INTERNET-DRAFT
Intended Status: Proposed Standard
Updates: 7176

Mingui Zhang
Huawei
Radia Perlman
Intel
Hongjun Zhai
ZTE
Muhammad Durrani
Mukhtiar Shaikh
Brocade
Sujay Gupta
IP Infusion

Expires: December 19, 2014
June 17, 2014

TRILL Active-Active Edge Using Multiple MAC Attachments

draft-zhang-trill-aa-multi-attach-04.txt

Abstract

TRILL active-active service provides end stations with flow level load balance and resilience against link failures at the edge of TRILL campuses.

This draft specifies a method in which member RBridges in an active-active edge RBridge group use their own nicknames as ingress RBridge nicknames to encapsulate frames from attached end systems. Thus, remote edge RBridges are required to keep multiple locations of one MAC address in one Data Label. Design goals of this specification are discussed in the document.

Status of this Memo

http://www.ietf.org/shadow.html

Table of Contents

1. Introduction

In the TRILL Active-Active Edge (AAE) topology, a Multi-Chassis Link Aggregation Group (MC-LAG) is used to connect multiple RBridges to a switch, vSwitch or multi-port end station. An endnode clump is attached to this switch or vSwitch. It’s required that data traffic within a specific VLAN from this endnode clump (including the multi-port end station) can be ingressed and egressed by any of these RBridges simultaneously. End systems in the clump can spread their traffic among these edge RBridges at the flow level. When a link fails, end systems keep using the rest of links in the MC-LAG without waiting for the convergence of TRILL, which provides resilience to link failures.

Since a packet from each endnode can be ingressed by any RBridge in the AAE group, a remote edge RBridge may observe multiple attachment points (i.e., egress RBridges) for this endnode identified by its MAC address and Data Label (VLAN or Fine Grained Label (FGL)). This issue is known as the "MAC flip-flopping". Three potential solutions arise to address this issue:

1) AAE member RBridges use a pseudonode nickname, instead of their own, as the ingress nickname for end systems attached to the MC-LAG. [CMT] falls within this category.

2) AAE member RBridges split work among themselves for which one will be responsible for which MAC addresses. A member RBridge will encapsulate the packet using its own nickname if it is responsible for the source MAC address. Otherwise, if the frame is known unicast, it encapsulates the packet using the nickname of the responsible RBridge; if the frame is multicast, it needs to redirect the packet to its responsible RBridge for encapsulation.

3) AAE member RBridges keep using their own nicknames. Remote edge RBridges are required to keep multiple points of attachment per MAC address and Data Label attached to the AAE.

The purpose of this document is to develop an approach based on solution 3. Although it focuses on exploring solution 3, the major design goals discussed here are common for all three AAE solutions. Through mirroring the scenarios studied in this draft, other potential solutions may benefit as well.

The main body of the document is organized as follows. Section 2 lists the acronyms and terminologies. Section 3 gives the overview model. Section 4 provides three options for incremental deployment. Section 5 describes how this approach meets the design goals.

2. Acronyms and Terminology

2.1. Acronyms

AAE: Active-Active Edge

Data Label: VLAN or FGL

ESADI: End Station Address Distribution Information [ESADI]

FGL: Fine Grained Label [RFC7172]

IS-IS: Intermediate System to Intermediate System [ISIS]

MC-LAG: Multi-Chassis Link Aggregation Group

TRILL: TRansparent Interconnection of Lots of Links [RFC6325]

vSwitch: A virtual switch such as a hypervisor that also simulates a bridge.

2.2. Terminology

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].

Familiarity with [RFC6325], [RFC6439] and [RFC7177] is assumed in this document.

3. Overview

```
                               +-----+
                               | RB4 |
                    +----------+-----+----------+
                    |                           |
                    |                           |
                    |       Rest of campus      |
                    |                           |
                    |                           |
                    +-+-----+--+-----+--+-----+-+
                      | RB1 |  | RB2 |  | RB3 |
                      +-----\  +-----+  /-----+
                              \   |   /
                                \ | /
                                 |||MC-LAG1
                                 |||
                                +---+
                                | B |
                                +---+
                             H1 H2 H3 H4: VLAN 10
```

Figure 3.1: An example topology for TRILL Active-Active Edge

Figure 3.1 shows an example network for TRILL Active-Active Edge. In this figure, endnodes (H1, H2, H3 and H4) are attached to a bridge (B) which communicates with multiple RBridges (RB1, RB2 and RB3) via the MC-LAG. Suppose RB4 is a ’remote’ RBridge out of the AAE group in the TRILL campus. This connection model is also applicable to the virtualized environment where the physical bridge can be replaced with a vSwitch while those bare metal hosts are replaced with virtual machines (VM).

For a packet received from their attached endnode clumps, a member RBridge of the AAE group always encapsulates it using its own nickname as the ingress nickname no matter whether it’s unicast or multicast.

The remote RBridge RB4 will see multiple attachments of one MAC from each of the end nodes.

4. Incremental Deployable Options

Three options are listed below to cope with incremental deployment scenarios. Among them, Option C can be hardware independent.

-- Option A

A new capability announcement would appear in LSPs. "I can cope with multiple attachments for an endnode". Only if all edge

RBridges announce this capability can the AAE group use this approach. For those legacy RBridges who are not capable of coping with multiple endnode attachments, new type TRILL switches will not establish connectivity with them so that they are isolated from these new type TRILL switches. Note only edge RBridges (those that are Appointed Forwarders [RFC6439]) need to be able to support this. It does not affect totally transit RBridges.

-- Option B

Each edge RBridge in the AAE group ingress data frames from any MC-LAG into a specific TRILL topology [TRILL-MT]. In this way, the topology ID is used as the discriminator of different locations of a specific MAC address at the remote RBridge. TRILL MAY reserve a list of topology IDs to be dedicated to AAE. RBridges that do not support this reserved list MUST NOT establish connectivity with edge RBridges in the AAE group.

-- Option C

As pointed out in Section 4.2.6 of [RFC6325] and Section 5.3 of [ESADI], one MAC address may be persistently claimed to be attached to multiple RBridges within the same Data Label in the TRILL ESADI LSPs. For this option, AAE member RBridges make use of TRILL ESADI protocol to distribute multiple attachments of a MAC address. Remote RBridges disable the data plane learning for such multi-attached MAC addresses.

4.1. Detail of Option C

An RBridge in an AAE MUST advertise all Data Labels enabled for all its attached MC-LAGs. This causes remote RBridges to disable the MAC learning via the TRILL Data packet decapsulation within these Data Labels for this RBridge. The advertisement of such Data Labels can be realized by allocating one reserved flag from the Interested VLANs and Spanning Tree Roots Sub-TLV (Section 2.3.6 of [RFC7176]) and one reserved flag from the Interested Labels and Spanning Tree Roots Sub-TLV (Section 2.3.8 of [RFC7176]). When this flag is set to 1, the originating IS is advertising Data Labels for MC-LAGs rather than plain LAN links. (See Section 7.2)

Whenever a MAC from the MC-LAG of this AAE is learned, it needs to be advertised via the ESADI protocol. In its TRILL ESADI LSPs, the originating IS needs to include the identifier of this AAE. Remote RBridges need to know all nicknames of RBridges in this AAE. This is achieved by listening to the "MC-LAG Group RBridges" TRILL APPsub-TLV defined in Section 5.3.2. MAC Reachability TLVs [RFC6165] are composed in a way that each TLV only contains MAC addresses of end nodes attached to a single MC-LAG. Each such TLV is enclosed in a TRILL APPsub-TLV defined as follows.

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type = MC-LAG-GROUP-MAC       | (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Length                      | (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+-+
|   MC-LAG System ID            (8 bytes)          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+-+
|   MAC-Reachability TLV        (7 + 6*n bytes)    |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+-+
```

o Type: MC-LAG Group MAC (TRILL APPsub-TLV type #TBD)

o Length: The MAC-Reachability TLV [RFC6165] is contained in the value field as a sub-TLV. The total number of bytes contained in the value field is given by 15+6*n.

o MC-LAG System ID: The System ID of the MC-LAG as specified in Section 5.3.2 of [802.1AX]. Here, it also serves as the identifier of the AAE.

o MAC-Reachability sub-TLV: The MC-LAG-GROUP-MAC APPsub-TLV value contains the MAC-Reachability TLV as a sub-TLV.

This MC-LAG-GROUP-MAC APPsub-TLV SHOULD be included in a GENINFO TLV [RFC6823] in the ESADI-LSP. There may be more than one occurrence of such TRILL APPsub-TLV in one ESADI-LSP fragment.

For those MAC addresses contained in an MC-LAG-GROUP-MAC APPsub-TLV, this document applies. Otherwise, [ESADI] applies. For example, an AAE member RBridge continues to enclose MAC addresses learned from TRILL Data packet decapsulation in MAC-Reachability TLV as per [RFC6165] and advertise them using the ESADI protocol.

When the remote RBridge learns MAC addresses contained in the MC-LAG-GROUP-MAC APPsub-TLV via the ESADI protocol, it always sends the packets destined to these MAC addresses to the closest one (the one to which the remote RBridge has the least cost forwarding path) of those RBridges in the AAE identified by the MC-LAG System ID in the MC-LAG-GROUP-MAC APPsub-TLV. If there are multiple such member RBridges, the ingress RBridge is required to select a unique one in a pseudo-random way as specified in Section 5.3 of [ESADI].

When another RBridge in the same AAE group receives an ESADI-LSP with the MC-LAG-GROUP-MAC APPsub-TLV, it also learns MAC addresses of those end nodes served by the corresponding MC-LAG. These MAC addresses SHOULD be learned as if those end nodes are locally attached to this RBridge itself.

An AAE member RBridge MUST use the MC-LAG-GROUP-MAC APPsub-TLV to advertise the MAC addresses learned from a plain local link (a non MC-LAG link) with Data Labels that happen to be covered by the Data Labels of any attached MC-LAG. The reason is that data plane learning within these Data Labels at the remote RBridge has been disabled for this RBridge.

4.2. Capability Flags TLV

The following Capability Flags TLV will be included in LSP as a TRILL APPsub-TLV of GENINFO-TLV.

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type = MULTI-MAC-ATTACH-CAP   | (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Length                      | (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|E|H| Reserved  |                 (1 byte)
+-+-+-+-+-+-+-+-+
```

o Type: Multi-MAC-Attach Capability (TRILL APPsub-TLV type #TBD)

o Length: Set to 1.

o E: When this bit is set, it indicates the originating IS acts as specified in Option C.

o H: When this bit is set, it indicates that the originating IS keeps multiple MAC attachments with fast path hardware at the data plane.

o Reserved: Reserved flags for future use. These MUST be sent as zero and ignored on receipt.

The Capability Flags TRILL APPsub-TLV is used to notify other RBridges whether the originating IS supports the capability indicated by the E and H bits. For example, if E bit is set, it indicates the originating IS will act as defined in Option C. That is, it will disable the data plane MAC learning for AAE RBridges within Data Labels advertised by them while waiting for the TRILL ESADI LSPs to distribute the {MAC, Nickname, Data Label} association. Meanwhile, this RBridge is able to act as an AAE RBridge. It’s required to advertise MAC addresses learned from MC-LAGs in TRILL ESADI LSPs using the MC-LAG-GROUP-MAC APPsub-TLV defined in Section 4.1. AAE RBridges supporting Options C won’t establish connectivity with remote edge RBridges unless this RBridge has advertised this Capability Flags TLV with E bit set.

Capability specification for Option B is out the scope of this document. It may be specified in documents for TRILL multi-topology [TRILL-MT].

5. Design Goals

How this specification meets the major design goals of AAE is explored in this section.

5.1. No MAC Flip-Floping (Normal Unicast Egress)

Since all RBridges talking with the AAE RBridges in the campus are able to keep multiple locations for one MAC address, a MAC address learned from one AAE member will not be overwritten by the same MAC address learned from another AAE member. Although multiple entries for this MAC address will be created, the remote RBridge is required to adhere to a unique one of the locations (see Section 4.1) for each MAC address rather than keep flip-flopping among them.

5.2. Regular Unicast/Multicast Ingress

MC-LAG guarantees that each frame will be sent upward to the AAE via exactly one uplink. RBridges in the AAE can simply follow the process per [RFC6325] to ingress the frame. For example, each RBridge uses its own nickname as the ingress nickname to encapsulate the packet. In such scenario, each RBridge takes for granted that it is the Appointed Forwarder for the VLANs enabled on the uplink of the MC-LAG.

5.3. Right Multicast Egress

A fundamental design goal of AAE is that there is no duplication or forwarding loop.

5.3.1. No Duplication (Single Exit Point)

When multi-destination packets for a specific Data Label are received from the campus, it’s important that exactly one RBridge out of the AAE group let through each multicast packet, therefore no duplication happens. Since AAE member RBridges support MC-LAG, they are able to utilize the hashing function of MC-LAG to determine the single exit point. If the output of the hashing function points to the port attached to the receiver RBridge itself (i.e., the packet should be egressed out of this node), it egresses this packet. Otherwise, the packet MUST be dropped.

5.3.2. No Echo (Split Horizon)

When a multicast frame originated from an MC-LAG is ingressed by an RBridge of an AAE group, forwarded across the TRILL network and then received by another RBridge in the same AAE group, it is important that this RBridge does not egress this frame back to this MC-LAG. Otherwise, it will cause a forwarding loop (echo). The well known 'split horizon' technique can be used to eliminate the echo issue.

RBridges in the AAE group need to split horizon based on the ingress RBridge nickname plus the VLAN of the TRILL Data packet. They need to set up per port filtering lists consists of the tuple of <ingress nickname, VLAN>. Packets with information matching with any entry of the filtering list MUST NOT be egressed out of that port. The information of such filters is obtained by listening to the following "MC-LAG Group RBridges" TRILL APPsub-TLV included in the GENINFO TLV in LSPs.

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Type = MC-LAG-GROUP-RBRIDGES  | (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Length                      | (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Sender Nickname             | (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+-+
|   MC-LAG System ID            (8 bytes)       |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...+-+-+
```

- o Type: MC-LAG Group RBridges (TRILL APPsub-TLV type #TBD)
- o Length: 10
- o Sender Nickname: The nickname of the originating IS.
- o MC-LAG System ID: The System ID of the MC-LAG as specified in Section 5.3.2 of [802.1AX].

All enabled VLANs MUST be consistent on all ports connected to an MC-LAG. So that the enabled VLANs need not to be included in the MC-LAG Group RBridges TRILL APPsub-TLV. They can be locally obtained from the port attached to that MC-LAG.

Through parsing an MC-LAG Group RBridges TRILL APPsub-TLV, the receiver RBridge discovers all other RBridges connected to the same MC-LAG. The Sender Nickname of the originating IS will be added into the filtering list of the port attached to the MC-LAG. For example, RB3 in Figure 3.1 will set up a filtering list looks like {<RB1, VLAN10>, <RB2, VLAN10>} on its port attached to MC-LAG1. According to split horizon, TRILL Data packets within VLAN10 ingressed by RB1 or RB2 will not be egressed out of this port.

When there are multiple MC-LAGs connected to the same RBridge, these MC-LAGs may have overlap VLANs. Customer may need hosts within these overlap VLANs to communicate with each other. In Appendix A, several scenarios are given to explain how hosts communicate within the overlap VLANs and how split horizon happens.

5.4. No Black-hole or Triangular Forwarding

If a sub-link of the MC-LAG fails while remote RBridges continue to send packets towards the failed port, a black-hole happens. If the AAE member RBridge with that failed port starts to redirect the packets to other member RBridges for delivery, triangular forwarding forms.

The member RBridge attached to the failed sub-link can make use of the ESADI protocol to flush those failure affected MAC addresses as defined in Section 5.2 of [ESADI]. After doing that, no packets will be sent towards the failed port, hence no black-hole will happen. Nor will the member RBridge need to redirect packets to other member RBridges, which may otherwise lead to the triangular forwarding.

5.5. Load Balance Towards the AAE

Since a remote RBridge can record multiple attachments of one MAC address, this remote RBridge can choose to spread the traffic towards the AAE members. Each of them is able to act as the egress point. By doing this, the forwarding paths may be not limited to the least cost Equal Cost Multiple Paths from the ingress RBridge to the AAE RBridges. The traffic load from the remote RBridge towards the AAE RBridges can be balanced based on a pseudo-random selection method (see Section 4.1).

Note that the load balance method adopted at the ingress RBridge is not to replace the load balance mechanism of MC-LAG. These two load spreading mechanisms should take effect separately.

5.6. Scalability

With option A, multiple attachments need to be recorded for a MAC address learned from AAE RBridges. More entries may be consumed in the MAC table. However, MAC addresses attached to an MC-LAG are only a small part of all MAC addresses in the whole TRILL campus. As a result, the extra space required by the multi-attached MAC addresses can be accommodated by RBridges’ unused MAC table space.

With option C, remote RBridges will keep the multiple attachments of a MAC address in the ESADI link state databases. While in the MAC table, an RBridge still establishes only one entry for each MAC address.

6. Security Considerations

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

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

For general TRILL security considerations, see [RFC6325].

7. IANA Considerations

7.1. TRILL APPsub-TLVs

IANA is requested to allocate three new types under the Generic Information TLV (#251) [RFC6823] for the TRILL APPsub-TLVs defined in Section 4.1, 4.2 and 5.3.2 of this document.

Reference: [ESADI] and [This document]

Type	Name	Reference
0	Reserved	
1	ESADI-PARAM	[ESADI]
2-254	Unassigned	
255	Reserved	
256	MC-LAG-GROUP-MAC	This document
257	MULTI-MAC-ATTACH-CAP	This document
258	MC-LAG-GROUP-RBRIDGES	This document
260-65534	Unassigned	
65535	Reserved	

7.2. Active Active Flags

IANA is requested to allocate two flag bits, as follows:

One flag bit appears in the "Interested VLANs and Spanning Tree Roots Sub-TLV".

References: [RFC7176], [ESADI] and [This document]

Bit	Mnemonic	Description	Reference

0	M4	IPv4 Multicast Router Attached	[RFC7176]
1	M6	IPv6 Multicast Router Attached	[RFC7176]
2	-	Unassigned	
3	ES	ESADI Participation	[ESADI]
4-15	-	(used for a VLAN ID)	[RFC7176]
16	AA	Enabled VLANs for Active-Active	This document
17-19	-	Unassigned	
20-31	-	(used for a VLAN ID)	[RFC7176]

One flag bit appears in the "Interested Labels and Spanning Tree Roots Sub-TLV".

References: [RFC7176], [ESADI] and [This document]

Bit	Mnemonic	Description	Reference
0	M4	IPv4 Multicast Router Attached	[RFC7176]
1	M6	IPv6 Multicast Router Attached	[RFC7176]
2	BM	Bit Map	[RFC7176]
3	ES	ESADI Participation	[ESADI]
4	AA	FGLs for Active-Active	This document
5-7	-	Unassigned	

Acknowledgements

Authors would like to thank the comments and suggestions from Erik Nordmark, Fangwei Hu and Liang Xia.

8. References

8.1. Normative References

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

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

[RFC6325] Perlman, R., Eastlake 3rd, D., Dutt, D., Gai, S., 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.

[RFC6823] Ginsberg, L., Previdi, S., and M. Shand, "Advertising

Generic Information in IS-IS", RFC 6823, December 20165

[RFC7172] D. Eastlake 3rd and M. Zhang and P. Agarwal and R. Perlman and D. Dutt, "Transparent Interconnection of Lots of Links (TRILL): Fine-Grained Labeling", RFC 7172, May 2014.

[RFC7176] D. Eastlake 3rd and T. Senevirathne and A. Ghanwani and D. Dutt and A. Banerjee, "Transparent Interconnection of Lots of Links (TRILL) Use of IS-IS", RFC7176, May 2014.

[RFC7177] D. Eastlake 3rd and R. Perlman and A. Ghanwani and H. Yang and V. Manral, "Transparent Interconnection of Lots of Links (TRILL): Adjacency", RFC 7177, May 2014.

[ESADI] H. Zhai, F. Hu, et al, "TRILL (Transparent Interconnection of Lots of Links): ESADI (End Station Address Distribution Information) Protocol", draft-ietf-trill-esadi-07.txt, April 2014, Submitted to IESG for Publication.

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

8.2. Informative References

[CMT] T. Senevirathne, J. Pathangi, et al, "Coordinated Multicast Trees (CMT)for TRILL", draft-ietf-trill-cmt-03.txt, April 2014, working in progress.

[TRILL-MT] D. Eastlake, M. Zhang, A. Banerjee, V. Manral, "TRILL: Multi-Topology", draft-eastlake-trill-multi-topology, work in progress.

[ISIS] ISO, "Intermediate system to Intermediate system routeing information exchange protocol for use in conjunction with the Protocol for providing the Connectionless-mode Network Service (ISO 8473)", ISO/IEC 10589:2002.

[RFC5310] Bhatia, M., Manral, V., Li, T., Atkinson, R., White, R., and M. Fanto, "IS-IS Generic Cryptographic Authentication", RFC 5310, February 2009.

Appendix A. Scenarios on Split Horizon

```
+------------------+   +------------------+   +------------------+
|        RB1       |   |        RB2       |   |        RB3       |
+------------------+   +------------------+   +------------------+
L1       L2       L3   L1       L2       L3   L1       L2       L3
VL10~20  VL15~25  VL15 VL10~20  VL15~25  VL15 VL10~20  VL15~25  VL15
MC-LAG1  MC-LAG2  LAN  MC-LAG1  MC-LAG2  LAN  MC-LAG1  MC-LAG2  LAN
B1       B2       B10  B1       B2       B20  B1       B2       B30
```

Figure A.1: An example topology to explain split horizon

Suppose RB1, RB2 and RB3 are the Active-Active group connecting MC-LAG1 and MC-LAG2. MC-LAG1 and MC-LAG2 are connected to B1 and B2 at their other ends. Suppose all these RBridges use port L1 to connect MC-LAG1 while they use port L2 to connect MC-LAG2. Assume all three L1 enable VLAN 10~20 while all three L2 enable VLAN 15~25. So that there is an overlap of VLAN 15~20. Customer needs hosts in these overlap VLANs to communicate with each other. That is, hosts attached to B1 in VLAN 15~20 need to communicate with hosts attached to B2 in VLAN 15~20. Assume the remote plain RBridge RB4 also has hosts attached in VLAN 15~20 which need to communicate with those hosts in these VLANs attached to B1 and B2.

Two major requirements:

1. Frames ingressed from RB1-L1-VLAN 15~20 MUST NOT be egressed out of ports RB2-L1 and RB3-L1. At the same time,

2. frames coming from B1-VLAN 15~20 should reach B2-VLAN 15~20.

RB3 stores the information for split horizon on its ports L1&L2. On L1: {<ingress_nickname_RB1, VLAN 10~20>, <ingress_nickname_RB2, VLAN 10~20>} and on L2: {<ingress_nickname_RB1, VLAN 15~25>, <ingress_nickname_RB2, VLAN 15~25>}.

Five clarification scenarios:

a. Suppose RB2/RB3 receives a TRILL multicast data packet with VLAN 15 and ingress nickname RB1. RB3 is the single exit point (selected out according to the hashing function of MC-LAG) for this packet. On ports L1&L2, RB3 has covered <ingress_nickname_RB1, VLAN 15>, so that RB3 will not egress this packet out of either L1 or L2. Here, _split horizon_ happens.

   Beforehand, RB1 obtains a native frame on port L1 from B1 in VLAN 15. RB1 judges it should be forwarded as a multicast frame across the TRILL campus. Also, RB1 replicates this frame without TRILL encapsulation and sends it out of port L2, so that B2 will get this frame.

b. Suppose RB2/RB3 receives a TRILL multicast data packet with VLAN 15 and ingress nickname RB4. RB3 is the single exit point. On ports L1&L2, since RB3 has not stored any tuple with ingress_ nickname_RB4, RB3 will decapsulate the packet and egress it out of both ports L1 and L2. So both B1 and B2 will receive the frame.

c. Suppose there is a plain LAN link port L3 on RB1, RB2 and RB3, connecting to B10, B20 and B30 respectively. These L3 ports happen to be configured with VLAN 15. On port L3, RB1 and RB3 stores no information of split horizon for AAE (since this port has not been configured to be in any MC-LAG). They will egress the packet ingressed out of RB1-L1 in VLAN 15.

d. If a packet is ingressed from RB1-L1 or RB1-L2 with VLAN 15, port RB1-L3 will not egress packets with ingress-nickname-RB1. RB1 needs to replicate this frame without encapsulation and sends it out of port L3.

e. If a packet is ingressed from RB1-L3, since RB1-L1 and RB1-L2 cannot egress packets with VLAN 15 and ingress-nickname-RB1, RB1 needs to replicate this frame without encapsulation and sends it out of port L1 and L2.

Author's Addresses

Mingui Zhang
Huawei Technologies
No.156 Beiqing Rd. Haidian District,
Beijing 100095 P.R. China

EMail: zhangmingui@huawei.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

Hongjun Zhai
ZTE Corporation
68 Zijinghua Road
Nanjing 200012 China

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

Muhammad Durrani
Brocade

EMail: mdurrani@brocade.com

Mukhtiar Shaikh
Brocade

EMail: mshaikh@brocade.com

Sujay Gupta
IP Infusion
Bangalore, India

EMail: sujayg@ipinfusion.com