

TRILL Working Group
Internet-Draft
Intended status: Standards Track
Expires: April 21, 2011

D. Bond
UNH-IOL
V. Manral
IP Infusion Inc.
October 18, 2010

R Bridges: Operations, Administration, and Maintenance (OAM) Support
draft-bond-trill-rbridge-oam-00

Abstract

The IETF has standardized R Bridges, devices that implement the TRILL protocol, a solution for transparent shortest-path frame routing in multi-hop networks with arbitrary topologies, using a link-state routing protocol technology and encapsulation with a hop-count. As R Bridges are deployed in real-world situations, operators will need tools for debugging problems that arise. This document specifies a set of R Bridge features for operations, administration, and maintenance purposes in R Bridge campuses. The features specified in this document include tools for traceroute, ping, and error reporting.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <http://datatracker.ietf.org/drafts/current/>.

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."

This Internet-Draft will expire on April 21, 2011.

Copyright Notice

Copyright (c) 2010 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<http://trustee.ietf.org/license-info>) in effect on the date of

publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	4
1.1. Requirements Language	4
2. Acronyms	5
3. TRILL OAM Option	6
4. RBridge Tools	9
4.1. Application Sub-Options RBridge Tools	10
4.1.1. RBridge Traceroute	10
4.1.1.1. Route Respond Traceroute	10
4.1.1.1.1. Multi-Destination Targets	12
4.1.1.1.2. Route Respond Traceroute Example	12
4.1.1.2. Hop Count Traceroute	14
4.1.1.2.1. Multi-Destination Targets	15
4.1.1.2.2. Hop Count Traceroute Example	16
4.1.2. RBridge Ping	17
4.1.2.1. Ping Example	19
4.2. Error Sub-Options RBridge Tools	20
4.2.1. Hop Count Zero Error	21
4.2.2. MTU Error	21
4.2.3. Generic Error	22
5. TRILL OAM Option Format	22
5.1. Code Values	23
5.2. Codes	23
5.2.1. Application Codes	23
5.2.1.1. Echo Request	24
5.2.1.2. Route Respond Request	25
5.2.1.3. Echo Reply	26
5.2.2. Error Codes	27
5.2.2.1. Hop Count Zero Error	27
5.2.2.2. MTU Error	28
5.2.2.3. Generic Error	29
5.2.2.3.1. Error Specifiers	30
5.2.3. Expansion Code	33
5.3. Type, Length, Value (TLV) Encodings	34
5.3.1. TLV Types	34
5.3.1.1. Padding	35
5.3.1.2. Next Hop Nickname	35
5.3.1.3. Incoming Port ID	36
5.3.1.4. Outgoing Port ID	36
5.3.1.5. Outgoing Port MTU	37

6. OAM Option vs. OAM Frame	37
7. Notes	38
8. Acknowledgments	39
9. IANA Considerations	39
10. Security Considerations	39
11. References	40
11.1. Normative References	40
11.2. Informative References	41

1. Introduction

The IETF has standardized RBridges, devices that implement the TRILL protocol, a solution for transparent shortest-path frame routing in multi-hop networks with arbitrary topologies, using a link-state routing protocol technology and encapsulation with a hop-count (RFCtrill [I-D.ietf-trill-rbridge-protocol]). As RBridges are deployed, operators will face problems that require tools for troubleshooting of connectivity issues in the network. By TRILL's design, every RBridge in a campus contains a link-state database that may be useful in troubleshooting. Implementers are encouraged to leverage this by providing a means for operators to view the link-state database; however, simply being able to view the link-state database is insufficient for the requirements of operations, administration, and maintenance (OAM).

The link-state database is insufficient as the only tool for a number of reasons. As described in RFCtrill [I-D.ietf-trill-rbridge-protocol] and RBridgeMIB [I-D.ietf-trill-rbridge-mib], RBridges should support SNMP, but SNMP and the link-state database do not provide all the facilities needed. While the control plane within an RBridge campus may be functioning successfully the data plane may not be. This motivates the need for OAM tools that allow an operator to test the data plane. Protocols such as IP, MPLS, and IEEE 802.1 have features where an operator can exercise the data plane (RFC 4443 [RFC4443], RFC 0792 [RFC0792], IEEE 802.1ag [IEEE.802-1ag]). There is a need for a similar set of tools in TRILL.

Likewise, there is a need for error reporting capabilities inside an RBridge campus. For instance, if a TRILL Inner.VLAN tag has an illegal value there should be a way for devices to report this. This would allow administrators of an RBridge campus to quickly locate a problem device in the network. This document specifies a set of RBridge features for operations, administration, and maintenance purposes in RBridge campuses along with a frame format through the use of a TRILL header option for future OAM features. The features specified in this document include tools for traceroute, ping, and error reporting. Other documents may specify additional features.

1.1. Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC 2119 [RFC2119].

2. Acronyms

- o BPDU - Bridge PDU
- o CHbH - Critical Hop-by-Hop
- o CItE - Critical Ingress-to-Egress
- o DA - Destination Address
- o DR - Designated Router
- o DRB - Designated RBridge
- o ECMP - Equal-Cost Multi-Path
- o ESADI - End Station Address Distribution Instance
- o FCS - Frame Check Sequence
- o ID - Identification
- o IEEE - Institute of Electrical and Electronics Engineers
- o IETF - Internet Engineering Task Force
- o IP - Internet Protocol
- o IS-IS - Intermediate System to Intermediate System
- o MAC - Media Access Control
- o MPLS - Multiprotocol Label Switching
- o MTU - Maximum Transmission Unit
- o OAM - Operations, Administration, and Maintenance
- o P2P - Point-to-point
- o PDU - Protocol Data Unit
- o RBridge - Routing Bridge
- o SA - Source Address
- o SNMP - Simple Network Management Protocol

- o TLV - Type, Length, Value
- o TRILL - TRAnsparent Interconnection of Lots of Links
- o VLAN - Virtual Local Area Network

3. TRILL OAM Option

To facilitate message passing as needed by OAM, a new TRILL OAM option is specified. The motivation behind choosing an option to transport OAM messages is specifically to exercise the data plane of the RBridge campus, since options appear in TRILL data frames. This option is a critical ingress-to-egress option, so that RBridges that do not implement the option will not accidentally treat the encapsulated data as valid data which should be processed as a normal TRILL data frame. In special cases the option may be marked as non-critical, such as if valid data is tagged with the OAM option for debugging as in the end of Section 4.1.1.1. When a TRILL data frame has the critical bit set high in the OAM option the encapsulated frame MUST be discarded after the OAM logic processes it. If a TRILL data frame has the critical bit set low in the OAM option the encapsulated frame MUST be treated normally after the OAM logic processes it.

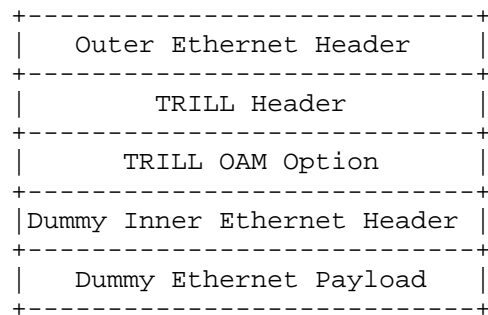
If, in contrast to using an option to transport the messages, a separate protocol data unit (PDU) were specified this new PDU might not follow the same path as the data. This OAM option is a TLV option with a common, fixed-sized initial part of the option value ([I-D.ietf-trill-rbridge-options]). This initial part contains a code that specifies a sub-option, and additional data may follow the initial part depending on this value. This section specifies the general usage of the option. Section 4 specifies some additional applications of the option. Section 5 specifies the format of the option on the wire.

There are two types of TRILL OAM messages: application and error-report. Application messages have code values ranging from 0 to 127. Error-report messages have code values ranging from 128 to 255. Frames with an error-report message MUST NOT be generated in response to frames with an error-report message. Implementations SHOULD rate limit the origination of error-report messages. As unknown unicast frames are sent as multi-destination message, sending unknown unicast frames with an error can lead to an amplification attack. As such special care and rate limiting needs to be done for error messages.

The specification of rate limiting is beyond the scope of this document. An RBridge SHOULD maintain counters for each type of error generated. Application frames such as traceroute or ping frames

generally contain a correctly formatted encapsulated Ethernet frame with a dummy payload. The TRILL OAM sub-option specifies what reaction the RBridge has to the application frame. Error frames, on the other hand, contain the error-causing frame or the initial part thereof.

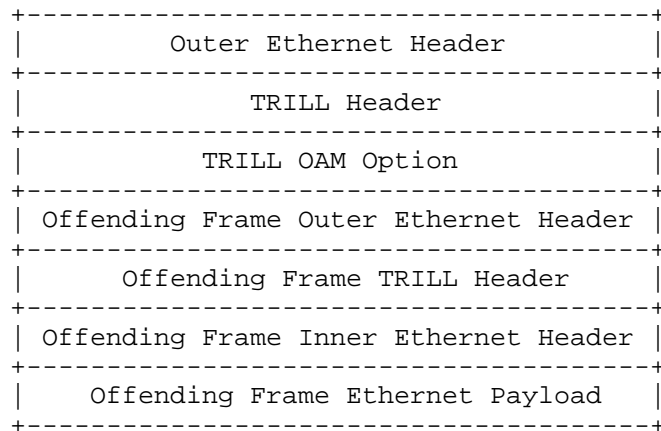
Both traceroute forms and ping use the following general layout with the TRILL OAM option being specific to the application. The fake data in certain applications can be real data:



Application Frame General Layout

Figure 1

The general layout of the TRILL OAM Error reporting frame appears below. The TRILL OAM Option is specific to the type of error being reported:



Error Frame General Layout

Figure 2

Frames with the TRILL OAM Option generated in response to another TRILL data frame MUST have fields set as follows unless otherwise specified:

Frame Type	Field	Value
Application or Error	Inner.MacSA	If the Inner.MacDA of the received frame is one of the MAC addresses of the RBridge generating the frame, the value MUST be that MAC address. Otherwise, it MUST be one of the RBridge's MAC addresses.
Application or Error	Inner.MacDA	The value MUST be the TRILL OAM unicast MAC address with a value of <TBD>. An egress RBridge MUST treat this MAC address as if it were one of its own MAC addresses. The Inner.MacDA MAY be other values as specified in subsequent sections.

Application or Error	Inner.VLAN ID	The value MUST be one of the VLANs the egress RBridge advertises connectivity on.
Application or Error	Ingress RBridge nickname	If the egress RBridge nickname of the received frame is a nickname of the RBridge generating the frame, then the value MUST be that nickname. Otherwise, it MUST be one of the RBridge's nicknames.
Application or Error	Egress RBridge nickname	The value MUST be the ingress RBridge nickname of the received frame. If the ingress RBridge nickname received is unknown the frame MUST be generated on the port the frame was received on with an Outer.MacDA and egress RBridge nickname of the RBridge that transmitted the invalid frame.
Error	Encapsulated Frame	The value MUST be N bytes of the frame which had the error where N is the minimum of the frame size and the MTU. This MUST include the TRILL header and MUST NOT include the link-layer header.
Error	M Bit	The value MUST be zero.
Application or Error	Inner.Priority	The value SHOULD be one less than the priority of the received frame, but not less than the lowest priority.

Table 1: Frame Field Values

RBridge campuses do not, in general, guarantee lossless transport of frames so a frame containing a TRILL OAM Option, possibly generated in response to some other frame, might be lost.

4. RBridge Tools

This section specifies a number of RBridge OAM tools. For classification purposes they are divided into two sections,

applications and error tools.

4.1. Application Sub-Options RBridge Tools

4.1.1. RBridge Traceroute

The ability to trace the path through the network that the data is taking is an invaluable debugging tool. RBridge traceroute provides this functionality through use of the TRILL OAM option (See Section 3). This specification specifies two types of an RBridge traceroute, each providing varying benefits and drawbacks.

4.1.1.1. Route Respond Traceroute

In a route-respond traceroute, the originating RBridge transmits one or more TRILL data frames with a TRILL OAM option. This option contains a code of a route-respond request. (See Section 5.2.1.2) The ingress RBridge MUST be the RBridge originating the frame. The route-respond traceroute is similar to the IP Option traceroute found in RFC 1393 [RFC1393].

When a traceroute is initiated, it is either targeting a known unicast target or a multi-destination target as specified by the operator. If the route-respond traceroute is for a known unicast target, the egress RBridge is the destination RBridge to which connectivity will be checked and the M bit MUST be zero. Otherwise, if the route-respond traceroute is for a multi-destination target, the egress RBridge is the distribution tree nickname for the traceroute. Multi-destination targets are handled the same as known unicast targets but require a small amount of additional logic as specified in Section 4.1.1.1.1.

The purpose of the traceroute is to confirm connectivity of the data plane, and therefore additional options such as a flow ID or a security option MAY be included. If an RBridge supports equal-cost multi-pathing (ECMP) or load balancing, the RBridge SHOULD allow operators to specify which flow the traceroute is assigned to. There is no need for all RBridges to use the same assignment method. Being able to specify the flow allows operators to test the path taken by data through the data plane. The purpose of the frame is to mimic a data frame that follows the same path through the data plane that a 'real' data frame would.

The route-respond request MAY have an arbitrary 32-bit unsigned integer sequence number to assist in matching reply messages to the request. In most circumstances a single route-respond request is needed to complete the trace but it might be desirable for a single RBridge to trace paths to multiple egress RBridges, or to trace

differing flows simultaneously. Assigning differing sequence numbers to each frame aids in matching which trace the reply belongs to.

The Inner.VLAN, Inner.MacSA, and Inner.MacDA SHOULD default to the values specified in Table 1. RBridges SHOULD provide the ability to change these values to assign the TRILL data frame to a flow. The payload of the frame is arbitrary and MAY contain any value. This value MAY have an influence on which flow the frame is assigned to.

RBridges implementing route-respond traceroute MAY issue a reply in response to this request. See Section 10 for reasoning on why some RBridges may choose not to respond to a request. If an RBridge chooses to respond to the request, the reply MUST consist of one TRILL data frame per request with a TRILL OAM option containing the code of an echo reply. The echo reply MUST have the same sequence number as the request being replied to.

For the reply the ingress RBridge field MUST be the reply-originating RBridge. The egress RBridge MUST be the request-originating RBridge. The Inner.VLAN, Inner.MacSA, and Inner.MacDA SHOULD default to the values specified in Table 1. The Outer.VLAN ID MUST be preserved. The M bit MUST be zero.

The replying RBridge MUST include its 16-bit port ID from the port on which the request was received in the incoming port field of the reply. It MUST also include its 16-bit port ID from the port on which the frame is forwarded. A port ID of 0xFFFF indicates the frame was consumed by the RBridge itself. Finally the reply MUST include the 16-bit nickname of the next hop RBridge the frame is being sent to. If the request is a multi-destination frame, this field MUST be set to the nickname of the RBridge the request frame was received from. This is the previous hop RBridge. This is to facilitate knowledge of a more precise path through the campus as seen in RFC 5837 [RFC5837].

The Internal Hop Count field is a field encoded in the echo reply option. It MUST be set to the value of the received TRILL data frame's TRILL hop-count. This allows the request-originating RBridge to order the replies received according to location in the path to the final egress RBridge. (See Section 5.2.1.3)

The advantage of this traceroute method is the request-originating RBridge only sends one frame. The disadvantage of it is that each transit RBridge implementing the OAM option needs to inspect the ingress to egress route-respond request option even though they are transit RBridges. Also, it is important to note the reply frame need not follow the same path though the campus. The reply messages are not meant to test the data plane.

An important note to make is that the end stations are not involved in this process. RBridge traceroutes are from RBridge to RBridge. While the frames sent may emulate data sent from ESa to ESb, the end stations are not, in fact, involved. The one exception, however, is an RBridge MAY be configured to tag frames it ingresses with a route-respond request option. This would facilitate debugging of real traffic. The route-respond request option tagged frame MUST be processed normally by the egress RBridge. This is achieved by having the ingress RBridge mark real traffic with a non-critical route response option. If an RBridge is configured to tag certain frames on ingress with a route-respond request, it MUST rate limit the number of such frames that it tags to avoid becoming overwhelming the network with OAM traffic.

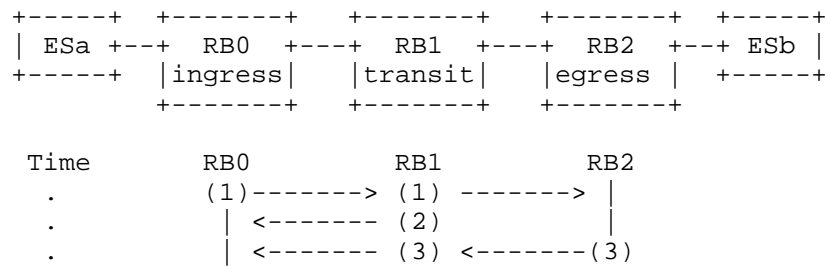
An important implementation consideration is that the transmitting RBridge MUST wait for a reply frame until a time-out occurs. At that time, the RBridge MUST assume the frame was lost, and this shall be indicated to the operator. The length of this time-out is not specified in this document.

4.1.1.1.1. Multi-Destination Targets

For multi-destination targets, it is important to note that at each branch in the tree the tagged frame will be replicated causing each RBridge in the tree to send a response. If all RBridges in the campus support the route-respond option, then the ingress RBridge will receive a reply from each of them less any RBridges pruned based on the Inner.VLAN. This is in contrast to a known unicast tagged frame where only the RBridges along the path from ingress to egress respond. The ingress RBridge can compile all of these replies, using the parent pointers located in the nexthop nickname field, into an output of the tree the traffic traversed. In the case that a non-valid distribution tree nickname is specified the traceroute frames should still be generated. The traceroute application MUST report any errors received due to the route-respond traceroute frames such as invalid nickname.

4.1.1.1.2. Route Respond Traceroute Example

Figure 3 contains a campus with three RBridges. Consider a route-respond traceroute from RB0 to RB2.



Route Respond Traceroute Example Topology

Figure 3

In this diagram RB0 transmits frame (1) destined to RB2. This frame has the route-respond request option. When RB1 receives this frame it forwards it to RB2 and it transmits an echo reply to RB0 in frame (2). When RB2 receives frame (1) it processes that frame and it transmits an echo reply to RB0 in frame (3). Some select fields for the frames are:

Frame #	Ingress RBridge	Egress RBridge	Option Code	Internal Hop Count	Option Sequence Number
(1) @ RB0	RB0	RB1	Route Respond Request	N/A	1
(1) @ RB1	RB0	RB1	Route Respond Request	N/A	1
(2) @ RB1	RB1	RB0	Echo Reply	N	1
(3)	RB2	RB0	Echo Reply	N-1	1

Table 2: Route Respond Traceroute Example Frames

For example, if the nicknames for RB0, RB1, and RB2 are 0x0001, 0x0002, and 0x0003 respectively, the console output from such a trace might be:

Route Respond Tracing

RBridge	Incoming Port Id	Outgoing Port Id	RBridge	NextHop	Nickname
0x0001	0xFFFF	0x0001		0x0002	
0x0002	0x0000	0x0001		0x0003	
0x0003	0x0000	0xFFFF		0x0000	

Table 3: Route Respond Traceroute Example Output

In this example, the first line of output is generated from local information, no route-respond frames are sent to generate it.

4.1.1.2. Hop Count Traceroute

In a hop-count traceroute, the originating RBridge starts by transmitting one TRILL data frame with a TRILL OAM option. This option contains a code of an echo request. (See Section 5.2.1.1) The ingress RBridge MUST be the RBridge originating the frame.

When a traceroute is initiated, it is either targeting a known unicast target or a multi-destination target as specified by the operator. If the hop-count traceroute is for a known unicast target, the egress RBridge is the destination RBridge to which connectivity will be checked and the M bit MUST be zero. Otherwise, if the hop-count traceroute is for a multi-destination target, the egress RBridge is the distribution tree nickname for the traceroute. Multi-destination targets are handled the same as known unicast targets but require a small amount of additional logic as specified in Section 4.1.1.2.1.

The first echo request frame transmitted MUST have a hop-count of zero. The RBridge will continue transmitting these echo requests, incrementing the hop-count by one each time until a hop-count error message is received from the destination. Each of these requests in turn will generate a hop-count error message until the destination is reached. If a transit RBridge decrements the hop-count by more than one it may transmit multiple hop-count error messages.

The purpose of the traceroute is to confirm connectivity of the data plane, and therefore additional options such as a flow ID or a security option MAY be included. If an RBridge supports equal-cost multi-pathing (ECMP) or load balancing, the RBridge SHOULD allow operators to specify which flow the traceroute is assigned to. There is no need for all RBridges to use the same assignment method. Being able to specify the flow allows operators to test the path taken by data through the data plane. The purpose of the frame is to mimic a data frame that follows the same path through the data plane that a

'real' data frame would.

The route-respond request MAY have an arbitrary 32-bit unsigned integer sequence number to assist in matching reply messages to the request. This is important for the hop-count traceroute since replies may return to the ingress RBridge in a different order than their matching requests were sent.

The Inner.VLAN, Inner.MacSA, and Inner.MacDA SHOULD default to the values specified in Table 1. RBridges SHOULD provide an option to change these values to assign the TRILL data frame to a flow. The payload of the frame is arbitrary and MAY contain any value. This value MAY have an influence on which flow the frame is assigned to.

The replying RBridge MUST include its 16-bit port ID from the port on which the hop-count error generating frame was received in the incoming port field of the reply. It MUST also include its 16-bit port ID from the port on which the frame would be forwarded if the frame did not have an hop-count error. A port ID of 0xFFFF indicates the frame was consumed by the RBridge itself. Finally the reply MUST include the 16-bit nickname of the next hop RBridge the frame would have been sent to if there were no error. If the request is a multi-destination frame, this field MUST be set to the nickname of the RBridge the frame was received from. This is the previous hop RBridge. This is to facilitate knowledge of a more precise path through the campus as seen in RFC 5837 [RFC5837].

The advantage of this traceroute method is the transit RBridges do not have to do any special processing of the frames until a hop-count error is detected, a condition they are required by the TRILL base protocol to at least detect. The disadvantage is the request-originating RBridge needs to transmit as many frames as there are hops between itself and the destination RBridge.

An important note to make is that the end stations are not involved in this process. RBridge traceroutes are from RBridge to RBridge. While the frames sent may emulate data sent from ESa to ESb, the end stations are not, in fact, involved.

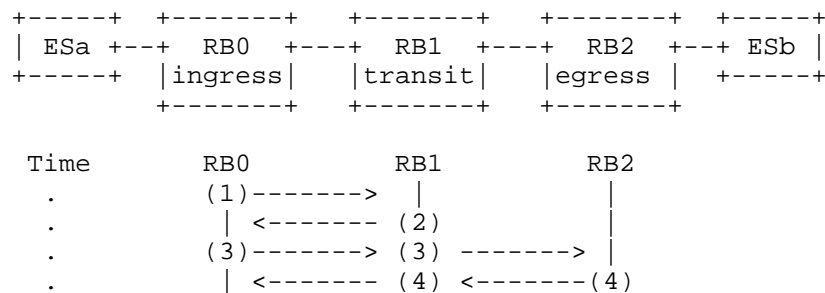
4.1.1.2.1. Multi-Destination Targets

For multi-destination targets, it is important to note that at each branch in the tree the tagged frame will be replicated causing each RBridge in the tree, possibly pruned by VLAN and/or multicast group, to send a response to the echo request. If all RBridges in the possibly pruned distribution tree support the echo request option, then the ingressing RBridge will receive a echo reply from each of them. This is in contrast to a known unicast tagged frame where only

the RBridges along the path from ingress to egress transmit the error report. The ingressing RBridge can compile all of these replies, using the parent pointers located in the nexthop nickname field, into an output of the tree the traffic traversed. In the case that a non-valid distribution tree nickname is specified the traceroute frames should still be generated. The traceroute application MUST report any errors received due to the hop-count traceroute frames such as invalid distribution tree nickname. RBridges receiving a multicast destination echo request MUST NOT transmit an echo reply if the multi-destination bit is set. Echo requests not used with the hop-count traceroute are pings, and pings are not valid to multi-destination traffic. In a hop-count traceroute devices will already be transmitting a hop-count error message and so there is no reason to transmit a double set of replies. A multi-destination hop-count traceroute does not stop when an echo reply is received. It stops when the transmitted hopcount reaches 0x3F.

4.1.1.2.2. Hop Count Traceroute Example

Figure 4 contains a campus with three RBridges. Consider a hop-count traceroute from RB0 to RB2.



Hop Count Traceroute Example Topology

Figure 4

In this diagram RB0 transmits frame (1) destined to RB2. This frame has the echo request option and a hop-count of 0. When RB1 receives this frame it drops it and transmits a hop-count-exceeded message, (2), to RB0. RB0 then transmits a frame, (3), with a hop-count of 1. RB1 decrements this hop-count by 1 to 0 and forwards it to RB2. RB2 drops frame (3) and transmits a hop-count-exceeded message, (4), to RB0. The traceroute is now complete.

Some select fields for the frames are:

Frame #	Ingress RBridge	Egress RBridge	Option Code	Option Sequence Number	Hop Count
(1)	RB0	RB2	Echo Request	1	0
(2)	RB1	RB0	Hop Count Error	1	N/A
(3) @ RB1	RB0	RB2	Echo Request	2	1
(3) @ RB2	RB0	RB2	Echo Request	2	0
(4) @ RB1	RB2	RB0	Hop Count Error	2	N/A
(4) @ RB0	RB2	RB0	Hop Count Error	2	N/A

Table 4: Hop Count Traceroute Example Frames

For example, if the nicknames for RB0, RB1, and RB2 are 0x0001, 0x0002, and 0x0003 respectively, the console output from such a trace might be:

Hop Count Tracing

RBridge	Incoming Port Id	Outgoing Port Id	RBridge	Nexthop Nickname
0x0001	0xFFFF	0x0001		0x0002
0x0002	0x0000	0x0001		0x0003
0x0003	0x0000	0xFFFF		0x0000

Table 5: Hop Count Traceroute Example Output

In this example, the first line of output is generated from local information, no hop-count frames are sent to generate it.

4.1.2. RBridge Ping

Ping is a tool for verifying RBridge connectivity. Like with an RBridge traceroute, the ping-originating RBridge transmits one or

more TRILL data frames with a TRILL OAM option. This option contains the code of an echo request (See Section 5.2.1.1). The ingress RBridge MUST be the RBridge-originating frame. The egress RBridge is the destination RBridge to which connectivity will be checked. The M bit MUST be zero.

As with RBridge traceroute, additional options such as a flow ID or a security option MAY be included. If an RBridge supports equal-cost multi-pathing (ECMP) or load balancing, the RBridge SHOULD allow operators to specify which flow the ping is assigned to. There is no need for all RBridges to use the same assignment method. This ping traffic, once again, will mimic real traffic through the network, like traceroute traffic as previously specified in Section 4.1.1.1.

The echo request MAY have an arbitrary 32-bit unsigned integer sequence number to assist in matching reply messages to the request. In most circumstances, a single echo request is needed to complete the ping but it might be desirable for a single RBridge to ping multiple egress RBridges, or trace differing flows simultaneously. Assigning differing sequence numbers to each frame aids in matching which trace the reply belongs to.

The Inner.VLAN, Inner.MacSA, and Inner.MacDA SHOULD default to the values specified in Table 1. RBridges SHOULD provide the ability to change these values as to assign the TRILL data frame to a flow. The payload of the frame is arbitrary and MAY contain any value. This value can have an influence on which flow the frame is assigned to.

RBridges implementing ping MAY issue a reply in response to this request. See Section 10 for reasoning on why some RBridges may choose not to respond to a request. If an RBridge chooses to respond to the request, the reply MUST consist of one TRILL data frame per request with a OAM option containing the code of an echo reply. The echo reply MUST have the same sequence number as the request being matched.

For the echo reply the ingress RBridge field MUST be the reply-originating RBridge's nickname. The egress RBridge MUST be the request-originating RBridge's nickname. The Inner.VLAN, Inner.MacSA, and Inner.MacDA SHOULD default to the values specified in Table 1. The Outer.VLAN ID MUST be preserved. The M bit MUST be zero.

The reply-originating RBridge MUST include its 16-bit port ID from the port on which the request was received in the incoming port field of the reply. It MUST also include its 16-bit port ID from the port on which the frame is forwarded. A port ID of 0xFFFF indicates the frame was consumed by the RBridge itself. The nickname field in the generated frame MUST be set to all zeros on transmission and ignored

on reception.

The Internal Hop Count field of the reply MUST be set to zero. The ping functionality does not use the Internal Hop Count field of the reply. (See Section 5.2.1.3)

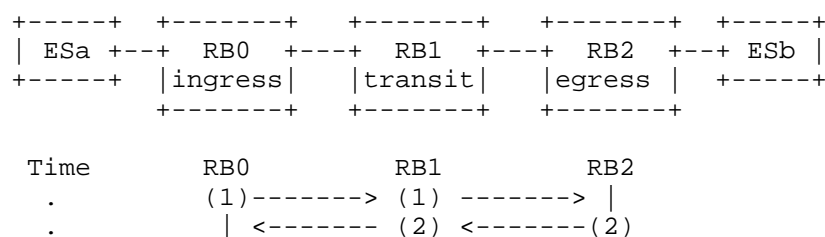
It is also important to note that the reply frame need not follow the same path though the campus. The reply messages are not meant to test the data plane.

End stations are not involved in this process. RBridge pings are from RBridge to RBridge. While the frames sent may emulate data sent from ESa to ESb, the end stations are not, in fact, involved. The one exception, however, is an RBridge MAY be configured to tag frames it ingresses with an echo request option. This would facilitate debugging of real traffic. The echo request option tagged frame MUST be processed normally by the egress RBridge. This is done by the ingress RBridge marking real traffic with a non-critical echo reply option. If an RBridge is configured to tag frames it ingresses with an echo request, it MUST rate limit how often it tags data being ingressed to prevent the network from becoming congested with OAM traffic.

An important implementation consideration is that the transmitting RBridge MUST wait for a reply frame until a time-out occurs. At that time, the RBridge MUST assume the frame was lost, and this shall be indicated to the operator. The length of this time-out is not specified in this document.

4.1.2.1. Ping Example

Figure 5 contains a campus with three RBridges. Consider a ping from RB0 to RB2.



Ping Example Topology

Figure 5

In this diagram RB0 transmits frame (1) destined to RB2. This frame has the echo request option. When RB1 receives this frame it forwards it to RB2. When RB2 receives this frame it transmits and echo reply frame (2) destined to RB0. RB1 receives this frame and forwards it to RB0.

Some select fields for the frames are:

Frame #	Ingress RBridge	Egress RBridge	Option Code	Option Sequence Number
(1)	RB0	RB2	Echo Request	1
(2)	RB2	RB0	Echo Reply	1

Table 6: Ping Example Frames

For example, if the nicknames for RB0, RB1, and RB2 are 0x0001, 0x0002, and 0x0003 respectively, the console output from such a ping might be:

Pinging

```

... from 0x0001 to 0x0003... 0x0003 is alive
... from 0x0001 to 0x0003... 0x0003 is alive
... from 0x0001 to 0x0003... 0x0003 is alive

```

Table 7: Ping Example Output

In this example, the ping was repeated three times with the sequence number being changed each time.

4.2. Error Sub-Options RBridge Tools

Errors can occur through the reception of TRILL data frames. For this purpose, the TRILL OAM Option has several error sub-options. These are generated due to various events as specified subsequently.

Each of these error sub-options is used in a similar fashion. When a TRILL data frame is received that triggers an error, an error notification frame MAY be generated. See Section 10 for reasoning on why some RBridges MAY choose not to report an error. This frame has a TRILL header and it contains, as its payload, the frame received

with the error. If the size of the received frame would cause the generated frame to exceed the campus-wide MTU, the payload MUST be truncated to the campus-wide MTU. The payload MUST include the TRILL header of the received frame and MUST NOT include the link-layer header. The generated reply MUST contain the error option specific to the error.

When the original ingress RBridge receives the error frame, at a minimum, the RBridge SHOULD update a counter specifying the number of error frames received for the causing error. The encapsulated frame MUST NOT be unencapsulated and transmitted. The RBridge SHOULD also keep a set of counters for errors reported by other RBridges.

4.2.1. Hop Count Zero Error

When a TRILL data frame is received with a hop-count of zero, an error notification frame MAY be generated. The generated reply MUST contain the hop-count zero error sub-option. If the received frame has the echo request option, the hop-count zero error option MUST have a sequence number matching the echo request. Otherwise, the sequence number MUST be set to zero. The incoming port ID MUST be the port ID the received frame arrived on. The outgoing port ID MUST be the port ID of the port the received frame would have been forwarded onto if the hop-count was not zero. Finally, the error frame MUST include the 16-bit nickname of the next hop RBridge the frame would have been sent to. If the request is a multi-destination frame, this field MUST be set to all zeros on transmission and ignored on reception. If the RBridge transmitting the request is the egress RBridge, this field MUST be set to 0x0000.

4.2.2. MTU Error

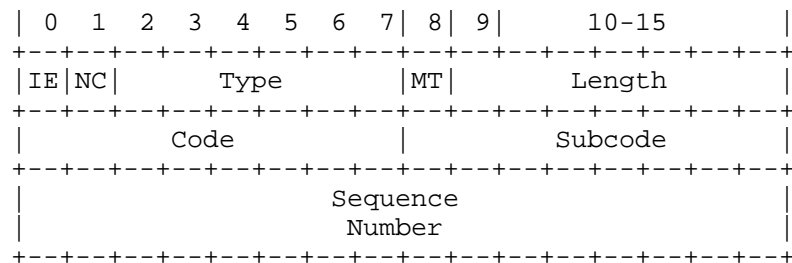
When a TRILL data frame is received with a payload that would exceed the MTU of the port the frame would otherwise be forwarded to, an error notification frame MAY be generated. The generated reply MUST contain the MTU error sub-option. The outgoing port MTU field MUST have the MTU of the port the frame would have otherwise been transmitted on. The incoming port ID MUST be the port ID the received frame arrived on. The outgoing port ID MUST be the port ID of the port the received frame would have been forwarded onto if the frame size was not too large. Finally, the error-report message MUST include the 16-bit nickname of the next hop RBridge the frame would have been sent to. If this is a multi-destination frame this field MUST be set to all zeros on transmission and ignored on reception. If the RBridge transmitting the request is the egress RBridge, this field MUST be set to 0x0000.

4.2.3. Generic Error

When a TRILL data frame is received with an error not already specified, an error notification frame is generated. The generated reply **MUST** contain the generic error sub-option. The sub-code **MUST** contain a code specifying the error encountered. The valid values are specified in Section 5.2.2.3.1. By way of note for future error code specifications, this generic error reporting feature is meant for errors occurring where no additional information needs to be communicated back to the ingressing RBridge.

5. TRILL OAM Option Format

This section specifies the format of the TRILL OAM Option on the wire.



TRILL OAM Option Common Initial Part

Figure 6

The option fields and flags are as follows:

- o Type: 0x02.
- o Length: The length of the option value in octets.
- o IE: **MUST** be one. This is an ingress to egress option.
- o NC: Varies depending on the code.
- o MT: **MUST** be zero. This is an immutable option.
- o Code: Specifies how this OAM option is to be interpreted. The value ranges from 0-255 inclusive and the code meanings are specified in Section 5.1

- o Subcode: Further specifies the code field. This allows for additional granularity specific to each code value. The value ranges from 0-255, inclusive and the meanings are specific to their code value.
- o Sequence Number: This field is used to sequence frames for certain tools. Not all tools utilize the sequence number field.

5.1. Code Values

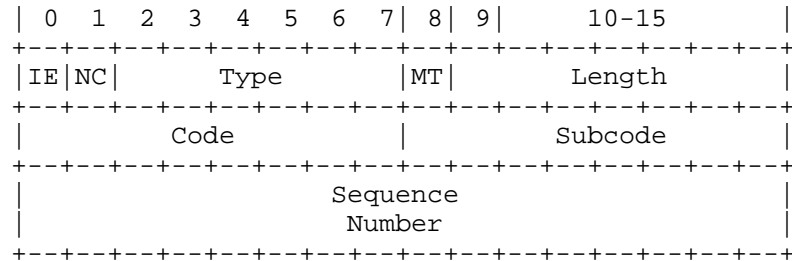
The code values are:

- o 0: Echo Request, See Section 5.2.1.1
- o 1: Route Respond Request, See Section 5.2.1.2
- o 2: Echo Reply, See Section 5.2.1.3
- o 3-122: Available for Allocation by IETF Review
- o 123-126: Reserved for Private Experimentation
- o 127: Application Expansion Value, See Section 5.2.3
- o 128: Hop Count Zero Error, See Section 5.2.2.1
- o 129: Generic Error, See Section 5.2.2.3
- o 130: MTU Error, See Section 5.2.2.2
- o 131-250: Available for Allocation by IETF Review
- o 251-254: Reserved for Private Experimentation
- o 255: Error Expansion Value, See Section 5.2.3

5.2. Codes

5.2.1. Application Codes

5.2.1.1. Echo Request



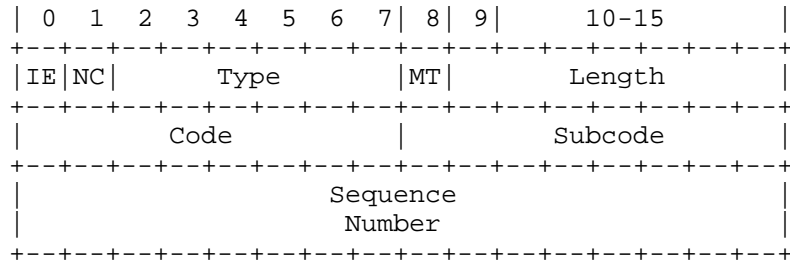
Echo Request

Figure 7

This option is used by ingress RBridges to request an echo reply from the egress RBridge. Further uses are specified in Section 4.1.1 and Section 4.1.2

- o Length: 6
- o IE: MUST be one. This is an ingress to egress option.
- o NC: Defaults to zero. The OAM option is normally a critical ingress-to-egress option but it MAY be a non-critical option if the encapsulated frame is real data that needs to be processed normally on egress.
- o MT: MUST be zero. This is an immutable option.
- o Code: MUST be 0.
- o Subcode: MUST be 0x00. This field is not used by this sub-option. It is set to zero on transmission and ignored on reception.
- o Sequence Number: An arbitrary 32-bit unsigned integer used to aid in matching reply messages to echo requests. MAY be zero.

5.2.1.2. Route Respond Request



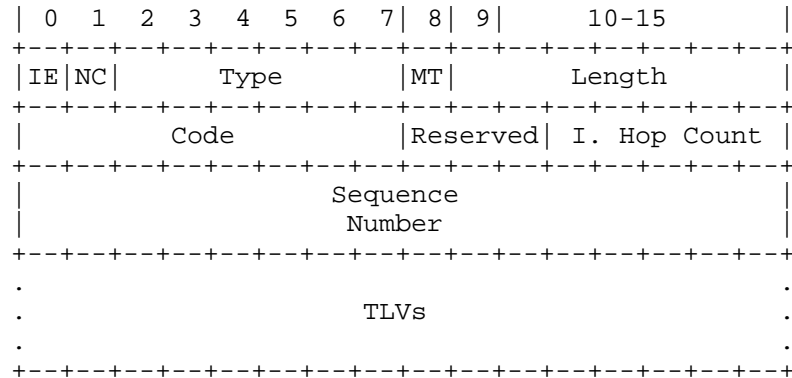
Route Respond Request Format

Figure 8

This option is used by ingress RBridges to trace a route through an RBridge campus. Further uses are specified in Section 4.1.1.

- o Length: 6
- o IE: MUST be one. This is an ingress to egress option.
- o NC: Defaults to zero. The OAM option is normally a critical ingress-to-egress option but it MAY be a non-critical option if the encapsulated frame is real data that needs to be processed normally on egress.
- o MT: MUST be zero. This is an immutable option.
- o Code: MUST be 1.
- o Subcode: MUST be 0x00. This field is not used by this sub-option. It is set to zero on transmission and ignored on reception.
- o Sequence Number: An arbitrary 32-bit unsigned integer used to aid in matching reply messages to echo requests. May be zero.

5.2.1.3. Echo Reply



Echo Reply Format

Figure 9

This option is used by egress RBridges to reply to an echo request from the ingress RBridge. Further uses are specified in Section 4.1.1 and Section 4.1.2.

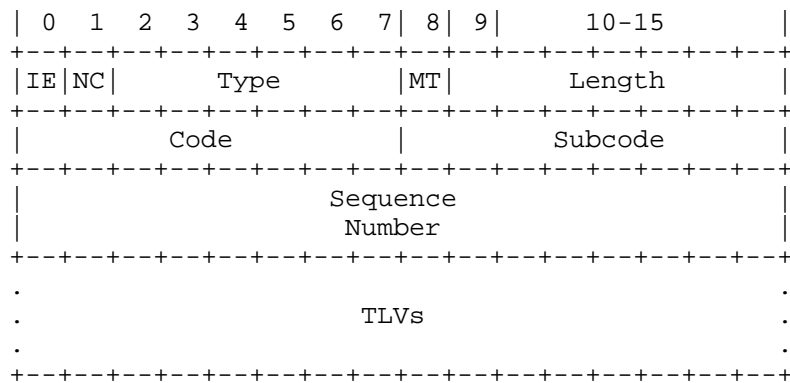
- o Length: 14
- o IE: MUST be one. This is an ingress to egress option.
- o NC: MUST be zero. This is a critical option
- o MT: MUST be zero. This is an immutable option.
- o Code: MUST be 2.
- o Reserved: A reserved field. Set to zero on transmission and ignored on reception.
- o Internal Hop Count: If the request being replied to was an echo request, this value MUST be zero on transmission and ignored on reception. If the request being replied to was a respond request, this value is a copy of the TRILL Hop Count value in the request. The reserved and internal hop-count fields combined occupy the subcode field of the TRILL OAM option.
- o Sequence Number: A 32-bit unsigned integer used to aid in matching reply messages to echo requests. This MUST match the request

being replied to.

- o TLVs: A set of type, length, value encoded fields as specified in Section 5.3. The next hop nickname, outgoing port ID, and incoming port ID TLVs are required.

5.2.2. Error Codes

5.2.2.1. Hop Count Zero Error



Hop Count Zero Error Format

Figure 10

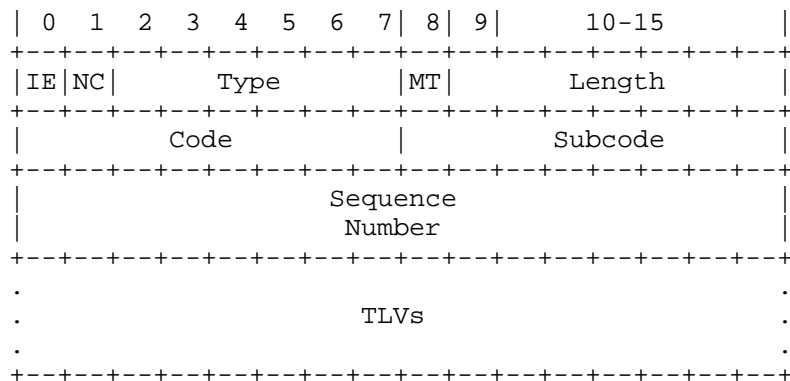
This option is used by egress or transit RBridges to signal that the TRILL hop-count field has reached zero.

- o Length: 14
- o IE: MUST be one. This is an ingress to egress option.
- o NC: MUST be zero. This is a critical option.
- o MT: MUST be zero. This is an immutable option.
- o Code: MUST be 128.
- o Subcode: MUST be 0x00. This field is not used by this sub-option. It is set to zero on transmission and ignored on reception.
- o Sequence Number: A 32-bit unsigned integer used to aid in matching reply messages to echo requests and route-respond requests. If

the frame whose hop-count dropped to zero contains the echo request option (See Section 5.2.1.1), this MUST match the sequence number echo request found in that option. If this is not in reply to a request, then the sequence number MUST be set to zero.

- o TLVs: A set of type, length, value encoded fields as specified in Section 5.3. The next hop nickname, outgoing port ID, and incoming port ID TLVs are required.

5.2.2.2. MTU Error



MTU Error Format

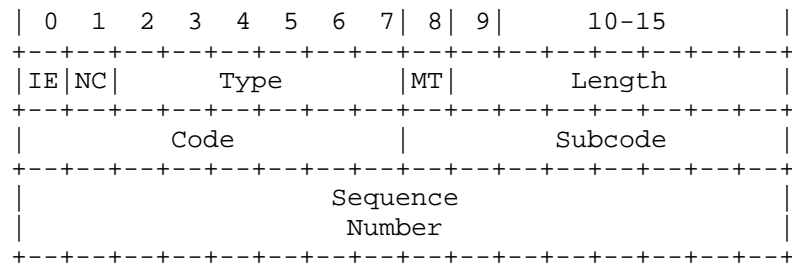
Figure 11

This option is used by a transit RBridge to indicate a TRILL data frame that exceeds the MTU of the outgoing port from which it was transmitted.

- o Length: 10
- o IE: MUST be one. This is an ingress to egress option.
- o NC: MUST be zero. This is a critical option.
- o MT: MUST be zero. This is an immutable option.
- o Code: MUST be 130.
- o Subcode: MUST be 0x00. This field is not used by this sub-option. It is set to zero on transmission and ignored on reception.

- o Sequence Number: This field is not used by this sub-option. It is set to zero on transmission and ignored on reception.
- o TLVs: A set of type, length, value encoded fields as specified in Section 5.3. The outgoing port MTU, next hop nickname, outgoing port ID, and incoming port ID TLVs are required.

5.2.2.3. Generic Error



Generic Format

Figure 12

This option is used by egress or transit RBridges to signal that a TRILL related frame has an error.

- o Length: 2
- o IE: MUST be one. This is an ingress to egress option.
- o NC: MUST be zero. This is a critical option.
- o MT: MUST be zero. This is an immutable option.
- o Code: MUST be 129.
- o Subcode: MUST be a specifier of the error discovered in the frame. The valid values are specified in Section 5.2.2.3.1
- o Sequence Number: This field is not used by this sub-option. It is set to zero on transmission and ignored on reception.

5.2.2.3.1. Error Specifiers

The sub-code values fall into three categories: errors, warnings, and comments. All sub-codes represent something out of the ordinary that has gone wrong, but certain ones are more important than others. Sub-codes that are classified as errors are the most severe with warning sub-codes being slightly less severe. These are by default enabled. Sub-codes classified as comments are minor and are by default disabled. They may be useful for operators debugging a network. All error generations are optional and therefore MAY be generated or not generated depending on security and implementation constraints.

The error specifiers sub-code values are:

Sub-codes

- o 0: Unknown Error: Indicates and an error has occurred.
- o 1: Corrupt Frame: Frame received with invalid FCS or that was not an 8-bit multiple in length. It may be impossible for a device to signal this if the low-level port hardware hides this from the software.
- o 2: Invalid Outer.MacDA: Indicates the MAC Address is a multicast address and the M bit is zero, the MAC Address is not a multicast address and the M bit is one, or the M bit is zero and the frame carried is an ESADI frame.
- o 3: Illegal Outer.VLAN: Indicates the Outer.VLAN ID is 0xFFFF.
- o 4: Invalid Outer.VLAN: Indicates the Outer.VLAN ID was not the designated VLAN ID.
- o 5: Unknown TRILL Version: Indicates the TRILL Version is unknown.
- o 6: Op-Length Exceeds Frame Length: Indicates the Op-Length says the options field extends beyond the end of the received frame length.
- o 8: Unknown Egress RBridge: Indicates the Egress RBridge in a received frame is unknown.
- o 9: Unknown Ingress RBridge: Indicates the Ingress RBridge in a received frame is unknown.
- o 10: Unsupported Critical Hop-by-hop Option: Indicates an unsupported critical hop-by-hop option was received.

- o 11: Unsupported Critical Ingress-to-Egress Option: Indicates an unsupported critical ingress-to-egress option was received.
- o 12-84: Available for allocated by IETF Review
- o 85: Reserved for Private Experimentation

Warning Sub-codes

- o 86: Illegal Inner.VLAN: Indicates the Inner.VLAN ID is 0xFFFF.
- o 87: Inner/Outer VLAN Priority Mismatch: Indicates the priority values in the inner and outer VLANs do not match.
- o 88: P2P Hello on TRILL Hello Link: Indicates a P2P Hello was received on a TRILL Hello Link.
- o 89: TRILL Hello on P2P Hello Link: Indicates a TRILL Hello was received on a P2P Hello Link.
- o 90: No Adjacency: Indicates a TRILL data frame was sent from an RBridge the receiving RBridge is not adjacent with.
- o 91: Encapsulated BPDU/VRP Frame: A TRILL Frame containing a BPDU or VRP frame was received.
- o 92: Invalid Mutability Flag: Indicates the mutability flag was set on a received CHbH Option.
- o 93: Invalid TLV Option Length: Indicates the option length field of a TLV option was between 121 and 127.
- o 94: Options Ordering Error: Indicates the TLV options are ordered incorrectly.
- o 95: Additional Flag TLV Zero: Indicates a problem in the additional Flag TLV.
- o 96: Configured Nickname Collision: Indicates an RBridge was detected in the campus with the same nickname (Configured or not).
- o 97: Multiple DRBs detected.
- o 98: Multiple appointed forwarders detected.
- o 99-169: Available for allocation by IETF Review

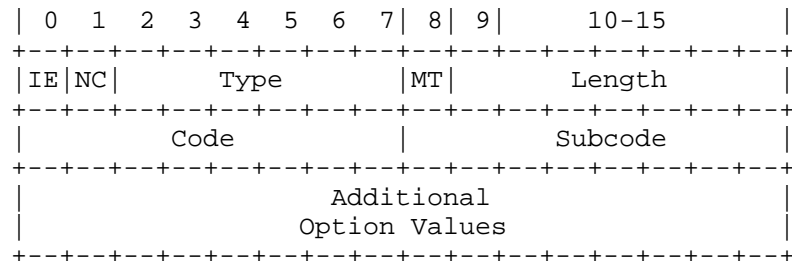
- o 170: Reserved for Private Experimentation

Comment Sub-codes

- o 171: Inner.VLAN C-Bit Set: Indicates the C-Bit in the Inner.VLAN is set.
- o 172: Unknown Inner.MacDA: Indicates the Inner.MacDA is unknown. This may occur if devices are configured to explicitly register end stations and an unknown Inner.MacDA occurs in a unicast TRILL data frame. This also only applies at egress and could indicate that the Inner.MacDA was a learned address that has timed out.
- o 173: Unknown Inner.MacSA: Indicates the Inner.MacSA is unknown. This may occur if devices are configured to explicitly register end stations and an unknown Inner.MacSA occurs in a TRILL data frame.
- o 174: Outer.VLAN C-Bit Set: Indicates the C-Bit in the Outer.VLAN is set for an Ethernet frame.
- o 175: Invalid Reserved Bits: Indicates the reserved bits are non-zero in a received frame.
- o 176: Invalid Nickname: Indicates a nickname in the reserved space of 0xFFC0 to 0xFFFF was received that is not implemented at the receiving RBridge.
- o 177: Unsupported Non-Critical Hop-by-hop Option: Indicates an unsupported non-critical hop-by-hop option was received. While sending a non-critical option to an unsupported device is not an error this could be used to support identification of devices needing an upgrade.
- o 178: Unsupported Non-Critical Ingress-to-Egress Option: Indicates an unsupported non-critical ingress-to-egress option was received. While sending a non-critical option to an unsupported device is not an error this could be used to support identification of devices needing an upgrade.
- o 179: Performance Exceeded: Indicates a frame was discarded due to performance problems such as a buffer overflow.
- o 180: Insufficient Hop Count: Indicates a frame was received with a hop-count that was insufficient to reach the destination.
- o 181-254: Available for allocation by IETF Review

- o 255: Reserved for Private Experimentation

5.2.3. Expansion Code



Expansion Code Format

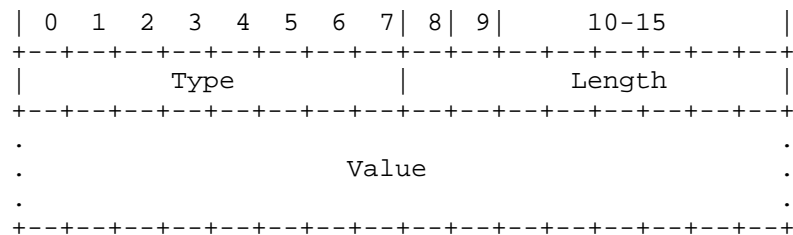
Figure 13

This option is used to specify additional TRILL OAM Option code space beyond the 255 values specified.

- o Length: The length of the option value in octets.
- o IE: MUST be one. This is an ingress-to-egress option.
- o NC: Varies depending on the code.
- o MT: MUST be zero. This is an immutable option.
- o Code: MUST BE 127 or 255.
- o Subcode: Further specifies the code field. This allows for additional granularity specific to each code value. The value ranges from 0-255 inclusive, and the meanings are specific to their code value.
- o Additional Option Values: Specify how this OAM option is to be interpreted just as the code value does in the TRILL OAM option. The value meanings are available for allocation by IETF Review. This field occupies the sequence number field of the common OAM option initial part.

5.3. Type, Length, Value (TLV) Encodings

To facilitate future interoperable expansion of the data carried in OAM sub-options some sub-options use a TLV encoding. These TLV sections consist of a list of type, length, value encoded data where the type signals to the RBridge how to interpret the value, and the length tells the RBridge the length of the value in bytes. The type and length are both 8 bit fields. A length of zero indicates the value is a UTF-8 string with a NULL ('\0') terminating byte.



TLV Format

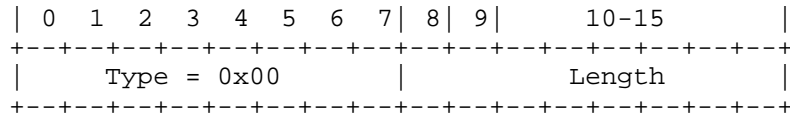
Figure 14

The type values are:

- o 0: Padding, See Section 5.3.1.1
- o 1: Next Hop Nickname, See Section 5.3.1.2
- o 2: Outgoing Port ID, See Section 5.3.1.4
- o 3: Incoming Port ID, See Section 5.3.1.3
- o 4: Outgoing Port MTU, See Section 5.3.1.5
- o 5-254: Available for allocation by IETF Review
- o 255: Reserved for Private Experimentation

5.3.1. TLV Types

5.3.1.1. Padding

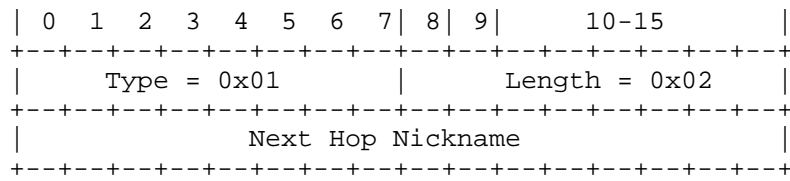


Padding Format

Figure 15

The padding TLV MAY appear in any TLV list to increase the length of the TRILL OAM sub-option to a multiple of 32-bits. If the length is zero the value MUST NOT be interpreted as a UTF-8 string and the value is instead interpreted as not present.

5.3.1.2. Next Hop Nickname

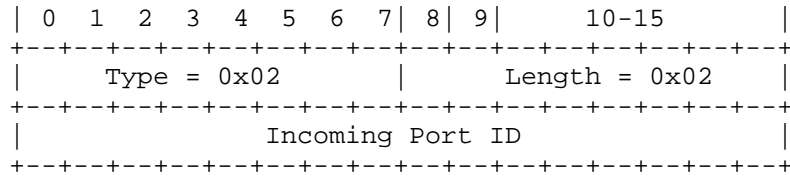


Next Hop Nickname Format

Figure 16

For traceroutes targeting known unicast destinations, hop-count errors, and MTU errors, this TLV MUST be the 16-bit nickname of the next hop RBridge the frame is being or would have been sent to. If the RBridge transmitting the TLV is the egress RBridge this field MUST be set to 0x0000. For traceroutes targeting multi-destination destinations, e.g. with the TRILL M bit high, this field contains the nickname of the RBridge the frame being responded to is from. For pings, this field MUST be set to all zeros on transmission and ignored on reception. For multi-destination hop-count errors this field contains the nickname of the RBridge the frame with the exceeded hop-count was sent from. For multi-destination MTU error traffic, this field MUST be set to all zeros on transmission and ignored on reception.

5.3.1.3. Incoming Port ID

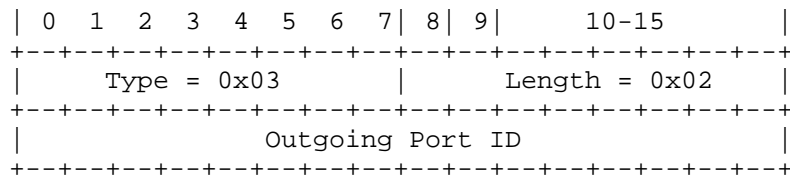


Incoming Port ID Format

Figure 17

This TLV MUST be set to the Port ID found in 'The Special VLANs and Flags sub-TLV' for the port the request being replied to was received on. ([I-D.ietf-isis-trill])

5.3.1.4. Outgoing Port ID

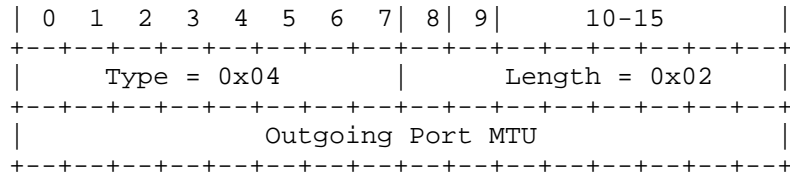


Outgoing Port ID Format

Figure 18

This TLV MUST be set to the Port ID found in 'The Special VLANs and Flags sub-TLV' for the port the frame is being forwarded on to (or would have been for an echo request/hop-count error). ([I-D.ietf-isis-trill]) If the request was consumed by the replying RBridge, the port ID MUST be 0xFFFF.

5.3.1.5. Outgoing Port MTU



Outgoing Port MTU Format

Figure 19

This TLV MUST be the MTU of the outgoing port specified in the outgoing port ID TLV.

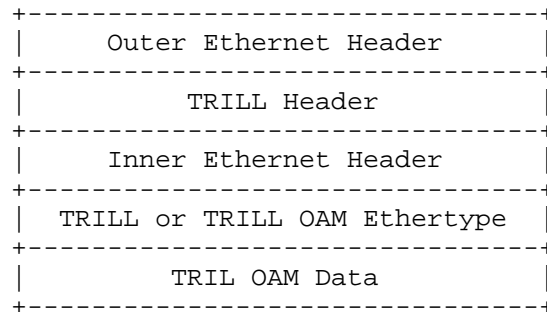
6. OAM Option vs. OAM Frame

During some offline discussion there was much debate on the use of the OAM option as presented in this draft. The problem with using an option is some ASIC implementations could slow path any TRILL data frame with an option length greater than zero by sending it to software. This means the OAM frame might not be handled by the same logic a regular data frame would be handled by.

The intention of this draft was to allow OAM frames to still take the fast path by using a CItE option. All the forwarding path would have to do is peak at the first two bits in the TRILL options to know it does not need to slow path this frame. For hop count traceroutes this is fine since the frame only needs to be sent to the software after it has hit an error. With the error reporting and ping mechanisms this is also not a problem since these tools are end-to-end. The one place this might be a problem is in the route-respond traceroute. In this case transit RBridges implementing the OAM option are expected to snoop the ingress-to-egress option. Fortunately in practice if a device kept the frame on the fast path and did not snoop the OAM option this would only cause the RBridge performing the traceroute to skip certain hops along the way as seen in IP traceroutes.

Another problem with using an OAM option is it limits the size of the OAM option to 120 bytes. In this presented draft this is fine since no TRILL OAM codes require a large amount of space but one can imagine more complicated applications defined later that need more bytes.

An alternative solution to an OAM option would be to use the encapsulated frame for OAM purposes. The basic idea can be seen in Figure 20. The idea is to not use an option and drop the 16 bits of the IE, NC, Type, MT, and Length fields seen in Figure 6. The one change required here is the TLV sections would require an additional total TLV length field. to indicate how long the TLV section is.



OAM Frame Format

Figure 20

The disadvantage of this type of solution is real data can no longer be tagged with the TRILL OAM option to debug problems in real time. Also this solution does not solve the requirement of route-respond traceroute frames needing to be snooped. With this in mind a future version of this draft will present both of these solutions in parallel and perhaps using an OAM/control header as presented in other drafts.

7. Notes

NOTE: This section contains some ideas and will be removed later.

For the sequence number field in the generic error which is currently not used perhaps this could contain a pointer to the offending field in the frame. Then again we don't need a 32-bit number for that.

The port-id use of 0xFFFF is not consistent with the -16 draft and would need to be reserved. Another option is to use a boolean to indicate this.

Itt might be nice to specify a IS-IS sub-TLV for port-id to ifname string mapping.

Perhaps we should specify advertisement of this documents options in ISIS TLVs.

Perhaps add a diagram for a multi-destination traceroute and for a error message

A more detailed requirements section would benefit this draft.

Traceroutes to specific multicast groups to test group pruning would be useful.

8. Acknowledgments

Many people have contributed to this work, including the following, in alphabetic order: Donald E. Eastlake 3rd, Anoop Ghanwani, Jeff Laird, and Marc Sklar

9. IANA Considerations

IANA will create four subregistries within the TRILL registry. A "TRILL OAM Option Code" subregistry that is initially populated as specified in Section 5.1. A "TRILL OAM Option Error Sub-Option Error Specifiers" subregistry that is initially populated as specified in Section 5.2.2.3.1. A "TRILL OAM Option Application Expansion Additional Option Values" and a "TRILL OAM Option Error Expansion Additional Option Values".

Additional values for these subregistries are allocated by IETF Review [RFC5226].

This draft also requires action to reserve the TRILL Header TLV Option Type 0x02 and of the TRILL OAM unicast MAC address.

10. Security Considerations

The nature of the TRILL OAM Option lends itself to security concerns. By providing information about the topology of a network, attackers can gain greater knowledge of a network in order to exploit the network. Passive attacks such as reading frames with the OAM option could be used to gain such knowledge or active attacks where an attacker mimics an RBridge can be used to probe the network. Authentication, data integrity, protection against replay attacks, and confidentiality for TRILL OAM frames may be provided using a to-be-specified TRILL Security Option. Using such a security option would mitigate both the passive and active attacks.

For instance, data origin authentication could be provided in the future using a security options in the TRILL Header by verifying a

hash using shared keys or a mechanism like SEND with CGA [RFC 3971]. To prevent against replay attacks rate limiting, sequence numbers as well as some nonce based mechanism could be provided. Confidentiality for TRILL OAM frames could be provided based on some future security option extension which encypts TRILL frames.

In a network where one does not wish to configure a security option, the threat of attackers is still present. For this reason, generation of any TRILL OAM Option frames is optional and SHOULD be configurable by an operator on a per RBridge basis. An RBridge MAY have this configurable on a per port basis. For instance, an operator SHOULD be able to disable route-respond traceroute reply messages or error-report message generation per port.

Another security threat is denial of service through use of OAM options. For this reason, RBridges MUST rate limit the generation of OAM option frames. For multi-destination frames, the frames MAY be discarded silently to prevent any DoS attacks in case of an errored packet such as an 'options not recognized' error message.

11. References

11.1. Normative References

- | | |
|-----------------------------------|---|
| [I-D.ietf-isis-layer2] | Banerjee, A. and D. Ward,
"Extensions to IS-IS for Layer-2
Systems",
draft-ietf-isis-layer2-07 (work in
progress), September 2010. |
| [I-D.ietf-isis-trill] | 3rd, D., Banerjee, A., Dutt, D.,
Perlman, R., and A. Ghanwani,
"TRILL Use of IS-IS",
draft-ietf-isis-trill-01 (work in
progress), August 2010. |
| [I-D.ietf-trill-rbridge-options] | 3rd, D., Ghanwani, A., and C.
Bestler, "RBridges: TRILL Header
Options", draft-ietf-trill-
rbridge-options-02 (work in
progress), July 2010. |
| [I-D.ietf-trill-rbridge-protocol] | 3rd, D., Dutt, D., Gai, S.,
Ghanwani, A., and R. Perlman,
"Rbridges: Base Protocol
Specification", draft-ietf-trill-
rbridge-protocol-16 (work in
progress), March 2010. |

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

11.2. Informative References

- [I-D.ietf-trill-rbridge-mib] Rijhsinghani, A. and K. Zebroze, "Definitions of Managed Objects for RBridges", draft-ietf-trill-rbridge-mib-01 (work in progress), September 2010.
- [IEEE.802-1ag] Institute of Electrical and Electronics Engineers, "IEEE Standard for Local and metropolitan area networks / Virtual Bridged Local Area Networks / Connectivity Fault Management", IEEE Standard 802.1Q, December 2007.
- [RFC0792] Postel, J., "Internet Control Message Protocol", STD 5, RFC 792, September 1981.
- [RFC1393] Malkin, G., "Traceroute Using an IP Option", RFC 1393, January 1993.
- [RFC4443] Conta, A., Deering, S., and M. Gupta, "Internet Control Message Protocol (ICMPv6) for the Internet Protocol Version 6 (IPv6) Specification", RFC 4443, March 2006.
- [RFC5226] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.
- [RFC5837] Atlas, A., Bonica, R., Pignataro, C., Shen, N., and JR. Rivers, "Extending ICMP for Interface and Next-Hop Identification", RFC 5837, April 2010.

Authors' Addresses

David Michael Bond
University of New Hampshire InterOperability Laboratory
121 Technology Drive Suite #2
Durham, New Hampshire 03824
US

Phone: +1-603-339-7575
EMail: david.bond@iol.unh.edu
URI: <http://mokon.net>

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

EMail: vishwas@ipinfusion.com

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

Donald Eastlake 3rd
Stellar Switches
Vishwas Manral
IP Infusion
Dave Ward
Juniper
Ayan Banerjee
Cisco
October 17, 2010

Expires: April 16, 2011

R Bridges: OAM and BFD Support for TRILL
<draft-eastlake-trill-rbridge-bfd-00.txt>

Abstract

This document specifies a general channel for sending OAM (Operations, Administration, and Maintenance) messages between R Bridges in a campus through an extension to the TRILL (Transparent Interconnection of Lots of Links) protocol. It further specifies use of this channel for the BFD (Bidirectional Forwarding Detection) protocol.

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.

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."

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

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

Table of Contents

1. Introduction.....	3
1.1 Terminology.....	3
1.2 Additional Acronyms.....	3
1.3 Acknowledgements.....	4
2. The TRILL OAM Message Channel.....	5
2.1 The OAM Message Inner Frame.....	5
2.1.1 Inner Ethernet Header.....	6
2.1.2 TRILL OAM Header.....	7
2.2 The TRILL Header for OAM Messages.....	8
2.3 OAM Message Ethernet Link Header.....	9
2.4 The TRILL OAM-Channel Bit Option.....	9
2.5 Processing TRILL OAM Messages.....	10
2.5.1 Processing the TRILL OAM Channel Header.....	10
2.5.2 Native TRILL-OAM Frames.....	11
3. TRILL BFD.....	12
3.1 Sessions and Initialization.....	12
3.2 TRILL BFD Control Protocol.....	12
3.2.1 One-Hop TRILL BFD Control.....	13
3.2.2 BFD Control Frame Processing.....	13
3.3 TRILL BFD Echo Protocol.....	14
3.3.1 BFD Echo Frame Processing.....	14
4. Management and Operations Considerations.....	16
5. Allocations Considerations.....	17
5.1 IANA Considerations.....	17
5.2 IEEE Registration Authority Considerations.....	18
6. Security Considerations.....	19
6.1 OAM Channel Security Considerations.....	19
6.2 BFD Security Considerations.....	19
7. Normative References.....	21
8. Informative References.....	21

1. Introduction

The TRILL IS-IS Hellos used between RBridges provide a basic neighbor and continuity check for TRILL links [RFCtrill]. However, failure detection by non-receipt of such Hellos is based on the holding time parameter which is typically set to a value over ten seconds and, in any case, has a minimum expressible value of one second.

Many applications, including voice over IP, may wish, with very high probability, to detect interruptions in continuity within a much shorter time period. In some cases physical layer failures can be detected very rapidly but this is not always possible, such as when there is a failure between two devices that are in turn between two RBridges, and there are many subtle failures possible at higher levels. For example, some forms of failure could affect unicast frames while still letting multicast frames through and all TRILL IS-IS frames, including Hellos, are multicast. Thus, a method of frequently testing continuity for the TRILL Data between neighbor RBridges is necessary for some applications.

Such continuity testing is one example of TRILL data plane Operations, Administration, and Maintenance (OAM) requirements. Various of such requirements can be met by a variety of protocols such as the Bidirectional Forwarding Detection (BFD) [RFC5880] [RFC5882] and [Y.1731].

This document specifies, in Section 2, a general channel for sending OAM messages between RBridges in a campus using extensions to the TRILL protocol and further specifies, in Section 3, use of this channel for the BFD protocol. TRILL BFD can be used to provide rapid detection of link continuity failure for TRILL Data frames.

1.1 Terminology

The terminology and acronyms of [RFCtrill] are used 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].

1.2 Additional Acronyms

The following acronyms are used in this document in addition to those defined in [RFCtrill]:

BFD - Bidirectional Forwarding Detection

MH - Multi-Hop

OAM - Operations, Administration, and Maintenance

OV - OAM (Message Channel) Version

SL - Silent

1.3 Acknowledgements

The authors would like to particularly thank David Katz, co-author of [RFC5880] and [RFC5882]. Some of the text in this document was adapted from those RFCs.

2. The TRILL OAM Message Channel

TRILL OAM messages are transmitted as TRILL Data frames. They are primarily identified as OAM messages by their Inner.MacDA and Inner.Ethertype. This Inner Ethertype is followed by a 32-bit TRILL OAM Header used to indicate the OAM protocol of the following OAM protocol specific data. A TRILL Header bit option is provided that may optionally be used to guarantee that frames sent over the TRILL OAM Message Channel cannot accidentally be forwarded to end stations, even by RBridges that are ignorant of the TRILL OAM Message Channel mechanism.

The diagram below shows the overall structure of a TRILL OAM Message Channel frame on a link between two RBridges:

Frame Structure	Section of This Document
+-----+ Outer Link Header	Section 2.3 if Ethernet Link
+-----+ TRILL Header	Section 2.2
+-----+ Inner Ethernet Header	Section 2.1.1
+-----+ TRILL OAM Channel Header	Section 2.1.2
+-----+ OAM Protocol Specific Payload	See specific OAM protocol
+-----+ Link Trailer (FCS if Ethernet)	
+-----+	

The Sections 2.1 and 2.2 below describe the Inner frame and TRILL Header for frames sent in the TRILL OAM Message Channel. As always, the Outer link header is whatever is needed to get a TRILL Data frame from one RBridge to the next, depends on link technology, and can change with each hop for multi-hop OAM messages. Section 2.3 describes the Outer link header for Ethernet. Section 2.4 goes into further detail on the OAM-Channel Bit Option. And Section 2.5 describes some details of TRILL OAM Message processing.

2.1 The OAM Message Inner Frame

The encapsulated Inner frame within A TRILL OAM Message Channel frame is as shown below.

Inner Ethernet Header:

```

+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Special Inner.MacDA                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+
|   Special Inner.MacDA cont.   |   Inner.MacSA   |
+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Inner.MacSA cont.                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+
|   Ethertype = C-Tag (0x8100)   |   Priority, VLAN ID   |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

TRILL OAM Channel Header:

```

+-----+-----+-----+-----+-----+-----+-----+-----+
|   TRILL-OAM Ethertype   |
+-----+-----+-----+-----+-----+-----+-----+-----+
|   Flags   |   OV   |   TRILL OAM Protocol   |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

OAM Protocol Specific Information:

```

+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     OAM Protocol Specific Data                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+
|   ...   |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

2.1.1 Inner Ethernet Header

The special Inner.MacDA is one of two values: OAM-RBridge-MAC if the OAM message is unicast or All-OAM-RBridges if the OAM message is multi-destination (see Section 6).

The Inner.MacSA is selected by the RBridge originating the OAM message. If it is a unicast MAC address, on decapsulation it will be learned as being attached to the ingress RBridge. If that learning is not desired, the Inner.MacSA MAY be set to All-OAM-RBridges. Address learning on decapsulation does not occur if the source MAC has the group bit on.

As with all TRILL encapsulated frames, a VLAN tag MUST be present. Use of a VLAN tag Ethertype other than 0x8100 is beyond the scope of this document. Recommendations for the frame priority are as follows:

- For one-hop known unicast OAM messages critical to network connectivity, such as one-hop BFD for rapid link failure detection in support of TRILL IS-IS, the RECOMMENDED priority is 7.
- For multi-hop known unicast OAM messages, the RECOMMENDED priority is 6.
- For multi-destination OAM messages, it is RECOMMENDED that the priority be no higher than 5.

Multi-destination TRILL OAM messages are VLAN scoped so the Inner.VLAN ID MUST be set to the VLAN of interest. To the extent that distribution tree pruning is in effect, such OAM messages will only reach RBridges advertising that they have appointed forwarder connectivity to that VLAN.

For known unicast OAM messages, if the message is one-hop it is RECOMMENDED that the Inner.VLAN ID be the Designated VLAN on that hop. For multi-hop unicast OAM messages, it is RECOMMENDED that the Inner.VLAN ID be the default VLAN 1.

2.1.2 TRILL OAM Header

After the TRILL OAM Ethertype (see Section 6) is a four-byte quantity with three sub-fields. The first, Flags, provides 16 bits of flags which, except as specified below, MUST be sent as zero, transparently copied by transit RBridges, and ignored on receipt. The next field, OV, gives the OAM Header version and MUST be zero. Lastly, a 12-bit field specified the particular TRILL OAM protocol to which the message applies. See Section 6 for IANA Considerations.

The flag bits are numbered from 0 to 15 as shown below.

```

    0  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|SL|MH|               Available Flags               |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
```

Bit 0, which is the high order bit in network order, is defined as the SL or Silent bit. If it is a one, it suppresses OAM Channel Error messages due to the use of an unknown version or OAM protocol (see Section 2.5.1). Bit 1 is the MH or Multi-Hop bit. It is used to inform the destination OAM protocol that the message was intended to be multi-hop (MH=1) or one-hop (MH=0).

The TRILL OAM Protocol field specifies the OAM protocol that the OAM Channel message relates to. Initial defined values are as listed below. See Section 6 for IANA Considerations.

Protocol	Name - Section of this Document
-----	-----
0x0001	OAM Channel Error - Section 2.5
0x0002	TRILL BFD Control - Section 3.2
0x0003	TRILL BFD Echo - Section 3.3

2.2 The TRILL Header for OAM Messages

After the Outer link header (which for Ethernet ends with the TRILL Ethertype) and before the encapsulated frame, the OAM message's TRILL Header appears as follows:

```

+-----+-----+-----+-----+-----+-----+-----+-----+
|V=0|0 0|M| Op-Len  | Hops=0x3F |
+-----+-----+-----+-----+-----+-----+-----+-----+
|      Egress Nickname      |      Ingress Nickname      |
+-----+-----+-----+-----+-----+-----+-----+-----+

```

The TRILL Header version V, MUST be zero, the M bit is set appropriately as the OAM message is known unicast (M=0) or multi-destination (M=1), and Op-Len is set appropriately for the length of the options area, if any, all as specified in [RFCtrill].

When a TRILL OAM message is originated, the hop count field is always set to the maximum value, 0x3F. For messages sent a known number of hops, particularly one-hop messages or neighbor echo messages, checking the Hops (Hop Count) field provides an additional validity check as discussed in [RFC5082].

The RBridge originating a TRILL OAM message places a nickname that it holds into the ingress nickname field.

There are several cases for the egress nickname field. If the OAM message is multi-destination, then the egress nickname designates the distribution tree to use. If the OAM message is a multi-hop unicast message, then the egress nickname is a nickname of the target RBridge; this includes the special case of an "echo" OAM message where the originator places its own nickname in both the ingress and egress nickname fields. If the OAM message is a one-hop unicast message, there are two possibilities for the egress nickname.

- o The egress nickname can be set to a nickname of the target neighbor RBridge. This will usually work well but there is a small chance that, due to a nickname transient, the frame will actually be delivered to some other RBridge in the campus. Due to this possibility, both here and in the multi-hop unicast case, if a TRILL OAM message is intended for a specific RBridge in the campus topology, it is RECOMMENDED that the OAM protocol specific data include the IS-IS SystemID of the target RBridge for an added check.
- o The special nickname Any-RBridge may be used. This will guarantee decapsulation at the immediate neighbor RBridge regardless of the state of nickname assignments. RBridges supporting the TRILL OAM Channel facility MUST recognize the Any-RBridge special nickname and accept TRILL Data frames having that value in the egress

nickname field as being sent to them as the egress.

2.3 OAM Message Ethernet Link Header

If the link on which a TRILL OAM frame is transmitted between neighbor RBridges is Ethernet, the link header follows the usual rules for a TRILL Data frame over Ethernet [RFCtrill]. In particular, the Outer.MacSA is the MAC address of the port from which the frame is sent. The Outer.MacDA is the MAC address of the next-hop RBridge port for unicast TRILL OAM messages or the All-RBridges multicast address for multi-destination TRILL OAM messages. If an Outer.VLAN tag is present, it must specify the Designated VLAN for that hop and the priority must be the same as in the Inner.VLAN tag.

2.4 The TRILL OAM-Channel Bit Option

A critical ingress-to-egress TRILL Header bit option, OAM-Channel, is specified associated with the TRILL OAM Channel facility. This option is NOT REQUIRED to appear in the TRILL Header in TRILL OAM message frames. It serves two functions, as follows:

- o An RBridge indicates that it supports the TRILL OAM Channel facility by advertising, in the link state database, its support for this bit option.
- o If this bit option is present in a TRILL OAM message frame, it guarantees that, if the inner frame is decapsulated by an RBridge that does not implement the TRILL OAM Channel it will be discarded rather than being locally flooded as a native frame out all ports for which that RBridge is appointed forwarder for the Inner.VLAN. However, if it is certain that all RBridges in the campus implement the TRILL OAM Channel or if the possible local flooding of the inner frame as specified above is acceptable, there is NO REQUIREMENT to include an options area or to set this particular option bit in the TRILL Header options area even if an options area is included.

As with any other critical ingress-to-egress option, if the bit options area is present and this bit option is set, then the summary CItE bit MUST be set at the top of the options area.

2.5 Processing TRILL OAM Messages

TRILL OAM messages are designed to look like and, to the extent practical, be processed as regular TRILL Data frames. On receiving a TRILL OAM frame, the initial tests on the Outer.MacDA, Outer Ethertype, TRILL Header V and Hop Count fields and the RPF check if the frame is multi-destination, are all performed as usual. The forwarding and/or decapsulation decisions are the same as for a regular TRILL Data frame with the exception that a RBridge implementing the TRILL OAM Channel MUST recognize the Any-RBridge egress nickname in unicast TRILL Data frames, decapsulating and not forwarding such frames if they meet other checks.

If the OAM-Channel critical ingress-to-egress bit option is present and the egressing RBridge does not implement the TRILL OAM Channel, the frame is discarded. If other options are present, they may affect processing or cause the frame to be discarded.

On decapsulation, the special Inner.MacDA values of OAM-RBridge-MAC (unicast) and All-OAM-RBridges (multicast) and/or the Inner Ethertype of TRILL-OAM MUST be recognized to trigger processing as a TRILL OAM message. If the decapsulating RBridge does not implement the TRILL OAM Channel, it will treat the frame as a regular TRILL Data frame and locally flood the decapsulated native frame out all ports where it is appointed forwarder for the Inner.VLAN.

2.5.1 Processing the TRILL OAM Channel Header

Knowing that it has a TRILL OAM Channel message, the egress RBridge looks at the OV (OAM Message Header version) and OAM Protocol fields.

If the OV field is non-zero or if the OAM Protocol field is a reserved value or a value unknown to the egress RBridge, the egress RBridge returns an OAM Channel Error frame unless the "SL" (Silent) flag is a one in the OAM message. An OAM Channel Error frame is a multi-hop unicast TRILL OAM Channel message with the ingress nickname set to the nickname of the RBridge detecting the error, and the egress nickname set to the value of the ingress nickname in the OAM message for which the error was detected. For the protocol specific data area, an OAM Channel Message Error frame has at least the first 256 bytes (or less if less are available) of the erroneous decapsulated OAM message starting with the Inner.MacDA. All RBridges implementing the TRILL OAM Message Channel MUST recognize the OAM Message Channel Error protocol value (0x001) and MUST NOT generate an OAM Message Channel Error message in response to a received OAM Message Channel Error frame, even if they always set the "SL" flag is all TRILL OAM messages they send so they would not normally expect to receive an OAM Channel Message Error frame.

If the OV field is zero and the processing RBridge recognizes the OAM Protocol value, it processes the message in accordance with that OAM protocol.

Errors within a recognized OAM Protocol are handled within that protocol and do not produce OAM Message Channel Error frames.

2.5.2 Native TRILL-OAM Frames

A TRILL OAM Message Channel frame MAY be generated, if provided for by the OAM protocol involved, as the result of the receipt by an RBridge of a native frame with the TRILL-OAM Ethertype. Such a native frame must meet the usual VLAN restrictions to be accepted by the ingress RBridge generating the TRILL OAM Message Channel frame. If the native frame's destination MAC address is not one of the special MAC destination addresses All-OAM-RBridges or OAM-RBridge-MAC, it MUST be changed to one of those two addresses before the frame is encapsulated.

The decapsulation and processing of a TRILL OAM Message Channel frame MAY, if provided for by the OAM protocol involved, result in the sending of a native frame with the TRILL-OAM Ethertype out one or more ports of the egress RBridge. The VLAN, and the MAC destination address, of the frame MAY be set to appropriate values before it is transmitted.

3. TRILL BFD

Using the TRILL OAM Message Channel facility, described in Section 2, TRILL supports one-hop and multi-hop BFD Control and neighbor BFD Echo as detailed below. Multi-destination BFD is beyond the scope of this document.

3.1 Sessions and Initialization

Within an RBridge campus, there will be only a single TRILL BFD Control session between two RBridges over a given interface visible to TRILL. This BFD session must be bound to this interface. As such, both sides of a session MUST take the "Active" role (sending initial BFD Control packets with a zero value of Your Discriminator), and any BFD packet from the remote machine with a zero value of Your Discriminator MUST be associated with the session bound to the remote system and interface.

Note that TRILL BFD provides OAM facilities for the TRILL Data plane. This is above whatever protocol is in use on a particular link, such as PPP [TrillPPP]. Link technology specific OAM protocols may be used on a link between neighbor RBridges, for example Continuity Fault Management [802.lag] if the link is Ethernet. But such link layer OAM and coordination between it and TRILL data plan layer OAM, such as TRILL BFD, is beyond the scope of this document.

If lower level mechanisms, such as link aggregation [802.1AX], are in use that present a single logical interface to TRILL IS-IS, only a single TRILL BFD session can be established to any other RBridge over this logical interface. However, link layer OAM could be run separately on each of the components of a link aggregation.

3.2 TRILL BFD Control Protocol

TRILL BFD Control frames are unicast TRILL OAM Message Channel frames as described in Section 2 above supplemented by the specifications below.

As a unicast message, the M bit in the TRILL Header is zero and the Inner.MacDA is OAM-RBridge-MAC. The TRILL OAM Protocol value is 0x002.

The protocol specific data associated with the TRILL BFD Control protocol is as shown below. See [RFC5880] for further information on the fields after the initial SystemIDs.

TRILL BFD Control Protocol Data:

```

+-----+
|                                     Target RBridge SystemID                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Target RBridge SystemID | Orig. RBridge SystemID |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Originating RBridge SystemID                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|Vers | Diag |Sta|P|F|C|A|D|M| Detect Mult | Length |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     My Discriminator                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Your Discriminator                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Desired Min TX Interval                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Required Min RX Interval                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
|                                     Required Min Echo RX Interval                                     |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
Optional Authentication Section:
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+
| Auth Type | Auth Len | Authentication Data... |
+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+-----+

```

3.2.1 One-Hop TRILL BFD Control

One-hop TRILL BFD Control is typically used in support of TRILL IS-IS to rapidly detect link failure. Such TRILL BFD frames SHOULD be sent with priority 7.

For neighbor RBridges RB1 and RB2, each RBridge sends one-hop TRILL BFD Control frames to the other only if TRILL IS-IS has detected bi-directional connectivity and both RBridges indicate support of TRILL BFD is enabled. The BFD Enabled TLV is used to indicate this as specified in [RFCbftlv]. The indication of TRILL BFD support with the BFD Enabled TLV overrides any indication of lack of support through failure to advertise support of the OAM-Channel TRILL Header bit option in the link state database.

3.2.2 BFD Control Frame Processing

The following tests SHOULD be performed on received TRILL BFD Control frames before generic BFD processing. (In some implementations, the TRILL Header may not be available to the TRILL BFD module in which case some of these check are not possible.)

Is the M bit in the TRILL Header non-zero? If so, discard the frame. TRILL support of multi-destination BFD Control is beyond the scope of this document.

If the MH OAM Header flag is zero, indicating one-hop, test that the TRILL Header hop count received was 0x3F (i.e., is 0x3E if it has already been decremented) and if it is any other value discard the frame. If the MH OAM flag is one, indicating multi-hop, test that the TRILL Header hop count received was not less than a configurable value that defaults to 0x30. If it is less, discard the frame.

Check that the target IS-IS SystemID in the OAM protocol data is your SystemID. If not, discard the frame.

If the MH OAM Header flag is zero, test that the originating SystemID is that of a neighbor RBridge. If not, discard the frame.

3.3 TRILL BFD Echo Protocol

A TRILL BFD Echo frame is a unicast TRILL OAM Message Channel frame, as specified in Section 2, which should be bounced back by an immediate neighbor because both the ingress and egress nicknames are set to a nickname of the originating RBridge. Normal TRILL Data frame forwarding will cause the frame to be returned.

TRILL BFD Echo frames SHOULD only be sent on a link if a TRILL BFD Control session has been established, TRILL BFD Echo support is indicated by the potentially echo responding RBridge, and the TRILL BFD Echo originating RBridge wishes to make use of this optional feature.

Since the originating RBridge is the RBridge that will be processing a returned Echo frame, the entire TRILL BFD Echo protocol specific data area is considered opaque and left to the discretion of the originating RBridge. Nevertheless, it is RECOMMENDED that this data include information by which the originating RBridge can authenticate the returned BFD Echo frame and confirm the neighbor that echoed the frame back. For example, it could include its own SystemID, the neighbor's SystemID, a session identifier and a sequence count as well as a Message Authentication Code.

3.3.1 BFD Echo Frame Processing

The following tests SHOULD be performed on returned TRILL BFD Echo frames before other processing. (In some implementations, the TRILL

Header may not be available to the TRILL BFD Echo module in which case these check are not possible.)

Is the M bit in the TRILL Header non-zero? If so, discard the frame. TRILL support of multi-destination BFD Echo is beyond the scope of this document.

The TRILL BFD Echo frame should have gone exactly two hops so test that the TRILL Header hop count as received was 0x3E (i.e., 0x3D if it has already been decremented) and if it is any other value discard the frame. (The value of the MH flag is ignored for TRILL BFD Echo protocol.)

4. Management and Operations Considerations

The TRILL BFD parameters at an RBridge are configurable... The default values are ... TBD.

It is required that the operator of an RBridge campus configure the rates at which TRILL BFD frames are transmitted on a link to avoid congestion (e.g., link, I/O, CPU) and false failure detection.

5. Allocations Considerations

The following subsection give IANA and IEEE Registration Authority Considerations.

5.1 IANA Considerations

In this section, the allocation procedures "Standards Action", "IETF Review", and "RFC Publication" are as specified in [RFC5226].

IANA hereby allocates a previously unassigned TRILL Nickname as follows:

Any-RBridge	TBD (0xFFCO suggested)
-------------	------------------------

IANA hereby allocates a previously unassigned TRILL Multicast address as follows:

All-OAM-RBridges	TBD (01-80-C2-00-00-43 suggested)
------------------	-----------------------------------

IANA hereby allocates a previously unassigned TRILL critical ingress-to-egress Bit Option as follows:

TBD	OAM-Option
-----	------------

IANA allocates the following block of 16 globally unique unicast MAC addresses for use with the TRILL protocol and creates a sub-registry in the TRILL Parameter Registry for these addresses:

00-00-5E-xx-xx-x0	- OAM-RBridge-MAC
00-00-5E-xx-xx-x1 to 00-00-5E-xx-xx-xF	- available for allocation (suggested 00-00-5E-00-03-00 through 00-00-5E-00-03-0F)

Allocation of unicast MAC values from the above block for TRILL use is based on IETF Review.

IANA creates an additional sub-registry in the TRILL Parameter Registry for TRILL OAM Protocols, with initial contents as follows:

Protocol -----	Use ---
0x000	Reserved
0x001	OAM Channel Error
0x002	BFD Control
0x003	BFD Echo
0x004-0x0FF	Available for allocation (1)
0x100-0xFF7	Available for allocation (2)
0xFF8-0xFFE	For Experimental use, will not be allocated
0xFFF	Reserved

(1) TRILL OAM protocol code points from 0x004 to 0x0FF require an IETF Standards Action for allocation.

(2) TRILL OAM protocol code points from 0x100 to 0xFF7 require RFC Publication to allocate a single value or IETF Review to allocate multiple values.

IANA creates an additional sub-registry in the TRILL Parameter Registry for TRILL OAM Header Flags with initial contents as follows:

Flag Bit -----	Mnemonic -----	Allocation -----
0	SL	Silent
1	MH	Multi-hop
2-15	-	Available for allocation

Allocation of TRILL OAM Header Flags is based on IETF Standards Action [RFC5226].

5.2 IEEE Registration Authority Considerations

The Ethertype <tbid> is assigned by the IEEE Registration Authority for TRILL-OAM.

6. Security Considerations

The following sections provide security considerations for the TRILL OAM Message Channel and for TRILL BFD.

See [RFCtrill] for general RBridge Security Considerations.

6.1 OAM Channel Security Considerations

-- TBD --

6.2 BFD Security Considerations

BFD Control frames can be secured by authentication mechanisms native to BFD [RFC5880].

If shared secret IS-IS authentication is not in effect for the Hellos exchanged by two neighbor RBridges then, by default, TRILL BFD between those RBridges is also unsecured.

If shared secret IS-IS authentication is in effect for the Hellos exchanged by two neighbor RBridges then, by default, TRILL BFD Control frames sent between those RBridges use BFD Keyed SHA1 authentication with keying material derived as follows:

```
HMAC-SHA1 ( ( "TRILL BFD Control" | smallerSystemID |
              largerSystemID ), IS-IS-key )
```

where HMAC-SHA1 is as specified in [RFC2104] (see also [RFC4634]), "TRILL BFD Control" is the seventeen byte US ASCII string indicated which is then concatenated with the SystemIDs of both of the neighbor RBridges sorted as unsigned 48-bit integers, and IS-IS-key is the secret keying material being used for IS-IS authentication on the link. In the Authentication Section of the BFD Control frame OAM protocol specific data, Auth Type would be 4, Auth Len would be 28, and Auth Key ID is zero. The RBridges MAY be configured to use other BFD security modes or keying material including configuration to use no security.

Authentication for TRILL BFD Echo SHOULD be provided but is a local implementation issue as BFD Echo frames are only authenticated by their sender when received in the form of Echo responses. However, if TRILL IS-IS and BFD Control are being authenticated to a neighbor and BFD Echo is in use, BFD Echo frames to be returned by that neighbor SHOULD be authenticated and such authenticate SHOULD use different keying material from other types of authentication. For example, it

could use keying material derived as follows:

```
HMAC-SHA1 ( ( "TRILL BFD Echo" | smallerSystemID | largerSystemID
              ), IS-IS-key )
```

7. Normative References

- [RFC2104] - Krawczyk, H., Bellare, M., and R. Canetti, "HMAC: Keyed-Hashing for Message Authentication", RFC 2104, February 1997.
- [RFC2119] - Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, March 1997
- [RFC5226] - Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", BCP 26, RFC 5226, May 2008.
- [RFC5880] - D. Katz, D. Ward, "Bidirectional Forwarding Detection (BFD)", June 2010.
- [RFC5882] - D. Katz, D. Ward, "Generic Application of Bidirectional Forwarding Detection (BFD)", June 2010.
- [RFCtrill] - R. Perlman, D. Eastlake, D. Dutt, S. Gai, and A. Ghanwani, "RBridges: Base Protocol Specification", draft-ietf-trill-rbridge-protocol-16.txt, in RFC Editor queue.
- [RFCbfdtlv] - C. Hopps, L. Ginsberg, "IS-IS BFD Enabled TLV", draft-ietf-isis-bfd-tlv-02.txt, work in progress, 4 January 2010.

8. Informative References

- [802.1AX] - IEEE, "IEEE Standard for Local and metropolitan area networks / Link Aggregation", 802.1AX-2008, 1 January 2008.
- [802.1ag] - IEEE, "IEEE Standard for Local and metropolitan area networks / Virtual Bridged Local Area Networks / Connectivity Fault Management", 802.1ag-2007, 17 December 2007.
- [RFC4634] - Eastlake 3rd, D. and T. Hansen, "US Secure Hash Algorithms (SHA and HMAC-SHA)", RFC 4634, July 2006.
- [RFC5082] - Gill, V., Heasley, J., Meyer, D., Savola, P., Ed., and C. Pignataro, "The Generalized TTL Security Mechanism (GTSM)", RFC 5082, October 2007
- [TrillPPP] - Carlson, J., "PPP TRILL Protocol Control Protocol", draft-ietf-pppext-trill-protocol-01.txt, work in progress, May 2010.
- [Y.1731] - ITU-T Recommendation Y.1731 (02/08), "OAM functions and mechanisms for Ethernet based networks", February 2008

Authors' Addresses

Donald Eastlake 3rd
Stellar Switches
155 Beaver Street
Milford, MA 01757 USA

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

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

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

Dave Ward
Juniper Networks
1194 N. Mathilda Ave.
Sunnyvale, CA 94089-1206
USA

Phone: +1-408-745-2000
EMail: dward@juniper.net

Ayan Banerjee
Cisco Systems
170 W. Tasman Drive
San Jose, CA 95138 USA

Phone: +1-408-525-8781
EmMail: ayabaner@cisco.com

Copyright, Disclaimer, and Additional IPR Provisions

Copyright (c) 2010 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 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.

TRILL Working Group
INTERNET-DRAFT
Intended status: Proposed Standard

Donald Eastlake 3rd
Stellar Switches
Anoop Ghanwani
Brocade
Caitlin Bestler
Quantum
Vishwas Manral
IP Infusion
October 24, 2010

Expires: April 23, 2010

RBridges: TRILL Header Options
<draft-ietf-trill-rbridge-options-03.txt>

Abstract

The TRILL base protocol specification [RFCtrill], specifies minimal hooks for TRILL Header options. This draft specifies the format for options and an initial set of options.

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."

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

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

Table of Contents

1. Introduction.....	3
1.1 Conventions used in this document.....	3
2. TRILL Header Options.....	4
2.1 RBridge Option Handling Requirements.....	5
2.2 No Critical Surprises.....	6
2.3 Options Format.....	6
2.3.1 Summary Bits and Bit Options.....	6
2.3.2 TLV Option Format.....	8
2.3.3 Marshalling of Options.....	9
2.4 Conflict of Options.....	9
3. Specific Bit Option.....	10
3.1 ECN Bit Option.....	10
4. Specific TLV Options.....	12
4.1 Flow ID TLV Option.....	12
4.2 Test/Pad Option.....	13
5. Additions to IS-IS.....	14
6. IANA Considerations.....	15
7. Security Considerations.....	15
8. References.....	16
8.1 Normative References.....	16
8.2 Informative References.....	16
Change History.....	17
Version 00 to 02.....	17
Version 02 to 03.....	17

1. Introduction

The base TRILL protocol specification [RFCtrill] provides a TRILL Header options feature and describes minimal hooks to safely support that feature. But it does not specify the structure of options, their ordering, nor the details of any particular options. This draft specifies that format and some initial options

Section 2 below describes the general principles of operation, format, and ordering of TRILL Header Options. Such options are of two kinds: bit encoded options and TLV (Type, Length, Value) encoded options.

Section 3 describes a specific bit option while Section 4 describes specific TLV encoded options.

1.1 Conventions used in this document

The terminology and acronyms defined in [RFCtrill] are used herein with the same meaning.

In this documents, "IP" refers to both IPv4 and IPv6.

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. TRILL Header Options

The TRILL Protocol includes an option feature in the TRILL Header (see [RFCtrill] Sections 3.5 and 3.8). The 5-bit Op-Length header field gives the length of the 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; else, the options area follow immediately after the Ingress Rbridge Nickname field in the TRILL Header. The options area consists of bit options possibly followed by TLV options. Each TLV option present is 32-bit aligned.

As described below, provision is made for both hop-by-hop options, which might affect any RBridge that receives a TRILL frame containing such an option, and ingress-to-egress options, which would only necessarily affect the RBridge(s) where a TRILL frame is decapsulated. Provision is also made for both "critical" and "non-critical" options. An RBridge receiving a frame with a critical option that affects it and that it does not implement MUST discard the frame as it is unsafe to process the frame without understanding the critical option. Non-critical options can be safely ignored.

Any option indicating a significant change in the way later parts of the frame are interpreted or structure MUST be a critical option. If such an option affect any fields that transit RBridges will examine, it MUST be a hop-by-hop critical option.

Options also have a "mutability" flag that has a different meaning for ingress-to-egress options and for hop-by-hop options.

For an ingress-to-egress option, the mutability flag indicates whether the value associated with the option can change at a transit RBridge (mutable options) or cannot so change (immutable options). For example, an ingress-to-egress security option could protect the value of an immutable ingress-to-egress option. But such a security option generally could not protect a mutable value as a transit RBridge could change that value but would not normally have the keys to recompute a signature or authentication code to take a changed value into account.

For a non-critical hop-by-hop option, the mutability flag indicates whether a transit RBridge that does not implement the option is permitted (mutable) or not permitted (immutable) to remove the option. A transit RBridge is never required to remove a hop-by-hop options that it does not implement.

For critical hop-by-hop options, the mutability flag is meaningless. If the RBridge does not implement the critical hop-by-hop option, it MUST drop the frame. If it does implement the critical hop-by-hop option, it will know whether or not it may/should/must remove it. For critical hop-by-hop options, the mutability flag is set to zero

("immutable") on transmission and ignored on receipt.

Note: Most RBridges implementations are expected to be optimized for simple and common cases of frame forwarding and processing. Although the hard limit on options length, their 32-bit alignment, and the presence of critical option summary bits as described below, are intended to assist in the efficient processing of frames with options, nevertheless the inclusion of options may cause frame processing using a "slow path" with inferior performance to "fast path" processing. Limited slow path throughput of such frames could cause them to be discarded.

2.1 RBridge Option Handling Requirements

The requirements given in this section are in addition to all option handling requirements in [RFCtrill].

All Rbridges MUST be able to detect whether there are any critical options present that are necessarily applicable to their processing of the frame as detailed below. If they do not implement all such critical options present, they MUST discard the frame.

Transit RBridges MUST transparently forward all immutable ingress-to-egress options in frames that they forward. Any changes made by a transit RBridge to a mutable ingress-to-egress option value MUST be a change permitted by the specification of that option.

In addition, a transit RBridge:

- o MAY add, if space is available, or remove, hop-by-hop options as specified for such hop-by-hop options;
- o MAY change the value and/or length of a mutable ingress-to-egress option as permitted by that option's specification and provided there is enough room if lengthening the option;
- o MUST adjust the length of the options area, including changing Op-Length in the TRILL header, as appropriate for any changes it has made in the options;
- o MUST NOT add, remove, or re-order ingress-to-egress options.
- o with regard to any non-critical hop-by-hop options that the transit RBridge does not implement, it MAY remove them if they are mutable but MUST transparently copy them when forwarding a frame if they are immutable.

2.2 No Critical Surprises

R Bridges advertise the ingress-to-egress options that they support in their IS-IS LSP and advertise the hop-by-hop options they support in the Hellos they send. An R Bridge is not required to support any options.

Unless an R Bridge advertises support for a critical option, it will not normally receive frames with that option.

An R Bridge SHOULD NOT add a critical option to a frame unless,

- for a critical hop-by-hop option, it has determined that the next hop R Bridge or R Bridges to which the frame will be sent support that option, or
- for a critical ingress-to-egress option, it has determined that the R Bridge or R Bridges that will egress the frame support that option.

"SHOULD NOT" is specified since there may be cases where it is acceptable for those frames to be discarded by the egress R Bridges that do not implement the option.

2.3 Options Format

If any options are present in a TRILL Header, as indicated by a non-zero Op-Length field, the first 32 bits of the options area consist of two summary bits and 30 option bits as described below. The remainder of the options area, if any, consists of TLV (Type Length Value) encoded options aligned on 32-bit boundaries. Section 2.3.2 specifies the format of an individual TLV option. Section 2.3.3 describes the marshalling of TLV options.

2.3.1 Summary Bits and Bit Options

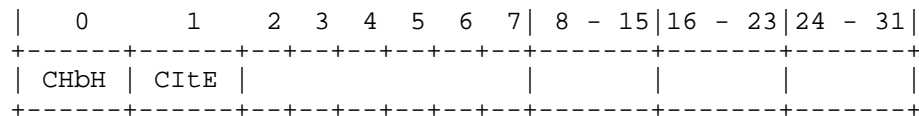


Figure 1: Options Area Initial 32 Bits

The top two bits of the options area, bits 0 and 1 above, are called summary bits and summarize the presence of critical options as follows:

If the CHbH (Critical Hop by Hop) bit is one, one or more critical

hop-by-hop options are present in the options area. Transit R Bridges that do not support all of the critical hop-by-hop options present, for example an R Bridge 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 R Bridge to forward, regardless of what options that R Bridge does or does not support. A transit R Bridge that supports none of the options present MUST transparently forward the options area when it forwards a frame, except that it MAY remove mutable hop-by-hop options.

If the CItE (Critical Ingress to Egress) bit is a one, one or more critical ingress-to-egress options are present in the options area. If it is zero, no such options are present. If either CHbH or CItE is non-zero, egress R Bridges that don't support all critical options present, for example an R Bridge 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 R Bridge to process, regardless of what options that R Bridge does or does not support.

The remaining 30 bits in the initial four octets of the options area are available for bit-encoded options. Any R Bridge adding an options area to a TRILL Header must set these 30 bits to zero except when permitted to set one or more of these bits as specified for an option that R Bridge implements. The 30 bits are categorized as follows:

Bits	Category

2- 7	Critical hop-by-hop bits
8-15	Non-critical hop-by-hop bits
16-23	Critical ingress-to-egress bits
24-31	Non-critical ingress-to-egress bits

All bit encoded options are considered mutable except the critical hop-by-hop options. Any transit R Bridge MUST transparently copy bits 16 through 31, except as permitted by an option implemented by that R Bridge, but MAY either copy or clear any of the bits from 8 through 15. Even if a transit R Bridge removes all TLV options from a TRILL Header when allowed to do so, it MUST NOT eliminate the options area in a forwarded frame if any of the 2 through 7 or 16 through 31 bits remain non-zero; however, if there are no TLV options and all of bits 2 through 31 are zero, then the summary bits will also be zero and the transit R Bridge may eliminate the Options area in the frame, setting Op-Length to zero.

2.3.2 TLV Option Format

TRILL Header options, other than bit options described above, are TLV encoded, with some flag bits in the Type and Length octets, in the format show in Figure 2.

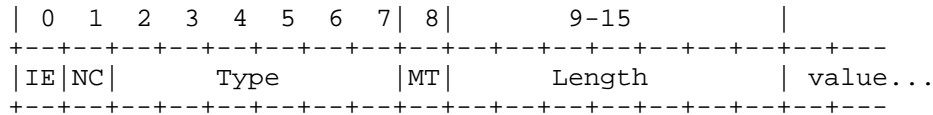


Figure 2. Option TLV Structure

The highest order bit of the first octet (IE) is zero for hop-by-hop options and one for ingress-to-egress options. Hop-by-hop options are potentially applicable to every RBridge that receives the frame. Ingress-to-egress options are only inserted at the ingress RBridge and are applicable at egress RBridges. Ingress-to-egress options MAY also be examined and acted upon by transit RBridges as specified in the particular option.

The second highest order bit of the first octet (NC) is zero for critical options and one for non-critical options.

The highest order bit of the second octet (MT) is zero for immutable options and one for mutable options. The IE, NC, Type, and MT fields themselves MUST NOT be changed even for a mutable option.

The bottom six bits of the first octet give the option Type code. The option Type may constrain the values of the IE, NC, and MT bits. For example, if the Type indicates a Flow ID option (see Section 4.1), then it MUST be marked as a hop-by-hop, non-critical, mutable option. If the IE, NC, or MT bits have a value not permitted by the option Type specification for an option that an RBridge must act on (any critical ingress-to-egress option at an egress RBridge and any critical hop-by-hop option), the RBridge MUST discard the frame. If these bits have a value not permitted by for the Type for an option that an RBridge may ignore (any ingress-to-egress option at a transit RBridge and any non-critical option), the RBridge MAY discard the frame. "MAY" is chosen in this case to minimize the checking burden.

The Length field is an unsigned quantity giving the length of the option value in octets. It gives the amount of option value data, if any, beyond the initial two Type and Length octets. The Length field MUST NOT be such that the option value extends beyond the end of the total options area as specified by the TRILL Header Op-Length. Thus, the value of Length can vary from zero to 118. The meaning of "Length" values of 119 through 127 is reserved and, when such values are noticed in a frame, the frame MUST be discarded.

2.3.3 Marshalling of Options

In a TRILL Header with options, those options start immediately after the Ingress RBridge Nickname and completely fill the options area.

TLV options start immediately after the initial four octets of option and summary bits and MUST appear in ascending order by the value of the nine high order bits of the Type and Length octets considered as an unsigned integer in network byte order. There MUST NOT be more than one option in a frame with any particular value of this nine high order bits. Thus the TLV options MUST be ordered as follows: (1) critical hop-by-hop options, (2) non-critical hop-by-hop options, (3) critical ingress-to-egress options, and (4) non-critical ingress-to-egress options. Frames that violate this paragraph are erroneous, will produce unspecified results, and MAY be discarded. "MAY" is chosen to minimize the format-checking burden on transit RBridges.

Options are 32-bit aligned. Should an option not consist of a multiple of four octets, the option is padded at the end up to the next multiple of four octets. These padding octets MUST be sent as zero and ignored on receipt.

If any options are present, those options, both flag and TLV, MUST be correctly summarized into the CHbH and CItE bits at the top of the initial four octets of the options area.

2.4 Conflict of Options

It is possible for options to conflict. Two or more options can be present in a frame that direct an RBridge processing the frame to do conflicting things or to change its interpretation of later parts of the frame in conflicting ways. Such conflicts are resolved by applying the following rules in the order given:

1. Any frame containing options that require mutually incompatible changes in way later parts of the frame are interpreted or structured MUST be discarded. (Such options will be critical options, normally hop-by-hop critical options.)
2. Critical options override non-critical options.
2. Within each of the two categories of critical and non-critical options, the option appearing first in lexical order in the frame always overrides an option appearing later in the frame. Thus a conflict between a bit option and a TLV option is always resolved in favor of the bit option. Bit options with lower bit numbers are considered to have occurred before bit options with higher bit numbers.

3. Specific Bit Option

The table below shows the state of TRILL Header bit option assignments. See Section 6 for IANA Considerations.

Bit	Purpose	Section

0-1	Summary	2.3
2-7	available for critical hop-by-hop options	
8-9	ECN	3.1
10-15	available for non-critical hop-by-hop options	
16-23	available for critical ingress-to-egress options	
24-31	available for non-critical ingress-to-egress options	

Table 1. Flag Options

3.1 ECN Bit Option

R Bridges may implement an ECN (Explicit Congestion Notification) option [RFC3168]. If implemented, it SHOULD be enabled by default but can be disabled on a per R Bridge basis by configuration.

R Bridges that do not implement this option or on which it is disabled simply (1) set bits 8 and 9 of the bit options area zero when they add an options area to a TRILL Header and (2) transparently copy those bits, if an options area is present, when they forward a frame with a TRILL Header.

An R Bridge that implements the ECN option does the following when that option is enabled:

- o When ingressing an IP frame that is ECN enabled, it MUST add an options area to the TRILL Header and copy the two ECN bits from the IP header into option bits 8 and 9.
- o When ingressing a frame for a non-IP protocol with a means of indicating ECN that is understood by the R Bridge, it MAY add an options area to the TRILL Header with the ECN bits set from the ingressed frame.
- o When forwarding a frame encountering congestion at an R Bridge, if an options area is present with option bits 8 and 9 indicating ECN-capable transport, the R Bridge MUST modify them to the congestion experienced value.
- o When egressing an IP frame, if the TRILL Header has an options area with option bits 8 and 9 non-zero, it copies those bits into the ECN bits in the IP header.
- o When egressing a non-IP protocol frame with a means of indicating ECN that is understood by the R Bridge, it MAY transfer the ECN information from the ECN bits in the options area to the egressed

native frame.

The following table is modified from [RFC3168] and shows the meaning of bit values in TRILL Header option bits 8 and 9, bits 6 and 7 in the IPv4 TOS Byte, and bits 6 and 7 in the IPv6 Traffic Class Octet:

Binary	Meaning
-----	-----
00	Not-ECT (Not ECN-Capable Transport)
01	ECT(1) (ECN-Capable Transport(1))
10	ECT(0) (ECN-Capable Transport(0))
11	CE (Congestion Experienced)

Table 2. ECN Bit Combinations

An RBridge detects congestion either by monitoring its own queue depths or from participation in a link-specific protocol. An RBridge implementing the ECN option MAY be configured to add congestion experienced marking using ECN to any frame with a TRILL Header that encounters congestion even if the frame was not previously marked as ECN-capable or did not have an options area.

4. Specific TLV Options

The table below shows the state of TRILL Header TLV option Type assignment. See Section 6 for IANA Considerations.

Type	Purpose	Section

0x00	reserved	
0x01	Flow ID	4.1
0x02-0x1F	available	
0x20	Test/Pad	4.2
0x21-0x3E	available	
0x3F	reserved	

Table 3. TLV Option Types

The following subsections specify particular TRILL TLV options.

4.1 Flow ID TLV Option

In connection with multi-pathing of frames, frames that are part of the same order dependent flow need to follow the same path for correct operation. Methods to determine flows are beyond the scope of the this document; however, it may be useful, once the flow of a frame has been determined, to preserve and transmit that information for use by subsequent RBridges.

This is a non-critical option. It is considered hop-by-hop because it can be added or changed by a transit RBridge and transit RBridges may wish to use it to make forwarding decisions. Because the ingress RBridge may know the most about a frame, it is expected that this option would most commonly be added at the ingress RBridge. Once in a frame, the option SHOULD NOT be removed or changed unless, for example, a campus is divided into regions such that different Flow IDs would make sense in different regions.

The value length of this option is fixed at 2 for efficiency. In a TRILL data frame with only this option, the size of the option plus the size of the initial 4 summary and flag option octets is such as to maintain 64-bit alignment of the encapsulated frame.

The option fields and flags are as follows:

- o Type is 0x01.
- o Length is 2. The data is an unsigned integer that is the Flow ID.
- o IE MUST be zero. This is a hop-by-hop option.

- o NC and MT MUST be one. This is a non-critical mutable option.

4.2 Test/Pad Option

This option is intended for testing and padding.

A specific meaning for this option with the critical flag set will not be defined so, in that form, it MUST always be treated as an unknown critical option. If the critical flag is not set, the option does nothing. In either case, it may be any length that will fit. Thus, for example, in the non-critical form, it can be used to cause the encapsulated frame starting right after the options area to be 64-bit aligned or for testing purposes.

- o Type is 0x20.
- o Length is variable. The value is ignored.
- o IE may be zero or one. This option has both hop-by-hop and ingress-to-egress versions.
- o NC is zero for the pad option and one for the test option.
 - + The non-critical version of this option does nothing.
 - + The critical version of this option MUST always be treated as an unknown critical option.
- o MT may be zero or one except that it must be zero if the other flags indicate the options is a critical hop-by-hop option. This option may be flagged as mutable or immutable.

5. Additions to IS-IS

RBridges use IS-IS PDUs to inform other RBridges which options they support. The specific IS-IS TLVs or sub-TLVs used to encode and advertise this information are specified in a separate document. Support for critical options MUST be advertised. Support for non-critical options MAY be advertised unless the specification of a particular non-critical option imposes a requirement higher than "MAY" for the advertising of that option by RBridges that implement it.

Rbridges indicate in their link state which ingress-to-egress TLV and bit options they support.

Rbridges indicate in their Hellos which hop-by-hop TLV and bit options they support.

6. IANA Considerations

IANA will create two subregistries within the TRILL registry. A "TRILL Header Bit Options" subregistry that is initially populated as specified in Table 1 in Section 3. And a "TRILL TLV Option Types" subregistry that is initially populated as specified in Table 3 in Section 4. References in both of those tables to sections of this document are to be replaced in the IANA subregistries by references to this document as an RFC.

New TRILL bit options and TLV option types are allocated by IETF Review [RFC5226].

7. Security Considerations

For general TRILL protocol security considerations, see [RFCtrill].

In order to facilitate authentication, options SHOULD be specified so they do not have alternative equivalent forms. Authentication of anything with alternative equivalent forms almost always requires canonicalization that an authenticating RBridge ignorant of the option would be unable to do and that may be complex and error prone even for an RBridge knowledgeable of the option. It is best for any option to have a unique encoding.

8. References

Normative and informative references for this document are given below.

8.1 Normative References

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

[RFC3168] - Ramakrishnan, K., Floyd, S., and D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", RFC 3168, September 2001.

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

[RFCtrill] - Perlman, R., D. Eastlake, D. Dutt, S. Gai, and A. Ghanwani, "RBrigdes: Base Protocol Specification", draft-ietf-trill-rbridge-protocol-16.txt, in RFC Editor's queue.

8.2 Informative References

None.

Change History

The sections below summarize changes between successive versions of this draft. RFC Editor: Please delete this section before publication.

Version 00 to 02

Change the requirement for TLV option ordering to be strictly ordered by the value of the top nine bits of their first two bytes so that the MT bit is included.

Specify meaning of mutability bit for hop-by-hop options.

Fix length of Flow ID Value at 2.

Require that options that may significantly affect the interpretation or format of subsequent parts of the frame be critical options.

Version 02 to 03

Move Test/Pad option into this document from the More Options draft and move the More Flags option from this document into the More Options draft.

Prohibit multiple occurrences of an option in a frame.

Authors' Addresses

Donald E. Eastlake 3rd
Stellar Switches
155 Beaver Street
Milford, MA 01757

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

Anoop Ghanwani
Brocade Communications Systems
1745 Technology Drive
San Jose, CA 95110 USA

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

Caitlin Bestler
Quantum
1650 Technology Drive , Suite 700
San Jose, CA 95110

Phone: +1-408-944-4000
email: cait@asomi.com

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

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

Copyright and IPR Provisions

Copyright (c) 2010 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 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.

