

Internet Engineering Task Force (IETF)
Request for Comments: 7782
Category: Standards Track
ISSN: 2070-1721

M. Zhang
Huawei
R. Perlman
EMC
H. Zhai
Astute Technology
M. Durrani
Cisco Systems
S. Gupta
IP Infusion
February 2016

Transparent Interconnection of Lots of Links (TRILL) Active-Active Edge Using Multiple MAC Attachments

Abstract

TRILL (Transparent Interconnection of Lots of Links) 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 document specifies a method by which member RBridges (also referred to as Routing Bridges or TRILL switches) 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 (who are not in the group) will see one host Media Access Control (MAC) address being associated with the multiple RBridges in the group. Such remote edge RBridges are required to maintain all those associations (i.e., MAC attachments) and to not flip-flop among them (as would occur prior to the implementation of this specification). The design goals of this specification are discussed herein.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at <http://www.rfc-editor.org/info/rfc7782>.

Copyright Notice

Copyright (c) 2016 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
2. Acronyms and Terminology	4
3. Overview	5
4. Incremental Deployable Options	6
4.1. Details of Option B	7
4.1.1. Advertising Data Labels for Active-Active Edge	7
4.1.2. Discovery of Active-Active Edge Members	8
4.1.3. Advertising Learned MAC Addresses	9
4.2. Extended RBridge Capability Flags APPsub-TLV	11
5. Meeting the Design Goals	12
5.1. No MAC Address Flip-Flopping (Normal Unicast Egress)	12
5.2. Regular Unicast/Multicast Ingress	12
5.3. Correct Multicast Egress	12
5.3.1. No Duplication (Single Exit Point)	12
5.3.2. No Echo (Split Horizon)	13
5.4. No Black-Hole or Triangular Forwarding	14
5.5. Load Balance towards the AAE	14
5.6. Scalability	14
6. E-L1FS Backward Compatibility	15
7. Security Considerations	15
8. IANA Considerations	16
8.1. TRILL APPsub-TLVs	16
8.2. Extended RBridge Capabilities Registry	16
8.3. Active-Active Flags	17
9. References	17
9.1. Normative References	17
9.2. Informative References	19
Appendix A. Scenarios for Split Horizon	20
Acknowledgments	21
Authors' Addresses	22

1. Introduction

As discussed in [RFC7379], in a TRILL (Transparent Interconnection of Lots of Links) Active-Active Edge (AAE) topology, a Local Active-Active Link Protocol (LAALP) -- for example, a Multi-Chassis Link Aggregation (MC-LAG) bundle -- is used to connect multiple R Bridges (Routing Bridges or TRILL switches) to multi-port Customer Equipment (CE), such as a switch, virtual switch (vSwitch), or multi-port end station. A set of end nodes is attached in the case of a switch or vSwitch. It is required that data traffic within a specific VLAN from this end node set (including the multi-port end-station case) can be ingressed and egressed by any of these R Bridges simultaneously. End systems in the set can spread their traffic among these edge R Bridges 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 end node can be ingressed by any R Bridge in the local AAE group, a remote edge R Bridge may observe multiple attachment points (i.e., egress R Bridges) for this end node. This issue is known as "MAC address flip-flopping"; see [RFC7379] for a discussion.

Per this document, AAE member R Bridges use their own nicknames to ingress frames into the TRILL campus. Remote edge R Bridges are required to keep multiple points of attachment per MAC address and Data Label (VLAN or Fine-Grained Label [RFC7172]) attached to the AAE. This addresses the MAC flip-flopping issue. Using this solution, as specified in this document, 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 document and the specification in [RFC7781] could be simultaneously deployed for different AAE groups in the same campus.

The main body of this document is organized as follows: Section 2 lists acronyms and terms. Section 3 describes the overview model. Section 4 provides options for incremental deployment. Section 5 describes how this approach meets the design goals. Section 6 discusses backward compatibility. Section 7 covers security considerations. Section 8 covers IANA considerations.

2. Acronyms and Terminology

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 Fine-Grained Label (FGL)

DRNI: Distributed Resilient Network Interconnect. A link aggregation specified in [802.1AX] that can provide an LAALP between (a) one, two, or three CEs and (b) two or three RBridges.

E-L1FS: Extended Level 1 Flooding Scope

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]

FS-LSP: Flooding Scope Link State Protocol Data Unit

IS: Intermediate System [IS-IS]

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

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

LSP: Link State PDU

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

PDU: Protocol Data Unit

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.

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

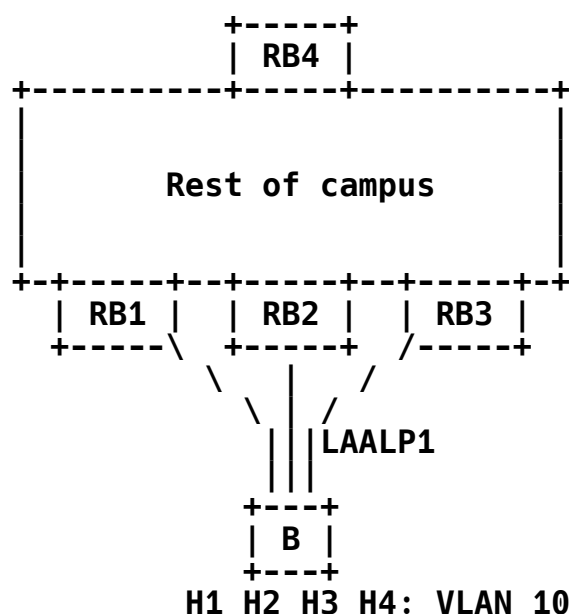


Figure 1: An Example Topology for TRILL Active-Active Edge

Figure 1 shows an example network for TRILL AAE (see also Figure 1 in [RFC7379]). In this figure, end nodes (H1, H2, H3, and H4) are attached to a bridge (B) that communicates with multiple R Bridges (RB1, RB2, and RB3) via the LAALP. Suppose that RB4 is a "remote" R Bridge 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 (VMs).

For a frame received from its attached end node sets, a member RBridge of the AAE group conforming to this document always encapsulates that frame using its own nickname as the ingress nickname, regardless of whether it is unicast or multicast.

With the two options specified below, even though remote RBridge RB4 will see multiple attachments for each MAC address from one of the end nodes, MAC address flip-flopping will not cause any problems.

4. Incremental Deployable Options

This section specifies two options. Option A requires new hardware support. Option B can be incrementally implemented throughout a TRILL campus with common existing TRILL "fast path" hardware. Further details on Option B 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 end node." This mode of operation is generally not supported by existing TRILL fast path hardware. Only if all edge R Bridges 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 R Bridges 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 R Bridges that would need to support this capability but do not support it. Further details for Option A are beyond the scope of this document, except that, as described 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 B.

Option B:

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 R Bridges within the same Data Label in the TRILL ESADI-LSPs. For Option B, AAE member R Bridges make use of the TRILL ESADI protocol to distribute multiple attachments of a MAC address. Remote R Bridges SHOULD disable 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. Details of Option B

With Option B, the receiving edge R Bridges MUST avoid flip-flop errors for MAC addresses learned from the TRILL Data packet decapsulation for the originating R Bridge within these Data Labels. It is RECOMMENDED that the receiving edge R Bridge disable data-plane MAC learning from TRILL Data packet decapsulation within those advertised Data Labels for the originating R Bridge, unless the receiving R Bridge also supports Option A. Alternative implementations that produce the same expected behavior, i.e., the receiving edge R Bridge does not flip-flop among multiple MAC attachments, are acceptable. For example, the confidence-level mechanism as specified in [RFC6325] can be used. Let the receiving edge R Bridge 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. The receiving edge R Bridge 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 a higher confidence value (e.g., 0x80) to override any alternative learning from the decapsulation of received TRILL Data packets [RFC6325].

4.1.1. Advertising Data Labels for Active-Active Edge

An R Bridge in an AAE group MUST participate in ESADI in Data Labels enabled for its attached LAALPs. This document further registers two data flags, which are used to advertise that the originating R Bridge supports and participates in an AAE. These two flags are allocated from the Interested VLANs Flag Bits that appear in the Interested VLANs and Spanning Tree Roots sub-TLV and the Interested Labels Flag Bits that appear in the Interested Labels and Spanning Tree Roots sub-TLV [RFC7176] (see Section 8.3). When these flags are set to 1, the originating R Bridge is advertising Data Labels for LAALPs rather than plain LAN links.

4.1.2. Discovery of Active-Active Edge Members

Remote edge RBridges need to discover RBridges in an AAE. This is achieved by listening to the following "AA LAALP Group RBridges" TRILL APPsub-TLV included in the TRILL GENINFO TLV in FS-LSPs [RFC7780]:

```

+---+---+---+---+---+---+---+---+---+---+
| Type = AA-LAALP-GROUP-RBRIDGES | (2 bytes)
+---+---+---+---+---+---+---+---+---+---+
| Length                           | (2 bytes)
+---+---+---+---+---+---+---+---+---+---+
| Sender Nickname                  | (2 bytes)
+---+---+---+---+---+---+---+---+---+---+
| LAALP ID Size |                  (1 byte)
+---+---+---+---+---+---+---+---+---+---+...+---+
| LAALP ID                        (k bytes) |
+---+---+---+---+---+---+---+---+---+---+...+---+

```

- o Type: AA LAALP Group RBridges (TRILL APPsub-TLV type 252)
- o Length: 3 + k
- o Sender Nickname: The nickname the originating RBridge will use as the ingress nickname. This field is useful because the originating RBridge might own multiple nicknames.
- o LAALP ID Size: The length, k, of the LAALP ID in bytes.
- o 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, as specified in Clause 6.3.2 of [802.1AX].

This APPsub-TLV is expected to rarely change, as it only does so in cases of the creation or elimination of an AAE group, or of link failure or restoration to the CE in such a group.

4.1.3. Advertising Learned MAC Addresses

Whenever MAC addresses from the LAALP of this AAE are learned through ingress or configuration, the originating RBridge **MUST** advertise these MAC addresses using the MAC-Reachability TLV [RFC6165] via the ESADI protocol [RFC7357]. The MAC-Reachability TLVs 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-GROUP-MAC | (2 bytes)
+---+---+---+---+---+---+---+---+---+---+
| Length | (2 bytes)
+---+---+---+---+---+---+---+---+---+---+
| LAALP ID Size | (1 byte)
+---+---+---+---+---+---+---+---+---+---+...+---+
| LAALP ID | (k bytes) |
+---+---+---+---+---+---+---+---+---+---+...+---+
| MAC-Reachability TLV | (7 + 6*n bytes) |
+---+---+---+---+---+---+---+---+---+---+...+---+

```

- o Type: AA LAALP Group MAC (TRILL APPsub-TLV type 253)
- 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 $k + 8 + 6*n$.
- o LAALP ID Size: The length, k , of the LAALP ID in bytes.
- o LAALP ID: The ID of the LAALP, which 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 of [802.1AX].
- o MAC-Reachability sub-TLV: The AA-LAALP-GROUP-MAC APPsub-TLV value contains the MAC-Reachability TLV as a sub-TLV (see [RFC6165]; n is the number of MAC addresses present). As specified in Section 2.2 of [RFC7356], the Type and Length fields of the MAC-Reachability TLV are encoded as unsigned 16-bit integers. The 1-byte unsigned confidence value, 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-TLVs 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 TLVs 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 one of them in a pseudorandom 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.

This APPsub-TLV changes whenever the MAC reachability situation for the LAALP changes.

4.2. Extended RBridge Capability Flags APPsub-TLV

The following Extended RBridge Capability Flags APPsub-TLV will be included in E-L1FS FS-LSP fragment zero [RFC7780] as an APPsub-TLV of the TRILL GENINFO TLV:

```

+-----+
| Type = EXTENDED-RBRIDGE-CAP | (2 bytes)
+-----+
| Length | (2 bytes)
+-----+
| Topology | (2 bytes)
+-----+
| E | H | Reserved
+-----+
| Reserved (continued)
+-----+

```

- o Type: Extended RBridge Capability (TRILL APPsub-TLV type 254)
- o Length: Set to 10.
- o Topology: Indicates the topology to which the capabilities apply. When this field is set to zero, either topologies are not in use or the capabilities apply to all topologies [TRILL-MT].
- o E: Bit 0 of the capability bits. When this bit is set, it indicates that the originating RBridge acts as specified in Option B above.
- o H: Bit 1 of the capability bits. When this bit is set, it indicates that the originating RBridge keeps multiple MAC attachments learned from TRILL Data packet decapsulation with fast path hardware; that is, it acts as specified in Option A above.
- o Reserved: Flags extending from bit 2 through bit 63 of the capability bits. 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 R Bridges as to whether the originating RBridge supports the capability indicated by the E and H bits. For example, if the E bit is set, it indicates that the originating RBridge will act as defined in Option B. That is, it will disable the MAC learning from TRILL Data packet decapsulation within Data Labels advertised by AAE R Bridges 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 is required that MAC addresses

learned from local LAALPs be advertised in TRILL ESADI-LSPs, using the AA-LAALP-GROUP-MAC APPsub-TLV, which is defined in Section 4.1.3. If an RBridge in an AAE group, as specified herein, observes 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 B, then the AAE group MUST fall back to active-standby mode.

This APPsub-TLV is expected to rarely change, as it only needs to be updated when RBridge capabilities change, e.g., due to an upgrade or reconfiguration.

5. Meeting the Design Goals

This section explores how this specification meets the major design goals of AAE.

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

Since all RBridges talking with the AAE RBridges in the campus are able to see multiple attachments 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 consistently use one of the attachments for each MAC address rather than flip-flopping among them (see Section 4.2.6 of [RFC6325] and Section 5.3 of [RFC7357]).

5.2. Regular Unicast/Multicast Ingress

LAALP guarantees that each frame will be sent to the AAE via exactly one uplink. RBridges in the AAE 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 is important that exactly one RBridge out of the AAE group let through each multi-destination

packet so that no duplication will happen. The LAALP will have defined its selection function (using hashing or an election algorithm) to designate 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 receiving RBridge itself (i.e., the packet should be egressed out of this node), the receiving RBridge MUST egress this packet for that AAE group. Otherwise, the packet MUST NOT be egressed for that AAE group. (For ports that lead to non-AAE links, the receiving RBridge determines whether to egress the packet or not, according to [RFC6325], which is updated by [RFC7172].)

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 receiving RBridge does not egress this frame back to this LAALP. Otherwise, it will cause a forwarding loop (echo). The well-known "split horizon" technique (as discussed in Section 2.2.1 of [RFC1058]) is used to eliminate the echo issue.

RBridges in the AAE group need to perform 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 consisting of the tuple of <ingress nickname, VLAN>. Packets with information matching any entry in the filtering list MUST NOT be egressed out of that port. The information for such filters is obtained by listening to the AA-LAALP-GROUP-RBRIDGES TRILL APPsub-TLVs, as defined in Section 4.1.2. Note that all enabled VLANs MUST be consistent on all ports connected to an LAALP. So, the enabled VLANs need not be included in these TRILL APPsub-TLVs. They can be locally obtained from the port attached to that LAALP. By parsing these APPsub-TLVs, the receiving RBridge discovers all other RBridges connected to the same LAALP. The Sender Nickname of the originating RBridge will be added to the filtering list of the port attached to the LAALP. For example, RB3 in Figure 1 will set up a filtering list that looks like {<RB1, VLAN 10>, <RB2, VLAN 10>} on its port attached to LAALP1. According to split horizon, TRILL Data packets within VLAN 10 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 VLANs that overlap. Here, a VLAN overlap means that this VLAN ID is enabled by multiple LAALPs. A customer may require that hosts within these overlapped VLANs communicate with each other.

Appendix A provides several scenarios to explain how hosts communicate within the overlapped 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 makes use of the ESADI protocol to flush those MAC addresses affected by the failure, as defined in Section 5.2 of [RFC7357]. After doing that, no packets will be sent towards the failed port, and hence no black-hole will happen. Nor will the member RBridge need to redirect packets to other member RBridges; thus, triangular forwarding is avoided.

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 path selection 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 pseudorandom selection method (see Section 4.1.3).

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 table memory space required by multi-attached MAC addresses can usually be accommodated in an RBridge's unused MAC table space.

With Option B, remote RBridges will keep the multiple attachments of a MAC address in the ESADI link-state databases, which are usually maintained by software. In the MAC table, which is normally implemented in hardware, an RBridge still establishes only one entry for each MAC address.

6. E-L1FS Backward Compatibility

The Extended TLVs defined in Sections 4.1.2, 4.1.3, and 4.2 of this document are to be used in an Extended Level 1 Flooding Scope (E-L1FS) PDU [RFC7356] [RFC7780]. For those RBridges that do not support E-L1FS, the EXTENDED-RBRIDGE-CAP TRILL APPsub-TLV will not be sent out either, and MAC multi-attach active-active is not supported.

7. Security Considerations

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

For extensions not transported by TRILL ESADI, RBridges may be configured to include the IS-IS Authentication TLV (10) in the IS-IS PDUs to use IS-IS security [RFC5304] [RFC5310].

Since currently deployed LAALPs [RFC7379] are proprietary, security over membership in, and internal management of, AAE groups is proprietary. In environments where the above authentication is not adopted, a rogue RBridge that insinuates itself into an AAE group can disrupt end-station traffic flowing into or out of that group. For example, if there are N RBridges in the group, it could typically control 1/Nth of the traffic flowing out of that group and a similar amount of unicast traffic flowing into that group.

For general TRILL security considerations, see [RFC6325].

8. IANA Considerations

8.1. TRILL APPsub-TLVs

IANA has allocated three new types under the TRILL GENINFO TLV [RFC7357] for the TRILL APPsub-TLVs defined in Sections 4.1.2, 4.1.3, and 4.2 of this document. The following entries have been 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
----	-----	-----
252	AA-LAALP-GROUP-RBRIDGES	RFC 7782
253	AA-LAALP-GROUP-MAC	RFC 7782
254	EXTENDED-RBRIDGE-CAP	RFC 7782

8.2. Extended RBridge Capabilities Registry

IANA has created a registry under the "Transparent Interconnection of Lots of Links (TRILL) Parameters" registry as follows:

Name: Extended RBridge Capabilities

Registration Procedure: Expert Review

Reference: RFC 7782

Bit	Mnemonic	Description	Reference
----	-----	-----	-----
0	E	Option B Support	RFC 7782
1	H	Option A Support	RFC 7782
2-63	-	Unassigned	

8.3. Active-Active Flags

IANA has allocated two flag bits, with mnemonic "AA", as follows:

One flag bit is allocated from the Interested VLANs Flag Bits.

Bit	Mnemonic	Description	Reference
---	-----	-----	-----
16	AA	VLANs for Active-Active	RFC 7782

One flag bit is allocated from the Interested Labels Flag Bits.

Bit	Mnemonic	Description	Reference
---	-----	-----	-----
4	AA	FGLs for Active-Active	RFC 7782

9. References

9.1. Normative References

- [802.1AX] IEEE, "IEEE Standard for Local and metropolitan area networks - Link Aggregation", IEEE Std 802.1AX-2014, DOI 10.1109/IEEESTD.2014.7055197, December 2014.
- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <<http://www.rfc-editor.org/info/rfc2119>>.
- [RFC6165] Banerjee, A. and D. Ward, "Extensions to IS-IS for Layer-2 Systems", RFC 6165, DOI 10.17487/RFC6165, April 2011, <<http://www.rfc-editor.org/info/rfc6165>>.
- [RFC6325] Perlman, R., Eastlake 3rd, D., Dutt, D., Gai, S., and A. Ghanwani, "Routing Bridges (Rbridges): Base Protocol Specification", RFC 6325, DOI 10.17487/RFC6325, July 2011, <<http://www.rfc-editor.org/info/rfc6325>>.
- [RFC6439] Perlman, R., Eastlake, D., Li, Y., Banerjee, A., and F. Hu, "Routing Bridges (Rbridges): Appointed Forwarders", RFC 6439, DOI 10.17487/RFC6439, November 2011, <<http://www.rfc-editor.org/info/rfc6439>>.
- [RFC7172] Eastlake 3rd, D., Zhang, M., Agarwal, P., Perlman, R., and D. Dutt, "Transparent Interconnection of Lots of Links (TRILL): Fine-Grained Labeling", RFC 7172, DOI 10.17487/RFC7172, May 2014, <<http://www.rfc-editor.org/info/rfc7172>>.

- [RFC7176] Eastlake 3rd, D., Senevirathne, T., Ghanwani, A., Dutt, D., and A. Banerjee, "Transparent Interconnection of Lots of Links (TRILL) Use of IS-IS", RFC 7176, DOI 10.17487/RFC7176, May 2014, <<http://www.rfc-editor.org/info/rfc7176>>.
- [RFC7177] Eastlake 3rd, D., Perlman, R., Ghanwani, A., Yang, H., and V. Manral, "Transparent Interconnection of Lots of Links (TRILL): Adjacency", RFC 7177, DOI 10.17487/RFC7177, May 2014, <<http://www.rfc-editor.org/info/rfc7177>>.
- [RFC7356] Ginsberg, L., Previdi, S., and Y. Yang, "IS-IS Flooding Scope Link State PDUs (LSPs)", RFC 7356, DOI 10.17487/RFC7356, September 2014, <<http://www.rfc-editor.org/info/rfc7356>>.
- [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, DOI 10.17487/RFC7357, September 2014, <<http://www.rfc-editor.org/info/rfc7357>>.
- [RFC7780] Eastlake 3rd, D., Zhang, M., Perlman, R., Banerjee, A., Ghanwani, A., and S. Gupta, "Transparent Interconnection of Lots of Links (TRILL): Clarifications, Corrections, and Updates", RFC 7780, DOI 10.17487/RFC7780, February 2016, <<http://www.rfc-editor.org/info/rfc7780>>.

9.2. Informative References

- [IS-IS] International Organization for Standardization, "Information technology -- Telecommunications and information exchange between systems -- Intermediate System to Intermediate System intra-domain routing information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode network service (ISO 8473)", ISO/IEC 10589:2002, Second Edition, November 2002.
- [RFC1058] Hedrick, C., "Routing Information Protocol", RFC 1058, DOI 10.17487/RFC1058, June 1988, <<http://www.rfc-editor.org/info/rfc1058>>.
- [RFC5304] Li, T. and R. Atkinson, "IS-IS Cryptographic Authentication", RFC 5304, DOI 10.17487/RFC5304, October 2008, <<http://www.rfc-editor.org/info/rfc5304>>.
- [RFC5310] Bhatia, M., Manral, V., Li, T., Atkinson, R., White, R., and M. Fanto, "IS-IS Generic Cryptographic Authentication", RFC 5310, DOI 10.17487/RFC5310, February 2009, <<http://www.rfc-editor.org/info/rfc5310>>.
- [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, DOI 10.17487/RFC7379, October 2014, <<http://www.rfc-editor.org/info/rfc7379>>.
- [RFC7781] Zhai, H., Senevirathne, T., Perlman, R., Zhang, M., and Y. Li, "Transparent Interconnection of Lots of Links (TRILL): Pseudo-Nickname for Active-Active Access", RFC 7781, DOI 10.17487/RFC7781, February 2016, <<http://www.rfc-editor.org/info/rfc7781>>.
- [TRILL-MT] Eastlake 3rd, D., Zhang, M., Banerjee, A., and V. Manral, "TRILL: Multi-Topology", Work in Progress, draft-ietf-trill-multi-topology-00, September 2015.

Appendix A. Scenarios for Split Horizon

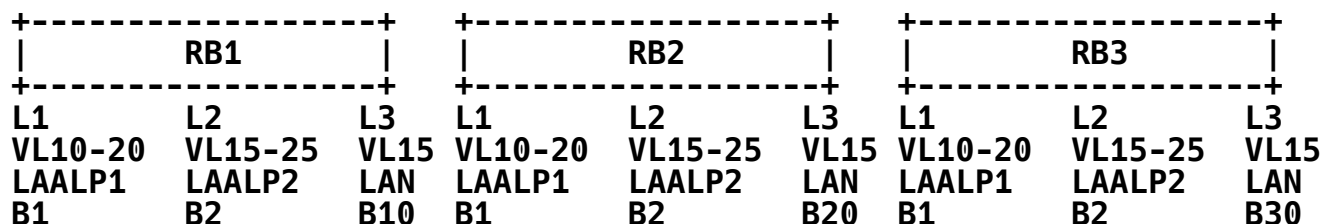


Figure 2: An Example Topology to Explain Split Horizon

Suppose that 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 that all these RBridges use port L1 to connect LAALP1 while they use port L2 to connect LAALP2. Assume that all three L1 ports enable VLANs 10-20 while all three L2 ports enable VLANs 15-25, so that there is an overlap of VLANs 15-20. A customer may require that hosts within these overlapped VLANs communicate with each other. That is, hosts attached to B1 in VLANs 15-20 need to communicate with hosts attached to B2 in VLANs 15-20. Assume that the remote plain RBridge RB4 also has hosts attached in VLANs 15-20 that need to communicate with those hosts in these VLANs attached to B1 and B2.

There are two major requirements:

1. Frames ingressed from RB1-L1-VLANs 15-20 MUST NOT be egressed out of ports RB2-L1 and RB3-L1.
2. At the same time, frames coming from B1-VLANs 15-20 should reach B2-VLANs 15-20.

RB3 stores the information for split horizon on its ports L1 and L2.

On L1: {<ingress_nickname_RB1, VLANs 10-20>, <ingress_nickname_RB2, VLANs 10-20>}.

On L2: {<ingress_nickname_RB1, VLANs 15-25>, <ingress_nickname_RB2, VLANs 15-25>}.

Five clarification scenarios follow:

- a. Suppose that RB2 or RB3 receives a TRILL multi-destination data packet with VLAN 15 and ingress_nickname_RB1. RB3 is the single exit point (selected 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 determines that 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 that RB2 or 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 that 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 store no information for 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 3 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 ports L1 and L2. (Also see paragraph 3 of Section 4.6.1.2 of [RFC6325].)

Acknowledgments

The authors would like to thank the following people for their comments and suggestions: Andrew Qu, Donald Eastlake, Erik Nordmark, Fangwei Hu, Liang Xia, Weiguo Hao, Yizhou Li, and Mukhtiar Shaikh.

Authors' Addresses

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

Email: zhangmingui@huawei.com

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

Email: radia@alum.mit.edu

Hongjun Zhai
Nanjing Astute Software Technology Co. Ltd
57 Andemen Avenue, Yuhuatai District
Nanjing, Jiangsu 210016
China

Email: honjun.zhai@tom.com

Muhammad Durrani
Cisco Systems
170 West Tasman Dr.
San Jose, CA 95134
United States

Email: mdurrani@cisco.com

Sujay Gupta
IP Infusion
RMZ Centennial
Mahadevapura Post
Bangalore 560048
India

Email: sujay.gupta@ipinfusion.com