Internet Engineering Task Force (IETF)
Request for Comments: 6325
Category: Standards Track
ISSN: 2070-1721

R. Perlman
Intel Labs
D. Eastlake 3rd
Huawei
D. Dutt
S. Gai
Cisco Systems
A. Ghanwani
Brocade

July 2011

Routing Bridges (RBridges): Base Protocol Specification

Abstract

Routing Bridges (RBridges) provide optimal pair-wise forwarding without configuration, safe forwarding even during periods of temporary loops, and support for multipathing of both unicast and multicast traffic. They achieve these goals using IS-IS routing and encapsulation of traffic with a header that includes a hop count.

RBridges are compatible with previous IEEE 802.1 customer bridges as well as IPv4 and IPv6 routers and end nodes. They are as invisible to current IP routers as bridges are and, like routers, they terminate the bridge spanning tree protocol.

The design supports VLANs and the optimization of the distribution of multi-destination frames based on VLAN ID and based on IP-derived multicast groups. It also allows unicast forwarding tables at transit RBridges to be sized according to the number of RBridges (rather than the number of end nodes), which allows their forwarding tables to be substantially smaller than in conventional customer bridges.

Status of This Memo

Copyright Notice

Table of Contents

Table of Figures

1. Introduction

In traditional IPv4 and IPv6 networks, each subnet has a unique prefix. Therefore, a node in multiple subnets has multiple IP addresses, typically one per interface. This also means that when an interface moves from one subnet to another, it changes its IP address. Administration of IP networks is complicated because IP routers require per-port subnet address configuration. Careful IP address management is required to avoid creating subnets that are sparsely populated, wasting addresses.

IEEE 802.1 bridges avoid these problems by transparently gluing many physical links into what appears to IP to be a single LAN [802.1D]. However, 802.1 bridge forwarding using the spanning tree protocol has some disadvantages:

- The spanning tree protocol works by blocking ports, limiting the number of forwarding links, and therefore creates bottlenecks by concentrating traffic onto selected links.

- Forwarding is not pair-wise shortest path, but is instead whatever path remains after the spanning tree eliminates redundant paths.

- The Ethernet header does not contain a hop count (or Time to Live (TTL)) field. This is dangerous when there are temporary loops such as when spanning tree messages are lost or components such as repeaters are added.

- VLANs can partition when the spanning tree reconfigures.

This document presents the design for RBridges (Routing Bridges [RBridges]) that implement the TRILL protocol and are poetically summarized below. Rbridges combine the advantages of bridges and routers and, as specified in this document, are the application of link state routing to the VLAN-aware customer bridging problem. With the exceptions discussed in this document, RBridges can incrementally replace IEEE [802.1Q-2005] or [802.1D] customer bridges.

While RBridges can be applied to a variety of link protocols, this specification focuses on IEEE [802.3] links. Use with other link types is expected to be covered in other documents.

The TRILL protocol, as specified herein, is designed to be a Local Area Network protocol and not designed with the goal of scaling beyond the size of existing bridged LANs. For further discussion of the problem domain addressed by RBridges, see [RFC5556].

1.1. Algorhyme V2, by Ray Perlner

I hope that we shall one day see
A graph more lovely than a tree.

A graph to boost efficiency
While still configuration-free.

A network where RBridges can
Route packets to their target LAN.

The paths they find, to our elation,
Are least cost paths to destination!

With packet hop counts we now see,
The network need not be loop-free!

RBridges work transparently,
Without a common spanning tree.

1.2. Normative Content and Precedence

The bulk of the normative material in this specification appears in Sections 1 through 4. In case of conflict between provisions in these four sections, the provision in the higher numbered section prevails.

1.3. Terminology and Notation in This Document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [RFC2119].

"TRILL" is the protocol specified herein while an "RBridge" is a device that implements that protocol. The second letter in Rbridge is case insensitive. Both Rbridge and RBridge are correct.

In this document, the term "link", unless otherwise qualified, means "bridged LAN", that is to say, the combination of one or more [802.3] links with zero or more bridges, hubs, repeaters, or the like. The term "simple link" or the like is used indicate a point-to-point or multi-access link with no included bridges or RBridges.

In this document, the term "port", unless otherwise qualified, includes physical, virtual [802.1AE], and pseudo [802.1X] ports. The term "physical port" or the like is used to indicate the physical point of connection between an RBridge and a link.

A "campus" is to RBridges as a "bridged LAN" is to bridges. An RBridge campus consists of a network of RBridges, bridges, hubs, repeaters, simple links, and the like and it is bounded by end stations and routers.

The term "spanning tree" in this document includes both classic spanning tree and rapid spanning tree (as in the Rapid Spanning Tree Protocol).

This document uses hexadecimal notation for MAC addresses. Two hexadecimal digits represent each octet (that is, 8-bit byte), giving the value of the octet as an unsigned integer. A hyphen separates successive octets. This document consistently uses IETF bit ordering, although the physical order of bit transmission within an octet on an IEEE [802.3] link is from the lowest order bit to the highest order bit, the reverse of IETF ordering.

1.4. Categories of Layer 2 Frames

In this document, Layer 2 frames are divided into five categories:

- Layer 2 control frames (such as Bridge PDUs (BPDUs))
- native frames (non-TRILL-encapsulated data frames)
- TRILL Data frames (TRILL-encapsulated data frames)
- TRILL control frames
- TRILL other frames

The way these five types of frames are distinguished is as follows:

- Layer 2 control frames are those with a multicast destination address in the range 01-80-C2-00-00-00 to 01-80-C2-00-00-0F or equal to 01-80-C2-00-00-21. RBridges MUST NOT encapsulate and forward such frames, though they MAY, unless otherwise specified in this document, perform the Layer 2 function (such as MAC-level security) of the control frame. Frames with a destination address of 01-80-C2-00-00-00 (BPDU) or 01-80-C2-00-00-21 (VLAN Registration Protocol) are called "high-level control frames" in this document. All other Layer 2 control frames are called "low-level control frames".

- Native frames are those that are not control frames and have an Ethertype other than "TRILL" or "L2-IS-IS" and have a destination MAC address that is not one of the 16 multicast addresses reserved for TRILL.

- TRILL Data frames have the Ethertype "TRILL". In addition, TRILL data frames, if multicast, have the multicast destination MAC address "All-RBridges".

- TRILL control frames have the Ethertype "L2-IS-IS". In addition, TRILL control frames, if multicast, have the multicast destination MAC addresses of "All-IS-IS-RBridges". (Note that ESADI frames look on the outside like TRILL data and are so handled but, when decapsulated, have the L2-IS-IS Ethertype.)

- TRILL other frames are those with any of the 16 multicast destination addresses reserved for TRILL other than All-RBridges and All-IS-IS-RBridges. RBridges conformant to this specification MUST discard TRILL other frames.

1.5. Acronyms

AllL1ISs - All Level 1 Intermediate Systems

AllL2ISs - All Level 2 Intermediate Systems

BPDU - Bridge PDU

CHbH - Critical Hop-by-Hop

CItE - Critical Ingress-to-Egress

CSNP - Complete Sequence Number PDU

DA - Destination Address

DR - Designated Router

DRB - Designated RBridge

EAP - Extensible Authentication Protocol

ECMP - Equal Cost Multipath

EISS - Extended Internal Sublayer Service

ESADI - End-Station Address Distribution Information

FCS - Frame Check Sequence

GARP - Generic Attribute Registration Protocol

GVRP - GARP VLAN Registration Protocol

IEEE - Institute of Electrical and Electronics Engineers

IGMP - Internet Group Management Protocol

IP - Internet Protocol

IS-IS - Intermediate System to Intermediate System

ISS - Internal Sublayer Service

LAN - Local Area Network

LSP - Link State PDU

MAC - Media Access Control

MLD - Multicast Listener Discovery

MRD - Multicast Router Discovery

MTU - Maximum Transmission Unit

MVRP - Multiple VLAN Registration Protocol

NSAP - Network Service Access Point

P2P - Point-to-point

PDU - Protocol Data Unit

PPP - Point-to-Point Protocol

RBridge - Routing Bridge

RPF - Reverse Path Forwarding

SA - Source Address

SNMP - Simple Network Management Protocol

SPF - Shortest Path First

TLV - Type, Length, Value

TRILL - TRansparent Interconnection of Lots of Links

VLAN - Virtual Local Area Network

VRP - VLAN Registration Protocol

2. RBridges

This section provides a high-level overview of RBridges, which implement the TRILL protocol, omitting some details. Sections 3 and 4 below provide more detailed specifications.

TRILL, as described in this document and with the exceptions discussed herein, provides [802.1Q-2005] VLAN-aware customer bridging service. As described below, TRILL is layered above the ports of an RBridge.

The RBridges specified by this document do not supply provider [802.1ad] or provider backbone [802.1ah] bridging or the like. The extension of TRILL to provide such provider services is left for future work that will be separately documented. However, provider or provider backbone bridges may be used to interconnect parts of an RBridge campus.

2.1. General Overview

RBridges run a link state protocol amongst themselves. This gives them enough information to compute pair-wise optimal paths for unicast, and calculate distribution trees for delivery of frames either to destinations whose location is unknown or to multicast/broadcast groups [RBridges] [RP1999].

To mitigate temporary loop issues, RBridges forward based on a header with a hop count. RBridges also specify the next hop RBridge as the frame destination when forwarding unicast frames across a shared-media link, which avoids spawning additional copies of frames during a temporary loop. A Reverse Path Forwarding Check and other checks are performed on multi-destination frames to further control potentially looping traffic (see Section 4.5.2).

The first RBridge that a unicast frame encounters in a campus, RB1, encapsulates the received frame with a TRILL header that specifies the last RBridge, RB2, where the frame is decapsulated. RB1 is known as the "ingress RBridge" and RB2 is known as the "egress RBridge". To save room in the TRILL header and simplify forwarding lookups, a dynamic nickname acquisition protocol is run among the RBridges to select 2-octet nicknames for RBridges, unique within the campus, which are an abbreviation for the IS-IS ID of the RBridge. The 2-octet nicknames are used to specify the ingress and egress RBridges in the TRILL header.

Multipathing of multi-destination frames through alternative distribution trees and ECMP (Equal Cost Multipath) of unicast frames are supported (see Appendix C).

Networks with a more mesh-like structure will benefit to a greater extent from the multipathing and optimal paths provided by TRILL than will more tree-like networks.

RBridges run a protocol on a link to elect a "Designated RBridge" (DRB). The TRILL-IS-IS election protocol on a link is a little different from the Layer 3 IS-IS [ISO10589] election protocol, because in TRILL it is essential that only one RBridge be elected DRB, whereas in Layer 3 IS-IS it is possible for multiple routers to be elected Designated Router (also known as Designated Intermediate System). As with an IS-IS router, the DRB may give a pseudonode name to the link, issue an LSP (Link State PDU) on behalf of the pseudonode, and issues CSNPs (Complete Sequence Number PDUs) on the link. Additionally, the DRB has some TRILL-specific duties, including specifying which VLAN will be the Designated VLAN used for communication between RBridges on that link (see Section 4.2.4.2).

The DRB either encapsulates/decapsulates all data traffic to/from the link, or, for load splitting, delegates this responsibility, for one or more VLANs, to other RBridges on the link. There must at all times be at most one RBridge on the link that encapsulates/decapsulates traffic for a particular VLAN. We will refer to the RBridge appointed to forward VLAN-x traffic on behalf of the link as the "appointed VLAN-x forwarder" (see Section 4.2.4.3). (Section 2.5 discusses VLANs further.)

Rbridges SHOULD support SNMPv3 [RFC3411]. The Rbridge MIB will be specified in a separate document. If IP service is available to an RBridge, it SHOULD support SNMPv3 over UDP over IPv4 [RFC3417] and IPv6 [RFC3419]; however, management can be used, within a campus, even for an RBridge that lacks an IP or other Layer 3 transport stack or which does not have a Layer 3 address, by transporting SNMP with Ethernet [RFC4789].

2.2. End-Station Addresses

An RBridge, RB1, that is the VLAN-x forwarder on any of its links MUST learn the location of VLAN-x end nodes, both on the links for which it is VLAN-x forwarder and on other links in the campus. RB1 learns the port, VLAN, and Layer 2 (MAC) addresses of end nodes on links for which it is VLAN-x forwarder from the source address of frames received, as bridges do (for example, see Section 8.7 of [802.1Q-2005]), or through configuration or a Layer 2 explicit registration protocol such as IEEE 802.11 association and authentication. RB1 learns the VLAN and Layer 2 address of distant VLAN-x end nodes, and the corresponding RBridge to which they are attached, by looking at the ingress RBridge nickname in the TRILL header and the VLAN and source MAC address of the inner frame of TRILL Data frames that it decapsulates.

Additionally, an RBridge that is the appointed VLAN-x forwarder on one or more links MAY use the End-Station Address Distribution Information (ESADI) protocol to announce some or all of the attached VLAN-x end nodes on those links.

The ESADI protocol could be used to announce end nodes that have been explicitly enrolled. Such information might be more authoritative than that learned from data frames being decapsulated onto the link. Also, the addresses enrolled and distributed in this way can be more secure for two reasons: (1) the enrollment might be authenticated (for example, by cryptographically based EAP methods via [802.1X]), and (2) the ESADI protocol also supports cryptographic authentication of its messages [RFC5304] [RFC5310] for more secure transmission.

If an end station is unplugged from one RBridge and plugged into another, then, depending on circumstances, frames addressed to that end station can be black-holed. That is, they can be sent just to the older RBridge that the end station used to be connected to until cached address information at some remote RBridge(s) times out, possibly for a number of minutes or longer. With the ESADI protocol, the link interruption from the unplugging can cause an immediate update to be sent.

Even if the ESADI protocol is used to announce or learn attached end nodes, RBridges MUST still learn from received native frames and decapsulated TRILL Data frames unless configured not to do so. Advertising end nodes using ESADI is optional, as is learning from these announcements.

(See Section 4.8 for further end-station address details.)

2.3. RBridge Encapsulation Architecture

The Layer 2 technology used to connect Rbridges may be either IEEE [802.3] or some other link technology such as PPP [RFC1661]. This is possible since the RBridge relay function is layered on top of the Layer 2 technologies. However, this document specifies only an IEEE 802.3 encapsulation.

Figure 1 shows two RBridges, RB1 and RB2, interconnected through an Ethernet cloud. The Ethernet cloud may include hubs, point-to-point or shared media, IEEE 802.1D bridges, or 802.1Q bridges.

```
                           ------------
                          /            \
                +-----+  /   Ethernet   \   +-----+
                | RB1 |----<                >---| RB2 |
                +-----+  \    Cloud     /   +-----+
                          \            /
                           ------------
```

Figure 1: Interconnected RBridges

Figure 2 shows the format of a TRILL data or ESADI frame traveling through the Ethernet cloud between RB1 and RB2.

```
                   +--------------------------------+
                   |     Outer Ethernet Header      |
                   +--------------------------------+
                   |          TRILL Header          |
                   +--------------------------------+
                   |     Inner Ethernet Header      |
                   +--------------------------------+
                   |        Ethernet Payload        |
                   +--------------------------------+
                   |         Ethernet FCS           |
                   +--------------------------------+
```

Figure 2: An Ethernet Encapsulated TRILL Frame

In the case of media different from Ethernet, the header specific to that media replaces the outer Ethernet header. For example, Figure 3 shows a TRILL encapsulation over PPP.

```
                   +--------------------------------+
                   |           PPP Header           |
                   +--------------------------------+
                   |          TRILL Header          |
                   +--------------------------------+
                   |     Inner Ethernet Header      |
                   +--------------------------------+
                   |        Ethernet Payload        |
                   +--------------------------------+
                   |            PPP FCS             |
                   +--------------------------------+
```

Figure 3: A PPP Encapsulated TRILL Frame

The outer header is link-specific and, although this document specifies only [802.3] links, other links are allowed.

In both cases, the inner Ethernet header and the Ethernet Payload come from the original frame and are encapsulated with a TRILL header as they travel between RBridges. Use of a TRILL header offers the following benefits:

1. loop mitigation through use of a hop count field;

2. elimination of the need for end-station VLAN and MAC address learning in transit RBridges;

3. direction of unicast frames towards the egress RBridge (this enables unicast forwarding tables of transit RBridges to be sized with the number of RBridges rather than the total number of end nodes); and

4. provision of a separate VLAN tag for forwarding traffic between RBridges, independent of the VLAN of the native frame.

When forwarding unicast frames between RBridges, the outer header has the MAC destination address of the next hop Rbridge, to avoid frame duplication if the inter-RBridge link is multi-access. This also enables multipathing of unicast, since the transmitting RBridge can specify the next hop. Having the outer header specify the transmitting RBridge as the source address ensures that any bridges inside the Ethernet cloud will not get confused, as they might be if multipathing is in use and they were to see the original source or ingress RBridge in the outer header.

2.4. Forwarding Overview

RBridges are true routers in the sense that, in the forwarding of a frame by a transit RBridge, the outer Layer 2 header is replaced at each hop with an appropriate Layer 2 header for the next hop, and a hop count is decreased. Despite these modifications of the outer Layer 2 header and the hop count in the TRILL header, the original encapsulated frame is preserved, including the original frame’s VLAN tag. See Section 4.6 for more details.

From a forwarding standpoint, transit frames may be classified into two categories: known-unicast and multi-destination. Layer 2 control frames and TRILL control and TRILL other frames are not transit frames, are not forwarded by RBridges, and are not included in these categories.

2.4.1. Known-Unicast

These frames have a unicast inner MAC destination address (Inner.MacDA) and are those for which the ingress RBridge knows the egress RBridge for the destination MAC address in the frame’s VLAN.

Such frames are forwarded Rbridge hop by Rbridge hop to their egress Rbridge.

2.4.2. Multi-Destination

These are frames that must be delivered to multiple destinations.

Multi-destination frames include the following:

1. unicast frames for which the location of the destination is unknown: the Inner.MacDA is unicast, but the ingress RBridge does not know its location in the frame’s VLAN.

2. multicast frames for which the Layer 2 destination address is derived from an IP multicast address: the Inner.MacDA is multicast, from the set of Layer 2 multicast addresses derived from IPv4 [RFC1112] or IPv6 [RFC2464] multicast addresses. These frames are handled somewhat differently in different subcases:

   2.1. IGMP [RFC3376] and MLD [RFC2710] multicast group membership reports

   2.2. IGMP [RFC3376] and MLD [RFC2710] queries and MRD [RFC4286] announcement messages

   2.3. other IP-derived Layer 2 multicast frames

3. multicast frames for which the Layer 2 destination address is not derived from an IP multicast address: the Inner.MacDA is multicast, and not from the set of Layer 2 multicast addresses derived from IPv4 or IPv6 multicast addresses.

4. broadcast frames: the Inner.MacDA is broadcast (FF-FF-FF-FF-FF-FF).

RBridges build distribution trees (see Section 4.5) and use these trees for forwarding multi-destination frames. Each distribution tree reaches all RBridges in the campus, is shared across all VLANs, and may be used for the distribution of a native frame that is in any VLAN. However, the distribution of any particular frame on a distribution tree is pruned in different ways for different cases to avoid unnecessary propagation of the frame.

2.5. RBridges and VLANs

A VLAN is a way to partition end nodes in a campus into different Layer 2 communities [802.1Q-2005]. Use of VLANs requires configuration. By default, the port of receipt determines the VLAN of a frame sent by an end station. End stations can also explicitly insert this information in a frame.

IEEE [802.1Q-2005] bridges can be configured to support multiple customer VLANs over a single simple link by inserting/removing a VLAN tag in the frame. VLAN tags used by TRILL have the same format as VLAN tags defined in IEEE [802.1Q-2005]. As shown in Figure 2, there are two places where such tags may be present in a TRILL-encapsulated frame sent over an IEEE [802.3] link: one in the outer header (Outer.VLAN) and one in the inner header (Inner.VLAN). Inner and outer VLANs are further discussed in Section 4.1.

RBridges enforce delivery of a native frame originating in a particular VLAN only to other links in the same VLAN; however, there are a few differences in the handling of VLANs between an RBridge campus and an 802.1 bridged LAN as described below.

(See Section 4.2.4 for further discussion of TRILL IS-IS operation on a link.)

2.5.1. Link VLAN Assumptions

Certain configurations of bridges may cause partitions of a VLAN on a link. For such configurations, a frame sent by one RBridge to a neighbor on that link might not arrive, if tagged with a VLAN that is partitioned due to bridge configuration.

TRILL requires at least one VLAN per link that gives full connectivity to all the RBridges on that link. The default VLAN is 1, though RBridges may be configured to use a different VLAN. The DRB dictates to the other RBridges which VLAN to use.

Since there will be only one appointed forwarder for any VLAN, say, VLAN-x, on a link, if bridges are configured to cause VLAN-x to be partitioned on a link, some VLAN-x end nodes on that link may be orphaned (unable to communicate with the rest of the campus).

It is possible for bridge and port configuration to cause VLAN mapping on a link (where a VLAN-x frame turns into a VLAN-y frame). TRILL detects this by inserting a copy of the outer VLAN into TRILL-Hello messages and checking it on receipt. If detected, it takes steps to ensure that there is at most a single appointed forwarder on the link, to avoid possible frame duplication or loops (see Section 4.4.5).

TRILL behaves as conservatively as possible, avoiding loops rather than avoiding partial connectivity. As a result, lack of connectivity may result from bridge or port misconfiguration.

2.6. RBridges and IEEE 802.1 Bridges

RBridge ports are, except as described below, layered on top of IEEE [802.1Q-2005] port facilities.

2.6.1. RBridge Ports and 802.1 Layering

RBridge ports make use of [802.1Q-2005] port VLAN and priority processing. In addition, they MAY implement other lower-level 802.1 protocols as well as protocols for the link in use, such as PAUSE (Annex 31B of [802.3]), port-based access control [802.1X], MAC security [802.1AE], or link aggregation [802.1AX].

However, RBridges do not use spanning tree and do not block ports as spanning tree does. Figure 4 shows a high-level diagram of an RBridge with one port connected to an IEEE 802.3 link. Single lines represent the flow of control information, double lines the flow of both frames and control information.

```
                          +-----------------------------------------
                          |                RBridge
                          |
                          |     Forwarding Engine, IS-IS, etc.
                          | Processing of native and TRILL frames
                          |
                          +----+---+--------++----------------------
                               |   |        ||         other ports...
                 +-------------+   |        ||
                 |                 |        ||
    +------------+-------------+   |        ||
    |         RBridge          |   |   +----++-------+ <- EISS
    |                          |   |   |             |
    | High-Level Control Frame |   |   | 802.1Q-2005 |
    |  Processing (BPDU, VRP)  |   |   |  Port VLAN  |
    |                          |   |   |  & Priority |
    +-----------++-------------+   |   |  Processing |
                ||                 |   |             |
      +---------++-----------------+---+-------------+ <-- ISS
      |                                              |
      |    802.1/802.3 Low-Level Control Frame       |
      |    Processing, Port/Link Control Logic       |
      |                                              |
      +-----------++---------------------------------+
                  ||
                  ||        +------------+
                  ||        | 802.3 PHY  |
                  |+--------+ (Physical  +--------- 802.3
                  +---------+ Interface) +--------- Link
                            |            |
                            +------------+
```

Figure 4: RBridge Port Model

The upper interface to the low-level port/link control logic corresponds to the Internal Sublayer Service (ISS) in [802.1Q-2005]. In RBridges, high-level control frames are processed above the ISS interface.

The upper interface to the port VLAN and priority processing corresponds to the Extended Internal Sublayer Service (EISS) in [802.1Q-2005]. In RBridges, native and TRILL frames are processed above the EISS interface and are subject to port VLAN and priority processing.

2.6.2. Incremental Deployment

Because RBridges are compatible with IEEE [802.1Q-2005] customer bridges, except as discussed in this document, a bridged LAN can be upgraded by incrementally replacing such bridges with RBridges. Bridges that have not yet been replaced are transparent to RBridge traffic. The physical links directly interconnected by such bridges, together with the bridges themselves, constitute bridged LANs. These bridged LANs appear to RBridges to be multi-access links.

If the bridges replaced by RBridges were default configuration bridges, then their RBridge replacements will not require configuration.

Because RBridges, as described in this document, only provide customer services, they cannot replace provider bridges or provider backbone bridges, just as a customer bridge can’t replace a provider bridge. However, such provider devices can be part of the bridged LAN between RBridges. Extension of TRILL to support provider services is left for future work and will be separately documented.

Of course, if the bridges replaced had any port level protocols enabled, such as port-based access control [802.1X] or MAC security [802.1AE], replacement RBridges would need the same port level protocols enabled and similarly configured. In addition, the replacement RBridges would have to support the same link type and link level protocols as the replaced bridges.

An RBridge campus will work best if all IEEE [802.1D] and [802.1Q-2005] bridges are replaced with RBridges, assuming the RBridges have the same speed and capacity as the bridges. However, there may be intermediate states, where only some bridges have been replaced by RBridges, with inferior performance.

See Appendix A for further discussion of incremental deployment.

3. Details of the TRILL Header

This section specifies the TRILL header. Section 4 below provides other RBridge design details.

3.1. TRILL Header Format

The TRILL header is shown in Figure 5 and is independent of the data link layer used. When that layer is IEEE [802.3], it is prefixed with the 16-bit TRILL Ethertype [RFC5342], making it 64-bit aligned. If Op-Length is a multiple of 64 bits, then 64-bit alignment is normally maintained for the content of an encapsulated frame.

```
                                +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
                                | V | R |M|Op-Length| Hop Count |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Egress RBridge Nickname     |  Ingress RBridge Nickname     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Options...
+-+-+-+-+-+-+-+-+-+-+-+-+-
```

Figure 5: TRILL Header

The header contains the following fields that are described in the sections referenced:

- V (Version): 2-bit unsigned integer. See Section 3.2.

- R (Reserved): 2 bits. See Section 3.3.

- M (Multi-destination): 1 bit. See Section 3.4.

- Op-Length (Options Length): 5-bit unsigned integer. See Section 3.5.

- Hop Count: 6-bit unsigned integer. See Section 3.6.

- Egress RBridge Nickname: 16-bit identifier. See Section 3.7.1.

- Ingress RBridge Nickname: 16-bit identifier. See Section 3.7.2.

- Options: present if Op-Length is non-zero. See Section 3.8.

3.2. Version (V)

Version (V) is a 2-bit field. Version zero of TRILL is specified in this document. An RBridge RB1 MUST check the V field in a received TRILL-encapsulated frame. If the V field has a value not recognized by RB1, then RB1 MUST silently discard the frame. The allocation of new TRILL Version numbers requires an IETF Standards Action.

3.3. Reserved (R)

The two R bits are reserved for future use in extensions to this version zero of the TRILL protocol. They MUST be set to zero when the TRILL header is added by an ingress RBridge, transparently copied but otherwise ignored by transit RBridges, and ignored by egress RBridges. The allocation of reserved TRILL header bits requires an IETF Standards Action.

3.4. Multi-destination (M)

The Multi-destination bit (see Section 2.4.2) indicates that the frame is to be delivered to a class of destination end stations via a distribution tree and that the egress RBridge nickname field specifies this tree. In particular:

- M = 0 (FALSE) - The egress RBridge nickname contains a nickname of the egress Rbridge for a known unicast MAC address.

- M = 1 (TRUE) - The egress RBridge nickname field contains a nickname that specifies a distribution tree. This nickname is selected by the ingress RBridge for a TRILL Data frame or by the source RBridge for a TRILL ESADI frame.

3.5. Op-Length

There are provisions to express in the TRILL header that a frame is using an optional capability and to encode information into the header in connection with that capability.

The Op-Length header field gives the length of the TRILL header options in units of 4 octets, which allows up to 124 octets of options area. If Op-Length is zero, there are no options present. If options are present, they follow immediately after the Ingress Rbridge Nickname field.

See Section 3.8 for more information on TRILL header options.

3.6. Hop Count

The Hop Count field is a 6-bit unsigned integer. An Rbridge drops frames received with a hop count of zero, otherwise it decrements the hop count. (This behavior is different from IPv4 and IPv6 in order to support the later addition of a traceroute-like facility that would be able to get a hop count exceeded from an egress RBridge.)

For known unicast frames, the ingress RBridge SHOULD set the Hop Count in excess of the number of RBridge hops it expects to the egress RBridge to allow for alternate routing later in the path.

For multi-destination frames, the Hop Count SHOULD be set by the ingress RBridge (or source RBridge for a TRILL ESADI frame) to at least the expected number of hops to the most distant RBridge. To accomplish this, RBridge RBn calculates, for each branch from RBn of the specified distribution tree rooted at RBi, the maximum number of hops in that branch.

Multi-destination frames are of particular danger because a loop involving one or more distribution tree forks could result in the rapid generation of multiple copies of the frame, even with the normal hop count mechanism. It is for this reason that multi-destination frames are subject to a stringent Reverse Path Forwarding Check and other checks as described in Section 4.5.2. As an optional additional traffic control measure, when forwarding a multi-destination frame onto a distribution tree branch, transit RBridge RBm MAY decrease the hop count by more than 1, unless decreasing the hop count by more than 1 would result in a hop count insufficient to reach all destinations in that branch of the tree rooted at RBi. Using a hop count close or equal to the minimum needed on multi-destination frames provides additional protection against problems with temporary loops when forwarding.

Although the RBridge MAY decrease the hop count of multi-destination frames by more than 1, under the circumstances described above, the RBridge forwarding a frame MUST decrease the hop count by at least 1, and discards the frame if it cannot do so because the hop count is 0. The option to decrease the hop count by more than 1 under the circumstances described above applies only to multi-destination frames, not to known unicast frames.

3.7. RBridge Nicknames

Nicknames are 16-bit dynamically assigned quantities that act as abbreviations for RBridges' IS-IS IDs to achieve a more compact encoding and can be used to specify potentially different trees with the same root. This assignment allows specifying up to 2**16 RBridges; however, the value 0x0000 is reserved to indicate that a nickname is not specified, the values 0xFFC0 through 0xFFFE are reserved for future specification, and the value 0xFFFF is permanently reserved. RBridges piggyback a nickname acquisition protocol on the link state protocol (see Section 3.7.3) to acquire one or more nicknames unique within the campus.

3.7.1. Egress RBridge Nickname

There are two cases for the contents of the egress RBridge nickname field, depending on the M bit (see Section 3.4). The nickname is filled in by the ingress RBridge for TRILL Data frames and by the source RBridge for TRILL ESADI frames.

- For known unicast TRILL Data frames, M == 0 and the egress RBridge nickname field specifies the egress RBridge; that is, it specifies the RBridge that needs to remove the TRILL encapsulation and forward the native frame. Once the egress nickname field is set, it MUST NOT be changed by any subsequent transit RBridge.

- For multi-destination TRILL Data frames and for TRILL ESADI frames, M == 1. The egress RBridge nickname field contains a nickname specifying the distribution tree selected to be used to forward the frame. This root nickname MUST NOT be changed by transit RBridges.

3.7.2. Ingress RBridge Nickname

The ingress RBridge nickname is set to a nickname of the ingress RBridge for TRILL Data frames and to a nickname of the source RBridge for TRILL ESADI frames. If the RBridge setting the ingress nickname has multiple nicknames, it SHOULD use the same nickname in the ingress field whenever it encapsulates a frame with any particular Inner.MacSA and Inner.VLAN value. This simplifies end node learning.

Once the ingress nickname field is set, it MUST NOT be changed by any subsequent transit RBridge.

3.7.3. RBridge Nickname Selection

The nickname selection protocol is piggybacked on TRILL IS-IS as follows:

- The nickname or nicknames being used by an RBridge are carried in an IS-IS TLV (type-length-value data element) along with a priority of use value [RFC6326]. Each RBridge chooses its own nickname or nicknames.

- Nickname values MAY be configured. An RBridge that has been configured with one or more nickname values will have priority for those nickname values over all Rbridges with non-configured nicknames.

- The nickname value 0x0000 and the values from 0xFFC0 through 0xFFFF are reserved and MUST NOT be selected by or configured for an RBridge. The value 0x0000 is used to indicate that a nickname is not known.

- The priority of use field reported with a nickname is an unsigned 8-bit value, where the most significant bit (0x80) indicates that the nickname value was configured. The bottom 7 bits have the default value 0x40, but MAY be configured to be some other value. Additionally, an RBridge MAY increase its priority after holding a nickname for some amount of time. However, the most significant bit of the priority MUST NOT be set unless the nickname value was configured.

- Once an RBridge has successfully acquired a nickname, it SHOULD attempt to reuse it in the case of a reboot.

- Each RBridge is responsible for ensuring that its nickname or each of its nicknames is unique. If RB1 chooses nickname x, and RB1 discovers, through receipt of an LSP for RB2 at any later time, that RB2 has also chosen x, then the RBridge or pseudonode with the numerically higher IS-IS ID (LAN ID) keeps the nickname, or if there is a tie in priority, the RBridge with the numerically higher IS-IS System ID keeps the nickname, and the other RBridge MUST select a new nickname. This can require an RBridge with a configured nickname to select a replacement nickname.

- To minimize the probability of nickname collisions, an RBridge selects a nickname randomly from the apparently available nicknames, based on its copy of the link state. This random selection can be by the RBridge hashing some of its parameters, e.g., SystemID, time and date, and other entropy sources, such as those given in [RFC4086], each time or by the RBridge using such hashing to create a seed and making any selections based on pseudo-random numbers generated from that seed [RFC4086]. The random numbers or seed and the algorithm used SHOULD make uniformly distributed selections over the available nicknames. Convergence to a nickname-collision-free campus is accelerated by selecting new nicknames only from those that appear to be available and by having the highest priority nickname involved in a nickname conflict retain its value. There is no reason for all Rbridges to use the same algorithm for selecting nicknames.

- If two RBridge campuses merge, then transient nickname collisions are possible. As soon as each RBridge receives the LSPs from the other RBridges, the RBridges that need to change nicknames select new nicknames that do not, to the best of their knowledge, collide with any existing nicknames. Some RBridges may need to change nicknames more than once before the situation is resolved.

- To minimize the probability of a new RBridge usurping a nickname already in use, an RBridge SHOULD wait to acquire the link state database from a neighbor before it announces any nicknames that were not configured.

- An RBridge by default has only a single nickname but MAY be configured to request multiple nicknames. Each such nickname would specify a shortest path tree with the RBridge as root but, since the tree number is used in tiebreaking when there are multiple equal cost paths (see Section 4.5.1), the trees for the different nicknames will likely utilize different links. Because of the potential tree computation load it imposes, this capability to request multiple nicknames for an RBridge should be used sparingly. For example, it should be used at a few RBridges that, because of campus topology, are particularly good places from which to calculate multiple different shortest path distribution trees. Such trees need separate nicknames so traffic can be multipathed across them.

o If it is desired for a pseudonode to be a tree root, the DRB MAY request one or more nicknames in the pseudonode LSP.

Every nickname in use in a campus identifies an RBridge (or pseudonode) and every nickname designates a distribution tree rooted at the RBridge (or pseudonode) it identifies. However, only a limited number of these potential distribution trees are actually computed by all the RBridges in a campus as discussed in Section 4.5.

3.8. TRILL Header Options

All Rbridges MUST be able to skip the number of 4-octet chunks indicated by the Op-Length field (see Section 3.5) in order to find the inner frame, since RBridges must be able to find the destination MAC address and VLAN tag in the inner frame. (Transit RBridges need such information to filter VLANs, IP multicast, and the like. Egress Rbridges need to find the inner header to correctly decapsulate and handle the inner frame.)

To ensure backward-compatible safe operation, when Op-Length is non-zero indicating that options are present, the top two bits of the first octet of the options area are specified as follows:

```
            +------+------+----+----+----+----+----+----+
            | CHbH | CItE |         Reserved            |
            +------+------+----+----+----+----+----+----+
```

Figure 6: Options Area Initial Flags Octet

If the CHbH (Critical Hop-by-Hop) bit is one, one or more critical hop-by-hop options are present. Transit RBridges that do not support all of the critical hop-by-hop options present, for example, an RBridge that supported no options, MUST drop the frame. If the CHbH bit is zero, the frame is safe, from the point of view of options processing, for a transit RBridge to forward, regardless of what options that RBridge does or does not support. A transit RBridge that supports none of the options present MUST transparently forward the options area when it forwards a frame.

If the CItE (Critical Ingress-to-Egress) bit is one, one or more critical ingress-to-egress options are present. If it is zero, no such options are present. If either CHbH or CItE is non-zero, egress RBridges that don’t support all critical options present, for example, an RBridge that supports no options, MUST drop the frame. If both CHbH and CItE are zero, the frame is safe, from the point of view of options, for any egress RBridge to process, regardless of what options that RBridge does or does not support.

Options, including the meaning of the bits labeled as Reserved in Figure 6, will be further specified in other documents and are expected to include provisions for hop-by-hop and ingress-to-egress options as well as critical and non-critical options.

Note: Most RBridge implementations are expected to be optimized for the simplest and most common cases of frame forwarding and processing. The inclusion of options may, and the inclusion of complex or lengthy options likely will, cause frame processing using a "slow path" with inferior performance to "fast path" processing. Limited slow path throughput may cause such frames to be discarded.

4. Other RBridge Design Details

Section 3 above specifies the TRILL header, while this section specifies other RBridge design details.

4.1. Ethernet Data Encapsulation

TRILL data and ESADI frames in transit on Ethernet links are encapsulated with an outer Ethernet header (see Figure 2). This outer header looks, to a bridge on the path between two RBridges, like the header of a regular Ethernet frame; therefore, bridges forward the frame as they normally would. To enable RBridges to distinguish such TRILL Data frames, a new TRILL Ethertype (see Section 7.2) is used in the outer header.

Figure 7 details a TRILL Data frame with an outer VLAN tag traveling on an Ethernet link as shown at the top of the figure, that is, between transit RBridges RB3 and RB4. The native frame originated at end station ESa, was encapsulated by ingress RBridge RB1, and will ultimately be decapsulated by egress RBridge RB2 and delivered to destination end station ESb. The encapsulation shown has the advantage, if TRILL options are absent or the length of such options is a multiple of 64 bits, of aligning the original Ethernet frame at a 64-bit boundary.

When a TRILL Data frame is carried over an Ethernet cloud, it has three pairs of addresses:

- o Outer Ethernet Header: Outer Destination MAC Address (Outer.MacDA) and Outer Source MAC Address (Outer.MacSA): These addresses are used to specify the next hop RBridge and the transmitting RBridge, respectively.

- o TRILL Header: Egress Nickname and Ingress Nickname. These specify nicknames of the egress and ingress RBridges, respectively, unless the frame is multi-destination, in which case the Egress Nickname specifies the distribution tree on which the frame is being sent.

- o Inner Ethernet Header: Inner Destination MAC Address (Inner.MacDA) and Inner Source MAC Address (Inner.MacSA): These addresses are as transmitted by the original end station, specifying, respectively, the destination and source of the inner frame.

A TRILL Data frame also potentially has two VLAN tags, as discussed in Sections 4.1.2 and 4.1.3 below, that can carry two different VLAN Identifiers and specify priority.

```
   Flow:
     +-----+  +-------+   +-------+       +-------+   +-------+  +-----+
     | ESa +--+  RB1  +---+  RB3  +-------+  RB4  +---+  RB2  +--+ESb |
     +-----+  |ingress|   |transit|   ^   |transit|   |egress |  +-----+
              +-------+   +-------+   |   +-------+   +-------+
                                      |
   Outer Ethernet Header:             |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |             Outer Destination MAC Address  (RB4)              |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      | Outer Destination MAC Address | Outer Source MAC Address      |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                Outer Source MAC Address  (RB3)                |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |Ethertype = C-Tag [802.1Q-2005]| Outer.VLAN Tag Information    |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   TRILL Header:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      | Ethertype = TRILL             | V | R |M|Op-Length| Hop Count |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      | Egress (RB2) Nickname         | Ingress (RB1) Nickname        |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   Inner Ethernet Header:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |             Inner Destination MAC Address  (ESb)              |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      | Inner Destination MAC Address | Inner Source MAC Address      |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                  Inner Source MAC Address  (ESa)              |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |Ethertype = C-Tag [802.1Q-2005]| Inner.VLAN Tag Information    |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   Payload:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      | Ethertype of Original Payload |                               |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+                               |
      |                                 Original Ethernet Payload    |
      |                                                               |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   Frame Check Sequence:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |               New FCS (Frame Check Sequence)                  |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 7: TRILL Data Encapsulation over Ethernet

4.1.1. VLAN Tag Information

A "VLAN Tag" (formerly known as a Q-tag), also known as a "C-tag" for customer tag, includes a VLAN ID and a priority field as shown in Figure 8. The "VLAN ID" may be zero, indicating that no VLAN is specified, just a priority, although such frames are called "priority tagged" rather than "VLAN tagged" [802.1Q-2005].

Use of [802.1ad] S-tags, also known as service tags, and use of stacked tags, are beyond the scope of this document.

```
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
| Priority  | C |                  VLAN ID                      |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
```

Figure 8: VLAN Tag Information

As recommended in [802.1Q-2005], Rbridges SHOULD be implemented so as to allow use of the full range of VLAN IDs from 0x001 through 0xFFE. Rbridges MAY support a smaller number of simultaneously active VLAN IDs. VLAN ID zero is the null VLAN identifier and indicates that no VLAN is specified while VLAN ID 0xFFF is reserved.

The VLAN ID 0xFFF MUST NOT be used. Rbridges MUST discard any frame they receive with an Outer.VLAN ID of 0xFFF. Rbridges MUST discard any frame for which they examine the Inner.VLAN ID and find it to be 0xFFF; such examination is required at all egress Rbridges that decapsulate a frame.

The "C" bit shown in Figure 8 is not used in the Inner.VLAN in TRILL. It MUST be set to zero there by ingress RBridges, transparently forwarded by transit RBridges, and is ignored by egress RBridges.

As specified in [802.1Q-2005], the priority field contains an unsigned value from 0 through 7 where 1 indicates the lowest priority, 7 the highest priority, and the default priority zero is considered to be higher than priority 1 but lower than priority 2. The [802.1ad] amendment to [802.1Q-2005] permits mapping some adjacent pairs of priority levels into a single priority level with and without drop eligibility. Ongoing work in IEEE 802.1 (802.1az, Appendix E) suggests the ability to configure "priority groups" that have a certain guaranteed bandwidth. RBridges ports MAY also implement such options. RBridges are not required to implement any particular number of distinct priority levels but may treat one or more adjacent priority levels in the same fashion.

Frames with the same source address, destination address, VLAN, and priority that are received on the same port as each other and are transmitted on the same port MUST be transmitted in the order received unless the RBridge classifies the frames into more fine-grained flows, in which case this ordering requirement applies to each such flow. Frames in the same VLAN with the same priority and received on the same port may be sent out different ports if multipathing is in effect. (See Appendix C.)

The C-Tag Ethertype [RFC5342] is 0x8100.

4.1.2. Inner VLAN Tag

The "Inner VLAN Tag Information" (Inner.VLAN) field contains the VLAN tag information associated with the native frame when it was ingressed or the VLAN tag information associated with a TRILL ESADI frame when that frame was created. When a TRILL frame passes through a transit RBridge, the Inner.VLAN MUST NOT be changed except when VLAN mapping is being intentionally performed within that RBridge.

When a native frame arrives at an RBridge, the associated VLAN ID and priority are determined as specified in [802.1Q-2005] (see Appendix D and [802.1Q-2005], Section 6.7). If the RBridge is an appointed forwarder for that VLAN and the delivery of the frame requires transmission to one or more other links, this ingress RBridge forms a TRILL Data frame with the associated VLAN ID and priority placed in the Inner.VLAN information.

The VLAN ID is required at the ingress Rbridge as one element in determining the appropriate egress Rbridge for a known unicast frame and is needed at the ingress and every transit Rbridge for multi-destination frames to correctly prune the distribution tree.

4.1.3. Outer VLAN Tag

TRILL frames sent by an RBridge, except for some TRILL-Hello frames, use an Outer.VLAN ID specified by the Designated RBridge (DRB) for the link onto which they are being sent, referred to as the Designated VLAN. For TRILL data and ESADI frames, the priority in the Outer.VLAN tag SHOULD be set to the priority in the Inner.VLAN tag.

TRILL frames forwarded by a transit RBridge use the priority present in the Inner.VLAN of the frame as received. TRILL Data frames are sent with the priority associated with the corresponding native frame when received (see Appendix D). TRILL IS-IS frames SHOULD be sent with priority 7.

Whether an Outer.VLAN tag actually appears on the wire when a TRILL frame is sent depends on the configuration of the RBridge port through which it is sent in the same way as the appearance of a VLAN tag on a frame sent by an [802.1Q-2005] bridge depends on the configuration of the bridge port (see Section 4.9.2).

4.1.4. Frame Check Sequence (FCS)

Each Ethernet frame has a single Frame Check Sequence (FCS) that is computed to cover the entire frame, for detecting frame corruption due to bit errors on a link. Thus, when a frame is encapsulated, the original FCS is not included but is discarded. Any received frame for which the FCS check fails SHOULD be discarded (this may not be possible in the case of cut through forwarding). The FCS normally changes on encapsulation, decapsulation, and every TRILL hop due to changes in the outer destination and source addresses, the decrementing of the hop count, etc.

Although the FCS is normally calculated just before transmission, it is desirable, when practical, for an FCS to accompany a frame within an RBridge after receipt. That FCS could then be dynamically updated to account for changes to the frame during Rbridge processing and used for transmission or checked against the FCS calculated for frame transmission. This optional, more continuous use of an FCS would be helpful in detecting some internal RBridge failures such as memory errors.

4.2. Link State Protocol (IS-IS)

TRILL uses an extension of IS-IS [ISO10589] [RFC1195] as its routing protocol. IS-IS has the following advantages:

- o It runs directly over Layer 2, so therefore it may be run without configuration (no IP addresses need to be assigned).

- o It is easy to extend by defining new TLV (type-length-value) data elements and sub-elements for carrying TRILL information.

This section describes TRILL use of IS-IS, except for the TRILL-Hello protocol, which is described in Section 4.4, and the MTU-probe and MTU-ack messages that are described in Section 4.3.

4.2.1. IS-IS RBridge Identity

Each RBridge has a unique 48-bit (6-octet) IS-IS System ID. This ID may be derived from any of the RBridge’s unique MAC addresses.

A pseudonode is assigned a 7-octet ID by the DRB that created it, by taking a 6-octet ID owned by the DRB, and appending another octet. The 6-octet ID used to form a pseudonode ID SHOULD be the DRB’s ID unless the DRB has to create IDs for pseudonodes for more than 255 links. The only constraint for correct operation is that the 7-octet ID be unique within the campus, and that the 7th octet be nonzero. An RBridge has a 7-octet ID consisting of its 6-octet system ID concatenated with a zero octet.

In this document, we use the term "IS-IS ID" to refer to the 7-octet quantity that can be either the ID of an RBridge or a pseudonode.

4.2.2. IS-IS Instances

TRILL implements a separate IS-IS instance from any used by Layer 3, that is, different from the one used by routers. Layer 3 IS-IS frames must be distinguished from TRILL IS-IS frames even when those Layer 3 IS-IS frames are transiting an RBridge campus.

Layer 3 IS-IS native frames have special multicast destination addresses specified for that purpose, such as AllL1ISs or AllL2ISs. When they are TRILL encapsulated, these multicast addresses appear as the Inner.MacDA and the Outer.MacDA will be the All-RBridges multicast address.

Within TRILL, there is an IS-IS instance across all Rbridges in the campus as described in Section 4.2.3. This instance uses TRILL IS-IS frames that are distinguished by having a different Ethertype "L2-IS-IS". Additionally, for TRILL IS-IS frames that are multicast, there is a distinct multicast destination address of All-IS-IS-RBridges. TRILL IS-IS frames do not have a TRILL header.

ESADI is a separate protocol from the IS-IS instance implemented by all the RBridges. There is a separate ESADI instance for each VLAN, and ESADI frames are encapsulated just like TRILL Data frames. After the TRILL header, the ESADI frame has an inner Ethernet header with the Inner.MacDA of "All-ESADI-RBridges" and the "L2-IS-IS" Ethertype followed by the ESADI frame.

4.2.3. TRILL IS-IS Frames

All Rbridges MUST participate in the TRILL IS-IS instance, which constitutes a single Level 1 IS-IS area using the fixed area address zero. TRILL IS-IS frames are never forwarded by an RBridge but are locally processed on receipt. (Such processing may cause the RBridge to send additional TRILL IS-IS frames.)

A TRILL IS-IS frame on an 802.3 link is structured as shown below. All such frames are Ethertype encoded. The RBridge port out of which such a frame is sent will strip the outer VLAN tag if configured to do so.

```
   Outer Ethernet Header:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |             All-IS-IS-RBridges Multicast Address              |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      | All-IS-IS-RBridges continued  | Source RBridge MAC Address    |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |             Source RBridge MAC Address continued              |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |Ethertype = C-Tag [802.1Q-2005]| Outer.VLAN Tag Information    |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |   L2-IS-IS Ethertype          |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   IS-IS Payload:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      | IS-IS Common Header, IS-IS PDU Specific Fields, IS-IS TLVs    |

   Frame Check Sequence:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                 FCS (Frame Check Sequence)                    |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 9: TRILL IS-IS Frame Format

The VLAN specified in the Outer.VLAN information will be the Designated VLAN for the link on which the frame is sent, except in the case of some TRILL Hellos.

4.2.4. TRILL Link Hellos, DRBs, and Appointed Forwarders

RBridges default to using TRILL Hellos unless, on a per-port basis, they are configured to use P2P Hellos. TRILL-Hello frames are specified in Section 4.4.

RBridges are normally configured to use P2P Hellos only when there are exactly two of them on a link. However, it can occur that RBridges are misconfigured as to which type of hello to use. This is safe but may cause lack of RBridge-to-RBridge connectivity. An RBridge port configured to use P2P Hellos ignores TRILL Hellos, and an RBridge port configured to use TRILL Hellos ignores P2P Hellos.

If any of the RBridge ports on a link is configured to use TRILL Hellos, one of such RBridge ports using TRILL Hellos is elected DRB (Designated RBridge) for the link. This election is based on configured priority (most significant field), and source MAC address, as communicated by TRILL-Hello frames. The DRB, as described in Section 4.2.4.2, designates the VLAN to be used on the link for inter-RBridge communication by the non-P2P RBridge ports and appoints itself or other RBridges on the link as appointed forwarder (see Section 4.2.4.3) for VLANs on the link.

4.2.4.1. P2P Hello Links

RBridge ports can be configured to use IS-IS P2P Hellos. This implies that the port is a point-to-point link to another RBridge. An RBridge MUST NOT provide any end-station (native frame) service on a port configured to use P2P Hellos.

As with Layer 3 IS-IS, such P2P ports do not participate in a DRB election. They send all frames VLAN tagged as being in the Desired Designated VLAN configured for the port, although this tag may be stripped if the port is so configured. Since all traffic through the port should be TRILL frames or Layer 2 control frames, such a port cannot be an appointed forwarder. RBridge P2P ports MUST use the IS-IS three-way handshake [RFC5303] so that extended circuit IDs are associated with the link for tie breaking purposes (see Section 4.5.2).

Even if all simple links in a network are physically point-to-point, if some of the nodes are bridges, the bridged LANs that include those bridges appear to be multi-access links to attached RBridges. This would necessitate using TRILL Hellos for proper operation in many cases.

While it is safe to erroneously configure ports as P2P, this may result in lack of connectivity.

4.2.4.2. Designated RBridge

TRILL IS-IS elects one RBridge for each LAN link to be the Designated RBridge (DRB), that is, to have special duties. The Designated RBridge:

- Chooses, for the link, and announces in its TRILL Hellos, the Designated VLAN ID to be used for inter-RBridge communication. This VLAN is used for all TRILL-encapsulated data and ESADI frames and TRILL IS-IS frames except some TRILL-Hello frames.

- If the link is represented in the IS-IS topology as a pseudonode, chooses a pseudonode ID and announces that in its TRILL Hellos and issues an LSP on behalf of the pseudonode.

- Issues CSNPs.

- For each VLAN-x appearing on the link, chooses an RBridge on the link to be the appointed VLAN-x forwarder (the DRB MAY choose itself to be the appointed VLAN-x forwarder for all or some of the VLANs).

- Before appointing a VLAN-x forwarder (including appointing itself), wait at least its Holding Time (to ensure it is the DRB).

- If configured to send TRILL-Hello frames, continues to send them on all its enabled VLANs that have been configured in the Announcing VLANs set of the DRB, which defaults to all enabled VLANs.

4.2.4.3. Appointed VLAN-x Forwarder

The appointed VLAN-x forwarder for a link is responsible for the following points. In connection with the loop avoidance points, when an appointed forwarder for a port is "inhibited", it drops any native frames it receives and does not transmit but instead drops any native frames it decapsulates, in the VLAN for which it is appointed.

- Loop avoidance:

  - Inhibiting itself for a time, configurable per port from zero to 30 seconds, which defaults to 30 seconds, after it sees a root bridge change on the link (see Section 4.9.3.2).

  - Inhibiting itself for VLAN-x, if it has received a Hello in which the sender asserts that it is appointed forwarder and that is either
    - received on VLAN-x (has VLAN-x as its Outer.VLAN) or
    - was originally sent on VLAN-x as indicated inside the body of the Hello.

  - Optionally, not decapsulating a frame from ingress RBridge RBm unless it has RBm’s LSP, and the root bridge on the link it is about to forward onto is not listed in RBm’s list of root bridges for VLAN-x. This is known as the "decapsulation check" or "root bridge collision check".

- Unless inhibited (see above), receiving VLAN-x native traffic from the link and forwarding it as appropriate.

- Receiving VLAN-x traffic for the link and, unless inhibited, transmitting it in native form after decapsulating it as appropriate.

- Learning the MAC address of local VLAN-x nodes by looking at the source address of VLAN-x frames from the link.

- Optionally learning the port of local VLAN-x nodes based on any sort of Layer 2 registration protocols, such as IEEE 802.11 association and authentication.

- Keeping track of the { egress RBridge, VLAN, MAC address } of distant VLAN-x end nodes, learned by looking at the fields { ingress RBridge, Inner.VLAN ID, Inner.MacSA } from VLAN-x frames being received for decapsulation onto the link.

- Optionally observe native IGMP [RFC3376], MLD [RFC2710], and MRD [RFC4286] frames to learn the presence of local multicast listeners and multicast routers.

- Optionally listening to TRILL ESADI messages for VLAN-x to learn { egress RBridge, VLAN-x, MAC address } triplets and the confidence level of such explicitly advertised end nodes.

- Optionally advertising VLAN-x end nodes, on links for which it is appointed VLAN-x forwarder, in ESADI messages.

- Sending TRILL-Hello frames on VLAN-x unless the Announcing VLANs set for the port has been configured to disable them.

- Listening to BPDUs on the common spanning tree to learn the root bridge, if any, for that link and to report in its LSP the complete set of root bridges seen on any of its links for which it is appointed forwarder for VLAN-x.

When an appointed forwarder observes that the DRB on a link has changed, it no longer considers itself appointed for that link until appointed by the new DRB.

4.2.4.4. TRILL LSP Information

The information items in the TRILL IS-IS LSP that are mentioned elsewhere in this document are listed below. Unless an item is stated in the list below to be optional, it MUST be included. Other items MAY be included unless their inclusion is prohibited elsewhere in this document. The actual encoding of this information and the IS-IS Type or sub-Type values for any new IS-IS TLV or sub-TLV data elements are specified in separate documents [RFC6165] [RFC6326].

1. The IS-IS IDs of neighbors (pseudonodes as well as RBridges) of RBridge RBn, and the cost of the link to each of those neighbors. RBridges MUST use the Extended IS Reachability TLV (#22, also

known as "wide metric" [RFC5305]) and MUST NOT use the IS Reachability TLV (#2, also known as "narrow metric"). To facilitate efficient operation without configuration and consistent with [802.1D], RBridges SHOULD, by default, set the cost of a link to the integer part of twenty trillion (20,000,000,000,000) divided by the RBridge port's bit rate but not more than 2**24-2 (16,777,214); for example, the cost for a link accessed by a 1Gbps port would default to 20,000. (Note that 2**24-1 has a special meaning in IS-IS and would exclude the link from SPF routes.) However, the link cost MAY, by default, be decreased for aggregated links and/or increased to not more than 2**24-2 if the link appears to be a bridged LAN. The tested MTU for the link (see Section 4.3) MAY be included via a sub-TLV.

2. The following information in connection with the nickname or each of the nicknames of RBridge RBn:

   2.1. The nickname value (2 octets).

   2.2. The unsigned 8-bit priority for RBn to have that nickname (see Section 3.7.3).

   2.3. The 16-bit unsigned priority of that nickname to becoming a distribution tree root.

3. The maximum TRILL Header Version supported by RBridge RBn.

4. The following information, in addition to the per-nickname tree root priority, in connection with distribution tree determination and announcement. (See Section 4.5 for further details on how this information is used.)

   4.1. An unsigned 16-bit number that is the number of trees all RBridges in the campus calculate if RBn has the highest priority tree root.

   4.2. A second unsigned 16-bit number that is the number of trees RBn would like to use.

   4.3. A third unsigned 16-bit number that is the maximum number of distribution trees that RBn is able to calculate.

   4.4. A first list of nicknames that are intended distribution trees for all RBridges in the campus to calculate.

   4.5. A second list of nicknames that are distribution trees RBn would like to use when ingressing multi-destination frames.

5. The list of VLAN IDs of VLANs directly connected to RBn for links on which RBn is the appointed forwarder for that VLAN. (Note: An RBridge may advertise that it is connected to additional VLANs in order to receive additional frames to support certain VLAN-based features beyond the scope of this specification as mentioned in Section 4.8.4 and in a separate document concerning VLAN mapping inside RBridges.) RBridges may associate advertised connectivity to different groups of VLANs with specific nicknames they hold. In addition, the LSP contains the following information on a per-VLAN basis:

   5.1. Per-VLAN Multicast Router attached flags: This is two bits of information that indicate whether there is an IPv4 and/or IPv6 multicast router attached to the Rbridge on that VLAN. An RBridge that does not do IP multicast control snooping MUST set both of these bits (see Section 4.5.4). This information is used because IGMP [RFC3376] and MLD [RFC2710] Membership Reports MUST be transmitted to all links with IP multicast routers, and SHOULD NOT be transmitted to links without such routers. Also, all frames for IP-derived multicast addresses MUST be transmitted to all links with IP multicast routers (within a VLAN), in addition to links from which an IP node has explicitly asked to join the group the frame is for, except for some IP multicast addresses that MUST be treated as broadcast.

   5.2. Per-VLAN mandatory announcement of the set of IDs of Root bridges for any of RBn’s links on which RBn is appointed forwarder for that VLAN. Where MSTP (Multiple Spanning Tree Protocol) is running on a link, this is the root bridge of the CIST (Common and Internal Spanning Tree). This is to quickly detect cases where two Layer 2 clouds accidentally get merged, and where there might otherwise temporarily be two DRBs for the same VLAN on the same link. (See Section 4.2.4.3.)

   5.3. Optionally, per-VLAN Layer 2 multicast addresses derived from IPv4 IGMP and IPv6 MLD notification messages received from attached end nodes on that VLAN, indicating the location of listeners for these multicast addresses (see Section 4.5.5).

   5.4. Per-VLAN ESADI protocol participation flag, priority, and holding time. If this flag is one, it indicates that the RBridge wishes to receive such TRILL ESADI frames (see Section 4.2.5.1).

   5.5. Per-VLAN appointed forwarder status lost counter (see Section 4.8.3).

6. Optionally, the largest TRILL IS-IS frame that the RBridge can handle using the originatingLSPBufferSize TLV #14 (see Section 4.3).

7. Optionally, a list of VLAN groups where address learning is shared across that VLAN group (see Section 4.8.4). Each VLAN group is a list of VLAN IDs, where the first VLAN ID listed in a group, if present, is the "primary" and the others are "secondary". This is to detect misconfiguration of features outside the scope of this document. RBridges that do not support features such as "shared VLAN learning" ignore this field.

8. Optionally, the Authentication TLV #10 (see Section 6).

4.2.5. The TRILL ESADI Protocol

RBridges that are the appointed VLAN-x forwarder for a link MAY participate in the TRILL ESADI protocol for that VLAN. But all transit RBridges MUST properly forward TRILL ESADI frames as if they were multicast TRILL Data frames. TRILL ESADI frames are structured like IS-IS frames but are always TRILL encapsulated on the wire as if they were TRILL Data frames.

Because of this forwarding, it appears to the ESADI protocol at an RBridge that it is directly connected by a shared virtual link to all other RBridges in the campus running ESADI for that VLAN. RBridges that do not implement the ESADI protocol or are not appointed forwarder for that VLAN do not decapsulate or locally process any TRILL ESADI frames they receive for that VLAN. In other words, these frames are transparently tunneled through transit RBridges. Such transit RBridges treat them exactly as multicast TRILL Data frames and no special processing is invoked due to such forwarding.

TRILL ESADI frames sent on an IEEE 802.3 link are structured as shown below. The outer VLAN tag will not be present if it was stripped by the port out of which the frame was sent.

```
   Outer Ethernet Header:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                 Next Hop Destination Address                   |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      | Next Hop Destination Address  | Sending RBridge MAC Address   |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |               Sending RBridge Port MAC Address                |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |Ethertype = C-Tag [802.1Q-2005]| Outer.VLAN Tag Information    |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   TRILL Header:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      | Ethertype = TRILL             | V | R |M|Op-Length| Hop Count |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      | Egress (Dist. Tree) Nickname  | Ingress (Origin) Nickname     |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   Inner Ethernet Header:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |             All-ESADI-RBridges Multicast Address              |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      | All-ESADI-RBridges continued  | Origin RBridge MAC Address    |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                Origin RBridge MAC Address continued           |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |Ethertype = C-Tag [802.1Q-2005]| Inner.VLAN Tag Information    |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      | Ethertype = L2-IS-IS          |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
   ESADI Payload (formatted as IS-IS):
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      | IS-IS Common Header, IS-IS PDU Specific Fields, IS-IS TLVs    |

   Frame Check Sequence:
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
      |                  FCS (Frame Check Sequence)                   |
      +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

Figure 10: TRILL ESADI Frame Format

The Next Hop Destination Address or Outer.MacDA is the All-RBridges multicast address. The VLAN specified in the Outer.VLAN information will always be the Designated VLAN for the link on which the frame is sent. The V and R fields will be zero while the M field will be one. The VLAN specified in the Inner.VLAN information will be the VLAN to which the ESADI frame applies. The Origin RBridge MAC Address or Inner.MacSA MUST be a globally unique MAC address owned by the

RBridge originating the ESADI frame, for example, any of its port MAC addresses, and each RBridge MUST use the same Inner.MacSA for all of the ESADI frames that RBridge originates.

4.2.5.1. TRILL ESADI Participation

An RBridge does not send any Hellos because of participation in the ESADI protocol. The information available in the TRILL IS-IS link state database is sufficient to determine the ESADI DRB on the virtual link for the ESADI protocol for each VLAN. In particular, the link state database information for each RBridge includes the VLANs, if any, for which that RBridge is participating in the ESADI protocol, its priority for being selected as DRB for the ESADI protocol for each of those VLANs, its holding time, and its IS-IS system ID for breaking ties in priority.

An RBridge need not perform any routing calculation because of participation in the ESADI protocol. Since all RBridges participating in ESADI for a particular VLAN appear to be connected to the same single virtual link, there are no routing decisions to be made. A participating RBridge merely transmits the ESADI frames it originates on this virtual link.

The ESADI DRB sends TRILL-ESADI-CSNP frames on the ESADI virtual link. For robustness, a participating RBridge that determines that some other RBridge should be ESADI DRB on such a virtual link but has not received or sent a TRILL-ESADI-CSNP in at least the ESADI DRB holding time MAY also send a TRILL-ESADI-CSNP on the virtual link. A participating RBridge that determines that no other RBridges are participating in the ESADI protocol for a particular VLAN SHOULD NOT send ESADI information or TRILL-ESADI-CSNPs on the virtual link for that VLAN.

4.2.5.2. TRILL ESADI Information

The information distributed with the ESADI protocol is the list of local end-station MAC addresses known to the originating RBridge and, for each such address, a one-octet unsigned "confidence" rating in the range 0-254 (see Section 4.8).

It is intended to optionally provide for VLAN ID translation within RBridges, as specified in [VLAN-MAPPING]. This includes translating TRILL ESADI frames. If TRILL ESADI frames could contain VLAN IDs in arbitrary internal locations, such translation would be impractical. Thus, TRILL ESADI frames MUST NOT contain the VLAN ID of the VLAN to which they apply in the body of the frame after the Inner.VLAN tag.

4.2.6. SPF, Forwarding, and Ambiguous Destinations

This section describes the logical result desired. Alternative implementation methods may be used as long as they produce the same forwarding behavior.

When building a forwarding table, an RBridge RB1 calculates shortest paths from itself as described in Appendix C.1 of [RFC1195]. Nicknames are added into the shortest path calculation as a final step, just as with an end node. If multiple RBridges, say, RBa and RBb, claim the same nickname, this is a transitory condition and one of RBa or RBb will defer and choose a new nickname. However, RB1 simply adds that nickname as if it were attached to both RBa and RBb, and uses its standard shortest path calculation to choose the next hop.

An ingress RBridge RB2 maps a native frame’s known unicast destination MAC address and VLAN into an egress RBridge nickname. If RB2 learns addresses only from the observation of received and decapsulated frames, then such MAC addresses cannot be duplicated within a VLAN in RB2 tables because more recent learned information, if of a higher or equal confidence, overwrites previous information and, if of a lower confidence, is ignored. However, duplicates of the same MAC within a VLAN can appear in ESADI data and between ESADI data and addresses learned from the observation of received and decapsulated frames, entered by manual configuration, or learned through Layer 2 registration protocols. If duplicate MAC addresses occur within a VLAN, RB2 sends frames to the MAC with the highest confidence. If confidences are also tied between the duplicates, for consistency it is suggested that RB2 direct all such frames (or all such frames in the same ECMP flow) toward the same egress RBridge; however, the use of other policies will not cause a network problem since transit RBridges do not examine the Inner.MacDA for known unicast frames.

4.3. Inter-RBridge Link MTU Size

There are two reasons why it is important to know what size of frame each inter-RBridge link in the campus can support:

1. RBridge RB1 must know the size of link state information messages it can generate that will be guaranteed to be forwardable across all inter-RBridge links in the campus.

2. If traffic engineering tools know which links support larger than minimally acceptable data packet sizes, paths can be computed that can support large data packets.

4.3.1. Determining Campus-Wide TRILL IS-IS MTU Size

In a stable campus, there must ultimately be agreement among all RBridges on the value of "Sz", the minimum acceptable inter-RBridge link size for the campus, for the proper operation of TRILL IS-IS. All RBridges MUST format their link state information messages to be in chunks of size no larger than what they believe Sz to be. Also, every RBridge RB1 SHOULD test each of its RBridge adjacencies, say, to RB2, to ensure that the RB1-RB2 link can forward packets of at least size Sz.

Sz has no direct effect on end stations and is not directly related to any end-station-to-end-station "path MTU". Methods of using Sz or any link MTU information gathered by TRILL IS-IS in the traffic engineering of routes or the determination of any path MTU is beyond the scope of this document. Native frames that, after TRILL encapsulation, exceed the MTU of a link on which they are sent will generally be discarded.

Sz is determined by having each RBridge (optionally) advertise, in its LSP, its assumption of the value of the campus-wide Sz. This LSP element is known in IS-IS as the originatingLSPBufferSize, TLV #14. The default and minimum value for Sz, and the implicitly advertised value of Sz if the TLV is absent, is 1470 octets. This length (which is also the maximum size of a TRILL-Hello) was chosen to make it extremely unlikely that a TRILL control frame, even with reasonable additional headers, tags, and/or encapsulation, would encounter MTU problems on an inter-RBridge link.

The campus-wide value of Sz is the smallest value of Sz advertised by any RBridge.

4.3.2. Testing Link MTU Size

There are two new TRILL IS-IS message types for use between pairs of RBridge neighbors to test the bidirectional packet size capacity of their connection. These messages are:

-- MTU-probe
-- MTU-ack

Both the MTU-probe and the MTU-ack are padded to the size being tested.

Sending of MTU-probes is optional; however, an RBridge RB2 that receives an MTU-probe from RB1 MUST respond with an MTU-ack padded to the same size as the MTU-probe. The MTU-probe MAY be multicast to

All-RBridges, or unicast to a specific RBridge. The MTU-ack is normally unicast to the source of the MTU-probe to which it responds but MAY be multicast to All-RBridges.

If RB1 fails to receive an MTU-ack to a probe of size X from RB2 after k tries (where k is a configurable parameter whose default is 3), then RB1 assumes the RB1-RB2 link cannot support size X. If X is not greater than Sz, then RB1 sets the "failed minimum MTU test" flag for RB2 in RB1’s Hello. If size X succeeds, and X > Sz, then RB1 advertises the largest tested X for each adjacency in the TRILL Hellos RB1 sends on that link, and RB1 MAY advertise X as an attribute of the link to RB2 in RB1’s LSP.

4.4. TRILL-Hello Protocol

The TRILL-Hello protocol is a little different from the Layer 3 IS-IS LAN Hello protocol and uses a new type of IS-IS message known as a TRILL-Hello.

4.4.1. TRILL-Hello Rationale

The reason for defining this new type of link in TRILL is that in Layer 3 IS-IS, the LAN Hello protocol may elect multiple Designated Routers (DRs) since, when choosing a DR, routers ignore other routers with whom they do not have 2-way connectivity. Also, Layer 3 IS-IS LAN Hellos are padded, to avoid forming adjacencies between neighbors that can’t speak the maximum-sized packet to each other. This means, in Layer 3 IS-IS, that neighbors that have connectivity to each other, but with an MTU on that connection less than what they perceive as maximum sized packets, will not see each other’s Hellos. The result is that routers might form cliques, resulting in the link turning into multiple pseudonodes.

This behavior is fine for Layer 3, but not for Layer 2, where loops may form if there are multiple DRBs. Therefore, the TRILL-Hello protocol is a little different from Layer 3 IS-IS’s LAN Hello protocol.

One other issue with TRILL-Hellos is to ensure that subsets of the information can appear in any single message, and be processable, in the spirit of IS-IS LSPs and CSNPs. TRILL-Hello frames, even though they are not padded, can become very large. An example where this might be the case is when some sort of backbone technology interconnects hundreds of TRILL sites over what would appear to TRILL to be a giant Ethernet, where the RBridges connected to that cloud will perceive that backbone to be a single link with hundreds of neighbors.

In TRILL (unlike in Layer 3 IS-IS), the DRB is selected based solely on priority and MAC address. In other words, if RB2 receives a TRILL-Hello from RB1 with higher (priority, MAC), RB2 defers to RB1 as DRB, regardless of whether RB1 lists RB2 in RB1’s TRILL-Hello.

Although the neighbor list in a TRILL-Hello does not influence the DRB election, it does determine what is announced in LSPs. RB1 only reports links to RBridges with which it has two-way connectivity. If RB1 is the DRB on a link, and for whatever reason (MTU mismatch, or one-way connectivity) RB1 and RB2 do not have two-way connectivity, then RB2 does not report a link to RB1 (or the pseudonode), and RB1 (or RB1 on behalf of the pseudonode) does not report a link to RB2.

4.4.2. TRILL-Hello Contents and Timing

The TRILL-Hello has a new IS-IS message type. It starts with the same fixed header as an IS-IS LAN Hello, which includes the 7-bit priority for the issuing RBridge to be DRB on that link. TRILL-Hellos are sent with the same timing as IS-IS LAN Hellos.

TRILL-Hello messages, including their Outer.MacDA and Outer.MacSA, but excluding any Outer.VLAN or other tags, MUST NOT exceed 1470 octets in length and SHOULD NOT be padded. The following information MUST appear in every TRILL-Hello. References to "TLV" may actually be a "sub-TLV" as specified in separate documents [RFC6165] [RFC6326].

1. The VLAN ID of the Designated VLAN for the link.

2. A copy of the Outer.VLAN ID with which the Hello was tagged on sending.

3. A 16-bit port ID assigned by the sending RBridge to the port the TRILL-Hello is sent on such that no two ports of that RBridge have the same port ID.

4. A nickname of the sending RBridge.

5. Two flags as follows:

   5.a. A flag that, if set, indicates that the sender has detected VLAN mapping on the link, within the past 2 of its Holding Times.

   5.b. A flag that, if set, indicates that the sender believes it is appointed forwarder for the VLAN and port on which the TRILL-Hello was sent.

The following information MAY appear:

1. The set of VLANs for which end-station service is enabled on the port.

2. Several flags as follows:

   2.a. A flag that, if set, indicates that the sender’s port was configured as an access port.

   2.b. A flag that, if set, indicates that the sender’s port was configured as a trunk port.

   2.c. A bypass pseudonode flag, as described below in this section.

3. If the sender is the DRB, the Rbridges (excluding itself) that it appoints as forwarders for that link and the VLANs for which it appoints them. As described below, this TLV is designed so that not all the appointment information need be included in each Hello. Its absence means that appointed forwarders should continue as previously assigned.

4. The TRILL neighbor list. This is a new TLV, not the same as the IS-IS Neighbor TLV, in order to accommodate fragmentation and reporting MTU on the link (see Section 4.4.2.1).

The Appointed Forwarders TLV specifies a range of VLANs and, within that range, specifies which Rbridge, if any, other than the DRB, is appointed forwarder for the VLANs in that range [RFC6326]. Appointing an RBridge as forwarder on a port for a VLAN that is not enabled on that port has no effect.

It is anticipated that many links between RBridges will be point-to-point, in which case using a pseudonode merely adds to the complexity. If the DRB specifies the bypass pseudonode bit in its TRILL-Hellos, the RBridges on the link just report their adjacencies as point-to-point. This has no effect on how LSPs are flooded on a link. It only affects what LSPs are generated.

For example, if RB1 and RB2 are the only RBridges on the link and RB1 is the DRB, then if RB1 creates a pseudonode that is used, there are 3 LSPs: for, say, RB1.25 (the pseudonode), RB1, and RB2, where RB1.25 reports connectivity to RB1 and RB2, and RB1 and RB2 each just say they are connected to RB1.25. Whereas if DRB RB1 sets the bypass pseudonode bit in its Hellos, then there will be only 2 LSPs: RB1 and RB2 each reporting connectivity to each other.

A DRB SHOULD set the bypass pseudonode bit for its links unless, for a particular link, it has seen at least two simultaneous adjacencies on the link at some point since it last rebooted.

4.4.2.1. TRILL Neighbor List

The new TRILL Neighbor TLV includes the following information for each neighbor it lists:

1. The neighbor’s MAC address.

2. MTU size to this neighbor as a 2-octet unsigned integer in units of 4-octet chunks. The value zero indicates that the MTU is untested.

3. A flag for "failed minimum MTU test".

To allow partial reporting of neighbors, the neighbor IDs MUST be sorted by ID. If a set of neighbors { X1, X2, X3, ... Xn } is reported in RB1’s Hello, then X1 < X2 < X3, ... < Xn. If RBridge RB2’s ID is between X1 and Xn, and does not appear in RB1’s Hello, then RB2 knows that RB1 has not heard RB2’s Hello.

Additionally there are two overall TRILL Neighbor List TLV flags: "the smallest ID I reported in this Hello is the smallest ID of any neighbor", and "the largest ID I reported in this Hello is the largest ID of any neighbor". If all the neighbors fit in RB1’s TRILL-Hello, both flags will be set.

If RB1 reports { X1, ... Xn } in its Hello, with the "smallest" flag set, and RB2’s ID is smaller than X1, then RB2 knows that RB1 has not heard RB2’s Hello. Similarly, if RB2’s ID is larger than Xn and the "largest" flag is set, then RB2 knows that RB1 has not heard RB2’s Hello.

To ensure that any RBridge RB2 can definitively determine whether RB1 can hear RB2, RB1’s neighbor list MUST eventually cover every possible range of IDs, that is, within a period that depends on RB1’s policy and not necessarily within any specific period such as the holding time. In other words, if X1 is the smallest ID reported in one of RB1’s neighbor lists, and the "smallest" flag is not set, then X1 MUST also appear as the largest ID reported in a different TRILL-Hello neighbor list. Or, fragments may overlap, as long as there is no gap, such that some range, say, between Xi and Xj, never appears in any fragment.

4.4.3. TRILL MTU-Probe and TRILL Hello VLAN Tagging

The MTU-probe mechanism is designed to determine the MTU for transmissions between RBridges. MTU-probes and probe acknowledgements are only sent on the Designated VLAN.

An RBridge RBn maintains for each port the same VLAN information as a customer IEEE [802.1Q-2005] bridge, including the set of VLANs enabled for output through that port (see Section 4.9.2). In addition, RBn maintains the following TRILL-specific VLAN parameters per port:

a) Desired Designated VLAN: the VLAN that RBn, if it is the DRB, will specify in its TRILL-Hellos as the VLAN to be used by all RBridges on the link to communicate all TRILL frames, except some TRILL-Hellos. This MUST be a VLAN enabled on RBn’s port. It defaults to the numerically lowest enabled VLAN ID, which is VLAN 1 for a default configuration RBridge.

b) Designated VLAN: the VLAN being used on the link for all TRILL frames except some TRILL Hellos. This is RBn’s Desired Designated VLAN if RBn believes it is the DRB or the Designated VLAN in the DRB’s Hellos if RBn is not the DRB.

c) Announcing VLANs set. This defaults to the enabled VLANs set on the port but may be configured to be a subset of the enabled VLANs.

d) Forwarding VLANs set: the set of VLANs for which an RBridge port is appointed VLAN forwarder on the port. This MUST contain only enabled VLANs for the port, possibly all enabled VLANs.

On each of its ports that is not configured to use P2P Hellos, an RBridge sends TRILL-Hellos Outer.VLAN tagged with each VLAN in a set of VLANs. This set depends on the RBridge’s DRB status and the above VLAN parameters. RBridges send TRILL Hellos Outer.VLAN tagged with the Designated VLAN, unless that VLAN is not enabled on the port. In addition, the DRB sends TRILL Hellos Outer.VLAN tagged with each enabled VLAN in its Announcing VLANs set. All non-DRB RBridges send TRILL-Hellos Outer.VLAN tagged with all enabled VLANs that are in the intersection of their Forwarding VLANs set and their Announcing VLANs set. More symbolically, TRILL-Hello frames, when sent, are sent as follows:

```
If sender is DRB
   intersection ( Enabled VLANs,
   union ( Designated VLAN, Announcing VLANs ) )
```

```
If sender is not DRB
   intersection ( Enabled VLANs,
   union ( Designated VLAN,
   intersection ( Forwarding VLANs, Announcing VLANs ) ) )
```

Configuring the Announcing VLANs set to be null minimizes the number of TRILL-Hellos. In that case, TRILL-Hellos are only tagged with the Designated VLAN. Great care should be taken in configuring an RBridge to not send TRILL Hellos on any VLAN where that RBridge is appointed forwarder as, under some circumstances, failure to send such Hellos can lead to loops.

The number of TRILL-Hellos is maximized, within this specification, by configuring the Announcing VLANs set to be the set of all enabled VLAN IDs, which is the default. In that case, the DRB will send TRILL-Hello frames tagged with all its Enabled VLAN tags; in addition, any non-DRB RBridge RBn will send TRILL-Hello frames tagged with the Designated VLAN, if enabled, and tagged with all VLANs for which RBn is an appointed forwarder. (It is possible to send even more TRILL-Hellos. In particular, non-DRB RBridges could send TRILL-Hellos on enabled VLANs for which they are not an appointed forwarder and which are not the Designated VLAN. This would cause no harm other than a further communications and processing burden.)

When an RBridge port comes up, until it has heard a TRILL-Hello from a higher priority RBridge, it considers itself to be DRB on that port and sends TRILL-Hellos on that basis. Similarly, even if it has at some time recognized some other RBridge on the link as DRB, if it receives no TRILL-Hellos on that port from an RBridge with higher priority as DRB for a long enough time, as specified by IS-IS, it will revert to believing itself DRB.

4.4.4. Multiple Ports on the Same Link

It is possible for an RBridge RB1 to have multiple ports to the same link. It is important for RB1 to recognize which of its ports are on the same link, so, for instance, if RB1 is appointed forwarder for VLAN A, RB1 knows that only one of its ports acts as appointed forwarder for VLAN A on that link.

RB1 detects this condition based on receiving TRILL-Hello messages with the same IS-IS pseudonode ID on multiple ports. RB1 might have one set of ports, say, { p1, p2, p3 } on one link, and another set of ports { p4, p5 } on a second link, and yet other ports, say, p6, p7, p8, that are each on distinct links. Let us call a set of ports on the same link a "port group".

If RB1 detects that a set of ports, say, { p1, p2, p3 }, is a port group on a link, then RB1 MUST ensure that it does not cause loops when it encapsulates and decapsulates traffic from/to that link. If RB1 is appointed forwarder for VLAN A on that Ethernet link, RB1 MUST encapsulate/decapsulate VLAN A on only one of the ports. However, if RB1 is appointed forwarder for more than one VLAN, RB1 MAY choose to load split among its ports, using one port for some set of VLANs, and another port for a disjoint set of VLANs.

If RB1 detects VLAN mapping occurring (see Section 4.4.5), then RB1 MUST NOT load split as appointed forwarder, and instead MUST act as appointed VLAN forwarder on that link on only one of its ports in the port group.

When forwarding TRILL-encapsulated multi-destination frames to/from a link on which RB1 has a port group, RB1 MAY choose to load split among its ports, provided that it does not duplicate frames, and provided that it keeps frames for the same flow on the same port. If RB1’s neighbor on that link, RB2, accepts multi-destination frames on that tree on that link from RB1, RB2 MUST accept the frame from any of RB2’s adjacencies to RB1 on that link.

If an RBridge has more than one port connected to a link and those ports have the same MAC address, they can be distinguished by the port ID contained in TRILL-Hellos.

4.4.5. VLAN Mapping within a Link

IEEE [802.1Q-2005] does not provide for bridges changing the C-tag VLAN ID for a tagged frame they receive, that is, mapping VLANs. Nevertheless, some bridge products provide this capability and, in any case, bridged LANs can be configured to display this behavior. For example, a bridge port can be configured to strip VLAN tags on output and send the resulting untagged frames onto a link leading to another bridge’s port configured to tag these frames with a different VLAN. Although each port’s configuration is legal under [802.1Q-2005], in the aggregate they perform manipulations not permitted on a single customer [802.1Q-2005] bridge. Since RBridge ports have the same VLAN capabilities as customer [802.1Q-2005] bridges, this can occur even in the absence of bridges. (VLAN mapping is referred to in IEEE 802.1 as "VLAN ID translation".)

RBridges include the Outer.VLAN ID inside every TRILL-Hello message. When a TRILL-Hello is received, RBridges compare this saved copy with the Outer.VLAN ID information associated with the received frame. If these differ and the VLAN ID inside the Hello is X and the Outer.VLAN is Y, it can be assumed that VLAN ID X is being mapped into VLAN ID Y.

When non-DRB RB2 detects VLAN mapping, based on receiving a TRILL-Hello where the VLAN tag in the body of the Hello differs from the one in the outer header, it sets a flag in all of its TRILL-Hellos for a period of two of its Holding Times since the last time RB2 detected VLAN mapping. When DRB RB1 is informed of VLAN mapping, either because of receiving a TRILL-Hello that has been VLAN-mapped, or because of seeing the "VLAN mapping detected" flag in a neighbor’s TRILL-Hello on the link, RB1 re-assigns VLAN forwarders to ensure there is only a single forwarder on the link for all VLANs.

4.5. Distribution Trees

RBridges use distribution trees to forward multi-destination frames (see Section 2.4.2). Distribution trees are bidirectional. Although a single tree is logically sufficient for the entire campus, the computation of additional distribution trees is warranted for the following reasons: it enables multipathing of multi-destination frames and enables the choice of a tree root closer to or, in the limit, identical with the ingress RBridge. Such a closer tree root improves the efficiency of the delivery of multi-destination frames that are being delivered to a subset of the links in the campus and reduces out-of-order delivery when a unicast address transitions between unknown and known. If applications are in use where occasional out-of-order unicast frames due to such transitions are a problem, the RBridge campus should be engineered to make sure they are of extremely low probability, such as by using the ESADI protocol, configuring addresses to eliminate unknown destination unicast frames, or using keep alive frames.

An additional level of flexibility is the ability of an RBridge to acquire multiple nicknames, and therefore have multiple trees rooted at the same RBridge. Since the tree number is used as a tiebreaker for equal cost paths, the different trees, even if rooted at the same RBridge, will likely utilize different equal cost paths.

How an ingress RBridge chooses the distribution tree or trees that it uses for multi-destination frames is beyond the scope of this document. However, for the reasons stated above, in the absence of other factors, a good choice is the tree whose root is least cost from the ingress RBridge and that is the default for an ingress RBridge that uses a single tree to distribute multi-destination frames.

RBridges will precompute all the trees that might be used, and keep state for Reverse Path Forwarding Check filters (see Section 4.5.2). Also, since the tree number is used as a tiebreaker, it is important for all RBridges to know:

- how many trees to compute
- which trees to compute
- what the tree number for each tree is
- which trees each ingress RBridge might choose (for building Reverse Path Forwarding Check filters)

Each RBridge advertises in its LSP a "tree root" priority for its nickname or for each of its nicknames if it has been configured to have more than one. This is a 16-bit unsigned integer that defaults, for an unconfigured RBridge, to 0x8000. Tree roots are ordered with highest numerical priority being highest priority, then with system ID of the RBridge (numerically higher = higher priority) as tiebreaker, and if that is equal, by the numerically higher nickname value, as an unsigned integer, having priority.

Each RBridge advertises in its LSP the maximum number of distribution trees that it can compute and the number of trees that it wants all RBridges in the campus to compute. The number of trees, k, that are computed for the campus is the number wanted by the RBridge RB1, which has the nickname with the highest "tree root" priority, but no more than the number of trees supported by the RBridge in the campus that supports the fewest trees. If RB1 does not specify the specific distribution tree roots as described below, then the k highest priority trees are the trees that will be computed by all RBridges. Note that some of these k highest priority trees might be rooted at the same RBridge, if that RBridge has multiple nicknames.

If an RBridge specifies the number of trees it can compute, or the number of trees it wants computed for the campus, as 0, it is treated as specifying them as 1. Thus, k defaults to 1.

In addition, the RBridge RB1 having the highest root priority nickname might explicitly advertise a set of s trees by providing a list of s nicknames. In that case, the first k of those s trees will be computed. If s is less than k, or if any of the s nicknames associated with the trees RB1 is advertising does not exist within the LSP database, then the RBridges still compute k trees, but the remaining trees they select are the highest priority trees, such that k trees are computed.

There are two exceptions to the above, which can cause fewer distribution trees to be computed, as follows:

- A nickname whose tree root priority is zero is not selected as a tree root based on priority, although it may be selected by being listed by the RBridge holding the highest priority tree root nickname. The one exception to this is that if all nicknames have priority zero, then the highest priority among them as determined by the tiebreakers is used as a tree root so that there is always guaranteed to be at least one distribution tree.

- As a transient condition, two or more identical nicknames can appear in the list of roots for trees to be computed. In such a case, it is useless to compute a tree for the nickname(s) that are about to be lost by the RBridges holding them. So a distribution tree is only computed for the instance of the nickname where the priority to hold that nickname value is highest, reducing the total number of trees computed. (It would also be of little use to go further down the priority ordered list of possible tree root nicknames to maintain the number of trees as the additional tree roots found this way would only be valid for a very brief nickname transition period.)

The k trees calculated for a campus are ordered and numbered from 1 to k. In addition to advertising the number k, RB1 might explicitly advertise a set of s trees by providing a list of s nicknames as described above.

- If s == k, then the trees are numbered in the order that RB1 advertises them.

- If s == 0, then the trees are numbered in order of decreasing priority. For example, if RB1 advertises only that k=2, then the highest priority tree is number 1 and the 2nd highest priority tree is number 2.

- If s < k, then those advertised by RB1 are numbered from 1 in the order advertised. Then the remainder are chosen by priority order from among the remaining possible trees with the numbering continuing. For example, if RB1 advertises k=4, advertises { Tx, Ty } as the nicknames of the root of the trees, and the campus-wide priority ordering of trees in decreasing order is Ty > Ta > Tc > Tb > Tx, the numbering will be as follows: Tx is 1 and Ty is 2 since that is the order they are advertised in by RB1. Then Ta is 3 and Tc is 4 because they are the highest priority trees that have not already been numbered.

4.5.1. Distribution Tree Calculation

RBridges do not use spanning tree to calculate distribution trees. Instead, distribution trees are calculated based on the link state information, selecting a particular RBridge nickname as the root. Each RBridge RBn independently calculates a tree rooted at RBi by performing the SPF (Shortest Path First) calculation with RBi as the root without requiring any additional exchange of information.

It is important, when building a distribution tree, that all RBridges choose the same links for that tree. Therefore, when there are equal cost paths for a particular tree, all RBridges need to use the same tiebreakers. It is also desirable to allow splitting of traffic on as many links as possible. For this reason, a simple tiebreaker such as "always choose the parent with lower ID" would not be desirable. Instead, TRILL uses the tree number as a parameter in the tiebreaking algorithm.

When building the tree number j, remember all possible equal cost parents for node N. After calculating the entire "tree" (actually, directed graph), for each node N, if N has "p" parents, then order the parents in ascending order according to the 7-octet IS-IS ID considered as an unsigned integer, and number them starting at zero. For tree j, choose N’s parent as choice j mod p.

Note that there might be multiple equal cost links between N and potential parent P that have no pseudonodes, because they are either point-to-point links or pseudonode-suppressed links. Such links will be treated as a single link for the purpose of tree building, because they all have the same parent P, whose IS-IS ID is "P.0".

In other words, the set of potential parents for N, for the tree rooted at R, consists of those that give equally minimal cost paths from N to R and that have distinct IS-IS IDs, based on what is reported in LSPs.

4.5.2. Multi-Destination Frame Checks

When a multi-destination TRILL-encapsulated frame is received by an RBridge, there are four checks performed, each of which may cause the frame to be discarded:

1. Tree Adjacency Check: Each RBridge RBn keeps a set of adjacencies ({ port, neighbor } pairs) for each distribution tree it is calculating. One of these adjacencies is toward the tree root RBi, and the others are toward the leaves. Once the adjacencies are chosen, it is irrelevant which ones are towards the root RBi and which are away from RBi. RBridges MUST drop a multi-destination frame that arrives at a port from an RBridge that is not an adjacency for the tree on which the frame is being distributed. Let’s suppose that RBn has calculated that adjacencies a, c, and f are in the RBi tree. A multi-destination frame for the distribution tree RBi is received only from one of the adjacencies a, c, or f (otherwise it is discarded) and forwarded to the other two adjacencies. Should RBn have multiple

ports on a link, a multi-destination frame it sends on one of these ports will be received by the others but will be discarded as an RBridge is not adjacent to itself.

2. RPF Check: Another technique used by RBridges for avoiding temporary multicast loops during topology changes is the Reverse Path Forwarding Check. It involves checking that a multi-destination frame, based on the tree and the ingress RBridge, arrives from the expected link. RBridges MUST drop multi-destination frames that fail the RPF check.

   To limit the amount of state necessary to perform the RPF check, each RBridge RB2 MUST announce which trees RB2 may choose when RB2 ingresses a multi-destination packet. When any RBridge, say, RB3, is computing the tree from nickname X, RB3 computes, for each RBridge RB2 that might act as ingress for tree X, the link on which RB3 should receive a packet from ingress RB2 on tree X, and note for that link that RB2 is a legal ingress RBridge for tree X.

   The information to determine which trees RB2 might choose is included in RB2’s LSP. Similarly to how the highest priority RBridge RB1 specifies the k trees that will be computed by all RBridges, RB2 specifies a number j, which is the total number of different trees RB2 might specify, and the specific trees RB2 might specify are a combination of specified trees and trees selected from highest priority trees. If RB2 specifies any trees that are not in the k trees as specified by RB1, they are ignored.

   The j potential ingress trees for RB2 are the ones with nicknames that RB2 has explicitly specified in "specified ingress tree nicknames" (and that are included in the k campus-wide trees selected by highest priority RBridge RB1), with the remainder (up to the maximum of {j,k}) being the highest priority of the k campus-wide trees.

   The default value for j is 1. The value 0 for j is special and means that RB2 can pick any of the k trees being computed for the campus.

3. Parallel Links Check: If the tree-building and tiebreaking for a particular multi-destination frame distribution tree selects a non-pseudonode link between RB1 and RB2, that "RB1-RB2 link" might actually consist of multiple links. These parallel links would be visible to RB1 and RB2, but not to the rest of the campus (because the links are not represented by pseudonodes). If this bundle of parallel links is included in a tree, it is important for RB1 and RB2 to decide which link to use, but is irrelevant to other RBridges, and therefore, the tiebreaking algorithm need not be

visible to any RBridges other than RB1 and RB2. In this case, RB1-RB2 adjacencies are ordered as follows, with the one "most preferred" adjacency being the one on which RB1 and RB2 transmit to and receive multi-destination frames from each other.

a) Most preferred are those established by P2P Hellos. Tiebreaking among those is based on preferring the one with the numerically highest Extended Circuit ID as associated with the adjacency by the RBridge with the highest System ID.

b) Next considered are those established through TRILL-Hello frames, with suppressed pseudonodes. Note that the pseudonode is suppressed in LSPs, but still appears in the TRILL-Hello, and therefore is available for this tiebreaking. Among these links, the one with the numerically largest pseudonode ID is preferred.

4. Port Group Check: If an RBridge has multiple ports attached to the same link, a multi-destination frame it is receiving will arrive on all of them. All but one received copy of such a frame MUST be discarded to avoid duplication. All such frames that are part of the same flow must be accepted on the same port to avoid re-ordering.

When a topology change occurs (including apparent changes during start up), an RBridge MUST adjust its input distribution tree filters no later than it adjusts its output forwarding.

4.5.3. Pruning the Distribution Tree

Each distribution tree SHOULD be pruned per VLAN, eliminating branches that have no potential receivers downstream. Multi-destination TRILL Data frames SHOULD only be forwarded on branches that are not pruned.

Further pruning SHOULD be done in two cases: (1) IGMP [RFC3376], MLD [RFC2710], and MRD [RFC4286] messages, where these are to be delivered only to links with IP multicast routers; and (2) other multicast frames derived from an IP multicast address that should be delivered only to links that have registered listeners, plus links that have IP multicast routers, except for IP multicast addresses that must be broadcast. Each of these cases is scoped per VLAN.

Let’s assume that RBridge RBn knows that adjacencies (a, c, and f) are in the nickname1 distribution tree. RBn marks pruning information for each of the adjacencies in the nickname1-tree. For each adjacency and for each tree, RBn marks:

- the set of VLANs reachable downstream,

- for each one of those VLANs, flags indicating whether there are IPv4 or IPv6 multicast routers downstream, and

- the set of Layer 2 multicast addresses derived from IP multicast groups for which there are receivers downstream.

4.5.4. Tree Distribution Optimization

RBridges MUST determine the VLAN associated with all native frames they ingress and properly enforce VLAN rules on the emission of native frames at egress RBridge ports according to how those ports are configured and designated as appointed forwarders. RBridges SHOULD also prune the distribution tree of multi-destination frames according to VLAN. But, since they are not required to do such pruning, they may receive TRILL data or ESADI frames that should have been VLAN pruned earlier in the tree distribution. They silently discard such frames. A campus may contain some Rbridges that prune distribution trees on VLAN and some that do not.

The situation is more complex for multicast. RBridges SHOULD analyze IP-derived native multicast frames, and learn and announce listeners and IP multicast routers for such frames as discussed in Section 4.7 below. And they SHOULD prune the distribution of IP-derived multicast frames based on such learning and announcements. But, they are not required to prune based on IP multicast listener and router attachment state. And, unlike VLANs, where VLAN attachment state of ports MUST be maintained and honored, RBridges are not required to maintain IP multicast listener and router attachment state.

An RBridge that does not examine native IGMP [RFC3376], MLD [RFC2710], or MRD [RFC4286] frames that it ingresses MUST advertise that it has IPv4 and IPv6 IP multicast routers attached for all the VLANs for which it is an appointed forwarder. It need not advertise any IP-derived multicast listeners. This will cause all IP-derived multicast traffic to be sent to this RBridge for those VLANs. It then egresses that traffic onto the links for which it is appointed forwarder where the VLAN of the traffic matches the VLAN for which it is appointed forwarder on that link. (This may cause the suppression of certain IGMP membership report messages from end stations, but that is not significant because any multicast traffic that such reports would be requesting will be sent to such end stations under these circumstances.)

A campus may contain a mixture of Rbridges with different levels of IP-derived multicast optimization. An RBridge may receive IP-derived multicast frames that should have been pruned earlier in the tree distribution. It silently discards such frames.

See also "Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches" [RFC4541].

4.5.5. Forwarding Using a Distribution Tree

With full optimization, forwarding a multi-destination data frame is done as follows. References to adjacencies below do not include the adjacency on which a frame was received.

- The RBridge RBn receives a multi-destination TRILL Data frame with inner VLAN-x and a TRILL header indicating that the selected tree is the nickname1 tree;

- if the source from which the frame was received is not one of the adjacencies in the nickname1 tree for the specified ingress RBridge, the frame is dropped (see Section 4.5.1);

- else, if the frame is an IGMP or MLD announcement message or an MRD query message, then the encapsulated frame is forwarded onto adjacencies in the nickname1 tree that indicate there are downstream VLAN-x IPv4 or IPv6 multicast routers as appropriate;

- else, if the frame is for a Layer 2 multicast address derived from an IP multicast group, but its IP address is not the range of IP multicast addresses that must be treated as broadcast, the frame is forwarded onto adjacencies in the nickname1 tree that indicate there are downstream VLAN-x IP multicast routers of the corresponding type (IPv4 or IPv6), as well as adjacencies that indicate there are downstream VLAN-x receivers for that group address;

- else (the inner frame is for a Layer 2 multicast address not derived from an IP multicast group or an unknown destination or broadcast or an IP multicast address that is required to be treated as broadcast), the frame is forwarded onto an adjacency if and only if that adjacency is in the nickname1 tree, and marked as reaching VLAN-x links.

For each link for which RBn is appointed forwarder, RBn additionally checks to see if it should decapsulate the frame and send it to the link in native form, or process the frame locally.

TRILL ESADI frames will be delivered only to RBridges that are appointed forwarders for their VLAN. Such frames will be multicast throughout the campus, like other non-IP-derived multicast data frames, on the distribution tree chosen by the RBridge that created the TRILL ESADI frame, and pruned according to the Inner.VLAN ID. Thus, all the RBridges that are appointed forwarders for a link in that VLAN receive them.

4.6. Frame Processing Behavior

This section describes RBridge behavior for all varieties of received frames, including how they are forwarded when appropriate. Section 4.6.1 covers native frames, Section 4.6.2 covers TRILL frames, and Section 4.6.3 covers Layer 2 control frames. Processing may be organized or sequenced in a different way than described here as long as the result is the same. See Section 1.4 for frame type definitions.

Corrupt frames, for example, frames that are not a multiple of 8 bits, are too short or long for the link protocol/hardware in use, or have a bad FCS are discarded on receipt by an RBridge port as they are discarded on receipt at an IEEE 802.1 bridge port.

Source address information ({ VLAN, Outer.MacSA, port }) is learned by default from any frame with a unicast source address (see Section 4.8).

4.6.1. Receipt of a Native Frame

If the port is configured as disabled or if end-station service is disabled on the port by configuring it as a trunk port or configuring it to use P2P Hellos, the frame is discarded.

The ingress Rbridge RB1 determines the VLAN ID for a native frame according to the same rules as IEEE [802.1Q-2005] bridges do (see Appendix D and Section 4.9.2). Once the VLAN is determined, if RB1 is not the appointed forwarder for that VLAN on the port where the frame was received or is inhibited, the frame is discarded. If it is appointed forwarder for that VLAN and is not inhibited (see Section 4.2.4.3), then the native frame is forwarded according to Section 4.6.1.1 if it is unicast and according to Section 4.6.1.2 if it is multicast or broadcast.

4.6.1.1. Native Unicast Case

If the destination MAC address of the native frame is a unicast address, the following steps are performed.

The Layer 2 destination address and VLAN are looked up in the ingress RBridge’s database of MAC addresses and VLANs to find the egress RBridge RBm or the local egress port or to discover that the destination is the receiving RBridge or is unknown. One of the following four cases will apply:

1. If the destination is the receiving RBridge, the frame is locally processed.

2. If the destination is known to be on the same link from which the native frame was received but is not the receiving RBridge, the RBridge silently discards the frame, since the destination should already have received it.

3. If the destination is known to be on a different local link for which RBm is the appointed forwarder, then RB1 converts the native frame to a TRILL Data frame with an Outer.MacDA of the next hop RBridge towards RBm, a TRILL header with M = 0, an ingress nickname for RB1, and the egress nickname for RBm. If ingress RB1 has multiple nicknames, it SHOULD use the same nickname in the ingress nickname field whenever it encapsulates a native frame from any particular source MAC address and VLAN. This simplifies end node learning. If RBm is RB1, processing then proceeds as in Section 4.6.2.4; otherwise, the Outer.MacSA is set to the MAC address of the RB1 port on the path to the next hop RBridge towards RBm and the frame is queued for transmission out of that port.

4. If a unicast destination MAC is unknown in the frame’s VLAN, RB1 handles the frame as described in Section 4.6.1.2 for a broadcast frame except that the Inner.MacDA is the original native frame’s unicast destination address.

4.6.1.2. Native Multicast and Broadcast Frames

If the RBridge has multiple ports attached to the same link, all but one received copy of a native multicast or broadcast frame is discarded to avoid duplication. All such frames that are part of the same flow must be accepted on the same port to avoid re-ordering.

If the frame is a native IGMP [RFC3376], MLD [RFC2710], or MRD [RFC4286] frame, then RB1 SHOULD analyze it, learn any group membership or IP multicast router presence indicated, and announce that information for the appropriate VLAN in its LSP (see Section 4.7).

For all multi-destination native frames, RB1 forwards the frame in native form to its links where it is appointed forwarder for the frame’s VLAN, subject to further pruning and inhibition. In addition, it converts the native frame to a TRILL Data frame with the All-RBridges multicast address as Outer.MacDA, a TRILL header with the multi-destination bit M = 1, the ingress nickname for RB1, and the egress nickname for the distribution tree it decides to use. It then forwards the frame on the pruned distribution tree (see Section 4.5) setting the Outer.MacSA of each copy sent to the MAC address of the RB1 port on which it is sent.

The default is for RB1 to write into the egress nickname field the nickname for a distribution tree, from the set of distribution trees RB1 has announced it might use, whose root is least cost from RB1. RB1 MAY choose different distribution trees for different frames if RB1 has been configured to path-split multicast. In that case, RB1 MUST select a tree by specifying a nickname that is a distribution tree root (see Section 4.5). Also, RB1 MUST select a nickname that RB1 has announced (in RB1’s own LSP) to be one of those that RB1 might use. The strategy RB1 uses to select distribution trees in multipathing multi-destination frames is beyond the scope of this document.

4.6.2. Receipt of a TRILL Frame

A TRILL frame either has the TRILL or L2-IS-IS Ethertype or has a multicast Outer.MacDA allocated to TRILL (see Section 7.2). The following tests are performed sequentially, and the first that matches controls the handling of the frame:

1. If the Outer.MacDA is All-IS-IS-RBridges and the Ethertype is L2-IS-IS, the frame is handled as described in Section 4.6.2.1.

2. If the Outer.MacDA is a multicast address allocated to TRILL other than All-RBridges, the frame is discarded.

3. If the Outer.MacDA is a unicast address other than the receiving Rbridge port MAC address, the frame is discarded. (Such discarded frames are most likely addressed to another RBridge on a multi-access link and that other Rbridge will handle them.)

4. If the Ethertype is not TRILL, the frame is discarded.

5. If the Version field in the TRILL header is greater than 0, the frame is discarded.

6. If the hop count is 0, the frame is discarded.

7. If the Outer.MacDA is multicast and the M bit is zero or if the Outer.MacDA is unicast and M bit is one, the frame is discarded.

8. By default, an RBridge MUST NOT forward TRILL-encapsulated frames from a neighbor with which it does not have a TRILL IS-IS adjacency. RBridges MAY be configured per port to accept these frames for forwarding in cases where it is known that a non-peering device (such as an end station) is configured to originate TRILL-encapsulated frames that can be safely forwarded.

9. The Inner.MacDA is then tested. If it is the All-ESADI-RBridges multicast address and RBn implements the ESADI protocol, processing proceeds as in Section 4.6.2.2 below. If it is any other address or RBn does not implement ESADI, processing proceeds as in Section 4.6.2.3.

4.6.2.1. TRILL Control Frames

The frame is processed by the TRILL IS-IS instance on RBn and is not forwarded.

4.6.2.2. TRILL ESADI Frames

If M == 0, the frame is silently discarded.

The egress nickname designates the distribution tree. The frame is forwarded as described in Section 4.6.2.5. In addition, if the forwarding Rbridge is an appointed forwarder for a link in the specified VLAN and implements the TRILL ESADI protocol and ESADI is enabled at the forwarding Rbridge for that VLAN, the inner frame is decapsulated and provided to that local ESADI protocol.

4.6.2.3. TRILL Data Frames

The M flag is then checked. If it is zero, processing continues as described in Section 4.6.2.4, if it is one, processing continues as described in Section 4.6.2.5.

4.6.2.4. Known Unicast TRILL Data Frames

The egress nickname in the TRILL header is examined, and if it is unknown or reserved, the frame is discarded.

If RBn is a transit RBridge, the hop count is decremented by one and the frame is forwarded to the next hop RBridge towards the egress RBridge. (The provision permitting RBridges to decrease the hop count by more than one under some circumstances (see Section 3.6) applies only to multi-destination frames, not to the known unicast frames considered in this subsection.) The Inner.VLAN is not examined by a transit RBridge when it forwards a known unicast TRILL Data frame. For the forwarded frame, the Outer.MacSA is the MAC address of the transmitting port, the Outer.MacDA is the unicast address of the next hop RBridge, and the VLAN is the Designated VLAN on the link onto which the frame is being transmitted.

If RBn is not a transit RBridge, that is, if the egress RBridge indicated is the RBridge performing the processing, the Inner.MacSA and Inner.VLAN ID are, by default, learned as associated with the ingress nickname unless that nickname is unknown or reserved or the Inner.MacSA is not unicast. Then the frame being forwarded is decapsulated to native form, and the following checks are performed:

- The Inner.MacDA is checked. If it is not unicast, the frame is discarded.

- If the Inner.MacDA corresponds to the RBridge doing the processing, the frame is locally delivered.

- The Inner.VLAN ID is checked. If it is 0x0 or 0xFFF, the frame is discarded.

- The Inner.MacDA and Inner.VLAN ID are looked up in RBn’s local address cache and the frame is then either sent onto the link containing the destination, if the RBridge is appointed forwarder for that link for the frame’s VLAN and is not inhibited (or discarded if it is inhibited), or processed as in one of the following two paragraphs.

A known unicast TRILL Data frame can arrive at the egress Rbridge only to find that the combination of Inner.MacDA and Inner.VLAN is not actually known by that RBridge. One way this can happen is that the address information may have timed out in the egress RBridge MAC address cache. In this case, the egress RBridge sends the native frame out on all links that are in the frame’s VLAN for which the RBridge is appointed forwarder and has not been inhibited, except that it MAY refrain from sending the frame on links where it knows there cannot be an end station with the destination MAC address, for example, the link port is configured as a trunk (see Section 4.9.1).

If, due to manual configuration or learning from Layer 2 registration, the destination MAC and VLAN appear in RBn’s local address cache for two or more links for which RBn is an uninhibited appointed forwarder for the frame’s VLAN, RBn sends the native frame on all such links.

4.6.2.5. Multi-Destination TRILL Data Frames

The egress and ingress nicknames in the TRILL header are examined and, if either is unknown or reserved, the frame is discarded.

The Outer.MacSA is checked and the frame discarded if it is not a tree adjacency for the tree indicated by the egress RBridge nickname on the port where the frame was received. The Reverse Path Forwarding Check is performed on the ingress and egress nicknames and the frame discarded if it fails. If there are multiple TRILL-Hello pseudonode suppressed parallel links to the previous hop RBridge, the frame is discarded if it has been received on the wrong one. If the RBridge has multiple ports connected to the link, the frame is discarded unless it was received on the right one. For more information on the checks in this paragraph, see Section 4.5.2.

If the Inner.VLAN ID of the frame is 0x0 or 0xFFF, the frame is discarded.

If the RBridge is an appointed forwarder for the Inner.VLAN ID of the frame, the Inner.MacSA and Inner.VLAN ID are, by default, learned as associated with the ingress nickname unless that nickname is unknown or reserved or the Inner.MacSA is not unicast. A copy of the frame is then decapsulated, sent in native form on those links in its VLAN for which the RBridge is appointed forwarder subject to additional pruning and inhibition as described in Section 4.2.4.3, and/or locally processed as appropriate.

The hop count is decreased (possibly by more than one; see Section 3.6), and the frame is forwarded down the tree specified by the egress RBridge nickname pruned as described in Section 4.5.

For the forwarded frame, the Outer.MacSA is set to that of the port on which the frame is being transmitted, the Outer.MacDA is the All-RBridges multicast address, and the VLAN is the Designated VLAN of the link on which the frame is being transmitted.

4.6.3. Receipt of a Layer 2 Control Frame

Low-level control frames received by an RBridge are handled within the port where they are received as described in Section 4.9.

There are two types of high-level control frames, distinguished by their destination address, which are handled as described in the sections referenced below.

Name	Section	Destination Address
BPDU	4.9.3	01-80-C2-00-00-00
VRP	4.9.4	01-80-C2-00-00-21

4.7. IGMP, MLD, and MRD Learning

Ingress RBridges SHOULD learn, based on native IGMP [RFC3376], MLD [RFC2710], and MRD [RFC4286] frames they receive in VLANs for which they are an uninhibited appointed forwarder, which IP-derived multicast messages should be forwarded onto which links. Such frames are also, in general, encapsulated as TRILL Data frames and distributed as described below and in Section 4.5.

An IGMP or MLD membership report received in native form from a link indicates a multicast group listener for that group on that link. An IGMP or MLD query or an MRD advertisement received in native form from a link indicates the presence of an IP multicast router on that link.

IP multicast group membership reports have to be sent throughout the campus and delivered to all IP multicast routers, distinguishing IPv4 and IPv6. All IP-derived multicast traffic must also be sent to all IP multicast routers for the same version of IP.

IP multicast data SHOULD only be sent on links where there is either an IP multicast router for that IP type (IPv4 or IPv6) or an IP multicast group listener for that IP-derived multicast MAC address, unless the IP multicast address is in the range required to be treated as broadcast.

RBridges do not need to announce themselves as listeners to the IPv4 All-Snoopers multicast group (the group used for MRD reports [RFC4286]), because the IPv4 multicast address for that group is in the range where all frames sent to that IP multicast address must be broadcast (see [RFC4541], Section 2.1.2). However, RBridges that are performing IPv6-derived multicast optimization MUST announce themselves as listeners to the IPv6 All-Snoopers multicast group.

See also "Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches" [RFC4541].

4.8. End-Station Address Details

RBridges have to learn the MAC addresses and VLANs of their locally attached end stations for link/VLAN pairs for which they are the appointed forwarder. Learning this enables them to do the following:

- o Forward the native form of incoming known unicast TRILL Data frames onto the correct link.

- Decide, for an incoming native unicast frame from a link, where the RBridge is the appointed forwarder for the frame’s VLAN, whether the frame is

  - known to have been destined for another end station on the same link, so the RBridge need do nothing, or

  - has to be converted to a TRILL Data frame and forwarded.

RBridges need to learn the MAC addresses, VLANs, and remote RBridges of remotely attached end stations for VLANs for which they and the remote RBridge are an appointed forwarder, so they can efficiently direct native frames they receive that are unicast to those addresses and VLANs.

4.8.1. Learning End-Station Addresses

There are five independent ways an RBridge can learn end-station addresses as follows:

1. From the observation of VLAN-x frames received on ports where it is appointed VLAN-x forwarder, learning the { source MAC, VLAN, port } triplet of received frames.

2. The { source MAC, VLAN, ingress RBridge nickname } triplet of any native frames that it decapsulates.

3. By Layer 2 registration protocols learning the { source MAC, VLAN, port } of end stations registering at a local port.

4. By running the TRILL ESADI protocol for one or more VLANs and thereby receiving remote address information and/or transmitting local address information.

5. By management configuration.

RBridges MUST implement capabilities 1 and 2 above. RBridges use these capabilities unless configured, for one or more particular VLANs and/or ports, not to learn from either received frames or from decapsulating native frames to be transmitted or both.

RBridges MAY implement capabilities 3 and 4 above. If capability 4 is implemented, the ESADI protocol is run only when the RBridge is configured to do so on a per-VLAN basis.

RBridges SHOULD implement capability 5.

Entries in the table of learned MAC and VLAN addresses and associated information also have a one-octet unsigned confidence level associated with each entry whose rationale is given below. Such information learned from the observation of data has a confidence of 0x20 unless configured to have a different confidence. This confidence level can be configured on a per-RBridge basis separately for information learned from local native frames and that learned from remotely originated encapsulated frames. Such information received via the TRILL ESADI protocol is accompanied by a confidence level in the range 0 to 254. Such information configured by management defaults to a confidence level of 255 but may be configured to have another value.

The table of learned MAC addresses includes (1) { confidence, VLAN, MAC address, local port } for addresses learned from local native frames and local registration protocols, (2) { confidence, VLAN, MAC address, egress RBridge nickname } for addresses learned from remote encapsulated frames and ESADI link state databases, and (3) additional information to implement timeout of learned addresses, statically configured addresses, and the like.

When a new address and related information learned from observing data frames are to be entered into the local database, there are three possibilities:

A. If this is a new { address, VLAN } pair, the information is entered accompanied by the confidence level.

B. If there is already an entry for this { address, VLAN } pair with the same accompanying delivery information, the confidence level in the local database is set to the maximum of its existing confidence level and the confidence level with which it is being learned. In addition, if the information is being learned with the same or a higher confidence level than its existing confidence level, timer information is reset.

C. If there is already an entry for this { address, VLAN } pair with different information, the learned information replaces the older information only if it is being learned with higher or equal confidence than that in the database entry. If it replaces older information, timer information is also reset.

4.8.2. Learning Confidence Level Rationale

The confidence level mechanism allows an RBridge campus manager to cause certain address learning sources to prevail over others. In a default configuration, without the optional ESADI protocol, addresses are only learned from observing local native frames and the decapsulation of received TRILL Data frames. Both of these sources default to confidence level 0x20 so, since learning at an equal or high confidence overrides previous learning, the learning in such a default case mimics default 802.1 bridge learning.

While RBridge campus management policies are beyond the scope of this document, here are some example types of policies that can be implemented with the confidence mechanism and the rationale for each:

- Locally received native frames might be considered more reliable than decapsulated frames received from remote parts of the campus. To stop MAC addresses learned from such local frames from being usurped by remotely received forged frames, the confidence in locally learned addresses could be increased or that in addresses learned from remotely sourced decapsulated frames decreased.

- MAC address information learned through a cryptographically authenticated Layer 2 registration protocol, such as 802.1X with a cryptographically based EAP method, might be considered more reliable than information learned through the mere observation of data frames. When such authenticated learned address information is transmitted via the ESADI protocol, the use of authentication in the TRILL ESADI LSP frames could make tampering with it in transit very difficult. As a result, it might be reasonable to announce such authenticated information via the ESADI protocol with a high confidence, so it would override any alternative learning from data observation.

Manually configured address information is generally considered static and so defaults to a confidence of 0xFF while no other source of such information can be configured to a confidence any higher than 0xFE. However, for other cases, such as where the manual configuration is just a starting point that the Rbridge campus manager wishes to be dynamically overridable, the confidence of such manually configured information may be configured to a lower value.

4.8.3. Forgetting End-Station Addresses

While RBridges need to learn end-station addresses as described above, it is equally important that they be able to forget such information. Otherwise, frames for end stations that have moved to a different part of the campus could be indefinitely black-holed by RBridges with stale information as to the link to which the end station is attached.

For end-station address information locally learned from frames received, the time out from the last time a native frame was received or decapsulated with the information conforms to the recommendations of [802.1D]. It is referred to as the "Ageing Time" and is configurable per RBridge with a range of from 10 seconds to 1,000,000 seconds and a default value of 300 seconds.

The situation is different for end-station address information acquired via the TRILL ESADI protocol. It is up to the originating RBridge to decide when to remove such information from its ESADI LSPs (or up to ESADI protocol timeouts if the originating RBridge becomes inaccessible).

When an RBridge ceases to be appointed forwarder for VLAN-x on a port, it forgets all end-station address information learned from the observation of VLAN-x native frames received on that port. It also increments a per-VLAN counter of the number of times it lost appointed forwarder status on one of its ports for that VLAN.

When, for all of its ports, RBridge RBn is no longer appointed forwarder for VLAN-x, it forgets all end-station address information learned from decapsulating VLAN-x native frames. Also, if RBn is participating in the TRILL ESADI protocol for VLAN-x, it ceases to so participate after sending a final LSP nulling out the end-station address information for the VLAN that it had been originating. In addition, all other RBridges that are VLAN-x forwarder on at least one of their ports notice that the link state data for RBn has changed to show that it no longer has a link on VLAN-x. In response, they forget all end-station address information they have learned from decapsulating VLAN-x frames that show RBn as the ingress RBridge.

When the appointed forwarder lost counter for RBridge RBn for VLAN-x is observed to increase via the TRILL IS-IS link state database but RBn continues to be an appointed forwarder for VLAN-x on at least one of its ports, every other RBridge that is an appointed forwarder for VLAN-x modifies the aging of all the addresses it has learned by decapsulating native frames in VLAN-x from ingress RBridge RBn as follows: the time remaining for each entry is adjusted to be no larger than a per-RBridge configuration parameter called (to correspond to [802.1D]) "Forward Delay". This parameter is in the range of 4 to 30 seconds with a default value of 15 seconds.

4.8.4. Shared VLAN Learning

RBridges can map VLAN IDs into a smaller number of identifiers for purposes of address learning, as [802.1Q-2005] bridges can. Then, when a lookup is done in learned address information, this identifier is used for matching in place of the VLAN ID. If the ID of the VLAN on which the address was learned is not retained, then there are the following consequences:

- o The RBridge no longer has the information needed to participate in the TRILL ESADI protocol for the VLANs whose ID is not being retained.

- o In cases where Section 4.8.3 above requires the discarding of learned address information based on a particular VLAN, when the VLAN ID is not available for entries under a shared VLAN identifier, instead the time remaining for each entry under that shared VLAN identifier is adjusted to be no longer than the RBridge's "Forward Delay".

Although outside the scope of this specification, there are some Layer 2 features in which a set of VLANs has shared learning, where one of the VLANs is the "primary" and the other VLANs in the group are "secondaries". An example of this is where traffic from different communities is separated using VLAN tags, and yet some resource (such as an IP router or DHCP server) is to be shared by all the communities. A method of implementing this feature is to give a VLAN tag, say, Z, to a link containing the shared resource, and have the other VLANs, say, A, C, and D, be part of the group { primary = Z, secondaries = A, C, D }. An RBridge, aware of this grouping, attached to one of the secondary VLANs in the group also claims to be attached to the primary VLAN. So an RBridge attached to A would claim to also be attached to Z. An RBridge attached to the primary would claim to be attached to all the VLANs in the group.

This document does not specify how VLAN groups might be used. Only RBridges that participate in a VLAN group will be configured to know about the VLAN group. However, to detect misconfiguration, an RBridge configured to know about a VLAN group SHOULD report the VLAN group in its LSP.

4.9. RBridge Ports

Section 4.9.1 below describes the several RBridge port configuration bits, Section 4.9.2 gives a logical port structure in terms of frame processing, and Sections 4.9.3 and 4.9.4 describe the handling of high-level control frames.

4.9.1. RBridge Port Configuration

There are four per-port configuration bits as follows:

- o Disable port bit. When this bit is set, all frames received or to be transmitted are discarded, with the possible exception of some Layer 2 control frames (see Section 1.4) that may be generated and transmitted or received and processed within the port. By default, ports are enabled.

- End-station service disable (trunk port) bit. When this bit is set, all native frames received on the port and all native frames that would have been sent on the port are discarded. (See Appendix B.) (Note that, for this document, "native frames" does not include Layer 2 control frames.) By default, ports are not restricted to being trunk ports.

  If a port with end-station service disabled reports, in a TRILL-Hello frame it sends out that port, which VLANs it provides end-station support for, it reports that there are none.

- TRILL traffic disable (access port) bit. If this bit is set, the goal is to avoid sending any TRILL frames, except TRILL-Hello frames, on the port since it is intended only for native end-station traffic. By default, ports are not restricted to being access ports. This bit is reported in TRILL-Hello frames. If RB1 is the DRB and has this bit set in its TRILL-Hello, the DRB still appoints VLAN forwarders. However, usually no pseudonode is reported, and none of the inter-RBridge links associated with that link are reported in LSPs.

  If the DRB RB1 does not have this bit set, but neighbor RB2 on the link does have the bit set, then RB1 does not appoint RB2 as appointed forwarder for any VLAN, and none of the RBridges (including the pseudonode) report RB2 as a neighbor in LSPs.

  In some cases even though the DRB has the "access port" flag set, the DRB MAY choose to create a pseudonode for the access port. In this case, the other RBridges report connectivity to the pseudonode in their LSP, but the DRB sets the "overload" flag in the pseudonode LSP.

- Use P2P Hellos bit. If this bit is set, Hellos sent on this port are IS-IS P2P Hellos. By default TRILL-Hellos are used. See Section 4.2.4.1 for more information on P2P links.

The dominance relationship of these four configuration bits is as follows, where configuration bits to the left dominate those to the right. That is to say, when any pair of bits is asserted, inconsistencies in behavior they mandate are resolved in favor of behavior mandated by the bit to the left of the other bit in this list.

```
    Disable > P2P > Access > Trunk
```

4.9.2. RBridge Port Structure

An RBridge port can be modeled as having a lower-level structure similar to that of an [802.1Q-2005] bridge port as shown in Figure 11. In this figure, the double lines represent the general flow of the frames and information while single lines represent information flow only. The dashed lines in connection with VRP (GVRP/MVRP) are to show that VRP support is optional. An actual RBridge port implementation may be structured in any way that provides the correct behavior.

```
                     +----------------------------------------------
                     |                RBridge
                     |
                     | Interport Forwarding, IS-IS.  Management, ...
                     |
                     +----++----------------------+-------------++--
                          ||                      |             ||
                    TRILL || Data                 |             ||
                          ||                   +--+---------+   ||
            +-------------++-----+             |   TRILL    |   ||
            |    Encapsulation   |      +------+ IS-IS Hello|   ||
            |    Decapsulation   |      |      | Processing |   ||
            |     Processing     |      |      +-----++-----+   ||
            +--------------------+      |            ||         ||
            |  RBridge Appointed +------+            ||         ||
        +---+   Forwarder and    |                   ||         ||
        |   |  Inhibition Logic  +==============\\   ||   //====++
        |   +---------+--------+-+   Native       \\ ++ //
        |             |        |     Frames         \++/
        |             |        |                     ||
   +----+-----+  +- - + - - +  |                     ||
   |  RBridge |  |  RBridge |  |                     || All TRILL and
   |   BPDU   |  |    VRP   |  |                     || Native Frames
   |Processing|  |Processing|  |                     ||
   +-----++---+  + - - -+- -+  |            +--------++--+ <- EISS
         ||             |      |            |   802.1Q   |
         ||            |       |            | Port VLAN  |
         ||             |      |            |and Priority|
         ||            |       |            | Processing |
     +---++-----------++------+-----------+-----------+ <-- ISS
     |        802.1/802.3 Low-Level Control Frame        |
     |        Processing, Port/Link Control Logic        |
     +------------++-------------------------------------+
                  ||
                  ||        +------------+
                  ||        | 802.3 PHY  |
                  ++========+ (Physical  +======== 802.3
                            | Interface) |         Link
                            +------------+
```

Figure 11: Detailed RBridge Port Model

Low-level control frames are handled in the lower-level port/link control logic in the same way as in an [802.1Q-2005] bridge port. This can optionally include a variety of 802.1 or link specific protocols such as PAUSE (Annex 31B of [802.3]), link layer discovery [802.1AB], link aggregation [802.1AX], MAC security [802.1AE], or port-based access control [802.1X]. While handled at a low level, these frames may affect higher-level processing. For example, a Layer 2 registration protocol may affect the confidence in learned addresses. The upper interface to this lower-level port control logic corresponds to the Internal Sublayer Service (ISS) in [802.1Q-2005].

High-level control frames (BPDUs and, if supported, VRP frames) are not VLAN tagged. Although they extend through the ISS interface, they are not subject to port VLAN processing. Behavior on receipt of a VLAN tagged BPDU or VLAN tagged VRP frame is unspecified. If a VRP is not implemented, then all VRP frames are discarded. Handling of BPDUs is described in Section 4.9.3. Handling of VRP frames is described in Section 4.9.4.

Frames other than Layer 2 control frames, that is, all TRILL and native frames, are subject to port VLAN and priority processing that is the same as for an [802.1Q-2005] bridge. The upper interface to the port VLAN and priority processing corresponds to the Extended Internal Sublayer Service (EISS) in [802.1Q-2005].

In this model, RBridge port processing below the EISS layer is identical to an [802.1Q-2005] bridge except for (1) the handling of high-level control frames and (2) that the discard of frames that have exceeded the Maximum Transit Delay is not mandatory but MAY be done.

As described in more detail elsewhere in this document, incoming native frames are only accepted if the RBridge is an uninhibited appointed forwarder for the frame’s VLAN, after which they are normally encapsulated and forwarded; outgoing native frames are usually obtained by decapsulation and are only output if the RBridge is an uninhibited appointed forwarder for the frame’s VLAN.

TRILL-Hellos, MTU-probes, and MTU-acks are handled per port and, like all TRILL IS-IS frames, are never forwarded. They can affect the appointed forwarder and inhibition logic as well as the RBridge’s LSP.

Except TRILL-Hellos, MTU-probes, and MTU-acks, all TRILL control as well as TRILL data and ESADI frames are passed up to higher-level RBridge processing on receipt and passed down for transmission on creation or forwarding. Note that these frames are never blocked due to the appointed forwarder and inhibition logic, which affects only native frames, but there are additional filters on some of them such as the Reverse Path Forwarding Check.

4.9.3. BPDU Handling

If RBridge campus topology were static, RBridges would simply be end stations from a bridging perspective, terminating but not otherwise interacting with spanning tree. However, there are reasons for RBridges to listen to and sometimes to transmit BPDUs as described below. Even when RBridges listen to and transmit BPDUs, this is a local RBridge port activity. The ports of a particular RBridge never interact so as to make the RBridge as a whole a spanning tree node.

4.9.3.1. Receipt of BPDUs

Rbridges MUST listen to spanning tree configuration BPDUs received on a port and keep track of the root bridge, if any, on that link. If MSTP is running on the link, this is the CIST root. This information is reported per VLAN by the RBridge in its LSP and may be used as described in Section 4.2.4.3. In addition, the receipt of spanning tree configuration BPDUs is used as an indication that a link is a bridged LAN, which can affect the RBridge transmission of BPDUs.

An RBridge MUST NOT encapsulate or forward any BPDU frame it receives.

RBridges discard any topology change BPDUs they receive, but note Section 4.9.3.3.

4.9.3.2. Root Bridge Changes

A change in the root bridge seen in the spanning tree BPDUs received at an RBridge port may indicate a change in bridged LAN topology, including the possibility of the merger of two bridged LANs or the like, without any physical indication at the port. During topology transients, bridges may go into pre-forwarding states that block TRILL-Hello frames. For these reasons, when an RBridge sees a root bridge change on a port for which it is appointed forwarder for one or more VLANs, it is inhibited for a period of time between zero and 30 seconds. (An inhibited appointed forwarder discards all native frames received from or that it would otherwise have sent to the link.) This time period is configurable per port and defaults to 30 seconds.

For example, consider two bridged LANs carrying multiple VLANs, each with various RBridge appointed forwarders. Should they become merged, due to a cable being plugged in or the like, those RBridges attached to the original bridged LAN with the lower priority root will see a root bridge change while those attached to the other original bridged LAN will not. Thus, all appointed forwarders in the lower priority set will be inhibited for a time period while things are sorted out by BPDUs within the merged bridged LAN and TRILL-Hello frames between the RBridges involved.

4.9.3.3. Transmission of BPDUs

When an RBridge ceases to be appointed forwarder for one or more VLANs out a particular port, it SHOULD, as long as it continues to receive spanning tree BPDUs on that port, send topology change BPDUs until it sees the topology change acknowledged in a spanning tree configuration BPDU.

RBridges MAY support a capability for sending spanning tree BPDUs for the purpose of attempting to force a bridged LAN to partition as discussed in Appendix A.3.3.

4.9.4. Dynamic VLAN Registration

Dynamic VLAN registration provides a means for bridges (and less commonly end stations) to request that VLANs be enabled or disabled on ports leading to the requestor. This is done by VLAN registration protocol (VRP) frames: GVRP or MVRP. RBridges MAY implement GVRP and/or MVRP as described below.

VRP frames are never encapsulated as TRILL frames between RBridges or forwarded in native form by an RBridge. If an RBridge does not implement a VRP, it discards any VRP frames received and sends none.

RBridge ports may have dynamically enabled VLANs. If an RBridge supports a VRP, the actual enablement of dynamic VLANs is determined by GVRP/MVRP frames received at the port as it would be for an [802.1Q-2005] / [802.1ak] bridge.

An RBridge that supports a VRP sends GVRP/MVRP frames as an [802.1Q-2005] / [802.1ak] bridge would send on each port that is not configured as an RBridge trunk port or P2P port. For this purpose, it sends VRP frames to request traffic in the VLANs for which it is appointed forwarder and in the Designated VLAN, unless the Designated VLAN is disabled on the port, and to not request traffic in any other VLAN.

5. RBridge Parameters

This section lists parameters for RBridges. It is expected that the TRILL MIB will include many of the items listed in this section plus additional Rbridge status and data including traffic and error counts.

The default value and range are given for parameters added by TRILL. Where a parameter is defined as a 16-bit unsigned integer and an explicit maximum is not given, that maximum is 2**16-1. For parameters imported from [802.1Q-2005], [802.1D], or IS-IS [ISO10589] [RFC1195], see those standards for default and range if not given here.

5.1. Per RBridge

The following parameters occur per RBridge:

- Number of nicknames, which defaults to 1 and may be configured in the range of 1 to 256.

- The desired number of distribution trees to be calculated by every RBridge in the campus and a desired number of distribution trees for the advertising RBridge to use, both of which are unsigned 16-bit integers that default to 1 (see Section 4.5).

- The maximum number of distribution trees the RBridge can compute. This is a 16-bit unsigned integer that is implementation and environment dependent and not subject to management configuration.

- Two lists of nicknames, one designating the distribution trees to be computed and one designating distribution trees to be used as discussed in Section 4.5. By default, these lists are empty.

- The parameters Ageing Timer and Forward Delay with the default and range specified in [802.1Q-2005].

- Two unsigned octets that are, respectively, the confidence in { MAC, VLAN, local port } triples learned from locally received native frames and the confidence in { MAC, VLAN, remote RBridge } triples learned from decapsulating frames. These each default to 0x20 and may each be configured to values from 0x00 to 0xFE.

- The desired minimum acceptable inter-RBridge link MTU for the campus, that is, originatingLSPBufferSize. This is a 16-bit unsigned integer number of octets that defaults to 1470 bytes, which is the minimum valid value. Any lower value being advertised by an RBridge is ignored.

- The number of failed MTU-probes before the RBridge concludes that a particular MTU is not supported, which defaults to 3 and may be configured between 1 and 255.

Static end-station address information and confidence in such end station information statically configured can also be configured with a default confidence of 0xFF and range of 0x00 to 0xFF. By default, there is no such static address information. The quantity of such information that can be configured is implementation dependent.

5.2. Per Nickname Per RBridge

The following is configuration information per nickname at each RBridge:

- Priority to hold the nickname, which defaults to 0x40 if no specific value has been configured or 0xC0 if it is configured (see Section 3.7.3).

- Nickname priority to be selected as a distribution tree root, a 16-bit unsigned integer that defaults to 0x8000.

- Nickname value, an unsigned 16-bit quantity that defaults to the configured value if configured, else to the last value held if the RBridge coming up after a reboot and that value is remembered, else to a random value; however, in all cases the reserved values 0x0000 and 0xFFC0 through 0xFFFF are excluded (see Section 3.7.3).

5.3. Per Port Per RBridge

An RBridge has the following per-port configuration parameters:

- The same parameters as an [802.1Q-2005] port in terms of C-VLAN IDs. In addition, there is an Announcing VLANs set that defaults to the enabled VLANs on the port (see Section 4.4.3) and ranges from the null set to the set of all legal VLAN IDs.

- The same parameters as an [802.1Q-2005] port in terms of frame priority code point mapping (see [802.1Q-2005]).

- The inhibition time for the port when it observed a change in the root bridge of an attached bridged LAN. This is in units of seconds, defaults to 30, and can be configured to any value from 0 to 30.

- The Desired Designated VLAN that the RBridge will advertise in its TRILL Hellos if it is the DRB for the link via that port. This defaults to the lowest VLAN ID enabled on the port and may be configured to any valid VLAN ID that is enabled on the port (0x001 through 0xFFE).

- Four per-port configuration bits: disable port (default 0 == enabled), disable end-station service (trunk, default 0 == enabled), access port (default 0 == not restricted to being an access port), and use P2P Hellos (default 0 == use TRILL Hellos). (See Section 4.9.1.)

- One bit per port such that, if the bit is set, it disables learning { MAC address, VLAN, port } triples from locally received native frames on the port. Default value is 0 == learning enabled.

- The priority of the RBridge to be DRB and its Holding Time via that port with defaults and range as specified in IS-IS [ISO10589] [RFC1195].

- A bit that, when set, enables the receipt of TRILL-encapsulated frames from an Outer.MacSA with which the RBridges does not have an IS-IS adjacency. Default value is 0 == disabled.

- Configuration for the optional send-BPDUs solution to the wiring closet topology problem as described in Appendix A.3.3. Default Bridge Address is the System ID of the RBridge with the lowest System ID. If RB1 and RB2 are part of a wiring closet topology, both need to be configured to know about this, and know the ID that should be used in the spanning tree protocol on the specified port.

5.4. Per VLAN Per RBridge

An RBridge has the following per-VLAN configuration parameters:

- Per-VLAN ESADI protocol participation flag, 7-bit priority, and Holding Time. Default participation flag is 0 == not participating. Default and range of priority and Holding Time as specified in IS-IS [ISO10589] [RFC1195].

- One bit per VLAN that, if set, disables learning { MAC address, VLAN, remote RBridge } triples from frames decapsulated in the VLAN. Defaults to 0 == learning enabled.

6. Security Considerations

Layer 2 bridging is not inherently secure. It is, for example, subject to spoofing of source addresses and bridging control messages. A goal for TRILL is that RBridges do not add new issues beyond those existing in current bridging technology.

Countermeasures are available such as to configure the TRILL IS-IS and ESADI protocols to use IS-IS security [RFC5304] [RFC5310] and ignore unauthenticated TRILL control and ESADI frames received. RBridges using IS-IS security will need configuration.

IEEE 802.1 port admission and link security mechanisms, such as [802.1X] and [802.1AE], can also be used. These are best thought of as being implemented below TRILL (see Section 4.9.2) and are outside the scope of TRILL (just as they are generally out of scope for bridging standards [802.1D] and 802.1Q); however, TRILL can make use of secure registration through the confidence level communicated in the optional TRILL ESADI protocol (see Section 4.8).

TRILL encapsulates native frames inside the RBridge campus while they are in transit between ingress RBridge and egress RBridge(s) as described in Sections 2.3 and 4.1. Thus, TRILL ignorant devices with firewall features that cannot be detected by RBridges as end stations will generally not be able to inspect the content of such frames for security checking purposes. This may render them ineffective. Layer 3 routers and hosts appear to RBridges to be end stations, and native frames will be decapsulated before being sent to such devices. Thus, they will not see the TRILL Ethertype. Firewall devices that do not appear to an RBridge to be an end station, for example, bridges with co-located firewalls, should be modified to understand TRILL encapsulation.

RBridges do not prevent nodes from impersonating other nodes, for instance, by issuing bogus ARP/ND replies. However, RBridges do not interfere with any schemes that would secure neighbor discovery.

6.1. VLAN Security Considerations

TRILL supports VLANs. These provide logical separation of traffic, but care should be taken in using VLANs for security purposes. To have reasonable assurance of such separation, all the RBridges and links (including bridged LANs) in a campus must be secured and configured so as to prohibit end stations from using dynamic VLAN registration frames or otherwise gaining access to any VLAN carrying traffic for which they are not authorized to read and/or inject.

Furthermore, if VLANs were used to keep some information off links where it might be observed in a bridged LAN, this will no longer work, in general, when bridges are replaced with RBridges; with encapsulation and a different outer VLAN tag, the data will travel the least cost transit path regardless of VLAN. Appropriate counter measures are to use end-to-end encryption or an appropriate TRILL security option should one be specified.

6.2. BPDU/Hello Denial-of-Service Considerations

The TRILL protocol requires that an appointed forwarder at an RBridge port be temporarily inhibited if it sees a TRILL-Hello from another RBridge claiming to be the appointed forwarder for the same VLAN or sees a root bridge change out that port. Thus, it would seem that forged BPDUs showing repeated root bridge changes and forged TRILL-Hello frames with the Appointed Forwarder flag set could represent a significant denial-of-service attack. However, the situation is not as bad as it seems.

The best defense against forged TRILL-Hello frames or other IS-IS messages is the use of IS-IS security [RFC5304] [RFC5310]. Rogue end stations would not normally have access to the required IS-IS keying material needed to forge authenticatible messages.

Authentication similar to IS-IS security is usually unavailable for BPDUs. However, it is also the case that in typical modern wired LANs, all the links are point-to-point. If you have an all-RBridged point-to-point campus, then the worst that an end-station can do by forging BPDUs or TRILL-Hello frames is to deny itself service. This could be either through falsely inhibiting the forwarding of native frames by the RBridge to which it is connected or by falsely activating the optional decapsulation check (see Section 4.2.4.3).

However, when an RBridge campus contains bridged LANs, those bridged LANs appear to any connected RBridges to be multi-access links. The forging of BPDUs by an end-station attached to such a bridged LAN could affect service to other end-stations attached to the same bridged LAN. Note that bridges never forward BPDUs but process them, although this processing may result in the issuance of further BPDUs. Thus, for an end-station to forge BPDUs to cause continuing changes in the root bridge as seen by an RBridge through intervening bridges would typically require it to cause root bridge thrashing throughout the bridged LAN that would be disruptive even in the absence of RBridges.

Some bridges can be configured to not send BPDUs and/or to ignore BPDUs on particular ports, and RBridges can be configured not to inhibit appointed forwarding on a port due to root bridge changes; however, such configuration should be used with caution as it can be unsafe.

7. Assignment Considerations

This section discuses IANA and IEEE 802 assignment considerations. See [RFC5226].

7.1. IANA Considerations

A new IANA registry has been created for TRILL Parameters with two subregistries as below.

The initial contents of the TRILL Nicknames subregistry are as follows:

0x0000 Reserved to indicate no nickname specified
0x0001-0xFFBF Dynamically allocated by the RBridges within each RBridge campus
0xFFC0-0xFFFE Available for allocation by RFC Required (single value) or IETF Review (single or multiple values)
0xFFFF Permanently reserved

The initial contents of the TRILL Multicast Address subregistry are as follows:

01-80-C2-00-00-40	Assigned as All-RBridges
01-80-C2-00-00-41	Assigned as All-IS-IS-RBridges
01-80-C2-00-00-42	Assigned as All-ESADI-RBridges
01-80-C2-00-00-43 to 01-80-C2-00-00-4F	Available for allocation by IETF Review

7.2. IEEE Registration Authority Considerations

The Ethertype 0x22F3 is assigned by the IEEE Registration Authority to the TRILL Protocol.

The Ethertype 0x22F4 is assigned by the IEEE Registration Authority for L2-IS-IS.

The block of 16 multicast MAC addresses from <01-80-C2-00-00-40> to <01-80-C2-00-00-4F> is assigned by the IEEE Registration Authority for IETF TRILL protocol use.

8. Normative References

[802.1ak] "IEEE Standard for Local and metropolitan area networks / Virtual Bridged Local Area Networks / Multiple Registration Protocol", IEEE Standard 802.1ak-2007, 22 June 2007.

[802.1D] "IEEE Standard for Local and metropolitan area networks / Media Access Control (MAC) Bridges", 802.1D-2004, 9 June 2004.

[802.1Q-2005]
"IEEE Standard for Local and metropolitan area networks / Virtual Bridged Local Area Networks", 802.1Q-2005, 19 May 2006.

[802.3] "IEEE Standard for Information technology / Telecommunications and information exchange between systems / Local and metropolitan area networks / Specific requirements Part 3: Carrier sense multiple access with collision detection (CSMA/CD) access method and physical layer specifications", 802.3-2008, 26 December 2008.

[ISO10589] ISO/IEC, "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.

[RFC1112] Deering, S., "Host extensions for IP multicasting", STD 5, RFC 1112, August 1989.

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

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

[RFC2464] Crawford, M., "Transmission of IPv6 Packets over Ethernet Networks", RFC 2464, December 1998.

[RFC2710] Deering, S., Fenner, W., and B. Haberman, "Multicast Listener Discovery (MLD) for IPv6", RFC 2710, October 1999.

[RFC3376] Cain, B., Deering, S., Kouvelas, I., Fenner, B., and A. Thyagarajan, "Internet Group Management Protocol, Version 3", RFC 3376, October 2002.

[RFC3417] Presuhn, R., Ed., "Transport Mappings for the Simple Network Management Protocol (SNMP)", STD 62, RFC 3417, December 2002.

[RFC3419] Daniele, M. and J. Schoenwaelder, "Textual Conventions for Transport Addresses", RFC 3419, December 2002.

[RFC4286] Haberman, B. and J. Martin, "Multicast Router Discovery", RFC 4286, December 2005.

[RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.

[RFC5303] Katz, D., Saluja, R., and D. Eastlake 3rd, "Three-Way Handshake for IS-IS Point-to-Point Adjacencies", RFC 5303, October 2008.

[RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, October 2008.

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

[RFC6326] Eastlake, D., Banerjee, A., Dutt, D., Perlman, R., and A. Ghanwani, "Transparent Interconnection of Lots of Links (TRILL) Use of IS-IS", RFC 6326, July 2011.

9. Informative References

[802.1AB] "IEEE Standard for Local and Metropolitan Networks / Station and Media Access Control Connectivity Discovery", 802.1AB-2009, 17 September 2009.

[802.1ad] "IEEE Standard for and metropolitan area networks / Virtual Bridged Local Area Networks / Provider Bridges", 802.1ad-2005, 26 May 2005.

[802.1ah] "IEEE Standard for Local and Metropolitan Area Networks / Virtual Bridged Local Area Networks / Provider Backbone Bridges", 802.1ah-2008, 1 January 2008.

[802.1aj] Virtual Bridged Local Area Networks / Two-Port Media Access Control (MAC) Relay, 802.1aj Draft 4.2, 24 September 2009.

[802.1AE] "IEEE Standard for Local and metropolitan area networks / Media Access Control (MAC) Security", 802.1AE-2006, 18 August 2006.

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

[802.1X] "IEEE Standard for Local and metropolitan area networks / Port Based Network Access Control", 802.1X-REV Draft 2.9, 3 September 2008.

[RBridges] Perlman, R., "RBridges: Transparent Routing", Proc. Infocom 2005, March 2004.

[RFC1661] Simpson, W., Ed., "The Point-to-Point Protocol (PPP)", STD 51, RFC 1661, July 1994.

[RFC3411] Harrington, D., Presuhn, R., and B. Wijnen, "An Architecture for Describing Simple Network Management Protocol (SNMP) Management Frameworks", STD 62, RFC 3411, December 2002.

[RFC4086] Eastlake 3rd, D., Schiller, J., and S. Crocker, "Randomness Requirements for Security", BCP 106, RFC 4086, June 2005.

[RFC4541] Christensen, M., Kimball, K., and F. Solensky, "Considerations for Internet Group Management Protocol (IGMP) and Multicast Listener Discovery (MLD) Snooping Switches", RFC 4541, May 2006.

[RFC4789] Schoenwaelder, J. and T. Jeffree, "Simple Network Management Protocol (SNMP) over IEEE 802 Networks", RFC 4789, November 2006.

[RFC5304] Li, T. and R. Atkinson, "IS-IS Cryptographic Authentication", RFC 5304, October 2008.

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

[RFC5342] Eastlake 3rd, D., "IANA Considerations and IETF Protocol Usage for IEEE 802 Parameters", BCP 141, RFC 5342, September 2008.

[RFC5556] Touch, J. and R. Perlman, "Transparent Interconnection of Lots of Links (TRILL): Problem and Applicability Statement", RFC 5556, May 2009.

[RP1999] Perlman, R., "Interconnection: Bridges, Routers, Switches, and Internetworking Protocols, 2nd Edition", Addison Wesley Longman, Chapter 3, 1999.

[VLAN-MAPPING]
Perlman, R., Dutt, D., Banerjee, A., Rijhsinghani, A., and D. Eastlake 3rd, "RBridges: Campus VLAN and Priority Regions", Work in Progress, April 2011.

Appendix A. Incremental Deployment Considerations

Some aspects of partial RBridge deployment are described below for link cost determination (Appendix A.1) and possible congestion due to appointed forwarder bottlenecks (Appendix A.2). A particular example of a problem related to the TRILL use of a single appointed forwarder per link per VLAN (the "wiring closet topology") is explored in detail in Appendix A.3.

A.1. Link Cost Determination

With an RBridged campus having no bridges or repeaters on the links between RBridges, the RBridges can accurately determine the number of physical hops involved in a path and the line speed of each hop, assuming this is reported by their port logic. With intervening devices, this is no longer possible. For example, as shown in Figure 12, the two bridges B1 and B2 can completely hide a slow link so that both Rbridges RB1 and RB2 incorrectly believe the link is faster.

```
+-----+        +----+        +----+        +-----+
|     |  Fast  |    |  Slow  |    |  Fast  |     |
| RB1 +--------+ B1 +--------+ B2 +--------+ RB2 |
|     |  Link  |    |  Link  |    |  Link  |     |
+-----+        +----+        +----+        +-----+
```

Figure 12: Link Cost of a Bridged Link

Even in the case of a single intervening bridge, two RBridges may know they are connected but each sees the link as a different speed from how it is seen by the other.

However, this problem is not unique to RBridges. Bridges can encounter similar situations due to links hidden by repeaters, and routers can encounter similar situations due to links hidden by bridges, repeaters, or Rbridges.

A.2. Appointed Forwarders and Bridged LANs

With partial RBridge deployment, the RBridges may partition a bridged LAN into a relatively small number of relatively large remnant bridged LANs, or possibly not partition it at all so a single bridged LAN remains. Such configuration can result in the following problem:

The requirement that native frames enter and leave a link via the link’s appointed forwarder for the VLAN of the frame can cause congestion or suboptimal routing. (Similar problems can occur within a bridged LAN due to the spanning tree algorithm.) The extent to which such a problem will occur is highly dependent on the network topology. For example, if a bridged LAN had a star-like structure with core bridges that connected only to other bridges and peripheral bridges that connected to end stations and are connected to core bridges, the replacement of all of the core bridges by RBridges without replacing the peripheral bridges would generally improve performance without inducing appointed forwarder congestion.

Solutions to this problem are discussed below and a particular example explored in Appendix A.3.

Inserting RBridges so that all the bridged portions of the LAN stay connected to each other and have multiple RBridge connections is generally the least efficient arrangement.

There are four techniques that may help if the problem above occurs and that can, to some extent, be used in combination:

1. Replace more IEEE 802.1 customer bridges with RBridges so as to minimize the size of the remnant bridged LANs between RBridges. This requires no configuration of the RBridges unless the bridges they replace required configuration.

2. Re-arrange network topology to minimize the problem. If the bridges and RBridges involved are configured, this may require changes in their configuration.

3. Configure the RBridges and bridges so that end stations on a remnant bridged LAN are separated into different VLANs that have different appointed forwarders. If the end stations were already assigned to different VLANs, this is straightforward (see Section 4.2.4.2). If the end stations were on the same VLAN and have to be split into different VLANs, this technique may lead to connectivity problems between end stations.

4. Configure the RBridges such that their ports that are connected to the bridged LAN send spanning tree configuration BPDUs (see Section A.3.3) in such a way as to force the partition of the bridged LAN. (Note: A spanning tree is never formed through an RBridge but always terminates at RBridge ports.) To use this technique, the RBridges must support this optional feature, and would need to be configured to use it, but the bridges involved would rarely have to be configured. This technique makes the bridged LAN unavailable for TRILL through traffic because the bridged LAN partitions.

Conversely to item 3 above, there may be bridged LANs that use VLANs, or use more VLANs than would otherwise be necessary, to support the Multiple Spanning Tree Protocol or otherwise reduce the congestion that can be caused by a single spanning tree. Replacing the IEEE 802.1 bridges in such LANs with RBridges may enable a reduction in or elimination of VLANs and configuration complexity.

A.3. Wiring Closet Topology

If 802.1 bridges are present and RBridges are not properly configured, the bridge spanning tree or the DRB may make inappropriate decisions. Below is a specific example of the more general problem that can occur when a bridged LAN is connected to multiple RBridges.

In cases where there are two (or more) groups of end nodes, each attached to a bridge (say, B1 and B2), and each bridge is attached to an RBridge (say, RB1 and RB2, respectively), with an additional link connecting B1 and B2 (see Figure 13), it may be desirable to have the B1-B2 link only as a backup in case one of RB1 or RB2 or one of the links B1-RB1 or B2-RB2 fails.

```
                  +-------------------------------+
                  |             |          |      |
                  |  Data    +-----+    +-----+   |
                  | Center  -| RB1 |----| RB2 |-  |
                  |          +-----+    +-----+   |
                  |             |          |      |
                  +-------------------------------+
                                |          |
                                |          |
                  +-------------------------------+
                  |             |          |      |
                  |          +----+     +----+    |
                  | Wiring   | B1 |-----| B2 |    |
                  | Closet   +----+     +----+    |
                  | Bridged                       |
                  | LAN                           |
                  +-------------------------------+
```

Figure 13: Wiring Closet Topology

For example, B1 and B2 may be in a wiring closet and it may be easy to provide a short, high-bandwidth, low-cost link between them while RB1 and RB2 are at a distant data center such that the RB1-B1 and RB2-B2 links are slower and more expensive.

Default behavior might be that one of RB1 or RB2 (say, RB1) would become DRB for the bridged LAN including B1 and B2 and appoint itself forwarder for the VLANs on that bridged LAN. As a result, RB1 would forward all traffic to/from the link, so end nodes attached to B2 would be connected to the campus via the path B2-B1-RB1, rather than the desired B2-RB2. This wastes the bandwidth of the B2-RB2 path and cuts available bandwidth between the end stations and the data center in half. The desired behavior would be to make use of both the RB1-B1 and RB2-B2 links.

Three solutions to this problem are described below.

A.3.1. The RBridge Solution

Of course, if B1 and B2 are replaced with RBridges, the right thing will happen without configuration (other than VLAN support), but this may not be immediately practical if bridges are being incrementally replaced by RBridges.

A.3.2. The VLAN Solution

If the end stations attached to B1 and B2 are already divided among a number of VLANs, RB1 and RB2 could be configured so that whichever becomes DRB for this link will appoint itself forwarder for some of these VLANs and appoint the other RBridge for the remaining VLANs. Should either of the RBridges fail or become disconnected, the other will have only itself to appoint as forwarder for all the VLANs.

If the end stations are all on a single VLAN, then it would be necessary to assign them between at least two VLANs to use this solution. This may lead to connectivity problems that might require further measures to rectify.

A.3.3. The Spanning Tree Solution

Another solution is to configure the relevant ports on RB1 and RB2 to be part of a "wiring closet group", with a configured per-RBridge port "Bridge Address" Bx (which may be RB1 or RB2’s System ID). Both RB1 and RB2 emit spanning tree BPDUs on their configured ports as highest priority root Bx. This causes the spanning tree to logically partition the bridged LAN as desired by blocking the B1-B2 link at one end or the other (unless one of the bridges is configured to also have highest priority and has a lower ID, which we consider to be a misconfiguration). With the B1-B2 link blocked, RB1 and RB2 cannot see each other’s TRILL-Hellos via that link and each acts as Designated RBridge and appointed forwarder for its respective partition. Of course, with this partition, no TRILL through traffic can flow through the RB1-B1-B2-RB2 path.

In the spanning tree configuration BPDU, the Root is "Bx" with highest priority, cost to Root is 0, Designated Bridge ID is "RB1" when RB1 transmits and "RB2" when RB2 transmits, and port ID is a value chosen independently by each of RB1 and RB2 to distinguish each of its own ports. The topology change flag is zero, and the topology change acknowledgement flag is set if and only if a topology change BPDU has been received on the port since the last configuration BPDU was transmitted on the port. (If RB1 and RB2 were actually bridges on the same shared medium with no bridges between them, the result would be that the one with the larger ID sees "better" BPDUs (because of the tiebreaker on the third field: the ID of the transmitting bridge), and would turn off its port.)

Should either RB1 or the RB1-B1 link or RB2 or the RB2-B2 link fail, the spanning tree algorithm will stop seeing one of the RBx roots and will unblock the B1-B2 link maintaining connectivity of all the end stations with the data center.

If the link RB1-B1-B2-RB2 is on the cut set of the campus and RB2 and RB1 have been configured to believe they are part of a wiring closet group, the campus becomes partitioned as the link is blocked.

A.3.4. Comparison of Solutions

Replacing all 802.1 customer bridges with RBridges is usually the best solution with the least amount of configuration required, possibly none.

The VLAN solution works well with a relatively small amount of configuration if the end stations are already divided among a number of VLANs. If they are not, it becomes more complex and problematic.

The spanning tree solution does quite well in this particular case. But it depends on both RB1 and RB2 having implemented the optional feature of being able to configure a port to emit spanning tree BPDUs as described in Appendix A.3.3 above. It also makes the bridged LAN whose partition is being forced unavailable for through traffic. Finally, while in this specific example it neatly breaks the link between the two bridges B1 and B2, if there were a more complex bridged LAN, instead of exactly two bridges, there is no guarantee that it would partition into roughly equal pieces. In such a case, you might end up with a highly unbalanced load on the RB1-B1 link and the RB2-B2 link although this is still better than using only one of these links exclusively.

Appendix B. Trunk and Access Port Configuration

Many modern bridged LANs are organized into a core and access model, The core bridges have only point-to-point links to other bridges while the access bridges connect to end stations, core bridges, and possibly other access bridges. It seems likely that some RBridge campuses will be organized in a similar fashion.

An RBridge port can be configured as a trunk port, that is, a link to another RBridge or RBridges, by configuring it to disable end-station support. There is no reason for such a port to have more than one VLAN enabled and in its Announcing Set on the port. Of course, the RBridge (or RBridges) to which it is connected must have the same VLAN enabled. There is no reason for this VLAN to be other than the default VLAN 1 unless the link is actually over carrier Ethernet or other facilities that only provide some other specific VLAN or the like. Such configuration minimizes wasted TRILL-Hellos and eliminates useless decapsulation and transmission of multi-destination traffic in native form onto the link (see Sections 4.2.4 and 4.9.1).

An RBridge access port would be expected to lead to a link with end stations and possibly one or more bridges. Such a link might also have more than one RBridge connected to it to provide more reliable service to the end stations. It would be a goal to minimize or eliminate transit traffic on such a link as it is intended for end-station native traffic. This can be accomplished by turning on the access port configuration bit for the RBridge port or ports connected to the link as further detailed in Section 4.9.1.

When designing RBridge configuration user interfaces, consideration should be given to making it convenient to configure ports as trunk and access ports.

Appendix C. Multipathing

Rbridges support multipathing of both known unicast and multi-destination traffic. Implementation of multipathing is optional.

Multi-destination traffic can be multipathed by using different distribution tree roots for different frames. For example, assume that in Figure 14 end stations attached to RBy are the source of various multicast streams each of which has multiple listeners attached to various of RB1 through RB9. Assuming equal bandwidth links, a distribution tree rooted at RBy will predominantly use the vertical links among RB1 through RB9 while one rooted at RBz will predominantly use the horizontal. If RBy chooses its nickname as the distribution tree root for half of this traffic and an RBz nickname as the root for the other half, it may be able to substantially increase the aggregate bandwidth by making use of both the vertical and horizontal links among RB1 through RB9.

Since the distribution trees an RBridge must calculate are the same for all RBridges and transit RBridges MUST respect the tree root specified by the ingress RBridge, a campus will operate correctly with a mix of RBridges some of which use different roots for different multi-destination frames they ingress and some of which use a single root for all such frames.

```
                              +---+
                              |RBy|---------------+
                              +---+               |
                             /  |  \              |
                           /    |    \            |
                         /      |      \          |
                      +---+   +---+   +---+       |
                      |RB1|---|RB2|---|RB3|       |
                      +---+   +---+   +---+\      |
                        |       |       |    \    |
                      +---+   +---+   +---+    \+---+
                      |RB4|---|RB5|---|RB6|-----|RBz|
                      +---+   +---+   +---+    /+---+
                        |       |       |    /
                      +---+   +---+   +---+/
                      |RB7|---|RB8|---|RB9|
                      +---+   +---+   +---+
```

Figure 14: Multi-Destination Multipath

Known unicast Equal Cost Multipathing (ECMP) can occur at an RBridge if, instead of using a tiebreaker criterion when building SPF paths, information is retained about ports through which equal cost paths are available. Different unicast frames can then be sent through those different ports and will be forwarded by equal cost paths. For example, in Figure 15, which shows only RBridges and omits any bridges present, there are three equal cost paths between RB1 and RB2 and two equal cost paths between RB2 and RB5. Thus, for traffic transiting this part of the campus from left to right, RB1 may be able to perform three way ECMP and RB2 may be able to perform two-way ECMP.

A transit RBridge receiving a known unicast frame forwards it towards the egress RBridge and is not concerned with whether it believes itself to be on any particular path from the ingress RBridge or a previous transit RBridge. Thus, a campus will operate correctly with a mix of RBridges some of which implement ECMP and some of which do not.

There are actually three possibilities for the parallel paths between RB1 and RB2 as follows:

1. If two or three of these paths have pseudonodes, then all three will be distinctly visible in the campus-wide link state and ECMP as described above is applicable.

2. If the paths use P2P Hellos or otherwise do not have pseudonodes, these three paths would appear as a single adjacency in the link state. In this case, multipathing across them would be an entirely local matter for RB1 and RB2. It can be freely done for known unicast frames but not for multi-destination frames as described in Section 4.5.2.

3. If and only if the three paths between RB1 and RB2 are single hop equal bandwidth links with no intervening bridges, then it would be permissible to combine them into one logical link through the [802.1AX] "link aggregation" feature. Rbridges MAY implement link aggregation since that feature operates below TRILL (see Section 4.9.2).

```
                               +----+       double line = 10 Gbps
                 -----      ===|RB3|---     single line = 1 Gbps
                /     \   //   +----+   \
            +----+     +----+           +----+
         ===|RB1|-----|RB2|           |RB5|===
            +----+     +----+           +----+
                \     /   \    +----+   //
                 -----     ----|RB4|===
                               +----+
```

Figure 15: Known Unicast Multipath

When multipathing is used, frames that follow different paths will be subject to different delays and may be re-ordered. While some traffic may be order/delay insensitive, typically most traffic consists of flows of frames where re-ordering within a flow is damaging. How to determine flows or what granularity flows should have is beyond the scope of this document. (This issue is discussed in [802.1AX].)

Appendix D. Determination of VLAN and Priority

A high-level, informative summary of how VLAN ID and priority are determined for incoming native frames, omitting some details, is given in the bulleted items below. For more detailed information, see [802.1Q-2005].

- When an untagged native frame arrives, an unconfigured RBridge associates the default priority zero and the VLAN ID 1 with it. It actually sets the VLAN for the untagged frame to be the "port VLAN ID" associated with that port. The port VLAN ID defaults to VLAN ID 1 but may be configured to be any other VLAN ID. An Rbridge may also be configured on a per-port basis to discard such frames or to associate a different priority code point with them. Determination of the VLAN ID associated with an incoming untagged non-control frame may also be made dependent on the Ethertype or NSAP (referred to in 802.1 as the Protocol) of the arriving frame, the source MAC address, or other local rules.

- When a priority tagged native frame arrives, an unconfigured RBridge associates with it both the port VLAN ID, which defaults to 1, and the priority code point provided in the priority tag in the frame. An Rbridge may be configured on a per-port basis to discard such frames or to associate them with a different VLAN ID as described in the point immediately above. It may also be configured to map the priority code point provided in the frame by specifying, for each of the eight possible values that might be in the frame, what actual priority code point will be associated with the frame by the RBridge.

- When a C-tagged (formerly called Q-tagged) native frame arrives, an unconfigured RBridge associates with it the VLAN ID and priority in the C-tag. An RBridge may be configured on a per-port per-VLAN basis to discard such frames. It may also be configured on a per-port basis to map the priority value as specified above for priority tagged frames.

In 802.1, the process of associating a priority code point with a frame, including mapping a priority provided in the frame to another priority, is referred to as priority "regeneration".

Appendix E. Support of IEEE 802.1Q-2005 Amendments

This informational appendix briefly comments on RBridge support for completed and in-process amendments to IEEE [802.1Q-2005]. There is no assurance that existing RBridge protocol specifications or existing bridges will support not yet specified future [802.1Q-2005] amendments just as there is no assurance that existing bridge protocol specifications or existing RBridges will support not yet specified future TRILL amendments.

The information below is frozen as of 25 October 2009. For the latest status, see the IEEE 802.1 working group (http://grouper.ieee.org/groups/802/1/).

E.1. Completed Amendments

802.1ad-2005 Provider Bridges - Sometimes called "Q-in-Q", because VLAN tags used to be called "Q-tags", 802.1ad specifies Provider Bridges that tunnel customer bridge traffic within service VLAN tags (S-tags). If the customer LAN is an RBridge campus, that traffic will be bridged by Provider Bridges. Customer bridge features involving Provider Bridge awareness, such as the ability to configure a customer bridge port to add an S-tag to a frame before sending it to a Provider Bridge, are below the EISS layer and can be supported in RBridge ports without modification to the TRILL protocol.

802.1ag-2007 Connectivity Fault Management (CFM) - This 802.1 feature is at least in part dependent on the symmetric path and other characteristics of spanning tree. The comments provided to the IETF TRILL working group by the IEEE 802.1 working group stated that "TRILL weakens the applicability of CFM".

802.1ak-2007 Multiple Registration Protocol - Supported to the extent described in Section 4.9.4.

802.1ah-2008 Provider Backbone Bridges - Sometimes called "MAC-in-MAC", 802.1ah provides for Provider Backbone Bridges that tunnel customer bridge traffic within different outer MAC addresses and using a tag (the "I-tag") to preserve the original MAC addresses and signal other information. If the customer LAN is an RBridge campus, that traffic will be bridged by Provider Backbone Bridges. Customer bridge features involving Provider Backbone Bridge awareness, such as the ability to configure a customer bridge port to add an I-tag to a frame before sending it to a Provider Backbone Bridge, are below the EISS layer and can be supported in RBridge ports without modification to the TRILL protocol.

802.1Qaw-2009 Management of Data-Driven and Data-Dependent Connectivity Fault - Amendment building on 802.1ag. See comments on 802.1ag-2007 above.

802.1Qay-2009 Provider Backbone Bridge Traffic Engineering - Amendment building on 802.1ah to configure traffic engineered routing. See comments on 802.1ah-2008 above.

E.2. In-Process Amendments

The following are amendments to IEEE [802.1Q-2005] that are in process. As such, the brief comments below are based on drafts and may be incorrect for later versions or any final amendment.

802.1aj Two-port MAC Relay [802.1aj] - This amendment specifies a MAC relay that will be transparent to RBridges. RBridges are compatible with IEEE 802.1aj devices as currently specified, in the same sense that IEEE 802.1Q-2005 bridges are compatible with such devices.

802.1aq Shortest Path Bridging - This amendment provides for improved routing in bridged LANs.

802.1Qat Stream Reservation Protocol - Modification to 802.1Q to support the 802.1 Timing and Synchronization. This protocol reserves resources for streams at supporting bridges.

802.1Qau Congestion Notification - It currently appears that modifications to RBridge behavior above the EISS level would be needed to support this amendment. Such modifications are beyond the scope of this document.

802.1Qav Forwarding and Queuing Enhancements for Time-Sensitive Streams - Modification to 802.1Q to support the 802.1 Timing and Synchronization protocol. This amendment specifies methods to support the resource reservations made through the 802.1Qat protocol (see above).

802.1Qaz Enhanced Transmission Selection - It appears that this amendment will be below the EISS layer and can be supported in RBridge ports without modification to the TRILL protocol.

802.1Qbb Priority-based Flow Control - Commonly called "per-priority pause", it appears that this amendment will be below the EISS layer and can be supported in RBridge ports without modification to the TRILL protocol.

802.1bc Remote Customer Service Interfaces. This is an extension to 802.1Q provider bridging. See 802.1ad-2005 above.

802.1Qbe Multiple Backbone Service Instance Identifier (I-SID) Registration Protocol (MIRP). This is an extension to 802.1Q provider backbone bridging. See 802.1ah-2008 above.

802.1Qbf Provider Backbone Bridge Traffic Engineering (PBB-TE) Infrastructure Segment Protection. This amendment extends 802.1Q to support certain types of failover between provider backbone bridges. See 802.1ah-2008 above.

Appendix F. Acknowledgements

Many people have contributed to this design, including the following, in alphabetic order:

Bernard Aboba, Alia Atlas, Ayan Banerjee, Caitlin Bestler, Suresh Boddapati, David Michael Bond, Stewart Bryant, Ross Callon, James Carlson, Pasi Eronen, Dino Farinacci, Adrian Farrell, Don Fedyk, Bill Fenner, Eric Gray, Sujay Gupta, Joel Halpern, Andrew Lange, Acee Lindem, Vishwas Manral, Peter McCann, Israel Meilik, David Melman, Nandakumar Natarajan, Erik Nordmark, Jeff Pickering, Tim Polk, Dan Romascanu, Sanjay Sane, Pekka Savola, Matthew R. Thomas, Joe Touch, Mark Townsley, Kate Zebrose.

Authors' Addresses

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

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

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

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

Dinesh G. Dutt
Cisco Systems
170 Tasman Drive
San Jose, CA 95134-1706 USA

Phone: +1-408-527-0955
EMail: ddutt@cisco.com

Silvano Gai
Cisco Systems
170 Tasman Drive
San Jose, CA 95134-1706 USA

EMail: silvano@ip6.com

Anoop Ghanwani
Brocade
130 Holger Way
San Jose, CA 95134 USA

Phone: +1-408-333-7149
EMail: anoop@alumni.duke.edu