

TRILL Working Group
INTERNET-DRAFT
Intended status: Proposed Standard
Updates: RFCtrill

Donald Eastlake 3rd
Huawei
Radia Perlman
Intel Labs
Anoop Ghanwani
Brocade
Dinesh G. Dutt
Cisco Systems
Vishwas Manral
IP Infusion
April 29, 2011

Expires: October 28, 2011

RBridges: Adjacency
<[draft-ietf-trill-adj-07.txt](#)>

Abstract

The IETF TRILL (TRansparent Interconnection of Lots of Links) protocol provides optimal pair-wise data forwarding without configuration, safe forwarding even during periods of temporary loops, and support for multipathing of both unicast and multicast traffic. TRILL accomplishes this by using IS-IS (Intermediate System to Intermediate System) link state routing and by encapsulating traffic using a header that includes a hop count. Devices that implement TRILL are called RBridges.

TRILL supports multi-access LAN (Local Area Network) links that can have multiple end stations and RBridges attached. This document describes four aspects of the TRILL LAN Hello protocol used on such links, particularly adjacency, designated RBridge selection, and MTU (Maximum Transmission Unit) and pseudonode procedures, with state machines. There is no change for IS-IS point-to-point Hellos used on links configured as point-to-point in TRILL.

Status of This Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of [BCP 78](#) and [BCP 79](#). Distribution of this document is unlimited. Comments should be sent to the TRILL working group mailing list <rbridge@postel.org>.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF), its areas, and its working groups. Note that other groups may also distribute working documents as Internet-Drafts.

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

material or to cite them other than as "work in progress."

D. Eastlake, et al

[Page 1]

The list of current Internet-Drafts can be accessed at
<http://www.ietf.org/1id-abstracts.html>

The list of Internet-Draft Shadow Directories can be accessed at
<http://www.ietf.org/shadow.html>

Acknowledgements

The authors of [RFCtrill], those listed in the Acknowledgements section of [RFCtrill] and the contributions of Jari Arkko, Ayan Banerjee, Les Ginsberg, Sujay Gupta, David Harrington, Erik Nordmark, Pete McCann, and Mike Shand, to this document, are hereby acknowledged.

Table of Contents

1. Introduction.....	4
1.1 Content and Precedence.....	4
1.2 Terminology and Acronyms.....	5
2. The TRILL Hello Environment and Purposes.....	6
2.1 Incrementally Replacing 802.1Q-2005 Bridges.....	6
2.2 Handling Native Frames.....	7
2.3 Zero or Minimal Configuration.....	8
2.4 MTU Robustness.....	8
2.5 Purposes of the TRILL Hello Protocol.....	8
3. Adjacency State Machinery.....	10
3.1 TRILL LAN Hellos, MTU Test, and VLANs.....	10
3.2 Adjacency Table Entries and States.....	10
3.3 Adjacency and Hello Events.....	11
3.4 Adjacency State Diagram and Table.....	13
3.5 Multiple Parallel Links.....	15
3.6 Insufficient Space in Adjacency Table.....	16
4. RBridge LAN Ports and DRB State.....	17
4.1 Port Table Entries and DRB Election State.....	17
4.2 DRB Election Events.....	18
4.2.1 DRB Election Details.....	18
4.2.2 Change in DRB.....	19
4.2.3 Change in Designated VLAN.....	19
4.3 State Table and Diagram.....	20
5. MTU Matching.....	22
6. Pseudonodes.....	23
7. TRILL Hello Reception and Transmission.....	24
7.1 Transmitting TRILL Hellos.....	24
7.1 Receiving TRILL Hellos.....	25
8. Multiple Ports on the Same Link.....	27
9. Security Considerations.....	27
10. IANA Considerations.....	27
11. References.....	28
11.1 Normative References.....	28
11.2 Informative References.....	28
Appendix: Change Record.....	30

1. Introduction

The IETF TRILL (TRansparent Interconnection of Lots of Links) protocol [[RFCtrill](#)] provides optimal pair-wise data frame forwarding without configuration, safe forwarding even during periods of temporary loops, and support for multipathing of both unicast and multicast traffic. TRILL accomplishes this by using [[IS-IS](#)] (Intermediate System to Intermediate System) link state routing and encapsulating traffic using a header that includes a hop count. The design supports VLANs (Virtual Local Area Networks) and optimization of the distribution of multi-destination frames based on VLANs and IP derived multicast groups. Devices that implement TRILL are called RBridges.

The purpose of this document is to improve the quality of the description of four aspects of the TRILL LAN (Local Area Network) Hello protocol that RBridges use on broadcast (LAN) links. It includes reference implementation details. Alternative implementations that interoperate on the wire are permitted. There is no change for IS-IS point-to-point Hellos used on links configured as point-to-point in TRILL.

The scope of this document is limited to the following aspects of the TRILL LAN Hello protocol:

- Adjacency formation
- DRB (Designated RBridge aka DIS (Designated Intermediate System)) election
- Rules for two-way and MTU (Maximum Transmission Unit) matching for advertisements
- Creation and use of pseudo-nodes

For other aspects of the TRILL base protocol see [[RFCtrill](#)].

1.1 Content and Precedence

[Section 2](#) below explains the rationale for the differences between the TRILL LAN Hello protocol and the Layer 3 IS-IS LAN Hello protocol [[IS-IS](#)] [[RFC1195](#)] in light of the environment for which the TRILL protocol is designed. It also describes the purposes of the TRILL LAN Hello protocol.

[Section 3](#) describes the adjacency state machine and its states and relevant events.

[Section 4](#) describes the Designated RBridge (DRB) election state machine for RBridge ports and its states and relevant events.

[Section 5](#) describes MTU testing and matching on a TRILL link.

[Section 6](#) discusses pseudonode creation and use.

[Section 7](#) provides more details on the reception and transmission of TRILL LAN Hellos.

[Section 8](#) discusses multiple ports from one RBridge on the same link.

In case of conflict between this document and [\[RFCtrill\]](#), this document prevails.

[1.2](#) Terminology and Acronyms

This document uses the acronyms defined in [\[RFCtrill\]](#) supplemented by the following additional acronym:

SNPA - Sub-Network Point of Attachment

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 [\[RFC2119\]](#).

2. The TRILL Hello Environment and Purposes

[IS-IS] has subnetwork independent functions and subnetwork dependent functions. Currently Layer 3 use of IS-IS supports two types of subnetworks: (1) point-to-point link subnetworks between routers and (2) general broadcast (LAN) subnetworks. Because of the differences between the environment of Layer 3 routers and the environment of TRILL RBridges, instead of the broadcast (LAN) subnetwork dependent functions encountered at Layer 3, which are specified in [[IS-IS](#)] [Section 8.4](#), the TRILL protocol uses modified subnetwork dependent functions for a LAN subnetwork. The environmental differences are described in [Sections 2.1](#) through [2.4](#) below followed by a summation, in [Section 2.5](#), of the purposes of the TRILL LAN Hello protocol.

2.1 Incrementally Replacing 802.1Q-2005 Bridges

RBridges can incrementally replace IEEE [802.1Q-2005] bridges. Thus RBridges need to provide similar services, including delivery of frames only to links in the frame's VLAN and priority queuing of frames, to the extent that multiple queues are implemented at any particular RBridge port.

RBridge ports are IEEE [802.1Q-2005] ports in terms of their frame VLAN and priority configuration and processing as described in Section 2.6 of [[RFCtrill](#)]. When a frame is received through an RBridge port, like a frame received through any [802.1Q-2005] port, it has an associated VLAN ID and frame priority. When a frame is presented to an [802.1Q-2005] port for queuing and transmission, it must be accompanied by a VLAN ID and frame priority, although whether the frame, if actually transmitted, will be VLAN tagged is determined by whether the port is configured to "strip VLAN tags" or not. Furthermore, in the general case, a broadcast (LAN) link between RBridges can be a VLAN-capable bridged LAN that may be configured to partition VLANs.

Because devices that restrict VLAN connectivity, such as bridged LANs or provider bridging equipment, can be part of the link between RBridges, TRILL Data and TRILL IS-IS frames between RBridges use the link's Designated VLAN. The Designated VLAN is dictated for a link by the elected Designated RBridge (equivalent to the Designated Intermediate System at Layer 3). Because TRILL Data frames flow between RBridges on a link only in the link's Designated VLAN, adjacency for routing calculations is based only on connectivity characteristics in that VLAN.

2.2 Handling Native Frames

Ordinary layer 3 data packets are already "tamed" when they are originated by an end station: they include a hop count and layer 3 source and destination address fields. Furthermore, for ordinary data packets, there is no requirement to preserve their outer layer 2 addressing and, at least if unicast, they are addressed to their first hop router. In contrast, RBridges running TRILL must accept, transport, and deliver untamed "native" frames (as defined in [Section 1.4](#) of [\[RFCtrill\]](#)). Native frames lack a TRILL hop count field. Native frames also have layer 2 addresses that indicate their source and are used as the basis for their forwarding. These layer 2 address must be preserved for delivery to the native frame's layer 2 destination. One resulting difference is that RBridge ports providing native frame service must receive in promiscuous MAC address mode while Layer 3 router ports typically receive in a regularly selective MAC address mode.

TRILL handles this by having, on the link where an end station originated a native frame, one RBridge "ingress" that native frame by adding a TRILL Header that includes a hop count, thus converting it to a TRILL Data frame. This augmented frame is then routed to one RBridge on the link having the destination end-station for the frame (or one RBridge on each such link if it is a multi-destination frame). Such final RBridges perform an "egress" function, removing the TRILL Header and delivering the original frame to its destination(s). (For the purposes of TRILL, a Layer 3 router is an end station.)

Care must be taken to avoid a loop that would involve egressing a native frame and then re-ingressing it because, while it is in native form, it would not be protected by a hop count. Such a loop could involve multiplication of the number of frames each time around and would likely saturate all links involved within milliseconds. For TRILL, safety against such loops for a link is more important than data connectivity on that link.

The primary TRILL defense mechanism against such loops, which is mandatory, is to assure that, as far as practically possible, there is only a single RBridge on each link that is in charge of ingressing and egressing native frames from and to that link. This is the Designated RBridge which is elected using TRILL LAN Hellos as further described in [Sections 2.5](#) and [4](#) below.

Because bridged LANs between RBridges can be configured in complex ways, including so as to pass frames in some VLANs in one direction only, and loop safety is so important, there are additional TRILL defenses against loops, where the looping traffic is in native format

for part of the loop, that are beyond the scope of this document.
These additional defenses have no effect on adjacency states or the

receipt or forwarding of TRILL Data frames, they only affect native frame ingress and egress.

[2.3](#) Zero or Minimal Configuration

RBridges are expected to provide service with zero configuration, except for services such as non-default VLAN or priority that require configuration when offered by [802.1Q-2005] bridges. This differs from Layer 3 routing where routers typically need to be configured as to the subnetworks connected to each port, etc., to provide service.

[2.4](#) MTU Robustness

TRILL IS-IS needs to be robust against links with reasonably restricted MTUs, including links that accommodate only classic Ethernet frames, despite the addition of reasonable headers such as VLAN tags. This is particularly true for TRILL LAN Hellos so as to assure that a unique DRB is elected.

TRILL will also be used inside data centers where it is not uncommon for all or most of the links and switches to support frames substantially larger than the classic Ethernet maximum. For example they may have an MTU adequate to comfortably handle Fiber Channel over Ethernet frames, for which T11 recommends a 2,500 byte MTU [[FCoE](#)]. It would be beneficial for an RBridge campus with such a larger MTU to be able to safely make use of it.

These needs are met by limiting the size of TRILL LAN Hellos and by the use of MTU testing as described below.

[2.5](#) Purposes of the TRILL Hello Protocol

There are three purposes for the TRILL Hello protocol as listed below along with a reference to the Section of this document in which each is discussed:

- a) To determine which RBridge neighbors have acceptable connectivity to be reported as part of the topology ([Section 3](#))
- b) To elect a unique Designated RBridge on the link ([Section 4](#))
- c) To determine the MTU with which it is possible to communicate with each RBridge neighbor ([Section 5](#))

In Layer 3 IS-IS all three of these functions are combined. Hellos may be padded to the maximum length (see [\[RFC3719\]](#), [Section 6](#)) so that a router neighbor is not even discovered if it is impossible to communicate with it using maximum sized packets. Also, even if Hellos from a neighbor R2 are received by R1, if connectivity to R2 is not 2-way (i.e., R2 does not list R1 in R2's Hello), then R1 does not consider R2 as a Designated Router candidate. Because of this logic, it is possible at Layer 3 for multiple Designated Routers to be elected on a LAN, with each representing the LAN as a pseudonode. It appears to the topology as if the LAN is now two or more separate LANs. Although this is surprising, it does not disrupt Layer 3 IS-IS.

In contrast, this behavior is not acceptable for TRILL, since in TRILL it is important that all RBridges on the link know about each other, and choose a single RBridge to be the DRB and to control the native frame ingress and egress on that link. Otherwise, multiple RBridges might encapsulate/decapsulate the same native frame, forming loops that are not protected by the hop count in the TRILL header as discussed above.

So, the TRILL Hello protocol is best understood by focusing on each of these functions separately.

One other issue with TRILL LAN 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. Thus the TRILL Hellos uses a different Neighbor TLV [\[RFCtisis\]](#) that lists neighbors seen for a range of MAC (SNPA) addresses.

3. Adjacency State Machinery

Each RBridge port has associated with it a port state, as discussed in [Section 4](#), and a table of zero or more adjacencies as discussed in this [Section 3](#). The states such adjacencies can have, the events that cause state changes, the actions associated with those state changes, and a state table and diagram are given below.

3.1 TRILL LAN Hellos, MTU Test, and VLANs

The determination of LSP reported adjacencies on links that are not configured as point-to-point is made using TRILL LAN Hellos (see also [Section 7](#)) and an optional MTU test. Appropriate TRILL LAN Hello exchange and the satisfaction of the MTU test, if the MTU test is enabled (see [Section 5](#)), is required for there to be an adjacency that will be reported in an LSP of the RBridge in question.

Because bridges acting as glue on the LAN might be configured in such a way that some VLANs are partitioned, it is necessary for RBridges to transmit Hellos with multiple VLAN tags. The conceptually simplest solution may have been to have all RBridges transmit up to 4,094 times as many Hellos, one with each legal VLAN ID enabled at each port, but this would obviously have deleterious performance implications. So, the TRILL protocol specifies that if RB1 knows it is not DRB, it transmits its Hellos on only a limited set of VLANs, and only an RBridge that believes itself to be DRB on a port "sprays" its TRILL Hellos on all of its enabled VLANs at a port (with the ability to configure to send on only a subset of those). The details are given in [[RFCtrill](#)] [Section 4.4.3](#).

If the MAC (SNPA) address of more than one RBridge port on a link are the same, all but one of such ports are put in the Suspended state (see [Section 4](#)) and do not participate in the link except to monitor whether they should stay suspended.

All TRILL LAN Hellos issued by an RBridge on a particular port MUST have the same source MAC address, priority, desired Designated VLAN, and Port ID regardless of the VLAN in which the Hello is sent. Of course, the priority and desired Designated VLAN can change on occasion, but then the new value must similarly be used in all TRILL Hellos on the port, regardless of VLAN.

3.2 Adjacency Table Entries and States

Each adjacency is in one of the following four states:

Down:

This is a virtual state for convenience in creating state diagrams and tables. It indicates that the adjacency is non-existent and there is no entry in the adjacency table for it.

Detect:

An adjacent neighbor has been detected either (1) not on the Designated VLAN or (2) on the Designated VLAN but neither 2-way connectivity nor the MTU of such connectivity has been confirmed.

2-Way:

2-way connectivity to the neighbor has been found on the Designated VLAN but MTU testing is enabled and has not yet confirmed that the connectivity meets the campus minimum MTU requirement.

Report:

There is 2-way connectivity to the neighbor on the Designated VLAN and either MTU testing has confirmed that the connectivity meets the campus minimum MTU requirement or MTU testing is not enabled. This connectivity will be reported in an LSP (with appropriate provision for the link pseudonode, if any, as described in [Section 6](#)).

For an adjacency in any of the three non-down states (Detect, 2-Way, and Report), there will be an adjacency table entry. That entry will give the state of the adjacency and will also include the information listed below.

- o The address of the neighbor (that is, its SNPA address, usually a 48-bit MAC address), the Port ID and the System ID in the received Hellos. Together, these three quantities uniquely identify the adjacency.
- o Exactly two Hello holding timers, each consisting of a 16-bit unsigned integer number of seconds: a Designated VLAN holding timer and a non-Designated VLAN holding timer.
- o The 7-bit unsigned priority of the neighbor to be DRB.
- o The VLAN that the neighbor RBridge wants to be the Designated VLAN on the link, called the desired Designated VLAN.

[3.3](#) Adjacency and Hello Events

The following events can change the state of an adjacency:

- A0. Receive a TRILL Hello whose source MAC address (SNPA) is equal to that of the port on which it is received. This is a special event that is handled as described immediately after this list of events. It does not appear in the state transition table or diagram.
- A1. Receive a TRILL Hello (other than an A0 event) on the Designated VLAN with a TRILL Neighbor TLV that explicitly lists the receiver's (SNPA) address
- A2. Receive a TRILL Hello (other than an A0 event) that either (1) is not on the Designated VLAN (any TRILL Neighbor TLV in such a Hello is ignored) or (2) is on the Designated VLAN but does not contain a TRILL Neighbor TLV covering an address range including the receiver's (SNPA) address
- A3. Receive a TRILL Hello (other than an A0 event) on the Designated VLAN with one or more TRILL Neighbor TLVs covering an address range including the receiver's (SNPA) address none of which lists the receiver
- A4. The expiration of one or both Hello holding timers results in them both being expired
- A5. The Designated VLAN Hello holding timer expires but the non-Designated VLAN Hello holding timer still has time left until it expires
- A6. MTU test successful
- A7. MTU test was successful but now fails
- A8. The RBridge port goes operationally down

For the special A0 event, the Hello is examined to determine if it is higher priority to be DRB than the port on which is received as described in [Section 4.2.1](#). If the Hello is of lower priority than the receiving port, it is discarded with no further action. If it is higher priority than the receiving port, then any adjacencies for that port are discarded (transitioned to the Down state) and the port is suspended as described in [Section 4.2](#).

The receipt of a TRILL LAN Hello with a source MAC (SNPA) address different from that of the receiving port, that is, the occurrence of events A1, A2, or A3, causes the following actions (except where the Hello would create a new adjacency table entry, the table is full, and the Hello is too low a priority to displace an existing entry as described in [Section 3.6](#)). The Designated VLAN used in these actions is the Designated VLAN dictated by the DRB determined without taking

the received TRILL LAN Hello into account (see [Section 4](#)).

- o If the receipt of the Hellos creates a new adjacency table entry, the neighbor RBridge MAC (SNPA) address, Port ID, and System ID are set from the Hello.
- o The appropriate Hello holding timer for the adjacency, depending on whether the Hello was received on the Designated VLAN or not, is set to the Holding Time field of the Hello. If the receipt of the Hello is creating a new adjacency table entry, the other timer is set to expired.
- o The priority of the neighbor RBridge to be DRB is set to the priority field of the Hello.
- o The VLAN that the neighbor RBridge wants to be the Designated VLAN on the link is set from the Hello.
- o If the creation of a new adjacency table entry or the priority update above changes the results of the DRB election on the link, the appropriate RBridge port event (D2 or D3) occurs, after the above actions, as described in [Section 4.2](#).
- o If there is no change in DRB but the neighbor Hello is from the DRB and has a changed Designated VLAN from the previous Hello received from the DRB, the result is a change in Designated VLAN for the link as specified in [Section 4.2.3](#).

An event A4 resulting in both Hello Holding timers for an adjacency being expired and the adjacency going Down may also result in an event D3 as described in [Section 4.2](#).

Concerning events A6 and A7, if MTU testing is not enabled, A6 is considered to occur immediately upon the adjacency entering the 2-Way state and A7 cannot occur.

See further TRILL LAN Hello receipt details in [Section 7](#).

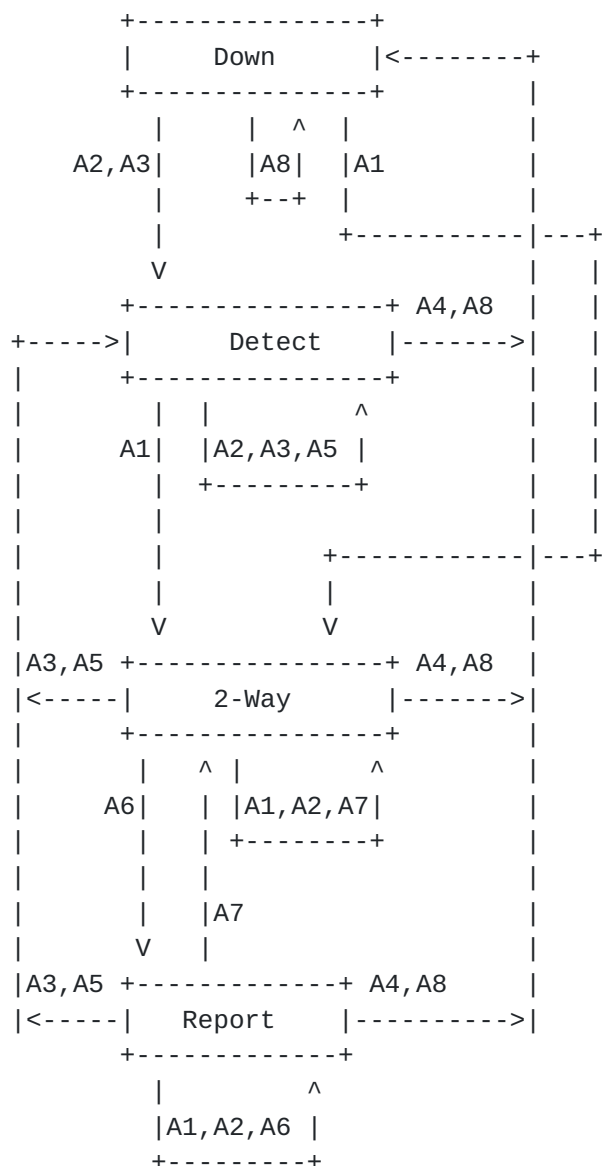
[3.4](#) Adjacency State Diagram and Table

The table below shows the transitions between the states defined above based on the events defined above:

Event	Down	Detect	2-Way	Report	
+-----+	+-----+	+-----+	+-----+	+-----+	+
A1	2-Way	2-Way	2-Way	Report	
A2	Detect	Detect	2-Way	Report	
A3	Detect	Detect	Detect	Detect	
A4	N/A	Down	Down	Down	
A5	N/A	Detect	Detect	Detect	
A6	N/A	N/A	Report	Report	
A7	N/A	N/A	2-Way	2-Way	
A8	Down	Down	Down	Down	

N/A indicates that the event to the left is Not Applicable in the state at the top of the column. These events affect only a single adjacency. The special A0 event transitions all adjacencies to Down, as explained immediately after the list of adjacency events above.

Below is the same information as that in the state table presented as a diagram:



3.5 Multiple Parallel Links

There can be multiple parallel adjacencies between neighbor RBridges that are visible to TRILL. (Multiple low level links that have been bonded together by technologies such as link aggregation [802.1AX] appear to TRILL as a single link over which only a single TRILL adjacency could be established.)

Any such links that have pseudonodes (see Section 6) are distinguished in the topology and such adjacencies, if they are in the Report state, appear in LSPs as per Section 6. However, there can be multiple parallel adjacencies without pseudonodes because they are

point-to-point adjacencies or LAN adjacencies for which a pseudonode is not being created. Such parallel, non-pseudonode adjacencies in

the Report state appear in LSPs as a single adjacency. The cost of such an adjacency MAY be adjusted downwards to account for the parallel paths. Multipathing across such parallel connections can be freely done for unicast TRILL Data traffic on a per flow basis but is restricted for multi-destination traffic, as described in [[RFCtrill](#)] [Section 4.5.2](#), point 3, and [Appendix C](#).

[3.6](#) Insufficient Space in Adjacency Table

If the receipt of a TRILL LAN Hello would create a new adjacency table entry, that is, would transition an adjacency out of the Down state, there may be no space for the new entry. In that case, the DRB election priority (see [Section 4.2.1](#)) of the new entry that would be created is compared with that priority for the existing entries. If the new entry is higher priority than the lowest priority existing entry, it replaces the lowest priority existing entry, which is transitioned to the Down state.

4. RBridge LAN Ports and DRB State

The information at an RBridge associated with each of its LAN ports includes the following:

- o Enablement bit, which defaults to enabled.
- o SNPA address (usually a 48-bit MAC address) of the port.
- o Port ID, used in TRILL Hellos sent on the port.
- o The Holding Time, used in TRILL Hellos sent on the port.
- o The Priority to be DRB, used in TRILL Hellos sent on the port.
- o The DRB status of the port, determined as specified below.
- o A 16-bit unsigned Suspension timer denominated in seconds
- o The desired Designated VLAN. The VLAN this RBridge wants to be the Designated VLAN for the link out this port, used in TRILL Hellos sent on the port.
- o A table of zero or more adjacencies (see [Section 3](#)).

4.1 Port Table Entries and DRB Election State

The TRILL equivalent of the DIS (Designated Intermediate System) on a link is the DRB or Designated RBridge. The DRB election state machinery is described below.

Each RBridge port is in one of the following four DRB states:

Down:

The port is operationally down. It might be administratively disabled or down at the link layer. In this state there will be no adjacency table entries for the port, and no TRILL Hellos or other IS-IS PDUs or TRILL Data frames are accepted or transmitted.

Suspended:

Operation of the port is suspended because there is a higher priority port on the link with the same MAC (SNPA) address. This is the same as the down state with the exception that TRILL Hellos are accepted for the sole purpose of determining whether to change the value of the Suspension timer for the port as described below.

DRB:

The port is DRB and can receive and transmit TRILL Data frames.

Not DRB:

The port is deferring to another port on the link, which it believes is DRB, but can still receive and transmit TRILL Data frames.

4.2 DRB Election Events

The following events can change the DRB state of a port:

- D1. Expiration of the suspension timer while the port is in the Suspended state or the enablement of the port.
- D2. Adjacency table for the port changes and there are now entries for one or more other RBridge ports on the link that appear to be higher priority to be DRB than the local port.
- D3. The port is not Down or Suspended and the adjacency table for the port changes so there are now no entries for other RBridge ports on the link that appear to be higher priority to be DRB than the local port.
- D4. Receipt of a TRILL Hello with the same MAC address (SNPA) as the receiving port and higher priority to be DRB as described for event A0.
- D5. The port becomes operationally down.

Event D1 is considered to occur on RBridge boot if the port is administratively and link layer enabled.

Event D4 causes the port to enter the Suspended state and all adjacencies for the port to be discarded (transitioned to the Down state). If it was in some state other than Suspended, the suspension timer is set to the Holding Time in the Hello causing event D4. If it was in the Suspended state, the suspension timer is set to the maximum of its current value and the Holding Time in the Hello causing event D4.

4.2.1 DRB Election Details

Events D2 and D3 constitute losing and winning the DRB election at the port, respectively.

The candidates for election are the local RBridge and all RBridges with which there is an adjacency on the port in an adjacency state other than Down state. The winner is the RBridge with highest priority to be DRB, as determined from the 7-bit priority field in that RBridge's Hellos received and the local port's priority to be DRB field, with MAC (SNPA) address as a tie breaker, Port ID as a secondary tie breaker, and System ID as a tertiary tie breaker. These fields are compared as unsigned integers with the larger magnitude being considered higher priority.

Resort to the secondary and tertiary tie breakers should only be necessary in rare circumstances when multiple ports have the same priority and MAC (SNPA) address and some of them are not yet suspended. For example, RB1, that has low priority to be DRB on the link, could receive Hellos from two other ports on the link that have the same MAC address as each other and are higher priority to be DRB. One of these two same-MAC ports will be suspended, cease sending Hellos, and the Hello from it received by RB1 will eventually time out. But in the meantime RB1 can use the tie breakers to determine which port is DRB and thus which port's Hello to believe for such purposes as setting the Designated VLAN on the link.

4.2.2 Change in DRB

Events D2 and D3 result from a change in the apparent DRB on the link. Unnecessary DRB changes should be avoided, especially on links offering native frame service, as a DRB change will generally cause a transient interruption to native frame service.

If a change in the DRB on the link changes the Designated VLAN on the link, the actions specified in [Section 4.2.3](#) are taken.

If an RBridge changes in either direction between being the Designated RBridge and not being the Designated RBridge at a port, this will generally change the VLANs on which Hellos are sent by that RBridge on that port as specified in Section 4.4.3 of [[RFCtrill](#)].

4.2.3 Change in Designated VLAN

Unnecessary changes in the Designated VLAN on a link should be avoided because a change in the Designated VLAN can cause a transient interruption to TRILL Data forwarding on the link. When practical, all RBridge ports on a link should be configured with the same desired Designated VLAN so that, in case the winner of the DRB election changes, for any reason, the Designated VLAN will remain the

same.

D. Eastlake, et al

[Page 19]

If an RBridge detects a change in Designated VLAN on a link, then, for all adjacency table entries for a port to that link, the RBridge takes the following steps in the order given:

- o The non-Designated VLAN Hello Holding timer is set to the maximum of its time to expiration and the current time to expiration of the Designated VLAN Hello Holding timer.
- o The Designated VLAN Hello Holding timer is then set to expired (if necessary) and an event A5 occurs for the adjacency (see [Section 3.3](#)).

If the Designated VLAN for a link changes, this will generally change the VLANs on which Hellos are sent by an RBridge port on that link as specified in Section 4.4.3 of [[RFCtrill](#)].

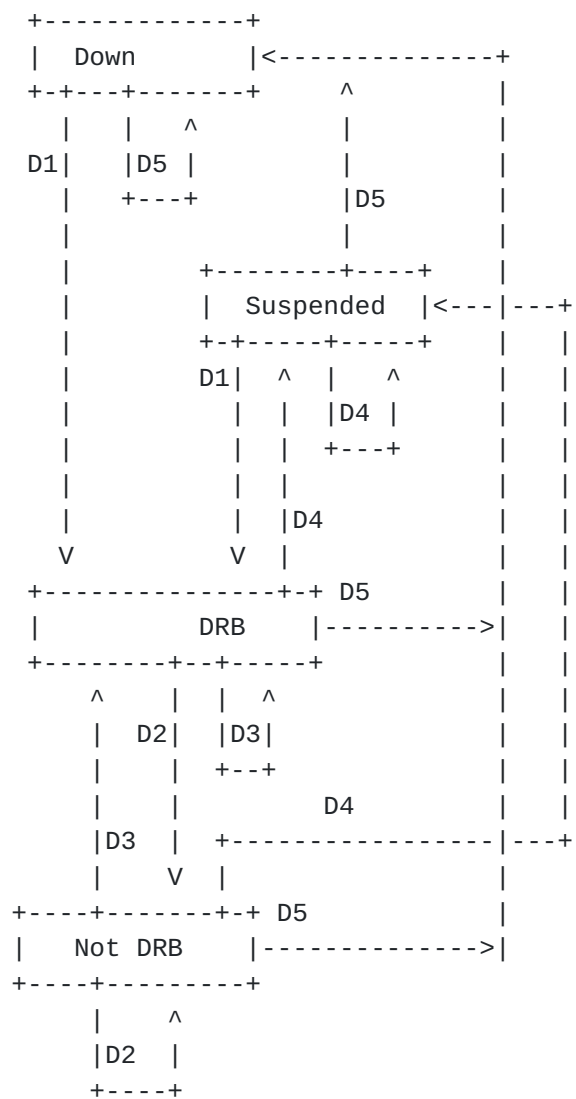
4.3 State Table and Diagram

The table below shows the transitions between the DRB states defined above based on the events defined above:

Event	Down	Suspend	DRB	Not DRB
+-----+	+-----+	+-----+	+-----+	+-----+
D1	DRB	DRB	N/A	N/A
D2	N/A	N/A	Not DRB	Not DRB
D3	N/A	N/A	DRB	DRB
D4	N/A	Suspend	Suspend	Suspend
D5	Down	Down	Down	Down

N/A indicates that the event to the left is Not Applicable in the state at the top of the column.

Below is the same information as in the state table presented as a diagram:



5. MTU Matching

The purpose of MTU testing is to ensure that the links used in the campus topology can pass TRILL IS-IS and Data frames at the RBridge campus MTU.

An RBridge, RB1, determines the desired campus link MTU by calculating the minimum of its `originatingL1LSPBufferSize` and the `originatingL1LSPBufferSize` of other RBridges in the campus, as advertised in the link state database, but not less than 1,470 bytes. Although `originatingL1LSPBufferSize` in Layer 3 [[IS-IS](#)] is limited to the range 512 to 1,492 bytes inclusive, in TRILL it is limited to the range 1,470 to 65,535 bytes inclusive.

Although MTU testing is optional, it is mandatory for an RBridge to respond to an MTU-probe PDU with an MTU-ack PDU [[RFCtrill](#)] [[RFCtisis](#)]. Use of multicast or unicast for MTU-probe and MTU-ack is an implementation choice. However, the burden on the link is generally minimized by multicasting MTU-probes when a response from all other RBridges on the link is desired, such as when initializing or re-confirming MTU, unicasting MTU-probes when a response from a single RBridge is desired, such as one that has just been detected on the link, and unicasting all MTU-ack frames.

RB1 can test the MTU size to RB2 as described in Section 4.3.2 of [[RFCtrill](#)]. For this purpose, MTU testing is only done in the Designated VLAN. An adjacency that fails the MTU test at the campus MTU will not enter the "Report" state or, if the adjacency is in that state, it leaves that state. Thus an adjacency failing the MTU test will not be reported by the RBridge performing the test. Since inclusion in least cost route computation requires the adjacency to be reported by both ends, as long as the MTU failure is noticed by the RBridge at either end of the adjacency, it will not be so used.

If it tests MTU, RB1 lists the largest size for which the MTU test succeeds or a flag indicating that it fails at the campus MTU, with the neighbor in RB1's TRILL Neighbor TLV and MAY report this with the adjacency in an Extended Reachability TLV in RB1's LSP. RB1 MAY choose to test MTU sizes greater than the desired campus MTU as well as the desired campus MTU.

Most types of TRILL IS-IS frames, such as LSPs, can make use of the campus MTU. The exceptions are TRILL Hellos, which must be kept small for loop safety, and the MTU PDUs, whose size must be adjusted appropriately for the tests being performed.

6. Pseudonodes

The Designated RBridge (DRB), determined as described above, controls whether a pseudonode will be used on a link.

If the DRB sets the bypass pseudonode bit in its TRILL LAN Hellos, the RBridges on the link (including the DRB) just directly report all their adjacencies on the LAN that are in the Report state. If the DRB does not set the bypass pseudonode bit in its TRILL Hellos, then (1) the DRB reports in its LSP its adjacency to the pseudonode, (2) the DRB sends LSPs on behalf of the pseudonode in which it reports adjacency to all other RBridges on the link where it sees that adjacency in the Report state, and (3) all other RBridges on the link report their adjacency to the pseudonode if they see their adjacency to the DRB as being in the Report state and do not report any other adjacencies on the link. Setting the bypass pseudonode bit has no effect on how LSPs are flooded on a link. It only affects what LSPs are generated.

It is anticipated that many links between RBridges will actually be point-to-point, in which case using a pseudonode merely adds to the complexity. For example, if RB1 and RB2 are the only RBridges on the link and RB1 is 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 in its Hellos if it has not seen at least two simultaneous adjacencies in the Report state since it last re-booted or was reset by network management.

7. TRILL Hello Reception and Transmission

This section provides further details on the receipt and transmission of TRILL LAN Hellos.

TRILL LAN Hellos, like all TRILL IS-IS frames, are primarily distinguished from Layer 3 IS-IS frames by being sent to the All-IS-IS-RBridges multicast address (01-80-C2-00-00-41). TRILL IS-IS frames also have the L2-IS-IS Ethertype (0x22F4) and are Ethertype encoded.

Although future extensions to TRILL may include use of Level 2 IS-IS, [RFCtrill] specifies TRILL using a single Level 1 Area with Area Address zero (see Section 4.2 of [RFCtisis]).

IS-IS Layer 3 routers are frequently connected to other Layer 3 routers that are part of a different routing domain. In that case, the externalDomain flag (see [IS-IS]) is normally set for the port through which such a connection is made. The setting of this flag to "true" causes no IS-IS PDUs to be sent out the port and any IS-IS PDUs received to be discarded, including Hellos. RBridges operate in a different environment where all neighbor RBridges merge into a single campus. For loop safety, RBridges do not implement the externalDomain flag or implement it with the fixed value "false". They send and receive TRILL LAN Hellos on every port that is not disabled or configured as point-to-point.

7.1 Transmitting TRILL Hellos

TRILL LAN Hellos are sent with the same timing as Layer 3 IS-IS LAN Hellos [IS-IS]; however, no Hellos are sent if a port is in the Suspended or Down states.

TRILL Hello PDUs SHOULD NOT be padded and MUST NOT be sent exceeding 1,470 octets; however, a received TRILL Hello longer than 1,470 octets is processed normally.

TRILL Hello PDU headers MUST conform to the following:

- o Maximum Area Addresses equal to 1.
- o Circuit Type equal to 1.

Each TRILL Hello MUST contain an Area Addresses TLV listing only the single Area zero, and an MT Port Capabilities TLV containing a VLAN-FLAGS sub-TLV [RFCtisis]. If a Protocols Supported TLV is present, it MUST list the TRILL NLPID (0xC0).

The TRILL Neighbor TLV sent in a Hello MUST show the neighbor

information, as sensed by the transmitting RBridge, for the VLAN on which the Hello is sent. Since implementations conformant to this document maintain such information on a per VLAN basis only for the Designated VLAN, such implementations only send the TRILL Neighbor TLV in TRILL Hellos on the Designated VLAN.

It is RECOMMENDED that, if there is sufficient room, a TRILL Neighbor TLV or TLVs, as described in Section 4.4.2.1 of [[RFCtrill](#)], covering the entire range of MAC addresses and listing all adjacencies with a non-zero Designated VLAN Hello Holding time, or an empty list of neighbors if there are no such adjacencies, be in TRILL Hellos sent on the Designated VLAN. If this is not possible, then TRILL Neighbor TLV's covering sub-ranges of MAC addresses should be sent so that the entire range is covered reasonably promptly. Delays in sending TRILL Neighbor TLVs will delay the advancement of adjacencies to the Report state and the discovery of some link failures. Rapid, for example sub-second, detection of link or node failures is best addressed with a protocol designed for that purpose, such as BFD [[RFC5880](#)], use of which with TRILL will be specified in a separate document.

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 its Holding Time. In other words, if X1 is the smallest reported in one of RB1's neighbor lists, and the "smallest" flag is not set, then X1 MUST appear in a different neighbor list as well, as the largest ID reported in that fragment. Or lists may overlap, as long as there is no gap, such that some range, say between Xi and Xj, never appears in any list.

A TRILL Hello MAY also contain any TLV permitted in a Layer 3 IS-IS Hello. TLVs that are unsupported/unknown are ignored.

7.1 Receiving TRILL Hellos

Assuming a frame has the All-IS-IS-RBridges multicast address and L2-IS-IS Ethertype, it will be examined to see if it appears to be an IS-IS PDU. If so, and it appears to be a LAN Hello PDU, the following tests are performed.

- o If the Circuit Type field is other than 1, the PDU is discarded.
- o If the PDU does not contain an Area Address TLV or it contains an Area Address TLV that is other than the single Area Address zero, it is discarded.

- o If the Hello includes a Protocols Supported TLV that does not list the TRILL NLPID (0xC0), it is discarded. It is acceptable if there is no Protocols Supported TLV present.
- o If the Hello does not contain an MT Port Capabilities TLV containing a VLAN-FLAGS sub-TLV [[RFCtisis](#)], it is discarded.
- o If the maximumAreaAddresses field of the PDU is not 1, it is discarded.
- o If IS-IS authentication is in use on the link and the PDU either has no Authentication TLV or validation of that Authentication TLV fails, it is discarded.

If none of the rules in the list above has been satisfied, and the frame is parseable, it is assumed to be a well-formed TRILL Hello received on the link. It is treated as an event A0, A1, A2, or A3 based on the criteria listed in [Section 3.3](#).

8. Multiple Ports on the Same Link

It is possible for an RBridge RB1 to have multiple ports on the same link that are not in the Suspended state. It is important for RB1 to recognize which of its ports are on the same link. RB1 can detect this condition based on receiving TRILL LAN Hello messages with the same LAN ID on multiple ports.

The DRB election is port-based (see [Section 4](#)) and only the Hellos from the elected port can perform certain functions such as dictating the Designated VLAN or whether a pseudonode will be used; however, the election also designates the RBridge with that port as DRB for the link. An RBridge may choose to load split some tasks among its ports on the link if it has more than one and it is safe to do so as described in Section 4.4.4 of [[RFCtrill](#)].

9. Security Considerations

This memo provides improved documentation of some aspects of the TRILL base protocol standard, particularly four aspects of the TRILL LAN Hello protocol. It does not change the security considerations of the TRILL base protocol. See Section 6 of [[RFCtrill](#)].

10. IANA Considerations

This document requires no IANA actions. RFC Editor: Please delete this section before publication.

11. References

Normative and Informational references for this document are listed below.

11.1 Normative References

- [IS-IS] - ISO/IEC 10589:2002, Second Edition, "Intermediate System to Intermediate System Intra-Domain Routing Exchange Protocol for use in Conjunction with the Protocol for Providing the Connectionless-mode Network Service (ISO 8473)", 2002.
- [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.
- [RFCtisis] - Eastlake, D., A. Banerjee, D. Dutt, R. Perlman, A. Ghanwani, "TRILL Use of IS-IS", [draft-ietf-isis-trill](#), work in progress.
- [RFCtrill] - Perlman, R., D. Eastlake, D. Dutt, S. Gai, and A. Ghanwani, "RBridges: Base Protocol Specification", [draft-ietf-trill-rbridge-protocol-16.txt](#), in RFC Editor's queue.

11.2 Informative References

- [802.1AX] - "IEEE Standard for Local and metropolitan area networks / Link Aggregation", 802.1AX-2008, 1 January 2008.
- [FCoE] - From [www.t11.org](#) discussion of "FCoE Max Size" generated from T11/09-251v1, 04/27/2009, "FCoE frame or FCoE PDU".
- [RFC3719] - Parker, J., Ed., "Recommendations for Interoperable Networks using Intermediate System to Intermediate System (IS-IS)", February 2004.
- [RFC5880] - Katz, D. and D. Ward, "Bidirectional Forwarding Detection (BFD)", [RFC 5880](#), June 2010.

Authors' Addresses

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

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

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

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

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

Phone: +1-408-333-7149
Email: anoop@brocade.com

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

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

Vishwas Manral
IP Infusion Inc.
1188 E. Arques Ave.
Sunnyvale, CA 94089 USA

Tel: +1-408-400-1900
email: vishwas@ipinfusion.com

Appendix: Change Record

This appendix summarizes changes between versions of this draft.

RFC Editor: Please delete this Appendix before publication.

From -01 to -02

1. Add a prefix letter "A" or "D" to the adjacency and DRB selection event numbers.
2. Some re-wording for clarification in [Section 6](#) concerning reporting pseudonode adjacencies.
3. Minor editorial changes.

From -02 to -03

1. Add coverage of the case where the DRB on a link stays the same but the Designated VLAN changes.
2. Clarify that, in all of the non-Down states, an RBridge port can receive and transmit TRILL Data frames.
3. Split out separate sub-sections on the details of the DRB election, on changes in the DRB, and on changes in the Designated VLAN.
4. Minor editorial changes.

From -03 to -04

1. Clarify exactly what adjacencies are reported in LSPs when bypass pseudonode is or is not set by the DRB.
2. Clarify that TRILL Hellos received in excess of 1470 bytes in length are still processed normally.
3. Items related to duplicate MAC addresses: Add the Suspended port state and the A0 event. Add a tertiary tie breaker for the DRB election to handle perverse cases of multiple ports with the same priority and MAC address on the same or different RBridges. Use the MAC address, Port ID, System ID triplet to identify adjacencies.

4. Add text recommending transmission of a TRILL Neighbor TLV covering the full range of MAC addresses (will a null neighbor list if there are none) in every TRILL Hello if there is room and prompt transmission of TRILL Neighbor TLV fragments covering that whole range, if the full TLV will not fit.
5. Add recommendation concerning use of unicast and multicast for MTU PDUs.
6. Minor editorial changes.

From -04 to -05

1. Eliminate the "Pre-DRB" port state. That state relates to one reason for the inhibition of Appointed Forwarders, is out of scope for this document, and this capability will be provided by other means in a separate document.
2. Clarify that a TRILL Hello need not contain a Protocols Supported TLV. But if one is present it must contain the TRILL NLPID.
3. Minor Editorial changes.

From -05 to -06

1. Swap Sections [7.1](#) and [7.2](#) so transmission appears before reception.
2. Minor changes so as to refer to "SNPA address" rather than just "SNPA".
3. Minor editorial changes and typo fixes.

From -06 to -07

Editorial changes resulting from IESG DISCUSSES and COMMENTS.

Copyright and IPR Provisions

Copyright (c) 2011 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License. The definitive version of an IETF Document is that published by, or under the auspices of, the IETF. Versions of IETF Documents that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of IETF Documents. The definitive version of these Legal Provisions is that published by, or under the auspices of, the IETF. Versions of these Legal Provisions that are published by third parties, including those that are translated into other languages, should not be considered to be definitive versions of these Legal Provisions. For the avoidance of doubt, each Contributor to the IETF Standards Process licenses each Contribution that he or she makes as part of the IETF Standards Process to the IETF Trust pursuant to the provisions of [RFC 5378](#). No language to the contrary, or terms, conditions or rights that differ from or are inconsistent with the rights and licenses granted under [RFC 5378](#), shall have any effect and shall be null and void, whether published or posted by such Contributor, or included with or in such Contribution.

