TRILL Working Group   H. Zhai
Internet-Draft   ZTE
Intended Status: Standards Track   T. Senevirathne
Expires: December 27, 2014   Cisco Systems
                                                  R. Perlman
                                                  Intel Labs
                                             D. Eastlake 3rd
                                                    M. Zhang
                                                       Y. Li
                                                      Huawei
                                               June 25, 2014

        RBridge: Pseudo-Nickname for Active-active Access
            draft-hu-trill-pseudonode-nickname-08

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. In this document, the edge RBridge (TRILL switch)
   group providing active-active access to such an end station can be
   represented as a Virtual RBridge. Based on the concept of Virtual
   RBridge along with its pseudo-nickname, this document facilitates the
   TRILL active-active access of such end stations.

Status of This Memo

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

Copyright and License Notice

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 switch.

   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 Local
   Area Network (LAN) segment). 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 Appointed Forwarder (AF) for the
   VLAN on the link [RFC6325] [RFC6439]. However, in some practical
   deployments, to increase the access bandwidth and reliability, an end
   station might multiply connect to several edge RBridges and treat all
   of the uplinks as a Multi-Chassis Link Aggregation (MC-LAG) bundle.
   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 for the end station.

   Traffic with the same VLAN and source MAC address but belonging to
   different flows might be sent by such an end station to different
   member RBridges of the AAE group, and then is ingressed into TRILL
   campus. When an 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.
   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 duplication egressing and loop of multi-destination frames may
   also disturb the end stations multiply connected to the member
   RBridges of an AAE group [AAProb].

   Edge RBridge groups, which can be represented as a Virtual RBridge
   (RBv) and assigned a pseudo-nickname, address the AAE issues of TRILL
   in this document. A member RBridge of such a group uses the pseudo-
   nickname, instead of its own nickname, as the ingress RBridge
   nickname when ingressing frames received on attached MC-LAG links.

   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 virtual RBridge and its pseudo-
nickname. Section 4 describes how edge RBridges constitute 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 and
forwarding loop, 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);
followed by Section 7, which describes the MAC information
synchronization among the member RBridges of an RBv. Section 8
discusses the protection against downlink failure at a member
RBridge; and Section 9 gives the necessary TLV extensions for AAE
RBridge group.

1.1. Terminology and Acronyms

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

CE - As in [CMT], Classic Ethernet device (end station or bridge).
The device can be either physical or virtual equipment.

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

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

RBv - Virtual RBridge, an alias of active-active edge RBridge group
in this document.

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

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

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, end stations (referred to as CEs 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 a Multi-Chassis
   Link Aggregation (MC-LAG) bundle. Then RB1, RB2, ... RBk constitute
   an Active-active Edge (AAE) RBridge group for CE1 in this MC-LAG.
   Even if a member RBridge or an uplink fails, CE1 can still get frame
   forwarding service from 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 MC-LAG bundle in the AAE group when it communicates with other
   end stations across the TRILL campus [AAProb].

```
                    ----------------------
                   |                      |
                   |    TRILL Campus      |
                   |                      |
                    ----------------------
                     |        |      |
               +-----+        |    +--------+
               |              |             |
           +------+      +------+      +------+
           |(RB1) |      |(RB2) |      | (RBk)|
           +------+      +------+      +------+
            |..|          |..|            |..|
            |  +----+     |  |            |  |
            |   +---|-----|--|----------+  |
            | +-|---|-----+  +-----------+ |
       MC-  | | |   +------------------+ | |
      LAG1-->(| | |)                    (| | |) <--MC-LAGn
           +-------+    .   .   .    +-------+
           | CE1   |                 | CEn   |
           +-------+                 +-------+
```
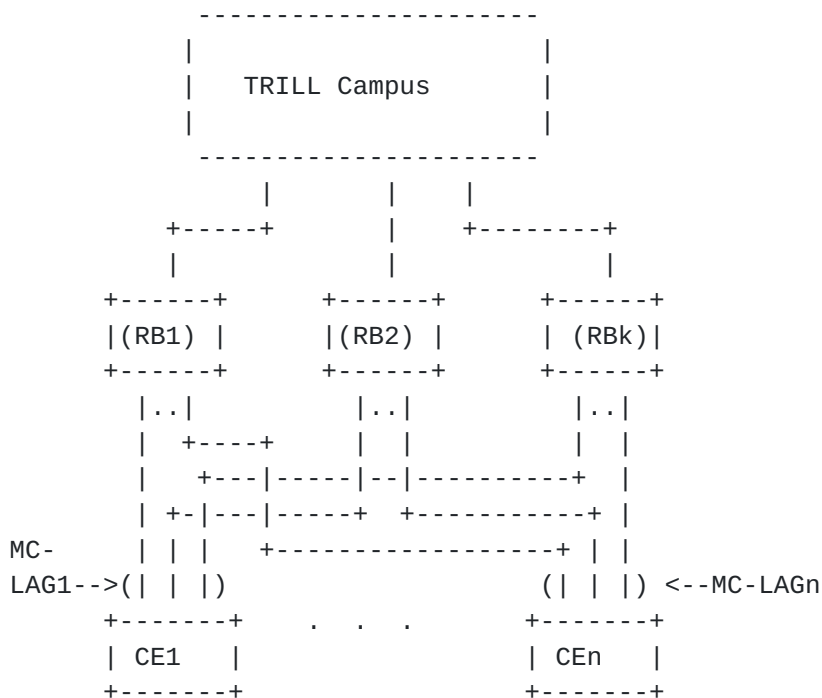
       Figure 1 Active-Active Connection to TRILL Edge RBridges

   By design, an MC-LAG (say MC-LAG1) 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 MC-LAG.
   So every member RBridge of MC-LAG1 thinks of itself as appointed
   forwarder for all VLANs enabled on an MC-LAG1 link and can
   ingress/egress frames simultaneously in these VLANs. The simultaneous
   flow-based ingressing/egressing may 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 [AAProb]); simultaneous ingressing of frames originated by CE1
for different flows in the same VLAN will result in MAC address flip-
flopping at remote egress RBridges (see Section 3.3 of [AAProb]). The
flip-flopping in turn causes packet re-ordering in reverse traffic.

Since the fact is true that 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]), the MAC flip-flopping issue should be solved based on the
assumption that the default learning is enabled at edge RBridges. So
this document specifies Virtual RBridge, together with its pseudo-
nickname, to fix these issues.

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 MC-LAG. More exactly, it
represents a group of end station service ports on the edge RBridges
and the end station service provided to the CE(s) on these ports,
through which the CE(s) is multiply attached to TRILL campus using
MC-LAG(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 MC-LAG 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 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 from RBvn when its
last port of 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 their path(s) to RBvn is not available
through RBi now.

When member RBridges receive native frames from 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 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 use its
own nickname as the ingress nickname.)

Since RBv is not a physical node and no TRILL frames are forwarded
between its ports via a local MC-LAG, pseudo-node LSP(s) MUST NOT be
created for an RBv. RBv cannot act as root when constructing
distribution trees for multi-cast 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 SHOULD be set to
0, and this nickname SHOULD NOT be listed in the "s" nicknames (see
Section 2.5 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 MC-LAGs, those edge RBridges should be considered as a
single RBv with a pseudo-nickname.

4. Member RBridges Auto-Discovery

Edge RBridges connected by CE(s) via MC-LAG(s) can automatically
discover each other with minimal configuration through exchange of
the MC-LAG(s) information.

From the perspective of edge RBridges, a CE that connects to edge
RBridges via an MC-LAG can be identified by the globally unique ID of
the MC-LAG (i.e., the MC-LAG System ID [802.1AX], also referred to as
MC-LAG ID in this document). On each of such edge RBridges, the
access port to such a CE is associated with an MC-LAG ID for the CE.
An MC-LAG is considered valid on an edge RBridge only if the RBridge
still has operational down-link to that MC-LAG. For such an edge
RBridge, it advertises a list of MC-LAG IDs for all the valid local
MC-LAGs to other edge RBridges via its TRILL IS-IS LSP(s). Based on
the MC-LAG IDs advertised by other edge 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 (i.e., 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 MC-LAG1 and MC-LAG2 respectively; CE3 and CE4
attach to RB3 and RB4 via MC-LAG3 and MC-LAG4 respectively. Assume
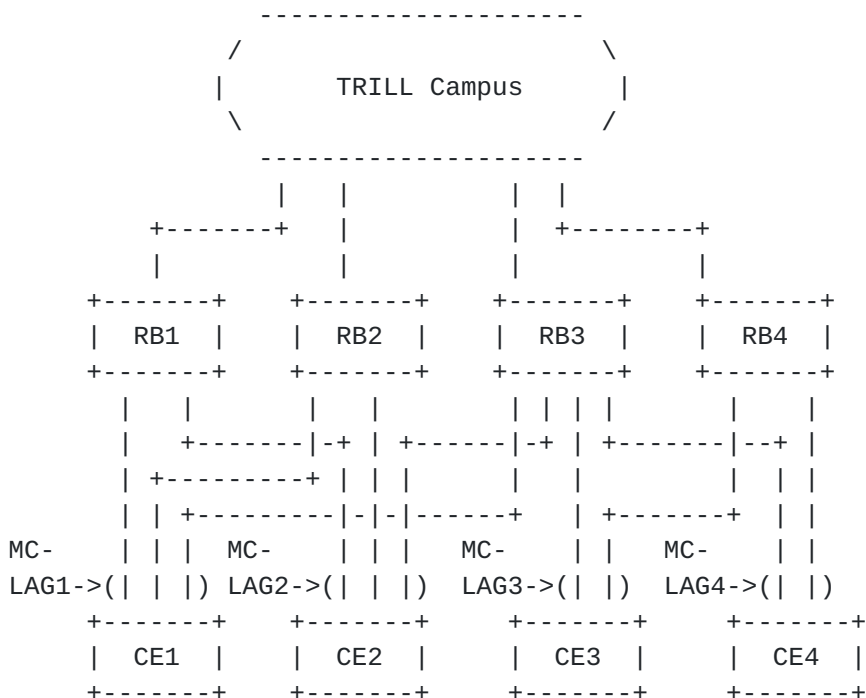MC-LAG3 is configured to occupy a Virtual RBridge by itself.

```
                  ---------------------
                 /                      \
                |      TRILL Campus      |
                 \                      /
                  ---------------------
                   |   |         |   |
           +-------+   |         |   +--------+
           |           |         |            |
       +-------+   +-------+   +-------+   +-------+
       |  RB1  |   |  RB2  |   |  RB3  |   |  RB4  |
       +-------+   +-------+   +-------+   +-------+
         |   |       |   |      | | | |       |   |
         |   +-------|-+ | +------|-+ | +-------|--+ |
         | +---------+ | | |      |   |         |  | |
         | | +---------|-|-|------+   | +-------+  | |
      MC-    | | |  MC-   | | |   MC-    | |  MC-     | |
    LAG1->(| | |) LAG2->(| | |)  LAG3->(| |)  LAG4->(| |)
         +-------+   +-------+      +-------+      +-------+
         |  CE1  |   |  CE2  |      |  CE3  |      |  CE4  |
         +-------+   +-------+      +-------+      +-------+
```

Figure 2   Different MC-LAGs to TRILL Campus

RB1 and RB2 advertise {MC-LAG1, MC-LAG2} in the MC-LAG Membership
sub-TLV (see Section 9.1 for more details) via their TRILL IS-IS LSPs
respectively; RB3 announces {MC-LAG1, MC-LAG2, MC-LAG3, MC-LAG4}; and
RB4 announces {MC-LAG3, MC-LAG4}, respectively.

An edge RBridge is called an MC-LAG related RBridge if it has at
least one MC-LAG configured on an access port. On receipt of the MC-
LAG Membership sub-TLVs, RBn ignores them if it is not an MC-LAG
related RBridge; otherwise, RBn SHOULD use the MC-LAG information
contained in the sub-TLVs, along with its own MC-LAG 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:

```
          MC-LAG ID   OE-flag    Set of edge RBridges
          ---------   --------   --------------------
          MC-LAG1     0          {RB1, RB2, RB3}
          MC-LAG2     0          {RB1, RB2, RB3}
          MC-LAG3     1          {RB3, RB4}
          MC-LAG4     0          {RB3, RB4}
```

Where the OE-flag indicates whether an MC-LAG is willing to share an
RBv with other MC-LAGs if they multiply attach to exact the same set
of edge RBridges as it. For an MC-LAG (for example MC-LAG3), if its
OE-flag is one, it means that MC-LAG3 does not want to share, so it
MUST Occupy an RBv Exclusively (OE).

Otherwise, the MC-LAG (for example MC-LAG1) will share an RBv with
other MC-LAGs if possible. By default, this flag is set zero. For an
MC-LAG, this flag is considered 1 only if any edge RBridge advertises
it as one (see Section 9.1).

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

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

Step 2: Sort the left valid MC-LAGs in the table in descending order
based on the number of RBridges in their associated set of multi-
homed RBridges.

Step 3: Take the valid MC-LAG (say MC-LAG_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 MC-LAGs in the table one by
one, pick up all the valid MC-LAGs that their sets of multi-homed
RBridges contain the same RBridges as that of MC-LAG_i and take them
out of the table. Then appoint RBv_i as the servicing RBv for those
MC-LAGs.

Step 5: Repeat Step 3-4 for the left MC-LAGs until all the valid
entries in the table has be associated with an RBv.

After performing the above steps, all the 4 RBridges know that MC-
LAG3 is served by an RBv, say RBv1, which has RB3 and RB4 as member
RBrdges; MC-LAG1 and MC-LAG2 are served by another RBv, say RBv2,
which has RB1, RB2 and RB3 as member RBridges; and MC-LAG4 is served

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

```
RBv     Serving MC-LAGs        Member RBridges
-----   -------------------    ---------------
RBv1    {MC-LAG3}              {RB3, RB4}
RBv2    {MC-LAG1, MC-LAG2}     {RB1, RB2, RB3}
RBv3    {MC-LAG4}              {RB3, RB4}
```

In each RBv, one of the member RBridges is elected as the DRB
(Designated RBridge) of the RBv. Then this RBridge picks up an
available nickname as the pseudo-nickname for the RBv and announce it
to all other member RBridges of the RBv via its TRILL IS-IS 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 (i.e., RBv), 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 MC-LAG Membership sub-TLVs by other
member RBridges of this RBv (see Section 9.1 for more details), the
vDRB select an available nickname as the pseudo-nickname for this RBv
and advertizes it to the other RBridges via its TRILL IS-IS LSP(s)
(see Section 9.2). 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 [RFC7180]. On receipt of the pseudo-
nickname advertised by the vDRB, all the other RBridges of that group
associate it with the MC-LAGs served by the RBv, and then download
the association to their data plane fast path logic.

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 selection nickname for the RBv. To help the
vDRB to do so, each MC-LAG related RBridge advertises a re-using
pseudo-nickname for each of its MC-LAGs in its MC-LAG Membership sub-
TLV if it has used such one for that MC-LAG recently. Although it is
up to the implementation of the vDRB as to how to treat the re-using
pseudo-nicknames, one suggestion is given as follows:

o If there are more than one available re-using pseudo-nickname that
   are reported by all the member RBridges of some MC-LAGs in this
   RBv, the available one that is reported by most of such MC-LAGs is
   chosen as the pseudo-nickname for this RBv. In the case that 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)
via its TRILL IS-IS LSP(s). After receiving the pseudo-nickname,
other RBridges of that RBv associate the nickname with their ports of
that RBv and download the association to their data plane fast path
logic.

5. **Distribution Trees and Designated Forwarder**

In an AAE group (i.e., an RBv), 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 ingress 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 egress it may cause local
CE(s) receiving duplication frames [AAProb]. 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 of frame for CEi.

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 use distribution trees to forward multi-
destination frames (although under some circumstances they can be
unicast as specified in [RFC7172]). RPF Check along with other
checking is used to avoid temporary multicast loops during topology
changes (Section 4.5.2 of [RFC6325]). RPF check mechanism only allows
a multi-destination frame ingressed by an RBridge RBi and forwarded
on a distribution tree Tx to arrive at another RBridge RBn on an
expected port. If arriving on other ports, 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 multicast
frames. From the view of other RBridges, these frames appear as if
they were ingressed by the RBv. When multicast frames of different
flows are ingressed by different member RBridges of an RBv and
forwarded along same a distribution tree, they may arrive at RBn from
different ports. Some of them will violate the RFC check principle at
RBn and be dropped, which may result in traffic disruption.

In an RBv, if different member RBridge uses different distribution
trees to ingress multi-destination frames, the RFC check violation
issue can be fixed. Coordinated Multicast Trees (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 calculate RFC check information for RBi
on those trees [CMT].

In this document, the approach proposed in [CMT] is used to fix the
RFC 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,
which has RB1 and RB2 as member RBridges. In VLAN x, the three CEs
can communicate with each other.

```
                   ---------------------
                  /                     \
                 |      TRILL Campus     |
                  \                     /
                   ----------------------
                      |           |
                   +----+        +------+
                      |           |
                +---------+          +--------+
                |   RB1   |          |   RB2   |
                | oooooooo|oooooooooooooooooo|ooooo    |
                +o--------+     RBv        +-----o--+
                  o|oooo|ooooooooooooooooooooooo|o|o   |
                   | +--|--------------------+ |    |
                   | |  +---------+ +----------+    |
                   | |  |         | |               |
                  (| |)<-MC-LAG1 (| |)<-MC-LAG2     |
              +-------+        +-------+      +-------+
              | CE1   |        | CE2   |      | CE3   |
              +-------+        +-------+      +-------+
```
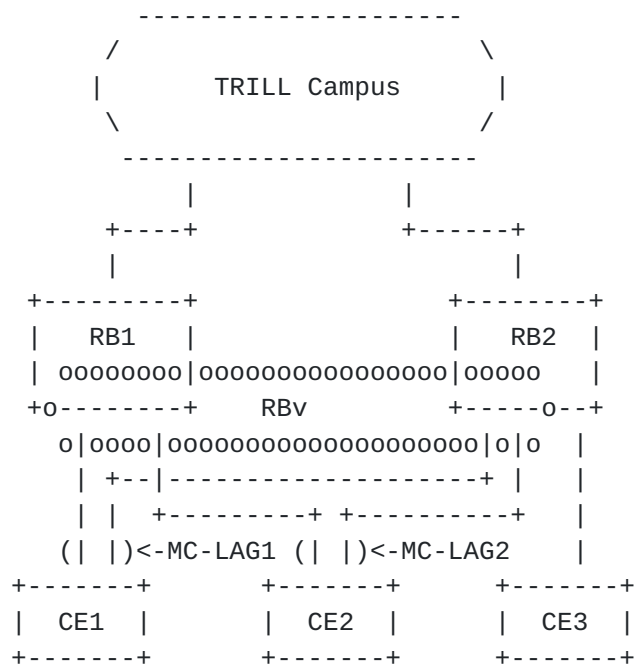
          Figure 3 A Topology with Multi-homed and Single-homed CEs

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 an FGL one), 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 duplication 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, Designated Forwarder (DF) for a VLAN is introduced to avoid the duplicate copies. The basic idea of 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 DF has an effect only on the egressing/replicating of multi-destination traffic, no effect on the ingressing of frames or forwarding/egressing of unicast frames. Furthermore, DF check is performed only for RBv ports, not on regular access ports.

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

Assuming there are m MC-LAGs and k member RBridges in an RBv; each MC-LAG is referred to as MC-LAGi where $0 <= i < m$, and each RBridge is referred to as RBj where $0 <= j < k-1$, DF election algorithm per VLAN is as follows:

Step 1: For MC-LAGi, sort all the RBridges in numerically ascending order based on (System IDj | MC-LAGi) mod k, where "System IDj" is the IS-IS System ID of RBj, "|" means concatenation, and MC-LAGi is the MC-LAG ID for MC-LAGi. In the case that some RBridges get the same result of the mod, these RBridges are sorted in numerically ascending order in the proper places of the result in the list by their System IDs.

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

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

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

For a multi-destination native frame of VLAN x received, if RBi is an MC-LAG attached RBridge, in addition to local replication of the frame to regular access port as per [RFC6325] (and [RFC7172] for FGL), it should also locally replicate the frame to the following RBv ports:

1) 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;

2) 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.

Furthermore, the frame MUST NOT be replicated back to the incoming port. For non-MC-LAG related RBridges or for non-RBv ports on an MC-LAG related RBridge, local replication is performed as per [RFC6325].

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 a multicast traffic in VLAN x to TRILL campus via a member RBridge (say RB1). The traffic is then TRILL-encapsulated by RB1 and delivered through 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 MC-LAG1. Then the traffic loops back to CE1 (see Section 3.2 of [AAProb]).

To fix the above issue, an ingress nickname filtering check is required by this document. The idea of this check is to check the ingress nickname of a multi-destination TRILL Data packet before egress 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 then 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 BUM (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, 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 further 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 MC-LAG 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 a 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 one local RBv, RBn
   is responsible for learning the source MAC address. The learned
   {Inner.MacSA, Data Label, ingress nickname} triplet SHOULD be shared
   within the AAE group (See 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 may 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 (Section 2.4.2.1 of [RFC7180]).
      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 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 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. That
   is to say, 1) if the port is associated with the same pseudo-
   nickname as the ingress nickname of the packet, the packet SHOULD
   be discarded; or 2) if RBn is not the DF for the packet's
   Inner.VLAN (or VLAN the packet's Inner.Label maps to) on the port,
   the packet SHOULD also be discarded; otherwise, it can be egressed
   out of the port.

7. MAC Information Synchronization in Edge Group

An edge RBridge, say RB1 in MC-LAG1, may have learned a MAC address
and 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
any other member RBridge of MC-LAG1, 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 MC-LAG1. RB1 synchronizes
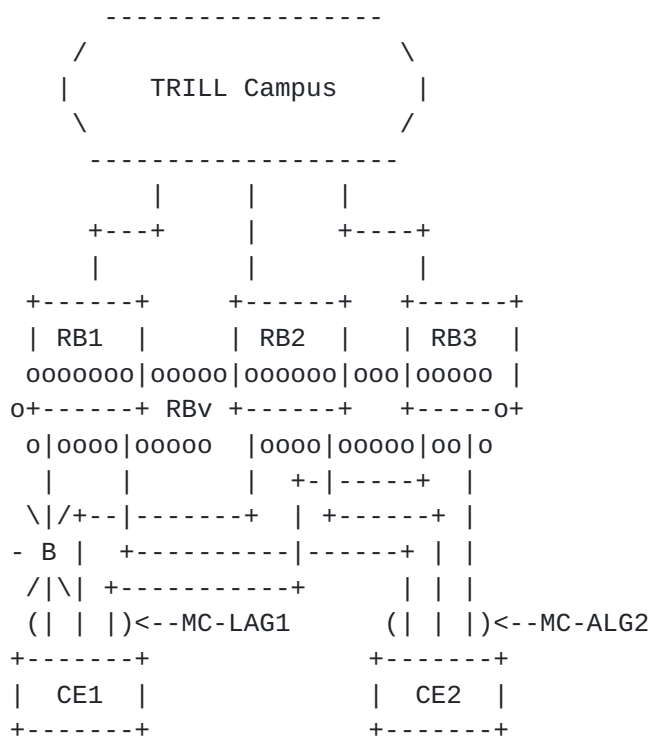the correspondence by using MAC-RI sub-TLV [RFC6165] in its ESADI
LSPs [ESADI].

On the other hand, RB2 has learned the MAC&VLAN of CE1 when CE1 sends

a frame to h1 through RB2. The returning traffic from h1 may go to
RB1. RB1 may have not CE1's MAC&VLAN stored even though it is in the
same MC-LAG for CE1 as RB2. Therefore it has to flood the traffic out
of its all access ports where it is appointed forwarder for the VLAN
(see Section 6.2.1). Such flooding is unnecessary since the returning
traffic is almost always expected and RB2 had learned the CE1's
MAC&VLAN information. To avoid that unnecessary flooding, RB2 SHOULD
share the MAC and VLAN (or MAC and FGL if the egress port is an FGL
port [RFC7172]) with other RBridges of MC-LAG1. RB2 synchronizes the
MAC and Data Label by enclosing the relative MAC-RI TLV with a pair
of boundary TRILL Appsub-TLVs for MC-LAG1 (see Section 9.3) in its
ESADI LSP [ESADI]. After receiving the enclosed MAC-RI TLVs, the
member RBridges of MAC-LAG1 (i.e., MAC-LAG1 related RBridges) treat
the MAC and Data Label as if it learned them locally on its member
port of MC-LAG1; the MC-LAG1 unrelated RBridges just ignore MC-LAG1's
information contained in the boundary sub-TLVs and treat the MAC and
Data Label per [ESADI]. Furthermore, in order to make the the MC-LAG1
unrelated RBridges know that the MAC/Data Label is reachable through
the RBv that provides service to MC-LAG1, 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 MC-LAG1 (see Section 5), the unicast traffic to CE1 might be
still 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, since 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), special
actions to protect against down-link failure for such multi-
desination packets is not needed.

```
                  -----------------
                 /                  \
                 |     TRILL Campus     |
                 \                  /
                  -------------------
                   |     |     |
                  +---+   |    +----+
                  |       |        |
           +------+    +------+   +------+
           | RB1  |    | RB2  |   | RB3  |
           ooooooo|ooooo|oooooo|ooo|ooooo |
           o+------+ RBv +------+   +-----o+
            o|oooo|ooooo  |oooo|ooooo|oo|o
             |    |       |  +-|-----+  |
             \|/+--|-------+  | +------+ |
             - B |   +----------|------+ | |
             /|\| +-----------+     | | |
             (| | |)<--MC-LAG1      (| | |)<--MC-ALG2
             +-------+              +-------+
             | 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 MC-LAG1 (for example RB2, since it is still
a member RBridge of MC-LAG1) 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 MC-LAG1 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

[9.1](). **MC-LAG Membership (LM) Sub-TLV**

This TLV is used by edge RBridge to announce its associated MC-LAG
information. It is defined as a sub-TLV of the Router Capability TLV
(#242) and the Multi-Topology-Aware Capability (MT-CAP) TLV (#144).
It has the following format:

```
    +-+-+-+-+-+-+-+-+
    |  Type= LM     |  (1 byte)
    +-+-+-+-+-+-+-+-+
    |  Length       |  (1 byte)
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+
    |  MC-LAG RECORD(1)                        |  (11 bytes)
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+
    .                                          .
    .                                          .
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+
    |  MC-LAG RECORD(n)                        |  (11 bytes)
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+
```

Figure 5 MC-LAG Membership Advertisement Sub-TLV

where each MC-LAG record has the following form:

```
    +--+-+-+-+-+-+-+-+
    |OE|    RESV    |                (1 byte)
    +--+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |  Re-using Pseudo-nickname      |  (2 bytes)
    +--+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+
    |  MC-LAG System ID                        |  (8 bytes)
    +--+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+
```

o LM (1 byte): Defines the type of this sub-TLV, #TBD.

o Length (1 byte): 11*n bytes, where there are n MC-LAG Records.

o OE (1 bit): an flag indicating whether or not the MC-LAG 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 MC-LAG, the values of this flag might conflict
   in the LSPs advertised by different member RBridges of that MC-
   LAG. In that case, the flag for that MC-LAG is considered as 1.

o RESV (7 bits): Transmitted as zero and ignored on receipt.

o Re-using Pseudo-nickname (2 bytes): In an MC-LAG record, it
   suggests the pseudo-nickname of the AAE group serving the MC-LAG.

If the MC-LAG 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 pseudo-nickname of an AAE group (see Section 4.2).

o MC-LAG System ID (8 bytes): The System ID of the MC-LAG as
   specified in Section 5.3.2 in [802.1AX].

On receipt of such a sub-TLV, if RBn is not an MC-LAG related edge
RBridge, it ignores the sub-TLV; otherwise, it parses the sub-TLV.
When new MC-LAGs 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" approach described in
Section 4.1.

9.2. PN-RBV sub-TLV

PN-RBv sub-TLV is used by a Designated RBridge of a Virtual RBridge
(vDRB) to appoint Pseudo-nickname for the MC-LAGs served by the RBv.
It is defined as a sub-TLV the Router Capability TLV (#242) and the
Multi-Topology-Aware Capability (MT-CAP) TLV (#144). It has the
following format:

```
    +-+-+-+-+-+-+-+-+
    | Type= PN_RBv  |  (1 byte)
    +-+-+-+-+-+-+-+-+
    | Length        |  (1 byte)
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |   RBv's Pseudo-Nickname      |  (2 bytes)
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+
    |   MC-LAG System ID (1)                      |  (8 bytes)
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+
    .                                            .
    .                                            .
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+
    |   MC-LAG System ID (n)                      |  (8 bytes)
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+
```

o PN_RBv (1 byte): Defines the type of this sub-TLV, #TBD.

o Length (1 byte): 2+8*n bytes, where there are n MC-LAG System IDs.

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

o MC-LAG System ID (8 bytes): The System ID of the MC-LAG as
   specified in Section 5.3.2 in [802.1AX].

On receipt of such a sub-TLV, if RBn is not an MC-LAG related edge
RBridge, it ignores the sub-TLV. Otherwise, if RBn is also a member
RBridge of the RBv identified by the list of MC-LAGs, it associates
the pseudo-nickname with the ports of these MC-LAGs and downloads the
association onto data plane fast path logic.

9.3. MAC-RI-MC-LAG Boundary sub-TLVs

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

```
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |Type =MAC-RI-MC-LAG-INFO-START | (2 byte)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | Length                        | (2 byte)
   +-+-+-+-+-+-+-+-+-+-+-+-+-...+-+-+-+-+-+-+
   | MC-LAG System ID                       | (8 bytes)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-...+-+-+-+-+-+-+
```

o MAC-RI-MC-LAG-INFO-START (1 byte): Defines the type of this sub-
   TLV, #TBD.

o Length (1 byte): 8.

o MC-LAG System ID (8 bytes): The System ID of the MC-LAG as
   specified in Section 5.3.2 in [802.1AX]. This ID identifies the
   MC-LAG for all MAC addresses contained in following MAC-RI TLVs
   until an MAC-RI-MC-LAG-INFO-END sub-TLV is encountered.

MAC-RI-MC-LAG-INFO-END sub-TLV is defined as follows:

```
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   |Type = MAC-RI-MC-LAG-INFO-END  | (2 byte)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   | Length                        | (2 byte)
   +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

o MAC-RI-MC-LAG-INFO-END (1 byte): Defines the type of this sub-TLV,
   #TBD.

o Length (1 byte): 0.

This pair of sub-TLVs can be carried multiple times in a message and
in multiple messages. When an MC-LAG related edge RBridge (say RBn)

wants to share with other edge RBridges the MAC addresses learned on
its local ports of different MC-LAGs, it uses one or more pairs of
such sub-TLVs for each of such MC-LAGs in its ESADI LSPs. Each
encloses the MAC-RI TLVs containing the MAC addresses learned from
the MC-LAG. Furthermore, if the MC-LAG 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 sub-TLVs, when the edge RBridge is not an
MC-LAG related one or cannot recognize such sub-TLVs, it ignores them
and continues to parse the enclosed MAC-RI TLVs per [ESADI].
Otherwise, the recipient parses the boundary sub-TLVs, and

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

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

10. OAM Frames

Attention must be paid when generating the OAM frames. To ensure the
response messages can return to the originating member RBridge of an
RBv, pseudo-nickname cannot be used as ingress nickname in TRILL OAM
messages, except that 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 ingress nickname in the response message.

Since OAM messages cannot be used by RBridges 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

It is important that the VLAN membership of all the RBridge ports in
an MC-LAG 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.

Furthermore, it is important that if any VLAN in an MC-LAG is being
mapped by edge RBridges to an FGL [RFC7172], that the mapping MUST be
same for all edge RBridge ports in the MC-LAG. 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.

12. Security Considerations

This draft does not introduce any extra security risks. For general
TRILL Security Considerations, see [RFC6325]. For ESADI Security
Considerations, see [ESADI].

13. IANA Considerations

IANA is requested to allocate code points for the 4 sub-TLVs defined
in Section 9.

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

16. References

16.1. Normative References

[CMT]   T. Senevirathne, J. Pathangi, and J. Hudson, "Coordinated
            Multicast Trees (CMT) for TRILL", draft-ietf-trill-cmt-
            01.txt Work in Progress, April 2014.

[ESADI]   H. Zhai, F. Hu, R. Perlman, D. Eastlake, "TRILL
            (Transparent Interconnection of Lots of Links): The ESADI
            (End Station Address Distribution Information) Protocol",
            draft-ietf-trill-esadi-09, June 2014.

[RFC1195]   R. Callon, "Use of OSI IS-IS for routing in TCP/IP and
            dual environments", RFC 1195, December 1990.

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

[RFC6325]   R. Perlman, D. Eastlake, D. Dutt, S. Gai, and A.
            Ghanwani, "Routing Bridges (RBridges): Base Protocol
            Specification", RFC 6325, July 2011. [RFC6165] Banerjee,
            A. and D. Ward, "Extensions to IS-IS for Layer-2 Systems",
            RFC 6165, April 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", RFC7176, May 2014.

[RFC7180]   D. Eastlake, M. Zhang, A. Ghanwani, V. Manral and A.
            Banerjee, "Transparent Interconnection of Lots of Links
            (TRILL): Clarifications, Corrections, and Updates",
            RFC7180, May 2014.

   [802.1AX] IEEE, "IEEE Standard for Local and Metropolitan Area/
              networks Link Aggregation", 802.1AX-2008, 1 January 2008.

16.2. **Informative References**

   [AAProb]   Y. Li, W. Hao, R. Perlman, J. Hudson and H. Zhai, "Problem
              Statement and Goals for Active-Active TRILL Edge", draft-
              ietf-trill-active-active-connection-prob-04, June 2014.

Authors' Addresses

   Hongjun Zhai
   ZTE Corporation
   68 Zijinghua Road, Yuhuatai District
   Nanjing, Jiangsu 210012
   China

   Phone: +86 25 52877345
   Email: zhai.hongjun@zte.com.cn

   Tissa Senevirathne
   Cisco Systems
   375 East Tasman Drive
   San Jose, CA 95134
   USA

   Phone: +1-408-853-2291
   Email: tsenevir@cisco.com

   Radia Perlman
   Intel Labs
   2200 Mission College Blvd
   Santa Clara, CA 95054-1549
   USA

   Phone: +1-408-765-8080
   Email: Radia@alum.mit.edu

   Donald Eastlake 3rd
   Huawei Technologies
   155 Beaver Street
   Milford, MA 01757
   USA

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

      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