   -----------------------
                       |             |
                  +----+             +------+
                  |                         |
             +---------+                +--------+
             |   RB1   |                |   RB2  |
             | oooooooo|oooooooooooooooo|ooooo   |
             +o--------+    RBv         +-----o--+
               o|oooo|oooooooooooooooooooo|o|o  |
                | +--|--------------------+ |   |
                | |  +---------+ +----------+   |
               (| |)<-LAALP1  (| |)<-LAALP2     |
            +-------+       +-------+      +-------+
            |  CE1  |       |  CE2  |      |  CE3  |
            +-------+       +-------+      +-------+
```

Figure 3  A Topology with Multi-homed and Single-homed CEs

When a remote RBridge (say RBn) sends a multi-destination TRILL Data packet in VLAN x (or the FGL that VLAN x maps to if the packet is

FGL), both RB1 and RB2 will receive it. As each of them thinks it is the appointed forwarder for VLAN x, without changes made for active-active connection support, they would both forward the frame to CE1/CE2. As a result, CE1/CE2 would receive duplicate copies of the frame through this RBv.

In another case, assume CE3 is single-homed to RB2. When it transmits a native multi-destination frame onto link CE3-RB2 in VLAN x, the frame can be locally replicated to the ports to CE1/CE2, and also encapsulated into TRILL Data packet and ingressed into TRILL campus. When the packet arrives at RB1 across the TRILL campus, it will be egressed to CE1/CE2 by RB1. Then CE1/CE2 receives duplicate copies from RB1 and RB2.

In this document, the Designated Forwarder (DF) for a VLAN is introduced to avoid the duplicate copies. The basic idea of the DF is to elect one RBridge per VLAN from an RBv to egress multi-destination TRILL Data traffic and replicate locally-received multi-destination native frames to the CEs served by the RBv.

Note that the DF has an effect only on the egressing/replicating of multi-destination traffic. It has no effect on the ingressing, forwarding, or egressing of unicast frames. Furthermore, the DF check is performed only for RBv ports, not on regular access ports.

Each RBridge in an RBv elects a DF using the same algorithm which guarantees the same RBridge elected as DF per VLAN by all members of the RBv.

Assuming there are m LAALPs and k member RBridges in an RBv; each LAALP is referred to as LAALPi where 0 <= i < m, and each RBridge is referred to as RBj where 0 <= j < k, the DF election algorithm per VLAN is as follows:

Step 1: For LAALPi, sort all the RBridges in numerically ascending order based on SHA-256(System IDj | LAALP IDi) considered as an unsigned intger, where SHA-256 is the hash function in [RFC6234], "System IDj" is the 6-byte IS-IS System ID of RBj, "|" means concatenation, and LAALP IDi is the LAALP ID for LAALPi. System ID and LAALP ID are considered as byte strings. In the case of a tie, the tied RBridges are sorted in numerically ascending order by their System IDs considered as unsigned integers.

Step 2: Each RBridge in the numerically sorted list is assigned a monotonically increasing number j, such that increasing number j corresponds to its position in the sorted list, i.e., the first RBridge (the one with the smallest SHA-256(System ID | LAALP ID)) is assigned zero and the last is assigned k-1.

Step 3: For each VLAN ID n, choose the RBridge whose number equals (n mod k) as the DF.

Step 4: Repeat Step 1-3 for the remaining LAALPs until there is a DF per VLAN per LAALP in the RBv.

For a multi-destination native frame of VLAN x received, if RBi is an LAALP attached RBridge, there are three cases where RBi replicates the multi-destination frame, as follows:

1) Local replication of the frame to regular (non-AAE) access ports as per [RFC6325] (and [RFC7172] for FGL).

2) RBv ports associated with the same pseudo-nickname as that of the incoming port, no matter whether RBi is the DF for the frame's VLAN on the outgoing ports except that the frame MUST NOT be replicated back to the incoming port. RBi cannot simply depend on the DF to forward the multi-destination frame back into the AAEs associated with pseudo-nickname as that would cause the source CE to get the frame back, which is a violation of basic Ethernet properties. The DF will not forward such a frame back into the AAE due to ingress nickname filtering as described in Section 5.3.

3) RBv ports on which RBi is the DF for the frame's VLAN while they are associated with different pseudo-nickname(s) to that of the incoming port.

For a multi-destination TRILL Data packet received, RBi MUST NOT egress it out of the RBv ports where it is not DF for the frame's Inner.VLAN (or for the VLAN corresponding to the Inner.Label if the packet is an FGL one). Otherwise, whether or not egressing it out of such ports is further subject to the filtering check result of the frame's ingress nickname on these ports (see Section 5.3).

### 5.3. Ingress Nickname Filtering

As shown in Figure 3, CE1 may send multi-destination traffic in VLAN x to TRILL campus via a member RBridge (say RB1). The traffic is then TRILL-encapsulated by RB1 and delivered through the TRILL campus to multi-destination receivers. RB2 may receive the traffic, and egress it back to CE1 if it is the DF for VLAN x on the port to LAALP1. Then the traffic loops back to CE1 (see Section 3.2 of [RFC7379).

To fix the above issue, an ingress nickname filtering check is required by this document. The idea is to check the ingress nickname of a multi-destination TRILL Data packet before egressing a copy of it out of an RBv port. If the ingress nickname matches the pseudo-

nickname of the RBv (associated with the port), the filtering check should fail and the copy MUST NOT be egressed out of that RBv port. Otherwise, the copy is egressed out of that port if it has also passed other checks, such as the appointed forwarder check in Section 4.6.2.5 of [RFC6325] and the DF check in Section 5.2.

Note that this ingress nickname filtering check has no effect on the multi-destination native frames received on access ports and replicated to other local ports (including RBv ports), since there is no ingress nickname associated with such frames. Furthermore, for the RBridge regular access ports, there is no pseudo-nickname associated with them; so no ingress nickname filtering check is required on those ports.

More details of data packet processing on RBv ports are given in the next section.

## 6. TRILL Traffic Processing

This section provides more details of native frame and TRILL Data packet processing as it relates to the RBv's pseudo-nickname.

### 6.1. Native Frames Ingressing

When RB1 receives a unicast native frame from one of its ports that has end-station service enabled, it processes the frame as described in Section 4.6.1.1 of [RFC6325] with the following exception.

o If the port is an RBv port, RB1 uses the RBv's pseudo-nickname, instead of one of its regular nickname(s) as the ingress nickname when doing TRILL encapsulation on the frame.

When RB1 receives a native multi-destination (Broadcast, Unknown unicast or Multicast) frame from one of its access ports (including regular access ports and RBv ports), it processes the frame as described in Section 4.6.1.2 of [RFC6325] with the following exceptions.

o If the incoming port is an RBv port, RB1 uses the RBv's pseudo-nickname, instead of one of its regular nickname(s) as the ingress nickname when doing TRILL encapsulation on the frame.

o For the copies of the frame replicated locally to RBv ports, there are two cases as follows:

    - If the outgoing port(s) is associated with the same pseudo-nickname as that of the incoming port but not with the same

LAALP as the incoming port, the copies are forwarded out of that outgoing port(s) after passing the appointed forwarder check for the frame's VLAN. That is to say, the copies are processed on such port(s) as Section 4.6.1.2 of [RFC6325].

- Else, the Designated Forwarder (DF) check is also made on the outgoing ports for the frame's VLAN after the appointed forwarder check. The copies are not output through the ports that failed the DF check (i.e., RB1 is not DF for the frame's VLAN on the ports); otherwise, the copies are forwarded out of the ports that pass the DF check (see Section 5.2).

For such a frame received, the MAC address information learned by observing it, together with the LAALP ID of the incoming port SHOULD be shared with other member RBridges in the group (see Section 7).

### 6.2. Egressing TRILL Data Packets

This section describes egress processing of the TRILL Data packets received on an RBv member RBridge (say RBn). Section 6.2.1 describes the egress processing of unicast TRILL Data packets and Section 6.2.2 specifies the multi-destination TRILL Data packets egressing.

#### 6.2.1. Unicast TRILL Data Packets

When receiving a unicast TRILL data packet, RBn checks the egress nickname in the TRILL header of the packet. If the egress nickname is one of RBn's regular nicknames, the packet is processed as defined in Section 4.6.2.4 of [RFC6325].

If the egress nickname is the pseudo-nickname of a local RBv, RBn is responsible for learning the source MAC address, unless data plane learning has been disabled. The learned {Inner.MacSA, Data Label, ingress nickname} triplet SHOULD be shared within the AAE group as described in Section 7.

Then the packet is de-capsulated to its native form. The Inner.MacDA and Data Label are looked up in RBn's local forwarding tables, and one of the three following cases will occur. RBn uses the first case that applies and ignores the remaining cases:

o If the destination end station identified by the Inner.MacDA and Data Label is on a local link, the native frame is sent onto that link with the VLAN from the Inner.VLAN or VLAN corresponding to the Inner.Label if the packet is FGL.

o Else if RBn can reach the destination through another member RBridge RBk, it tunnels the native frame to RBk by re-

encapsulating it into a unicast TRILL Data packet and sends it to RBk. RBn uses RBk's regular nickname, instead of the pseudo-nickname as the egress nickname for the re-encapsulation, and the ingress nickname remains unchanged (somewhat similar to Section 2.4.2.1 of [rfc7180bis]). If the hop count value of the packet is too small for it to reach RBk safely, RBn SHOULD increase that value properly in doing the re-encapsulation. (NOTE: When receiving that re-encapsulated TRILL Data packet, as the egress nickname of the packet is RBk's regular nickname rather than the pseudo-nickname of a local RBv, RBk will process it as Section 4.6.2.4 of [RFC6325], and will not re-forward it to another RBridge.)

o Else, RBn does not know how to reach the destination; it sends the native frame out of all the local ports on which it is appointed forwarder for the Inner.VLAN (or appointed forwarder for the VLAN into which the Inner.Label maps on that port for FGL TRILL Data packet [RFC7172]).

### 6.2.2. Multi-Destination TRILL Data Packets

When RB1 receives a multi-destination TRILL Data Packet, it checks and processes the packet as described in Section 4.6.2.5 of [RFC6325] with the following exception.

o On each RBv port where RBn is the appointed forwarder for the packet's Inner.VLAN (or for the VLAN to which the packet's Inner.Label maps on that port if it is an FGL TRILL Data packet), the Designated Forwarder check (see Section 5.2) and the Ingress Nickname Filtering check (see Section 5.3) are further performed. For such an RBv port, if either the DF check or the filtering check fails, the frame MUST NOT be egressed out of that port. Otherwise, it can be egressed out of that port.

## 7. MAC Information Synchronization in Edge Group

An edge RBridge, say RB1 in LAALP1, may have learned a { MAC address, Data Label } to nickname correspondence for a remote host h1 when h1 sends a packet to CE1. The returning traffic from CE1 may go to another member RBridge of LAALP1, for example RB2. RB2 may not have that correspondence stored. Therefore it has to do the flooding for unknown unicast. Such flooding is unnecessary since the returning traffic is almost always expected and RB1 had learned the address correspondence. To avoid the unnecessary flooding, RB1 SHOULD share the correspondence with other RBridges of LAALP1. RB1 synchronizes the correspondence by using the MAC-RI sub-TLV [RFC6165] in its ESADI-LSPs [RFC7357].

On the other hand, RB2 has learned the MAC address and Data Label of CE1 when CE1 sends a frame to h1 through RB2. The returning traffic from h1 may go to RB1. RB1 may not have CE1's MAC address and Data Label stored even though it is in the same LAALP for CE1 as RB2. Therefore it has to flood the traffic out of all its access ports where it is appointed forwarder for the VLAN (see Section 6.2.1) or the VLAN the FGL maps to on that port if the packet is FGL. Such flooding is unnecessary since the returning traffic is almost always expected and RB2 had learned the CE1's MAC and Data Label information. To avoid that unnecessary flooding, RB2 SHOULD share the MAC address and Data Label with other RBridges of LAALP1. RB2 synchronizes the MAC address and Data Label by enclosing the relative MAC-RI TLV within a pair of boundary TRILL APPsub-TLVs for LAALP1 (see Section 9.3) in its ESADI-LSP [RFC7357]. After receiving the enclosed MAC-RI TLVs, the member RBridges of LAALP1 (i.e., LAALP1 related RBridges) treat the MAC address and Data Label as if it was learned by them locally on their member port of LAALP1; the LAALP1 unrelated RBridges just ignore LAALP1's boundary APPsub-TLVs and treat the MAC address and Data Label as specified in [RFC7357]. Furthermore, in order to make the LAALP1 unrelated RBridges know that the MAC and Data Label is reachable through the RBv that provides service to LAALP1, the Topology-id/Nickname field of the MAC-RI TLV SHOULD carry the pseudo-nickname of the RBv rather than zero or one of the originating RBridge's (i.e., RB2's) regular nicknames.

## 8. Member Link Failure in RBv

As shown in Figure 4, suppose the link RB1-CE1 fails. Although a new RBv will be formed by RB2 and RB3 to provide active-active service for LAALP1 (see Section 5), the unicast traffic to CE1 might still be forwarded to RB1 before the remote RBridge learns CE1 is attached to the new RBv. That traffic might be disrupted by the link failure. Section 8.1 discusses the failure protection in this scenario.

However, for multi-destination TRILL Data packets, they can reach all member RBridges of the new RBv and be egressed to CE1 by either RB2 or RB3 (i.e., the new DF for the traffic's Inner.VLAN or the VLAN the packet's Inner.Label maps to in the new RBv). Although there might be a transient hang time between failure and the establishment of the new RBv, special actions to protect against downlink failure for such multi-destination packets is not needed.

```
                   ------------------
                 /                    \
                |     TRILL Campus     |
                 \                    /
                  --------------------
                      |     |     |
                  +---+     |     +----+
                  |         |          |
              +------+     +------+   +------+
              | RB1  |     | RB2  |   | RB3  |
              ooooooo|ooooo|oooooo|ooo|ooooo |
             o+------+ RBv +------+   +-----o+
              o|oooo|ooooooo|oooo|ooooo|oo|o
               |    |       |  +-|-----+  |
              \|/+--|-------+  | +------+ |
             - B |  +----------|------+ | |
              /|\| +-----------+      | | |
              (| | |)<--LAALP1       (| | |)<--LAALP2
             +-------+              +-------+
             |  CE1  |              |  CE2  |
             +-------+              +-------+
   B - Failed Link or Link bundle
```

Figure 4  A Topology with Multi-homed and Single-homed CEs

## 8.1. Link Protection for Unicast Frame Egressing

When the link CE1-RB1 fails, RB1 loses its direct connection to CE1. The MAC entry through the failed link to CE1 is removed from RB1's local forwarding table immediately. Another MAC entry learned from another member RBridge of LAALP1 (for example RB2, since it is still a member RBridge of LAALP1) is installed into RB1's forwarding table (see Section 9.3).  In that new entry, RB2 (identified by one of its regular nicknames) is the egress RBridge for CE1's MAC address. Then when a TRILL Data packet to CE1 is delivered to RB1, it can be tunneled to RB2 after being re-encapsulated (ingress nickname remains unchanged and egress nickname is replaced by RB2's regular nickname) based on the above installed MAC entry (see bullet 2 in Section 6.2.1). Then RB2 receives the frame and egresses it to CE1.

After the failure recovery, RB1 learns that it can reach CE1 via link CE1-RB1 again by observing CE1's native frames or from the MAC information synchronization by member RBridge(s) of LAALP1 described in Section 7, then it restores the MAC entry to its previous one and downloads it to its data plane fast path logic.

## 9. TLV Extensions for Edge RBridge Group

The following subsections specify the APPsub-TLVs needed to support pseudo-nickname edge groups.

### 9.1. PN-LAALP-Membership APPsub-TLV

This APPsub-TLV is used by an edge RBridge to announce its associated pseudo-nickname LAALP information. It is defined as a sub-TLV of the TRILL GENINFO TLV [RFC7357] and is distributed in E-L1FS FS-LSPs [rfc7180bis]. It has the following format:

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Type = PN-LAALP-Membership   |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Length                       |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+
|  LAALP RECORD(1)                            |  (variable)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+
.                                             .
.                                             .
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+
|  LAALP RECORD(n)                            |  (variable)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+
```

Figure 5 PN-LAALP-Membership Advertisement APPsub-TLV

where each LAALP RECORD has the following form:

```
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 ..
+--+-+-+-+-+-+-+-+-+
|OE|     RESV     |                  (1 byte)
+--+-+-+-+-+-+-+-+-+
|  Size           |                  (1 byte)
+--+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Re-using Pseudo-nickname      |  (2 bytes)
+--+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+
|  LAALP ID                                     |  (variable)
+--+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+
```

o PN-LAALP-Membership (2 bytes): Defines the type of this sub-TLV, #tbd1.

o Length (2 bytes): the sum of the lengths of the LAALP RECORDs.

o OE (1 bit): a flag indicating whether or not the LAALP wants to occupy an RBv by itself; 1 for occupying by itself (or Occupying Exclusively (OE)). By default, it is set to 0 on transmit. This bit is used for edge RBridge group auto-discovery (see Section 4.1). For any one LAALP, the values of this flag might conflict in

the LSPs advertised by different member RBridges of that LAALP. In that case, the flag for that LAALP is considered as 1.

- RESV (7 bits): MUST be transmitted as zero and ignored on receipt.

- Size (1 byte): Size of remaining part of LAALP RECORD (2 plus length of the LAALP ID).

- Re-using Pseudo-nickname (2 bytes): Suggested pseudo-nickname of the AAE group serving the LAALP. If the LAALP is not served by any AAE group, this field MUST be set to zero. It is used by the originating RBridge to help the vDRB to reuse the previous pseudo-nickname of an AAE group (see Section 4.2).

- LAALP ID (variable): The ID of the LAALP. See Section 9.4.

On receipt of such an APPsub-TLV, if RBn is not an LAALP related edge RBridge, it ignores the sub-TLV; otherwise, it parses the sub-TLV. When new LAALPs are found or old ones are withdrawn compared to its old copy, and they are also configured on RBn, it triggers RBn to perform the "Member RBridges Auto-Discovery" procedure described in Section 4.1.

## 9.2. PN-RBv APPsub-TLV

The PN-RBv APPsub-TLV is used by a Designated RBridge of a Virtual RBridge (vDRB) to dictate the pseudo-nickname for the LAALPs served by the RBv. It is defined as a sub-TLV of TRILL GENINFO TLV [RFC7357] and is distributed in E-L1FS FS-LSP [rfc7180bis]. It has the following format:

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type = PN-RBv                 |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Length                        |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| RBv's Pseudo-Nickname         |  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| LAALP ID Size |  (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+
| LAALP ID (1)                                  |  (variable)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+
.                                               .
.                                               .
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+
| LAALP ID (n)                                  |  (variable)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+
```

- o PN-RBv (2 bytes): Defines the type of this sub-TLV, #tbd2.

- o Length (2 bytes): 3+n*k bytes, where there are n LAALP IDs, each of size k bytes. k is found in the LLALP ID Size field below. If Length is not 3 plus an integer time k, the sub-TLV is corrupt and MUST be ignored.

- o RBv's Pseudo-Nickname (2 bytes): The appointed pseudo-nickname for the RBv that serves for the LAALPs listed in the following fields.

- o LAALP ID Size (1 byte): The size of each of the following LAALP IDs in this sub-TLV. 8 if the LAALPs listed are MC-LAGs or DRNI (Section 6.3.2 in [802.1AX]). The value in this field is the k that appears in the formula for Length above.

- o LAALP ID (LAAP ID Size bytes): The ID of the LAALP. See Section 9.4.

This sub-TLV may occur multiple times with the same RBv pseudo-nickname with the meaning that all of the LAALPs listed are identified by that pseudo-nickname. For example, if there are LAALP IDs of different length, then the LAALP IDs of each size would have to be listed in a separate sub-TLV.

Since a PN-RBv APPsub-TLV is distributed as part of the application link state, using the E-L1FS scope [rfc7180bis], changes in contents or withdrawal or creation of a PN-RBv APPsub-TLV is accomplished by the Designated RBridge updating and flooding an E-L1FS PDU.

On receipt of such a sub-TLV, if RBn is not an LAALP related edge RBridge, it ignores the sub-TLV. Otherwise, if RBn is also a member RBridge of the RBv identified by the list of LAALPs, it associates the pseudo-nickname with the ports of these LAALPs and downloads the association to data plane fast path logic. At the same time, RBn claims RBv pseudo-nickname across the campus and announces RBv as its child on the corresponding tree or trees using the Affinity sub-TLV [RFC7176] [CMT].

## 9.3. PN-MAC-RI-LAALP Boundary APPsub-TLVs

In this document, two APPsub-TLVs are used as boundary APPsub-TLVs for edge RBridge to enclose the MAC-RI TLV(s) containing the MAC address information leant form local port of an LAALP when this RBridge wants to share the information with other edge RBridges. They are defined as TRILL APPsub-TLVs [RFC7357]. The PN-MAC-RI-LAALP-INFO-START APPsub-TLV has the following format:

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|Type=PN-MAC-RI-LAALP-INFO-START| (2 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Length                        | (2 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-...+-+-+-+-+-+-+
| LAALP ID                                 | (variable)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-...+-+-+-+-+-+-+
```

o PN-MAC-RI-LAALP-INFO-START (2 bytes): Defines the type of this APPsub-TLV, #tbd3.

o Length (2 bytes): the size of the following LAALP ID. 8 if the LAALP listed is an MAC-LAG or DRNI.

o LAALP ID (variable): The ID of the LAALP (see Section 9.4).

PN-MAC-RI-LAALP-INFO-END APPsub-TLV is defined as follows:

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type=PN-MAC-RI-LAALP-INFO-END | (2 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Length                        | (2 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

o PN-MAC-RI-LAALP-INFO-END (2 bytes): Defines the type of this sub-TLV, #tbd4.

o Length (2 bytes): 0.

This pair of APPsub-TLVs can be carried multiple times in an ESADI LSP and in multiple ESADI-LSPs. When an LAALP related edge RBridge (say RBn) wants to share with other edge RBridges the MAC addresses learned on its local ports of different LAALPs, it uses one or more pairs of such APPsub-TLVs for each of such LAALPs in its ESADI-LSPs. Each encloses the MAC-RI TLVs containing the MAC addresses learned from a specific LAALP. Furthermore, if the LAALP is served by a local RBv, the value of Topology ID/Nickname field in the relative MAC-RI TLVs SHOULD be the pseudo-nickname of the RBv rather than one of the RBn's regular nickname or zero. Then on receipt of such a MAC-RI TLV, remote RBridges know that the contained MAC addresses are reachable through the RBv.

On receipt of such boundary APPsub-TLVs, when the edge RBridge is not an LAALP related one or cannot recognize such sub-TLVs, it ignores them and continues to parse the enclosed MAC-RI TLVs per [RFC7357]. Otherwise, the recipient parses the boundary APPsub-TLVs. The PN-MAC-RI-LAALP-INFO-START / PN-MAC-RI-LAALP-INFO-END pair MUST occur within one TRILL GENINFO TLV. If an END is encountered without any previous

START in the ESADI-LSP, the END APPsub-TLV is ignored. If, after encountering a START, the end of the ESADI-LSP is reached without encountering an END, then the end of the ESADI-LSP is treated as if it were a PN-MAC-RI-LAALP-INFO-END. The boundary APPsub-TLVs and TLVs between them are handled as follows:

1) If the edge RBridge is configured with the contained LAALP and the LAALP is also enabled locally, it treats all the MAC addresses, contained in the following MC-RI TLVs enclosed by the corresponding pair of boundary APPsub-TLVs, as if they were learned from its local port of that LAALP;

2) Else, it ignores these boundary APPsub-TLVs and continues to parse the following MAC-RI TLVs per [RFC7357] until another pair of boundary APPsub-TLVs is encountered.

### 9.4. LAALP IDs

The LAALP ID identifies an AAE RBridge Group in the TRILL campus and thus MUST be unique across the campus. In all of the APPsub-TLVs specified above, the length of the LAALP ID can be determined from a size field. If that length is 8 bytes, the LAALP ID is an MC-LAG or DRNI identifier as specified in Section 6.3.2 in [802.1AX]. The meaning and structure of LAALP IDs of other lengths is reserved and may be specified in future documents.

## 10. OAM Packets

Attention must be paid when generating OAM packets. To ensure the response messages can return to the originating member RBridge of an RBv, pseudo-nickname cannot be used as the ingress nickname in TRILL OAM messages, except in the response to an OAM message that has that RBv's pseudo-nickname as egress nickname. For example, assume RB1 is a member RBridge of RBvi, RB1 cannot use RBvi's pseudo-nickname as the ingress nickname when originating OAM messages; otherwise the responses to the messages may be delivered to another member RBridge of RBvi rather than RB1. But when RB1 responds to the OAM message with RBvi's pseudo-nickname as egress nickname, it can use that pseudo-nickname as the ingress nickname in the response message.

Since RBridges cannot use OAM messages for the learning of MAC addresses (Section 3.2.1 of [RFC7174]), it will not lead to MAC address flip-flopping at a remote RBridge even though RB1 uses its regular nicknames as ingress nicknames in its TRILL OAM messages while uses RBvi's pseudo-nickname in its TRILL Data packets.

## 11. Configuration Consistency

The VLAN membership of all the RBridge ports in an LAALP MUST be the same. Any inconsistencies in VLAN membership may result in packet loss or non-shortest paths.

Take Figure 1 for example, suppose RB1 configures VLAN1 and VLAN2 for the link CE1-RB1, while RB2 only configures VLAN1 for the CE1-RB2 link. Both RB1 and RB2 use the same ingress nickname RBv for all frames originating from CE1. Hence, a remote RBridge RBx will learn that CE1's MAC address in VLAN2 is originating from RBv. As a result, on the returning path, remote RBridge RBx may deliver VLAN2 traffic to RB2. However, RB2 does not have VLAN2 configured on CE1-RB2 link and hence the frame may be dropped or has to be redirected to RB1 if RB2 knows RB1 can reach CE1 in VLAN2.

How LAALP implementations maintain consistent VLAN configuration on the TRILL switch LAALP ports is out of scope for the TRILL protocol. However, considering the consequences that might cause by the inconsistency, TRILL switches MUST disable the ports connected to an LAALP with inconsistent VLAN configuration.

It is important that if any VLAN in an LAALP is being mapped by edge RBridges to an FGL [RFC7172], that the mapping MUST be same for all edge RBridge ports in the LAALP. Otherwise, for example, unicast FGL TRILL Data packets from remote RBridges may get mapped into different VLANs depending on which edge RBridge receives and egresses them.

It is important that RBridges in an AAE group not be configured to assert the OE bit if any RBridge in the group does not implement it. Since, as stated in [RFC7379], the RBridges in an AAE edge group are expected to be from the same vendor, due to the proprietary nature of deployed LAALPs, this will normally follow automatically from all of the RBridge in an AAE edge group supporting or all not supporting OE.

## 12. Security Considerations

Authenticity for contents transported in IS-IS PDUs is enforced using regular IS-IS security mechanism [IS-IS] [RFC5310].

For security considerations pertain to extensions transported by TRILL ESADI, see the Security Considerations section in [RFC7357].

Since currently deployed LAALPs [RFC7379] are proprietary, security over membership in and internal management of active-active edge groups is proprietary. If authentication is not used, a rogue RBridge that insinuates itself into an active-active edge group can disrupt end station traffic flowing into or out of that group. For example, if there are N RBridges in the group, it could typically control

1/Nth of the traffic flowing out of that group and a similar amount of unicast traffic flowing into that group. For multi-destination traffic flowing into that group, it could control all that was in a VLAN for which it was DF and it can exercise substantial control over the DF election by changing its own System ID.

For general TRILL Security Considerations, see [RFC6325].

## 13. IANA Considerations

IANA is requested to allocate code points tbd1, tbd2, tbd3 and tbd4 from the range below 255 for the 4 TRILL APPsub-TLVs specified in Section 9 and add them to the TRILL APPsub-TLV Types registry as follows:

| Type | Name | Reference |
|---|---|---|
| tbd1 | PN-LAALP-Membership | [this document] |
| tbd2 | PN-RBv | [this document] |
| tbd3 | PN-MAC-RI-LAALP-INFO-START | [this document] |
| tbd4 | PN-MAC-RI-LAALP-INFO-END | [this document] |

## 14. Acknowledgments

We would like to thank Mingjiang Chen for his contributions to this document. Additionally, we would like to thank Erik Nordmark, Les Ginsberg, Ayan Banerjee, Dinesh Dutt, Anoop Ghanwani, Janardhanan Pathang, Jon Hudson and Fangwei Hu for their good questions and comments.

## 15. Contributing Authors

Weiguo Hao
Huawei Technologies
101 Software Avenue,
Nanjing 210012
China

Phone: +86-25-56623144
Email: haoweiguo@huawei.com

Donald E. Eastlake, III
Huawei Technologies
155 Beaver Street
Milford, MA 01757 USA

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

## 16. References

### 16.1. Normative References

[CMT] T. Senevirathne, J. Pathangi, and J. Hudson, "Coordinated Multicast Trees (CMT) for TRILL", draft-ietf-trill-cmt Work in Progress.

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

[RFC5310] M. Bhatia, V. Manral, T. Li, et al, "IS-IS Generic Cryptographic Authentication", RFC 5310, February 2009.

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

[RFC6234] Eastlake 3rd, D. and T. Hansen, "US Secure Hash Algorithms (SHA and SHA-based HMAC and HKDF)", RFC 6234, May 2011.

[RFC6325] R. Perlman, D. Eastlake, D. Dutt, S. Gai, and A. Ghanwani, "Routing Bridges (RBridges): Base Protocol Specification", RFC 6325, July 2011.

[RFC6439] Perlman, R., Eastlake, D., Li, Y., Banerjee, A., and F. Hu, "Routing Bridges (RBridges): Appointed Forwarders", 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.

[RFC7176] D. Eastlake, A. Banerjee, A. Ghanwani, and R. Perlman, "Transparent Interconnection of Lots of Links (TRILL) Use of IS-IS", RFC 7176, May 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.

[RFC7356] Ginsberg, L., Previdi, S., and Y. Yang, "IS-IS Flooding Scope Link State PDUs (LSPs)", RFC 7356, September 2014.

[rfc7180bis] D. Eastlake, et al., draft-ietf-trill-rfc7180bis, work in progress.

[802.1AX] IEEE, "IEEE Standard for Local and Metropolitan Area/ networks Link Aggregation", IEEE Std 802.1AX-2014, 24 December 2014.

## 16.2. Informative References

[IS-IS] ISO/IEC 10589:2002, Second Edition, "Information technology -- Telecommunications and information exchange between systems -- Intermediate System to Intermediate System intra-domain routeing information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode network service (ISO 8473)", 2002.

[RFC7174] Salam, S., Senevirathne, T., Aldrin, S., and D. Eastlake 3rd, "Transparent Interconnection of Lots of Links (TRILL) Operations, Administration, and Maintenance (OAM) Framework", RFC 7174, May 2014.

[RFC7379] Li, Y., Hao, W., Perlman, R., Hudson, J., and H. Zhai, "Problem Statement and Goals for Active-Active Connection at the Transparent Interconnection of Lots of Links (TRILL) Edge", RFC 7379, October 2014.

[MultiAttach] Zhang, M., et al, "TRILL Active-Active Edge Using Multiple MAC Attachments", draft-ietf-trill-aa-multi-attach, Work in Progress.

Authors' Addresses

Hongjun Zhai
Jinling Institute of Technology
99 Hongjing Avenue, Jiangning District
Nanjing, Jiangsu 211169
China

Email: honjun.zhai@tom.com

Tissa Senevirathne
Consultant

Email: tsenevir@gmail.com

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

Email: Radia@alum.mit.edu

Mingui Zhang
Huawei Technologies
Huawei Building, No.156 Beiqing Rd.
Beijing, Beijing 100095
China

Email: zhangmingui@huawei.com

Yizhou Li
Huawei Technologies
101 Software Avenue,
Nanjing 210012
China

Phone: +86-25-56625409
Email: liyizhou@huawei.com