

TRILL Working Group
Internet-Draft
Intended status: Standards Track
Expires: August 18, 2014

H. Zhai
ZTE Corporation
T. Senevirathne
Cisco Systems
R. Perlman
Intel Labs
D. Eastlake 3rd
M. Zhang
Huawei Technologies
February 14, 2014

RBridge: Pseudo-Nickname
draft-hu-trill-pseudonode-nickname-06

Abstract

The IETF TRILL (Transparent Interconnection of Lots of Links) protocol provides loop free connectivity to Local Area Network (LAN) via choice of an Appoint Forwarder (AF) for a set of VLANs and to end node via point-to-point link. The active-active access, as an extension access of legacy network or end-device to TRILL campus has not been specified. This document gives the concept of Virtual RBridge(RBv), and based on it, specifies the active-active access of a legacy network or an end node via a switch that multi-homes to TRILL campus.

Status of This Memo

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

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

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

This Internet-Draft will expire on August 18, 2014.

Copyright Notice

Copyright (c) 2014 IETF Trust and the persons identified as the document authors. All rights reserved.

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

Table of Contents

1.	Introduction	3
1.1.	Terminology and Acronyms	4
1.2.	Contributors	4
2.	Overview	4
3.	Concept of Virtual RBridge and Pseudo-nickname	5
3.1.	VLAN-x Appointed Forwarder for Member Interfaces in RBv .	6
3.2.	Announcing Pseudo-Nickname of RBv	6
4.	Acquisition of RBv's Pseudo-nickname	7
4.1.	Picking up RBridges for Different RBvs	7
5.	Distribution Trees for Member RBridges in RBv	9
6.	Frame Processing	10
6.1.	Native Frames Ingressing	10
6.2.	TRILL Data Frames Egressing	11
6.2.1.	Unicast TRILL Data Frames	11
6.2.2.	Multi-Destination TRILL Data Frames	12
7.	Member Link Failure in RBv	12
7.1.	Link Protection for Unicast Frame Egressing	13
8.	TLV Extensions for RBv	14
8.1.	LAG Membership (LM) Sub-TLV	14
8.2.	PN-RBv sub-TLV	16
9.	OAM Frames	16
10.	Configuration Consistency	16
11.	IANA Considerations	17
12.	Security Considerations	17
13.	Acknowledgements	17
14.	Normative References	17
Appendix A.	Rationale for MAC Sharing among Member RBridges . .	18
	Authors' Addresses	19

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 [[RFC1195](#)] link state routing and encapsulating traffic using a header that includes a hop count. The design supports VLANs and optimization of the distribution of multi-destination frames based on VLANs and IP derived multicast groups. Devices that implement TRILL are called RBridges or TRILL Switches.

In the current TRILL protocol, an end node can be attached to 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, only one of the RBridges is permitted to provide forwarding service for end station traffic in each VLAN (Virtual LAN). That RBridge is referred as Appointed Forwarder (AF) for that VLAN on the link [6325][6439]. However, in some practical deployments, to increase the access bandwidth and reliability, an end node might be multi-homed to several edge RBridges and treat all of the uplinks as a single Multi-Chassis Link Aggregation (MC-LAG) bundle. In this case, it's required that data traffic within a specific VLAN from this end node can be ingressed and egressed by any of these RBridges simultaneously. These RBridges constitute an Active-Active Edge (AAE) RBridge Group.

Since a packet from each end node can be ingressed by any RBridge in the AAE group, a remote RBridge may observe multiple attachment points (i.e., egress RBridges) for this endnode which is identified by its MAC address. This issue is known as the "MAC flip-flopping" issue. In addition to this issue, other issues such as echo of multi-destination frames originated from the MC-LAG or duplication egressing of multi-destination from campus might be encountered by the AAE group (see [Section 5](#) for more details).

In this document, the concept of a Virtual RBridge(RBv) group, together with its Pseudo-nickname, is introduced to address the AAE issues in the scope of TRILL. For a member RBridge of such a group, it uses the pseudo-nickname of the RBv, instead of its own nickname, as the ingress RBridge nickname when ingressing frames into TRILL campus. So, in such a AAE Group, even if there are multiple RBridges providing frame forwarding service for an end node simultaneously, the ingress RBridge nickname associated to that end node's MAC address(s) still remains unchanged in remote RBridges' forwarding tables.

This document is organized as follows. [Section 2](#) is problem statement, which describes why virtual RBridge and its pseudo-nickname are required. [Section 3](#) gives the concept of virtual RBridge. [Section 4](#) discusses how a RBv gets its pseudo-nickname. [Section 5](#) describes the consideration for pseudo-nickname used in ingress multi-destination frames. [Section 6](#) covers processing of data frame traffic when considering pseudo-nickname. [Section 7](#) discusses the processing of MC_LAG links failure. And [Section 8](#) gives TLV extensions needed by RBv.

Familiarity with [\[RFC6325\]](#) is assumed in this document.

[1.1.](#) Terminology and Acronyms

This document uses the acronyms defined in [\[RFC6325\]](#) and the following additional acronym:

AF - Appointed Forwarder

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 [\[RFC2119\]](#).

When used in lower case, these words convey their typical use in common language, and are not to be interpreted as described in [\[RFC2119\]](#).

[1.2.](#) Contributors

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, and Jon Hudson for their good questions and comments.

[2.](#) Overview

In order to improve the reliability of connection to a TRILL network, multi-homing technique may be employed by a legacy device which can be a switch or end host. Take Figure 1 as an example, switch SW1 multi-homed to the TRILL network by connecting to RB1 and RB2 with respective links. Then the end station S1 can continue to get frame forwarding service from the TRILL network even if one of its up-links (e.g., SW1-RB1) fails.

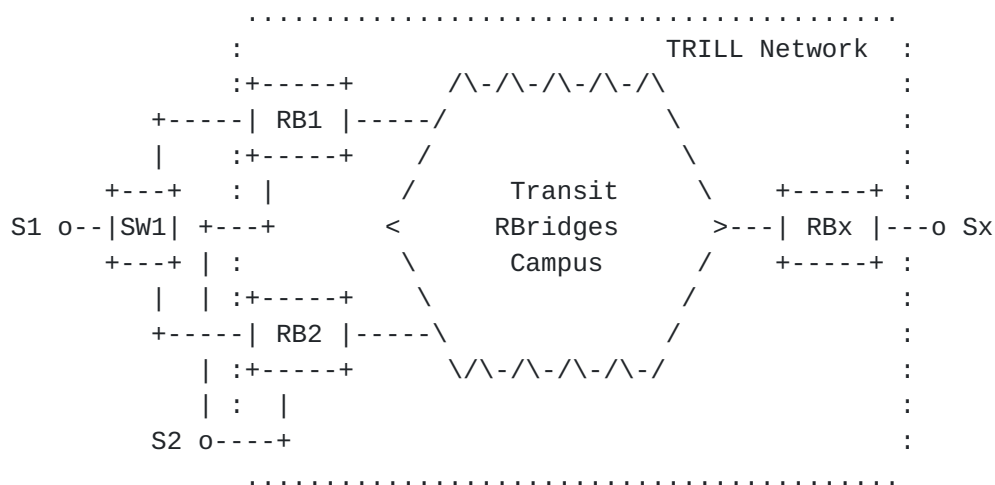


Figure 1 Multi-homed to TRILL Network

SW1 may treat the two links as a link bundle, so that the two links form active-active load sharing model instead of the previous active-standby model. That is to say, in Figure 1, two R Bridges (i.e., RB1 and RB2) provides frame forwarding service to S1 simultaneously in a VLAN. As stated previously, simultaneous frame forwarding may result in frame duplication, loops and the flip-flopping of the ingress RBridge associated to the MAC of S1 in remote R Bridges' (e.g., RBx) forwarding tables. The flip-flopping in turn causes packet disorder in reverse traffic and worsens the traffic disruption. Therefore, the concept of Virtual RBridge, together with its nickname, is introduced in the following section to fix these issues.

3. Concept of Virtual RBridge and Pseudo-nickname

A Virtual RBridge (RBv) represents a group of different end station service ports on different edge R Bridges. After joining RBv, such a RBridge port is called a member port of RBv, and such a RBridge becomes a member RBridge of RBv. An RBv is identified by its virtual nickname in TRILL campus, and this nickname is also referred to as pseudo-nickname in this document.

After joining a RBv, a member RBridge will announce its connection to RBv by including the information of that RBv, e.g., the pseudo-nickname of RBv, in its self-originated LSP. From such LSPs, other R Bridges that are not members of the RBv believe those member R Bridges are connected to RBv.

When a native frame from an end station S1 is received from such a port, the member RBridge encapsulates the frame with the RBv's nickname, instead of its own nickname, as the ingress nickname. When

the destination RBridges receive and de-capsulate this frame, they will learn that S1 is reachable through RBv.

For a member RBridge, it MUST move out of a RBv and clear the RBv's information from its self-originated LSPs when it loses all of its member ports of the RBv, due to port failure, configuration, etc.

NOTE1: In the multi-homing scenario of same a RBv, it is RECOMMENDED that all devices multi-homed to that RBv SHOULD have operational links to all the member RBridges of that RBv unless one or more of the links failed or administratively down.

3.1. VLAN-x Appointed Forwarder for Member Interfaces in RBv

Since member RBridges in RBv cannot see each others' Hellos on their member ports in the multi-homing scenario, then each RBridge becomes Designated RBridge (DRB) for that port and appoints itself as AF for all VLANs.

This extended AF framework allows:

- o Detection and protection against mis-configuration at the edge, e.g., on the device SW1 the two interfaces are not configured as multi-homing then RB1 and RB2 work in an unexpected active-standby mode rather than expected active-active mode for S1 or
- o Avoidance of loops in the event that S1 and S2 were connected by a native Ethernet Link. In this event, RB1's Hellos originated on link RB1-SW1 will be forwarded by S1 through the Ethernet Link to S2 then received by RB2, and vice versa. Therefore, RB1 and RB2 work in an active-standby mode for S1 (or S2) in any VLAN to avoid potential forwarding loops.

3.2. Announcing Pseudo-Nickname of RBv

Each member RBridge advertises the RBv's pseudo-nickname using the nickname sub-TLV [[rfc6326bis](#)], along with its regular nickname(s), in its LSPs. For a member RBridge, when its last member port is disconnected to RBv, it MUST leave from RBv and clear RBv's pseudo-nickname from its update LSPs.

RBv's pseudo-nickname is ignored when determining the distribution tree root for the campus. The tree root priority of RBv's nickname SHOULD be set to 0, and this nickname SHOULD NOT be listed in the "s" nicknames by the RBridge holding the highest priority tree root nickname.

4. Acquisition of RBv's Pseudo-nickname

In active-active connection scenario, a device is typically connected to multiple edge RBridges via a link bundle. From the perspective of the edge RBridges, the device can be identified by a globally unique identifier; and this identifier is called Link Aggregation Group Identifier (LAG-ID) in this document.

For an edge RBridge, if it has one or more operational ports through which a device multi-homed to it, it MUST announce that LAG-ID of the device to all other edge RBridges via RBridge Channel messages [RBChannel]. Based on the LAG-IDs received from other edge RBridges, edge RBridges can pick up, from TRILL campus, all the edge RBridges that can join same a RBv (See [Section 4.1](#) for more details) and elect one of them as the Designated RBridge (DRB) for that RBv. That DRB is responsible for appointing an available pseudo-nickname for that RBv.

4.1. Picking up RBridges for Different RBvs

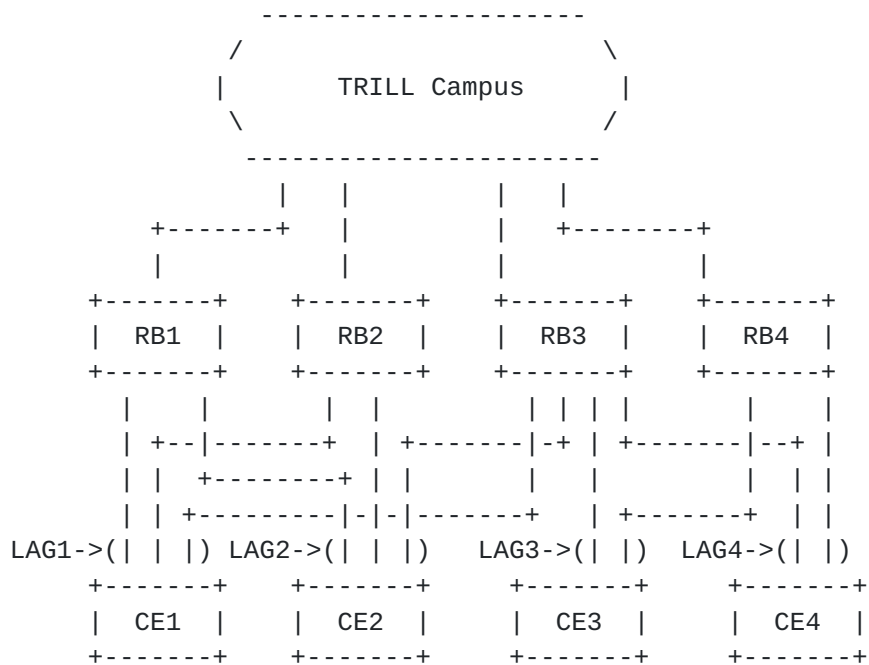


Figure 2 Different LAGs to TRILL Campus

For each edge RBridge with available multi-homed devices connected, it MUST announce a list of LAG-IDs of all of those devices to all other edge RBridges via RBridge Channel message (See [Section 8.1](#) for more details). Take Figure 2 as an example, RB1 and RB2 announce {LAG1, LAG2} in their lists respectively; RB3 announces {LAG1, LAG2, LAG3, LAG4}; and RB4 announces {LAG3, LAG4}, respectively.

Based on the LAG-IDs contained in these lists, each RBridge can know which set of RBridges each LAG is multi-homed to. For example, all the 4 RBridges know the information as follows:

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

In the above table, there might be some LAGs that multi-homes only to one single RBridge due to mis-configuration or link failure, etc. Those LAGs are considered as invalid entries. Then each of the relative edge RBridges performs the following approach to pick up which valid LAGs can be served by same a RBv.

Step 1: Take all the valid LAGs that have their OE-flags (Occupying Exclusively a RBv) set 1 out of the table and create a RBv per such LAG.

Step 2: Sort the left valid 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 LAG (say 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 remainder valid LAGs in the table one by one, pick up all the valid LAGs that their sets of multi-homed RBridges contain the same RBridges as that of LAG_i and take the LAGs out of the table. Then appoint RBv_i as those LAGs' servicing RBv.

Step 5: Repeat Step 3-4 for the left LAGs in the table.

For the example given in Figure 2, after performing the above steps, all the 4 RBridges know that LAG3 is served by a RBv, say RBv1, which has RB3 and RB4 as member RBridges; LAG1 and LAG2 are served by another RBv, say RBv2, which has RB1, RB2 and RB3 as member RBridges; and LAG4 is served by RBv3, which has RB3 and RB4 as member RBridges, shown as follows:

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

In each RBv, one of its member RBridges is elected as DRB. The winner is the member RBridge with the maximum device nickname in this RBv. Then this DRB picks up an available nickname as this RBv's pseudo-nickname and announce it to all other member RBridges in this RBv via RBridge Channel message (Refer [Section 8.3](#) for more details).

If possible, the DRB SHOULD attempt to reuse the RBv's previous pseudo-nickname to avoid traffic disruption caused by pseudo-nickname changing. If there is no such a previous nickname available, the DRB will acquire a new available nickname from TRILL campus and announce it as the RBv's pseudo-nickname among the RBv's member RBridges.

5. Distribution Trees for Member RBridges in RBv

In TRILL, RBridges use distribution trees to forward multi-destination frames. In the TRILL header of the multi-destination frames, the ingress nickname identifies the ingress RBridge and the egress nickname specifies the root of the chosen distribution tree. After receiving a multi-destination TRILL data frame, RBn performs Reverse Path Forwarding (RPF) check on the multi-destination frame to avoid temporary multicast loops during topology changes.

RPF specifies that a multi-destination TRILL data frame ingressed by an RBridge and forwarded on a distribution tree can only be received by RBn on an expected port. If the frame is not received from that port, it MUST be dropped.

However, member RBridges use RBv's pseudo-nickname other than their own nicknames as the ingress nickname when they forward unicast or non-unicast native frames. Therefore, when these TRILL data frames arrive at RBn, they will be treated as frames ingressed by the same RBridge, i.e., RBv. If they are multi-destination frames and the same distribution tree is chosen by different member RBridges to forward these frames, they may travel on the tree and arrive at RBn on different ports. Then the RPF check is violated, and some of the frames reaching the RBridge on unexpected ports will be dropped by RBn.

[CMT] proposes to assign different distribution trees for each member RBridge to fix the above RPF check issue, and makes use of the Affinity sub-TLV defined in [[rfc6326bis](#)] to achieve this kind of assignment. When the distribution tree Tx is assigned to member RBridge RBn in a RBv, RBn is called Tx's Designated Forwarder and Tx is called the Assigned Tree of RBn in the scope of this RBv.

This document assumes the approach proposed in [[CMT](#)] is supported by member RBridges of RBv.

To avoid duplication traffic being egressed through RBv to a multi-homing end-device, multi-destination TRILL traffic arriving at member RBridges of RBv on a tree (say Tx), only the Tx's Designated Forwarder is allowed to egress it to the device.

When RBn receives a native frame on its member ports of RBv, and decides to ingress it as a multi-destination frame (for example, the native frame is a broadcast frame or the destination is unknown), RBn can only choose one of its Assigned Trees to distribute the TRILL-encapsulated frame.

Since different member RBridge has different Assigned Trees and acts as Designated Forwarder on different trees in the scope of its RBv, the multi-destination frames ingressed from a MC_LAG by one member RBridge will not be egressed back to the MC_LAG by another member RBridges in the scope. That is to say, no echo of multi-destination traffic occurs in RBv.

When a member RBridge joins in or leaves from a virtual RBridge group, the assignment of distribution trees may change. That change has been discussed in [\[CMT\]](#) and is beyond the scope of this document.

6. Frame Processing

Although, there are five types of Layer 2 frames in [\[RFC6325\]](#), e.g., native frame, TRILL data frame, TRILL control frames, etc., pseudo-nickname of RBv is only used for native frame and TRILL data frame in this specification.

6.1. Native Frames Ingressing

When RB1 receives a native frame on one of its valid member ports of RBv, it uses the pseudo-nickname of RBv, instead of its own nickname, as ingress nickname, if it is the appointed forwarder for the VLAN of the frame on the port. If the frame is not received on a member port, RB1 MUST NOT use RBv's pseudo-nickname as ingress nickname when doing TRILL-encapsulation on the frame. Otherwise, the reverse traffic may be forwarded to another member RBridge that does not connect to the link containing the destination, which may cause the traffic disruption.

If the above native frame is ingressed by RB1 as a multi-destination TRILL data frame, e.g., its destination is unknown to RB1 or it is non-unicast frame, RB1 can only choose one of its assigned distribution trees in the RBv to distribute the TRILL-encapsulated frame [\[CMT\]](#). If not so, the multi-destination TRILL data frame will fail RPF check on another RBridge and be dropped.

Furthermore, for such a frame, its source MAC address information ({ VLAN, Outer.MacSA, port }) is learned by default if its source address is unicast. Then the learned information is shared with other member RBridges of the RBv (See [Appendix A](#) for more details for the information sharing).

6.2. TRILL Data Frames Egressing

6.2.1. Unicast TRILL Data Frames

When receiving a unicast TRILL data frame, RBn checks the egress nickname in the TRILL header of the frame. If the egress nickname is one of RBn's own nicknames, the frame is processed as defined in in [\[RFC6325\]](#).

If the egress nickname is RBv's pseudo-nickname and RBn is a member RBridge of the RBv, RBn is responsible to learn the source MAC address. If the learned { Inner.MacSA, Inner.VLAN ID, ingress nickname } triplet is a new one or it updates a previously learned one, this triplet SHOULD be shared with other member RBridges of the RBv (See [Appendix A](#) for more details for the triplet sharing).

Then the frame is de-capsulated to its native form. The Inner.MacDA and Inner.VLAN ID are looked up in RBn's local forwarding address cache, and one of the three following cases occurs:

- o If the destination end station identified by the Inner.MacDA and Inner.VLAN ID is on a local link to RBv, this frame is egressed onto that link.
- o Else if RBn can reach the destination through another member RBridge RBk, it tunnels the frame to RBk [\[ClearCorrect\]](#) by re-encapsulating the native frame into a unicast TRILL data frame. RBn uses RBk's own nickname, instead of RBv's pseudo-nickname as the egress nickname for the re-encapsulation, and remains the ingress nickname remains unchanged. If the hop count value of the frame is too small for the frame to reach RBk safely, RBn SHOULD increase that value properly in doing the re-encapsulation. [NOTE: When receiving that re-encapsulated TRILL frame, as the egress nickname of the frame is RBk's own nickname rather than the RBv's pseudo-nickname, RBk will process it as [Section 4.6.2.4 in \[RFC6325\]](#), 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 its member ports of RBv on which it is appointed forwarder for the Inner.VLAN.

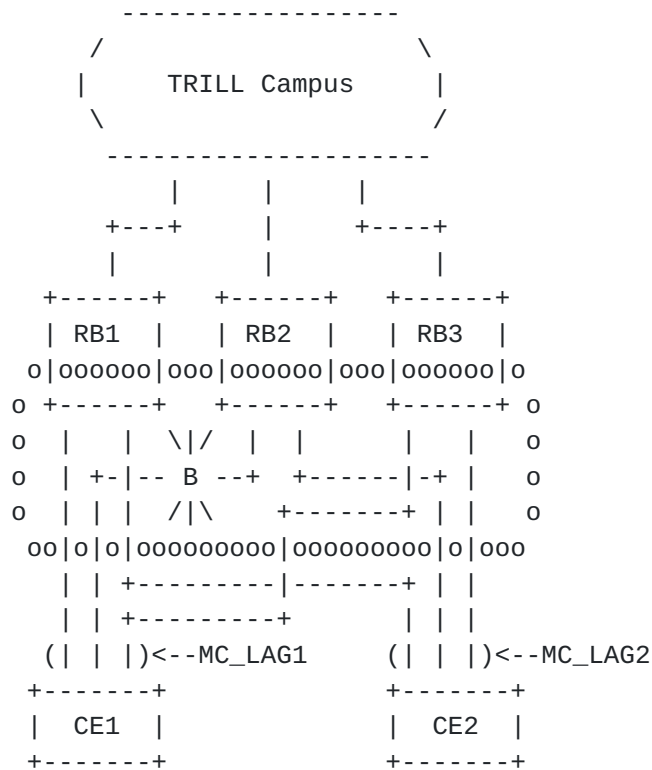
6.2.2. Multi-Destination TRILL Data Frames

To avoid multiple copies of a multi-destination TRILL data frames in VLAN x are egressed to same a MC_LAG, when receiving such a frame on tree Tx by multiple member RBridges of this MC_LAG, only the Designated Forwarder of Tx in the MC_LAG is permitted to egress the frame to this MC_LAG.

That is to say, if RBn is the the Designated Forwarder of Tx in MC_LAGi, a copy of the frame is de-capsulated by RBn into its native form, and sent to MC_LAGi. The { Inner.MacSA, Inner.VLAN ID, ingress nickname } triplet is also learned based on the de-capsulation. If the learned triplet is a new one or updates the previously learned one, it SHOULD be shared among the members RBridges of the virtual RBridge group (See [Appendix A](#) for more details for the triplet sharing).

7. Member Link Failure in RBv

As shown in Figure 3, suppose the link RB2-CE1 fails. Both unicast frames and multi-destination frames cannot be sent from RB2 to CE1. [Section 7.1](#) discusses the failure protection for unicast frames egressing.



B - Failed Link or Link bundle

Figure 3 Member Link Failure in LAG1

7.1. Link Protection for Unicast Frame Egressing

When the link CE1-RB2 fails, RB2 loses its direct connection to CE1. The MAC entry through the failed link to CE1 is removed from RB2's local forwarding table immediately. Another MAC entry through another member RBridge (say RB1) that has local link to CE1 is installed into RB2's forwarding table only if RB2 is still a member RBridge of RBv. Then when the TRILL data frames to CE1 are delivered to RB2, they can be re-encapsulated (ingress nickname remains unchanged and egress nickname is replaced with RB1's nickname) by RB2 and forwarded based on the above installed MAC entry. The member RBridge who receives the redirected frames will egress them to CE1.

When the failure recovers, RB2 will be aware that it can reach CE1 by observing CE1's native frames. Then RB2 installs the MAC entry for link CE1-RB2.

8. TLV Extensions for RBv

8.1. LAG Membership (LM) Sub-TLV

We propose to use LM sub-TLV to advertise the state of the RBridges' MC_LAG membership. There are following 3 different events, as follows:

- o Membership Add
- o Membership Withdrawal
- o Membership Refresh

```

+---+---+---+---+---+
| Type= LM          |          (1 byte)
+---+---+---+---+---+
|   Length          |          (1 byte)
+---+---+---+---+---+---+---+---+---+---+---+---+
|   RBv Nickname    |          (2 bytes)
+---+---+---+---+---+---+---+---+---+---+---+---+
| RESV              |OC |          (1 byte)
+---+---+---+---+---+---+---+---+---+---+---+---+
|                   LAG-ID (1)    | (10 bytes)
+---+---+---+---+---+---+---+---+---+---+---+---+
.
.
+---+---+---+---+---+---+---+---+---+---+---+---+
|                   LAG-ID (n)    |
+---+---+---+---+---+---+---+---+---+---+---+---+

```

Figure 4 Edge Membership advertisement sub-TLV

where each LAG_ID record is of the following form:

```

+---+---+---+---+---+
|   RESV            |OE|          (1 byte)
+---+---+---+---+---+---+---+---+---+---+---+---+
| Virtual Network ID(VNID)        | (3 bytes)
+---+---+---+---+---+---+---+---+---+---+---+---+
|                   LAG ID        | (6 bytes)
+---+---+---+---+---+---+---+---+---+---+---+---+

```

- o LM (1 byte): Defines the type of Edge Membership sub-TLV.
- o Length (1 byte): Defines the length of this sub-TLV which should be greater than 3.

- o RBV Nickname (2 bytes): 2 byte nickname of the RBv. By default, this field is zero. Otherwise, it indicates the pseudo-nickname that the originator of the TLV considers the RBv has used, which providing information for the DRB to reuse the RBv's previous pseudo-nickname.
- o RESV (6 bits): Transmitted as zero and ignored on receipt.
- o OE(1 bit): an flag indicating whether the end-device identified by the combination of the VNID and LAG-ID needs to Occupy a RBv exclusively or can share a RBv with other end-devices; 1 for occupying exclusively, and 0 for sharing. By default, it is set to 0.
- o OC (2 bits): Define the operation code.
 - * 00: Add (LAG-IDs in this sub-TLV are new and will trigger the process of picking up member RBridge for a RBv and the Designated Forwarder election on the relative edge RBridges).
 - * 01: Withdrawal (LAG-IDs in this sub-TLV do not have an active links from the announcing RBridge for RBv, the process of picking up member RBridge for a RBv and Designated Forwarder election MUST be triggered on the relative edge RBridges).
 - * 10: Refresh (LAG-IDs in this sub-TLV are being refreshed and no state change from the perspective of the announcing RBridge).
 - * 11: Reserved and currently unused.
- o VNID(24 bits): an identifier of an Virtual Network where the end-device populated. By default, this field is set to zero.
- o LAG-ID (2 bytes): an unsigned positive integer that uniquely identifies an end device multi-homed to the RBv. This ID along with the VNIT is globally meaningful in the scope of the TRILL campus. For convenience, this ID can be one of the MAC addresses of the end-device..

When receiving such a sub-TLV, if the RBridge has no membership for the listed LAGs in the RBv, it ignores the sub-TLV. If it has the membership, receiving such a sub-TLV where the operation code is 00 or 01 will triggers it to re-calculate the Designated Forwarder on each tree for the listed LAGs.

8.2. PN-RBv sub-TLV

The DRB employs PN-RBv sub-TLV to announce the RBv's pseudo-nickname, along with all the LAGs serviced by this RBv, to other relative edge RBridges.

The format of this sub-TLV is as follows, where the LAG-ID Record has the same format as the Record of LM Sub-TLV.

```

+---+---+---+---+
| Type= PN_RBv | (1 byte)
+---+---+---+---+
| Length | (1 byte)
+---+---+---+---+---+---+---+---+---+
| RBv's Pseudo-Nickname | (2 bytes)
+---+---+---+---+---+---+---+---+---+
| RESV | (1 byte)
+---+---+---+---+---+---+---+---+---+
| LAG-ID RECORD (1) | (10 bytes)
+---+---+---+---+---+---+---+---+---+
.
.
+---+---+---+---+---+---+---+---+---+
| LAG-ID RECORD (n) |
+---+---+---+---+---+---+---+---+---+

```

LAG-ID RECORDs list all the end-devices to which the RBv identified by the pseudo-nickname provides services.

After receiving such a sub-TLV, if the receipt RBridge has membership for at least one of the listed LAGs and accepts the DRB membership of the originator of the TLV, it uses the RBv identified by the pseudo-nickname to service the end-devices identified by some of the listed LAGs and multi-homed to it. Otherwise, the received sub-TLV is ignored.

9. OAM Frames

Attention must be paid when generating the OAM frames. When an OAM frame is generated with the ingress nickname of RBv, the originator RBridge's nickname MUST be included in the OAM message to ensure the response is returned to the originating member of the RBv group.

10. Configuration Consistency

It is important that VLAN membership of member ports of end switch SW1 is consistent across all of the member ports in the point-point

scenario. 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 SW1-RB1, while RB2 only configures VLAN1 for the SW1-RB2 link. Both RB1 and RB2 use the same ingress nickname RBv for all frames originating from S1. Hence, a remote RBridge RBx will learn that MAC addresses from S1 on VLAN2 are originating from RBv. As a result, on the returning path, RBx may deliver VLAN2 traffic to RB2. However, RB2 does not have VLAN2 configured on SW1-RB2 link and hence the frame may be dropped or has to be redirected to RB1 if RB2 knows RB1 can reach S1 in VLAN2.

11. IANA Considerations

TBD.

12. Security Considerations

TBD.

13. Acknowledgements

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, and Jon Hudson for their good questions and comments.

14. Normative References

- [CMT] Senevirathne, T., Pathangi, J., and J. Hudson, "Coordinated Multicast Trees (CMT)for TRILL", [draft-ietf-trill-cmt-00.txt](#) Work in Progress, April 2012.
- [ClearCorrect] Eastlake 3rd, D., Zhang, M., Ghanwani, A., Manral, V., and A. Banerjee, "TRILL: Clarifications, Corrections, and Updates", [draft-ietf-trill-clear-correct-06.txt](#) In RFC Editing Queue, July 2012.
- [RFC1195] Callon, R., "Use of OSI IS-IS for routing in TCP/IP and dual environments", [RFC 1195](#), December 1990.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.

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

[rfc6326bis]

Eastlake 3rd, D., Banerjee, A., Ghanwani, A., and R. Perlman, "Transparent Interconnection of Lots of Links (TRILL) Use of IS-IS", [draft-eastlake-isis-rfc6326bis-07.txt](#) Work in Progress, March 2012.

[Appendix A](#). Rationale for MAC Sharing among Member RBridges

With the introduction of virtual RBridge, MAC flip-flopping problem in LAN or LAG is resolved. However, in order to forward traffic effectively, member RBridges should share some of their learned MAC addresses with each other. For example, see Figure 5 shown below.

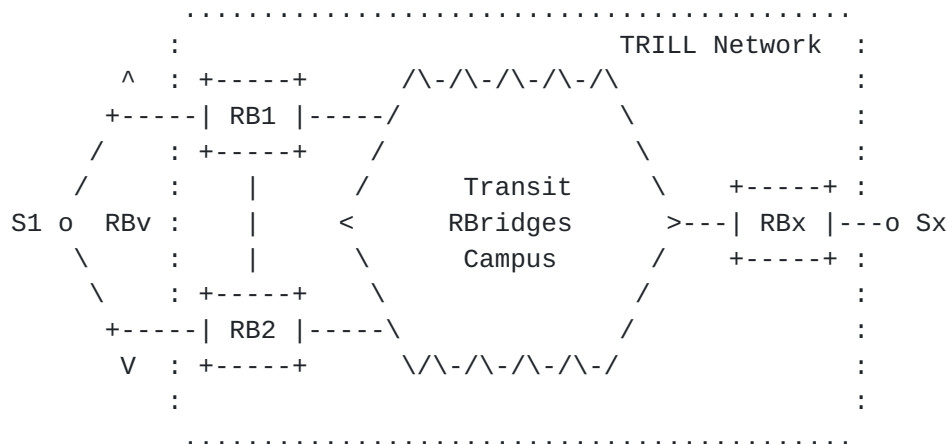


Figure 5 RBv in LAG scenario

Take Figure 5 as an example, the VLAN-x native frames from S1 to Sx will enter TRILL campus via one member RBridge of the RBv (say RB1). RB1 learns the location of S1 in VLAN-x. However, RBx may deliver the reverse traffic to RB2 if it thinks the shortest path to RBv is through RB2. If RB2 has not learned the location of S1 in VLAN-x from the MAC sharing, RB2 has to transmit the reverse traffic to S1 as unknown unicast.

Thus, the learned MAC addresses of attached end stations on one member RBridge SHOULD be shared with rest of the member RBridges in the same RBv. With these information shared, when RB2 receives

reverse frames, it can determine how to forward them to S1. For example, it can redirect them to RB1 if link RB2-S1 fails.

Since RBx always delivers the reverse traffic to RBv via RB2, RB2 egresses the traffic and learns the location of Sx. But RB1 will not know where Sx is, if RB2 does not share this information with RB1. As a result, RB1 has to treat the traffic from S1 to Sx as traffic with unknown destination and flood it in TRILL, which adds additional forwarding burden on the TRILL network.

Therefore, in addition to local attached end station MAC addresses, the learned remote MAC addresses should also be shared among all member RBridges of a RBv. With such information shared, RB1 can treat the traffic to Sx as known destination traffic and unicast it to RBx.

The design for above MAC sharing is currently beyond the scope of this document.

Authors' Addresses

Hongjun Zhai
ZTE Corperation
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

