TRILL Working Group
INTERNET-DRAFT
Intended status: Proposed Standard

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

Expires: April 23, 2010

October 24, 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/1id-abstracts.html

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

Table of Contents

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:

- MAY add, if space is available, or remove, hop-by-hop options as specified for such hop-by-hop options;
- 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;
- 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;
- MUST NOT add, remove, or re-order ingress-to-egress options.
- 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

RBridges 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 RBridge is not required to support any options.

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

An RBridge SHOULD NOT add a critical option to a frame unless,
- for a critical hop-by-hop option, it has determined that the next hop RBridge or RBridges to which the frame will be sent support that option, or
- for a critical ingress-to-egress option, it has determined that the RBridge or RBridges 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 RBridges 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

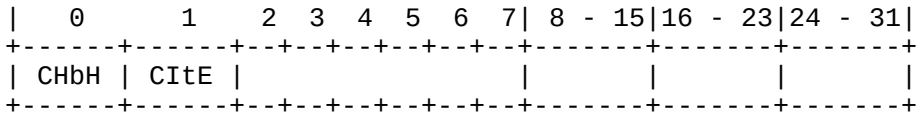

Figure 1: Options Area Initial 32 Bits

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 RBridges that do not support all of the critical hop-by-hop options present, for example an RBridge that supported no options, MUST drop the frame. If the CHbH bit is zero, the frame is safe, from the point of view of options processing, for a transit RBridge to forward, regardless of what options that RBridge does or does not support. A transit RBridge that supports none of the options present MUST transparently forward the options area when it forwards a frame, 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 RBridges that don't support all critical options present, for example an RBridge that supports no options, MUST drop the frame. If both CHbH and CItE are zero, the frame is safe, from the point of view of options, for any egress RBridge to process, regardless of what options that RBridge does or does not support.

The remaining 30 bits in the initial four octets of the options area are available for bit-encoded options. Any RBridge 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 RBridge 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 RBridge MUST transparently copy bits 16 through 31, except as permitted by an option implemented by that RBridge, but MAY either copy or clear any of the bits from 8 through 15. Even if a transit RBridge 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 RBridge 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.

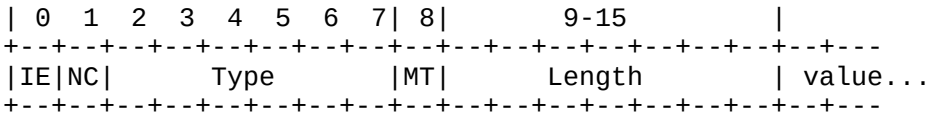

Figure 2. Option TLV Structure

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

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

RBridges 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 RBridge that implements the ECN option does the following when that option is enabled:

- 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.
- When ingressing a frame for a non-IP protocol with a means of indicating ECN that is understood by the RBridge, it MAY add an options area to the TRILL Header with the ECN bits set from the ingressed frame.
- When forwarding a frame encountering congestion at an RBridge, if an options area is present with option bits 8 and 9 indicating ECN-capable transport, the RBridge MUST modify them to the congestion experienced value.
- 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.
- When egressing a non-IP protocol frame with a means of indicating ECN that is understood by the RBridge, 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:

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

- 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 staring right after the options area to be 64-bit aligned or for testing purposes.

- Type is 0x20.
- Length is variable. The value is ignored.
- IE may be zero or one. This option has both hop-by-hop and ingress-to-egress versions.
- 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.
- 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, "RBridges: 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