INTERNET-DRAFT
Intended Status: Proposed Standard

Mingui Zhang
Huawei
Radia Perlman
EMC
Hongjun Zhai
JIT
Muhammad Durrani
Brocade
Sujay Gupta
IP Infusion

Expires: August 9, 2015
February 5, 2015

TRILL Active-Active Edge Using Multiple MAC Attachments
draft-ietf-trill-aa-multi-attach-03.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 as described in RFC 7379.

This draft specifies a method by 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.

Table of Contents

1. Introduction

As discussed in [RFC7379], in a TRILL Active-Active Edge (AAE) topology, a Local Active-Active Link Protocol (LAALP), for example, a Multi-Chassis Link Aggregation Group (MC-LAG), is used to connect multiple RBridges to multi-port Customer Equipment (CE), such as a switch, vSwitch or a multi-port end station. An endnode clump is attached in the case of switch or vSwitch. It is required that data traffic within a specific VLAN from this endnode clump (including the multi-port end station case) 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 remaining links in the LAALP without waiting for the convergence of TRILL, which provides resilience to link failures.

Since a frame 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 pseudo-nickname, instead of their own, as the ingress nickname for end systems attached to the LAALP. [PN] falls within this category.

2) AAE member RBridges split work among themselves as to which one will be responsible for which MAC addresses. A member RBridge will encapsulate the frame using its own nickname if it is responsible for the source MAC address. Otherwise, if the frame is known unicast, it encapsulates the frame using the nickname of the responsible RBridge; if the frame is multi-destination, it needs to tunnel the native frame to its responsible RBridge for encapsulation, for example using [ChannelTunnel].

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 specify 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. The use of any of these solutions in an AAE group does not prohibit the use of other solutions in other AAE groups in the same TRILL campus. For example, the specification in this draft and the specification in [PN] could be simultaneously deployed for different AAE groups in the same campus.

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 options for incremental deployment. Section 5 describes how this approach meets the design goals. The Sections after Section 5 cover security, IANA, and some backwards compatibility considerations.

2. Acronyms and Terminology

2.1. Acronyms and Terms

AAE: Active-Active Edge

Campus: a TRILL network consisting of TRILL switches, links, and possibly bridges bounded by end stations and IP routers. For TRILL, there is no "academic" implication in the name "campus"

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

Data Label: VLAN or FGL

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

Edge RBridge: An RBridge providing end station service on one or more of its ports.

ESADI: End Station Address Distribution Information [RFC7357]

FGL: Fine Grained Label [RFC7172]

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

LAALP: As in [RFC7379], Local Active-Active Link Protocol. Any protocol similar to MC-LAG (or DRNI) that runs in a distributed fashions on a CE, the links from that CE to a set of edge group RBridges, and on those RBridges.

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

RBridge: A device implementing the TRILL protocol.

TRILL: TRansparent Interconnection of Lots of Links or Tunneled Routing in the Link Layer [RFC6325] [RFC7177].

TRILL switch: An alternative name for an RBridge.

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

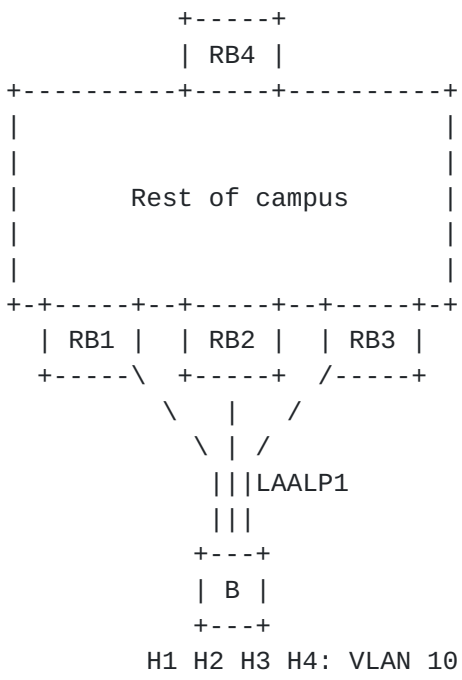


Figure 3.1: An example topology for TRILL Active-Active Edge

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 that communicates with multiple RBridges (RB1, RB2 and RB3) via the LAALP. Suppose RB4 is a 'remote' RBridge not in 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 frame received from its attached endnode clumps, a member

RBridge of the AAE group conforming to this document always encapsulates that frame using its own nickname as the ingress nickname no matter whether it's unicast or multicast.

The remote RBridge RB4 will see multiple attachments for each MAC from one of the end-nodes. Although this could cause problems if RB4 is learning remote end station attachments from the data plane, we specify a solution below ("Option C").

4. Incremental Deployable Options

Three options are listed below to handle incremental deployment scenarios. Among them, Option C can be incrementally implemented throughout a TRILL campus with common existing TRILL fast path hardware. Further details on Option C are given in Section 4.1.

-- Option A

A new capability announcement would appear in LSPs: "I can cope with data plane learning of multiple attachments for an endnode". This mode of operation is generally not supported by existing TRILL fast path hardware. Only if all edge RBridges to which the group has data connectivity and that are interested in any of the Data Labels in which the AAE is interested announce this capability can the AAE group safely use this approach. If all such RBridges do not announce this "Option A" capability, then a fallback would be needed such as reverting from active-active to active-standby operation or isolating the RBridge that would need to support this capability and do not support it. Further details for Options A are beyond the scope of this document except that in Section 4.2 a bit is reserved to indicate support for Option A because a remote RBridge supporting Option A is compatible with an AAE group using Option C.

-- Option B

Each edge RBridge in the AAE group ingresses frames from any LAALP 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 could reserve a list of topology IDs to be dedicated to AAE. A variety of fallbacks might be needed for RBridges that do not support multi-topology or do not support a needed topology. Further details for this Options B are beyond the scope of this document.

-- Option C

As pointed out in Section 4.2.6 of [RFC6325] and Section 5.3 of

[RFC7357], one MAC address may be persistently claimed to be attached to multiple RBridges within the same Data Label in the TRILL ESADI-LSPs. For Option C, AAE member RBridges make use of the TRILL ESADI protocol to distribute multiple attachments of a MAC address. Remote RBridges SHOULD disable the data plane MAC learning for such multi-attached MAC addresses from TRILL Data packet decapsulation unless they also support Option A. The ability to configure an RBridge to disable data plane learning is provided by the base TRILL protocol [RFC6325].

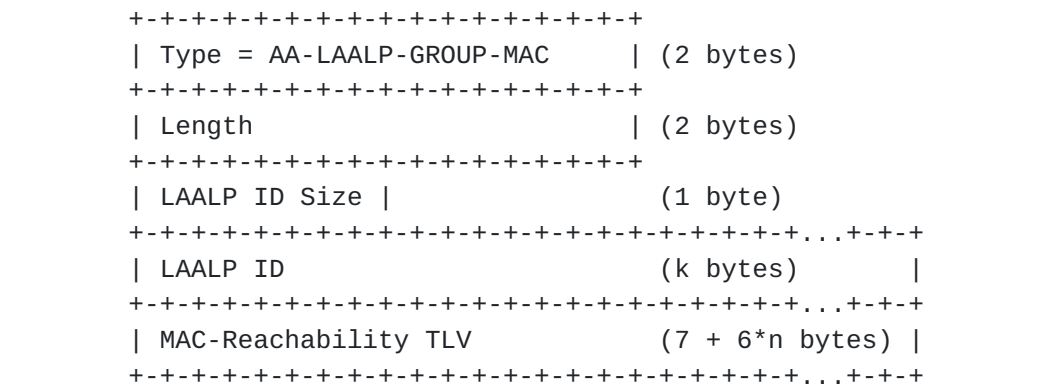
4.1. Detail of Option C

With Option C, an RBridge in an AAE group MUST advertise all Data Labels enabled for all its attached LAALPs and participate in ESADI for those Data Labels. Receiver edge RBridges MUST avoid flip-flop errors in MAC learned from the TRILL Data packet decapsulation for the originating RBridge within these Data Labels. It's RECOMMENDED that the receiver edge RBridge disable the data plane MAC learning from TRILL Data packet decapsulation within those advertised Data Labels for the originating RBridge unless the receiver RBridge also supports Option A. However, alternative implementations MAY be used to produce the same expected behavior. A promising way is to make use of the confidence level mechanism [RFC6325]. For example, let the receiver edge RBridge give a prevailing confidence value (e.g., 0x21) to the first MAC attachment learned from the data plane over others from the TRILL Data packet decapsulation. So the receiver edge RBridge will stick to this MAC attachment until it is overridden by one learned from the ESADI protocol [RFC7357]. The MAC attachment learned from ESADI is set to have higher confidence value (e.g., 0x80) to override any alternative learning from the decapsulation of received TRILL Data packets [RFC6325].

The advertisement of enabled Data Labels for an LAALP 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 (RBridge) is advertising Data Labels for LAALPs rather than plain LAN links. (See Section 8.3)

Whenever a MAC from the LAALP of this AAE is learned through ingress or configuration, it MUST be advertised via the ESADI protocol [RFC7357]. In its TRILL ESADI-LSPs, the originating RBridge 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 "AA LAALP Group RBridges" TRILL APPsub-TLV defined in Section 5.3.2. The MAC Reachability TLVs [RFC6165] are composed in a way that each TLV only contains MAC addresses of end-nodes attached to a

single LAALP. Each such TLV is enclosed in a TRILL APPsub-TLV defined as follows.



- Type: AA LAALP Grouped MAC (TRILL APPsub-TLV type tbd1)

- 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 k+8+6*n.

- LAALP ID Size: The length k of the LAALP ID in bytes.

- LAALP ID: The ID of the LAALP that is k bytes long. Here, it also serves as the identifier of the AAE. If the LAALP is an MC-LAG (or DRNI), it is the 8 byte ID as specified in Clause 6.3.2 in [802.1AX].

- MAC-Reachability sub-TLV: The AA-LAALP-GROUP-MAC APPsub-TLV value contains the MAC-Reachability TLV as a sub-TLV. As specified in Section 2.2 in [RFC7356], the type and length fields of the MAC-Reachability TLV are encoded as unsigned 16 bit integers. The one octet unsigned Confidence along with these TLVs SHOULD be set to prevail over those MAC addresses learned from TRILL Data decapsulation by remote edge RBridges.

This AA-LAALP-GROUP-MAC APPsub-TLV MUST be included in a TRILL GENINFO TLV [RFC7357] 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 AA-LAALP-GROUP-MAC APPsub-TLV, this document applies. Otherwise, [RFC7357] 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 AA-LAALP-GROUP-MAC APPsub-TLV via the ESADI protocol [RFC7357], it sends

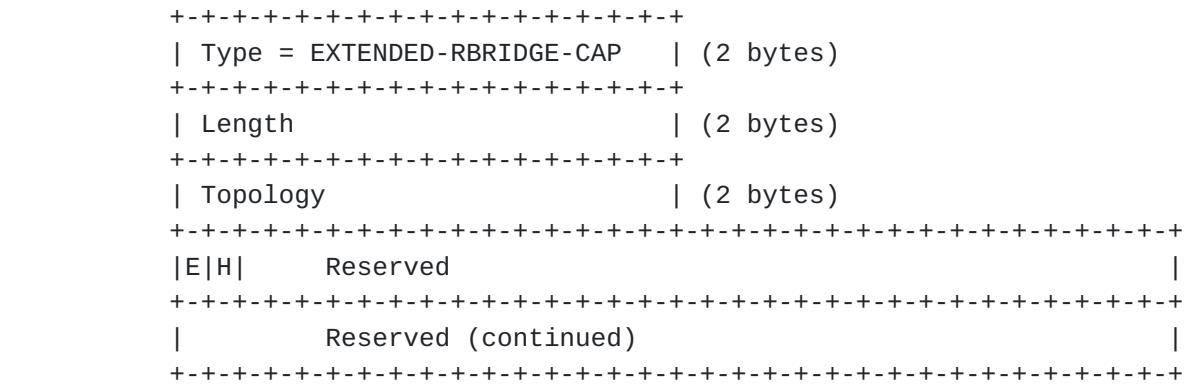
the packets destined to these MAC addresses to the closest one (the one to which the remote RBridge has the least cost forwarding path) of those RBridges in the AAE identified by the LAALP ID in the AA-LAALP-GROUP-MAC APPsub-TLV. If there are multiple equal least cost member RBridges, the ingress RBridge is required to select a unique one in a pseudo-random way as specified in Section 5.3 of [RFC7357].

When another RBridge in the same AAE group receives an ESADI-LSP with the AA-LAALP-GROUP-MAC APPsub-TLV, it also learns MAC addresses of those end-nodes served by the corresponding LAALP. 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 AA-LAALP-GROUP-MAC APPsub-TLV to advertise in ESADI the MAC addresses learned from a plain local link (a non LAALP link) with Data Labels that happen to be covered by the Data Labels of any attached LAALP. The reason is that MAC learning from TRILL Data packet decapsulation within these Data Labels at the remote edge RBridge has normally been disabled for this RBridge.

4.2. Extended RBridge Capability Flags APPsub-TLV

The following Extended RBridge Capability Flags APPsub-TLV will be included in an E-L1FS FS-LSP fragment zero [RFC7180bis] as an APPsub-TLV of the TRILL GENINFO-TLV.



- Type: Extended RBridge Capability (TRILL APPsub-TLV type tbd2)

- Length: Set to 8.

- Topology: Indicates the topology to which the capabilities apply. When this field is set to zero, this implies that the capabilities apply to all topologies or topologies are not in use [TRILL-MT].

- E: Bit 0 of the capability bits. When this bit is set, it indicates the originating IS acts as specified in Option C above.

- H: Bit 1 of the capability bits. When this bit is set, it indicates that the originating IS keeps multiple MAC attachments learned from TRILL Data packet decapsulation with fast path hardware, that is, it acts as specified in Option A above.

- Reserved: Flags extending from bit 2 through bit 63 of the capability fits reserved for future use. These MUST be sent as zero and ignored on receipt.

The Extended RBridge 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 MAC learning from TRILL Data packet decapsulation within Data Labels advertised by AAE RBridges 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 local LAALPs in TRILL ESADI-LSPs using the AA-LAALP-GROUP-MAC APPsub-TLV defined in Section 4.1. If the RBridge in an AAE group as specified herein observe a remote RBridge interested in one or more of that AAE group's Data Labels and the remote RBridge does not support, as indicated by its extended capabilities, either Option A or Option C, then the AAE group MUST fall back to active-standby mode.

Capability specification for Option B is out the scope of this document.

5. Meeting the Design Goals

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

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

Since all RBridges talking with the AAE RBridges in the campus are able to see multiple locations for one MAC address in ESADI [RFC7357], 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, for return traffic 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

LAALP 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 frame. In such a scenario, each RBridge takes for granted that it is the Appointed Forwarder for the VLANs enabled on the uplink of the LAALP.

5.3. Correct Multicast Egress

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

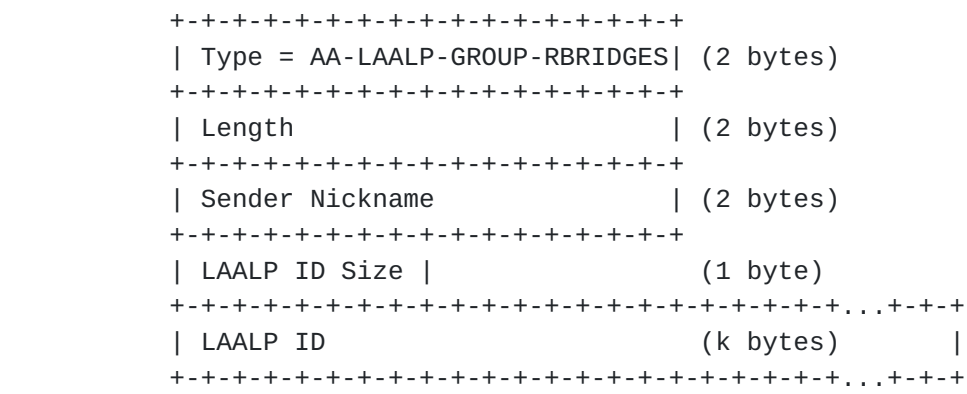
5.3.1. No Duplication (Single Exit Point)

When multi-destination TRILL Data 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 multi-destination packet so no duplication will happen. The LAALP will have defined its selection function (using hashing or election algorithm) to designated a forwarder for a multi-destination frame. Since AAE member RBridges support the LAALP, they are able to utilize that selection function to determine the single exit point. If the output of the selection function points to the port attached to the receiver RBridge itself (i.e., the packet should be egressed out of this node), it MUST egress this packet for that AAE group. Otherwise, the packet MUST NOT be egressed for that AAE group. (It is output or not as specified in [RFC6325] updated by [RFC7172] for ports that lead to non-AAE links.)

5.3.2. No Echo (Split Horizon)

When a multi-destination frame originated from an LAALP is ingressed by an RBridge of an AAE group, distributed to 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 LAALP. 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 "LAALP Group RBridges" APPsub-TLV included in the TRILL GENINFO TLV in FS-LSPs [RFC7180bis].



- Type: AA LAALP Grouped RBridges (TRILL APPsub-TLV type tbd3)

- Length: 3+k

- Sender Nickname: The nickname the originating IS will use as the ingress nickname. This field is useful because the originating IS might own multiple nicknames.

- LAALP ID Size: The length k of the LAALP ID in bytes.

- LAALP ID: The ID of the LAALP which is k bytes long. If the LAALP is an MC-LAG or DRNI, it is the 8-byte ID specified in Clause 6.3.2 in [802.1AX].

All enabled VLANs MUST be consistent on all ports connected to an LAALP. So the enabled VLANs need not be included in the AA-LAALP-GROUP-RBRIDGES TRILL APPsub-TLV. They can be locally obtained from the port attached to that LAALP.

Through parsing AA-LAALP-GROUP-RBRIDGES TRILL APPsub-TLVs, the receiver RBridge discovers all other RBridges connected to the same LAALP. The Sender Nickname of the originating IS will be added into the filtering list of the port attached to the LAALP. For example, RB3 in Figure 3.1 will set up a filtering list looks like {<RB1, VLAN10>, <RB2, VLAN10>} on its port attached to LAALP1. 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 LAALPs connected to the same RBridge, these LAALPs 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 LAALP 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 occurs.

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 [RFC7357]. 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 triangular forwarding.

5.5. Load Balance Towards the AAE

Since a remote RBridge can see multiple attachments of one MAC address in ESADI, this remote RBridge can choose to spread the traffic towards the AAE members on a per flow basis. Each of them is able to act as the egress point. In doing this, the forwarding paths need not be 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 a remote ingress RBridge is not to replace the load balance mechanism of LAALP. 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 learning table. However, MAC addresses attached to an LAALP are usually 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 usually 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 that are usually maintained by software. While in the MAC table that is normally implemented in hardware, an RBridge still establishes only one entry for each MAC address.

6. E-L1FS Backwards Compatibility

The Extended TLVs defined in Section 4 and 5 are to be used in an Extended Level 1 Flooding Scope (E-L1FS [RFC7356] [RFC7180bis]) PDU.

For those RBridges that do not support E-L1FS, the EXTENDED-RBRIDGE-CAP TRILL APPsub-TLV will not be sent out either and and MAC multi-attach active-active is not supported.

7. 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 transported by TRILL ESADI, see the Security Considerations section in [RFC7357].

For general TRILL security considerations, see [RFC6325].

8. IANA Considerations

8.1. TRILL APPsub-TLVs

IANA is requested to allocate three new types under the TRILL GENINFO TLV [RFC7357] for the TRILL APPsub-TLVs defined in Section 4.1, 4.2 and 5.3.2 of this document. The following entries are added to the "TRILL APPsub-TLV Types under IS-IS TLV 251 Application Identifier 1" Registry on the TRILL Parameters IANA web page.

Type	Name	Reference
tbd1[252]	AA-LAALP-GROUP-MAC	[This document]
tbd2[253]	EXTENDED-RBRIDGE-CAP	[This document]
tbd3[254]	AA-LAALP-GROUP-RBRIDGES	[This document]

8.2. Extended RBridge Capabilities Registry

IANA is requested to create a registry under the TRILL Parameters registry as follows:

Name: Extended RBridge Capabilities

Registration Procedure: Expert Review

Reference: [this document]

Bit	Mnemonic	Description	Reference
0	E	Option C Support	[this document]
1	H	Option A Support	[this document]
2-63	-	Unassigned	

8.3 Active Active Flags

IANA is requested to allocate two flag bits, with mnemonic "AA", as follows:

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

Bit	Mnemonic	Description	Reference
0	M4	IPv4 Multicast Router Attached	[RFC7176]
1	M6	IPv6 Multicast Router Attached	[RFC7176]
2	-	Unassigned	
3	ES	ESADI Participation	[RFC7357]
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".

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	[RFC7357]
4	AA	FGLs for Active-Active	[This document]
5-7	-	Unassigned	

9. Acknowledgements

Authors would like to thank the comments and suggestions from Andrew Qu, Donald Eastlake, Erik Nordmark, Fangwei Hu, Liang Xia, Weiguo Hao, Yizhou Li and Mukhtiar Shaikh.

10. References

10.1. 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.

[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.

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

[RFC7357] Zhai, H., Hu, F., Perlman, R., Eastlake 3rd, D., and O. Stokes, "Transparent Interconnection of Lots of Links (TRILL): End Station Address Distribution Information (ESADI) Protocol", RFC 7357, September 2014.

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

[RFC7180bis] D. Eastlake, M. Zhang, et al, "TRILL: Clarifications, Corrections, and Updates", draft-eastlake-trill-rfc7180bis, work in progress.

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

10.2. Informative References

[PN] H. Zhai, T. Senevirathne, et al, "TRILL: Pseudo-Nickname for Active-active Access", draft-ietf-trill-pseudonode-nickname, work in progress.

[ChannelTunnel] Eastlake, D. and Y. Li, "TRILL: RBridge Channel Tunnel Protocol", draft-ietf-trill-channel-tunnel, work 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 for Split Horizon

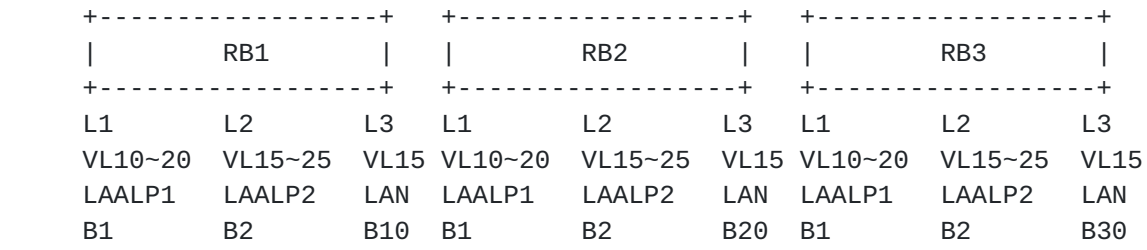


Figure A.1: An example topology to explain split horizon

Suppose RB1, RB2 and RB3 are the Active-Active group connecting LAALP1 and LAALP2. LAALP1 and LAALP2 are connected to B1 and B2 at their other ends. Suppose all these RBridges use port L1 to connect LAALP1 while they use port L2 to connect LAALP2. 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. The 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 and 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 multi-destination data packet with VLAN 15 and ingress nickname RB1. RB3 is the single exit point (selected out according to the hashing function of LAALP) for this packet. On ports L1 and 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 multi-destination packet 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 multi-destination data packet with VLAN 15 and ingress nickname RB4. RB3 is the single exit point. On ports L1 and 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, RB2 and RB3 stores no information of split horizon for AAE (since this port has not been configured to be in any LAALP). They will egress the packet ingressed from 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. This kind of 'bounce' behavior for multi-destination frames is just as specified in paragraph 2 of Section 4.6.1.2 of [RFC6325].

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. (Also see paragraph 2 of Section 4.6.1.2 of [RFC6325].)

Author's Addresses

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

EMail: zhangmingui@huawei.com

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

EMail: radia@alum.mit.edu

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

EMail: honjun.zhai@tom.com

Muhammad Durrani
Brocade
130 Holger Way
San Jose, CA 95134

EMail: mdurrani@brocade.com

Sujay Gupta
IP Infusion,
RMZ Centennial
Mahadevapura Post
Bangalore - 560048
India

EMail: sujay.gupta@ipinfusion.com