TRILL Working Group D. Bond
Internet-Draft UNH-IOL
Intended status: Standards Track V. Manral
Expires: April 21, 2011 IP Infusion Inc.
October 18, 2010

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

Abstract

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. As RBridges are deployed in real-world situations, operators will need tools for debugging problems that arise. This document specifies a set of RBridge features for operations, administration, and maintenance purposes in RBridge campuses. The features specified in this document include tools for traceroute, ping, and error reporting.

Status of This Memo

Copyright Notice

Table of Contents

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

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

- TLV - Type, Length, Value
- TRILL - TRansparent Interconnection of Lots of Links
- 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 critial 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:

```
+-----------------------------+
|   Outer Ethernet Header     |
+-----------------------------+
|        TRILL Header         |
+-----------------------------+
|      TRILL OAM Option       |
+-----------------------------+
|Dummy Inner Ethernet Header |
+-----------------------------+
|   Dummy Ethernet Payload   |
+-----------------------------+
```

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:

```
+--------------------------------------+
|         Outer Ethernet Header        |
+--------------------------------------+
|              TRILL Header            |
+--------------------------------------+
|            TRILL OAM Option          |
+--------------------------------------+
| Offending Frame Outer Ethernet Header |
+--------------------------------------+
|      Offending Frame TRILL Header    |
+--------------------------------------+
| Offending Frame Inner Ethernet Header |
+--------------------------------------+
|    Offending Frame Ethernet Payload  |
+--------------------------------------+
```

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.

```
            +-----+  +-------+   +-------+   +-------+  +-----+
            | ESa +--+  RB0  +---+  RB1  +---+  RB2  +--+ ESb |
            +-----+  |ingress|   |transit|   |egress |  +-----+
                     +-------+   +-------+   +-------+

             Time       RB0         RB1         RB2
              .         (1)-------> (1) -------> |
              .          | <------- (2)          |
              .          | <------- (3) <-------(3)
```

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 then 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-orginating 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.

```
            +-----+  +-------+   +-------+   +-------+  +-----+
            | ESa +--+  RB0  +---+  RB1  +---+  RB2  +--+ ESb |
            +-----+  |ingress|   |transit|   |egress |  +-----+
                     +-------+   +-------+   +-------+

             Time       RB0         RB1         RB2
              .         (1)-------> |           |
              .          | <------- (2)         |
              .         (3)-------> (3) -------> |
              .          | <------- (4) <-------(4)
```

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.

```
+-----+  +-------+   +-------+   +-------+  +-----+
| ESa +--+  RB0  +---+  RB1  +---+  RB2  +--+ ESb |
+-----+  |ingress|   |transit|   |egress |  +-----+
         +-------+   +-------+   +-------+

 Time       RB0         RB1         RB2
  .        (1)-------> (1) -------> |
  .         | <------- (2) <-------(2)
```

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.

```
| 0  1  2  3  4  5  6  7| 8| 9|     10-15       |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|IE|NC|      Type       |MT|       Length       |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|          Code         |        Subcode        |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|                    Sequence                   |
|                     Number                    |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
```

TRILL OAM Option Common Initial Part

Figure 6

The option fields and flags are as follows:

- Type: 0x02.

- Length: The length of the option value in octets.

- IE: MUST be one. This is an ingress to egress option.

- NC: Varies depending on the code.

- MT: MUST be zero. This is an immutable option.

- 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

- 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.

- 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:

- 0: Echo Request, See Section 5.2.1.1

- 1: Route Respond Request, See Section 5.2.1.2

- 2: Echo Reply, See Section 5.2.1.3

- 3-122: Available for Allocation by IETF Review

- 123-126: Reserved for Private Experimentation

- 127: Application Expansion Value, See Section 5.2.3

- 128: Hop Count Zero Error, See Section 5.2.2.1

- 129: Generic Error, See Section 5.2.2.3

- 130: MTU Error, See Section 5.2.2.2

- 131-250: Available for Allocation by IETF Review

- 251-254: Reserved for Private Experimentation

- 255: Error Expansion Value, See Section 5.2.3

5.2. Codes

5.2.1. Application Codes

5.2.1.1. Echo Request

```
| 0  1  2  3  4  5  6  7| 8| 9|     10-15       |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|IE|NC|      Type       |MT|       Length       |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|         Code          |        Subcode        |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|                    Sequence                   |
|                     Number                    |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
```

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

- Length: 6
- IE: MUST be one. This is an ingress to egress option.
- NC: Defaults to zero. The OAM option is normally a critical ingress-to-egress option but it MAY be a non-critial option if the encapsulated frame is real data that needs to be processed normally on egress.
- MT: MUST be zero. This is an immutable option.
- Code: MUST be 0.
- Subcode: MUST be 0x00. This field is not used by this sub-option. It is set to zero on transmission and ignored on reception.
- 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

```
   | 0  1  2  3  4  5  6  7| 8| 9|     10-15       |
   +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
   |IE|NC|      Type       |MT|       Length       |
   +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
   |         Code          |        Subcode        |
   +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
   |                    Sequence                   |
   |                     Number                    |
   +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
```

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.

- Length: 6

- IE: MUST be one. This is an ingress to egress option.

- NC: Defaults to zero. The OAM option is normally a critical ingress-to-egress option but it MAY be a non-critial option if the encapsulated frame is real data that needs to be processed normally on egress.

- MT: MUST be zero. This is an immutable option.

- Code: MUST be 1.

- Subcode: MUST be 0x00. This field is not used by this sub-option. It is set to zero on transmission and ignored on reception.

- 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

```
             | 0  1  2  3  4  5  6  7| 8| 9|     10-15       |
             +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
             |IE|NC|      Type       |MT|       Length       |
             +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
             |         Code          |Reserved| I. Hop Count |
             +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
             |                   Sequence                    |
             |                    Number                     |
             +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
             .                                               .
             .                     TLVs                      .
             .                                               .
             +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
```

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.

- Length: 14

- IE: MUST an one. This is an ingress to egress option.

- NC: MUST be zero. This is a critical option

- MT: MUST be zero. This is an immutable option.

- Code: MUST be 2.

- Reserved: A reserved field. Set to zero on transmission and ignored on reception.

- 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.

- 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

```
| 0  1  2  3  4  5  6  7| 8| 9|     10-15       |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|IE|NC|      Type       |MT|       Length       |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|         Code          |        Subcode        |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|                   Sequence                    |
|                    Number                     |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
.                                               .
.                     TLVs                      .
.                                               .
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
```

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.

- 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

```
| 0  1  2  3  4  5  6  7| 8| 9|     10-15       |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|IE|NC|      Type       |MT|       Length       |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|         Code          |        Subcode        |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|                    Sequence                   |
|                     Number                    |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
.                                               .
.                     TLVs                      .
.                                               .
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
```

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.

- Length: 10

- IE: MUST be one. This is an ingress to egress option.

- NC: MUST be zero. This is a critical option.

- MT: MUST be zero. This is an immutable option.

- Code: MUST be 130.

- 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

```
             | 0  1  2  3  4  5  6  7| 8| 9|     10-15       |
             +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
             |IE|NC|      Type       |MT|       Length       |
             +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
             |         Code          |         Subcode       |
             +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
             |                   Sequence                    |
             |                    Number                     |
             +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
```

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 then 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

- 0: Unknown Error: Indicates and an error has occurred.

- 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.

- 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.

- 3: Illegal Outer.VLAN: Indicates the Outer.VLAN ID is 0xFFF.

- 4: Invalid Outer.VLAN: Indicates the Outer.VLAN ID was not the designated VLAN ID.

- 5: Unknown TRILL Version: Indicates the TRILL Version is unknown.

- 6: Op-Length Exceeds Frame Length: Indicates the Op-Length says the options field extends beyond the end of the received frame length.

- 8: Unknown Egress RBridge: Indicates the Egress RBridge in a received frame is unknown.

- 9: Unknown Ingress RBridge: Indicates the Ingress RBridge in a received frame is unknown.

- 10: Unsupported Critical Hop-by-hop Option: Indicates an unsupported critical hop-by-hop option was received.

- 11: Unsupported Critical Ingress-to-Egress Option: Indicates an unsupported critical ingress-to-egress option was received.

- 12-84: Available for allocated by IETF Review

- 85: Reserved for Private Experimentation

Warning Sub-codes

- 86: Illegal Inner.VLAN: Indicates the Inner.VLAN ID is 0xFFF.

- 87: Inner/Outer VLAN Priority Mismatch: Indicates the priority values in the inner and outer VLANs do not match.

- 88: P2P Hello on TRILL Hello Link: Indicates a P2P Hello was received on a TRILL Hello Link.

- 89: TRILL Hello on P2P Hello Link: Indicates a TRILL Hello was received on a P2P Hello Link.

- 90: No Adjacency: Indicates a TRILL data frame was sent from an RBridge the receiving RBridge is not adjacent with.

- 91: Encapsulated BPDU/VRP Frame: A TRILL Frame containing a BPDU or VRP frame was received.

- 92: Invalid Mutability Flag: Indicates the mutability flag was set on a received CHbH Option.

- 93: Invalid TLV Option Length: Indicates the option length field of a TLV option was between 121 and 127.

- 94: Options Ordering Error: Indicates the TLV options are ordered incorrectly.

- 95: Additional Flag TLV Zero: Indicates a problem in the additional Flag TLV.

- 96: Configured Nickname Collision: Indicates an RBridge was detected in the campus with the same nickname (Configured or not).

- 97: Multiple DRBs detected.

- 98: Multiple appointed forwarders detected.

- 99-169: Available for allocation by IETF Review

- 170: Reserved for Private Experimentation

Comment Sub-codes

- 171: Inner.VLAN C-Bit Set: Indicates the C-Bit in the Inner.VLAN is set.

- 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.

- 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.

- 174: Outer.VLAN C-Bit Set: Indicates the C-Bit in the Outer.VLAN is set for an Ethernet frame.

- 175: Invalid Reserved Bits: Indicates the reserved bits are non-zero in a received frame.

- 176: Invalid Nickname: Indicates a nickname in the reserved space of 0xFFC0 to 0xFFFF was received that is not implemented at the receiving RBridge.

- 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.

- 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.

- 179: Performance Exceeded: Indicates a frame was discarded due to performance problems such as a buffer overflow.

- 180: Insufficient Hop Count: Indicates a frame was received with a hop-count that was insufficient to reach the destination.

- 181-254: Available for allocation by IETF Review

- 255: Reserved for Private Experimentation

5.2.3. Expansion Code

```
| 0  1  2  3  4  5  6  7| 8| 9|     10-15       |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|IE|NC|      Type       |MT|       Length       |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|         Code          |         Subcode       |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|                   Additional                  |
|                 Option Values                 |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
```

Expansion Code Format

Figure 13

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

- Length: The length of the option value in octets.

- IE: MUST be one. This is an ingress-to-egress option.

- NC: Varies depending on the code.

- MT: MUST be zero. This is an immutable option.

- Code: MUST BE 127 or 255.

- 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.

- 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.

```
| 0  1  2  3  4  5  6  7| 8| 9|     10-15       |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|         Type          |         Length        |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
.                                               .
.                     Value                     .
.                                               .
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
```

TLV Format

Figure 14

The type values are:

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

5.3.1. TLV Types

5.3.1.1. Padding

```
             | 0  1  2  3  4  5  6  7| 8| 9|     10-15       |
             +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
             |     Type = 0x00       |          Length       |
             +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
```

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

```
             | 0  1  2  3  4  5  6  7| 8| 9|     10-15       |
             +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
             |     Type = 0x01       |      Length = 0x02    |
             +--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
             |              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

```
| 0  1  2  3  4  5  6  7| 8| 9|     10-15       |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|     Type = 0x02       |      Length = 0x02    |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|               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

```
| 0  1  2  3  4  5  6  7| 8| 9|     10-15       |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|     Type = 0x03       |      Length = 0x02    |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|               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

```
| 0  1  2  3  4  5  6  7| 8| 9|     10-15       |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|     Type = 0x04       |      Length = 0x02    |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|               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.

```
+--------------------------------+
|     Outer Ethernet Header      |
+--------------------------------+
|          TRILL Header          |
+--------------------------------+
|     Inner Ethernet Header      |
+--------------------------------+
|  TRILL or TRILL OAM Ethertype  |
+--------------------------------+
|         TRIL OAM Data          |
+--------------------------------+
```

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 atacks 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. Zebrose, "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 Stadard for Local and metropolitian 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