TRILL Working Group
INTERNET-DRAFT
Intended status: Proposed Standard

Expires: September 20, 2016

Donald Eastlake
Huawei
Bob Briscoe
Simula Research Lab
March 21, 2016

TRILL: ECN (Explicit Congestion Notification) Support <draft-eastlake-trill-ecn-support-00.txt>

Abstract

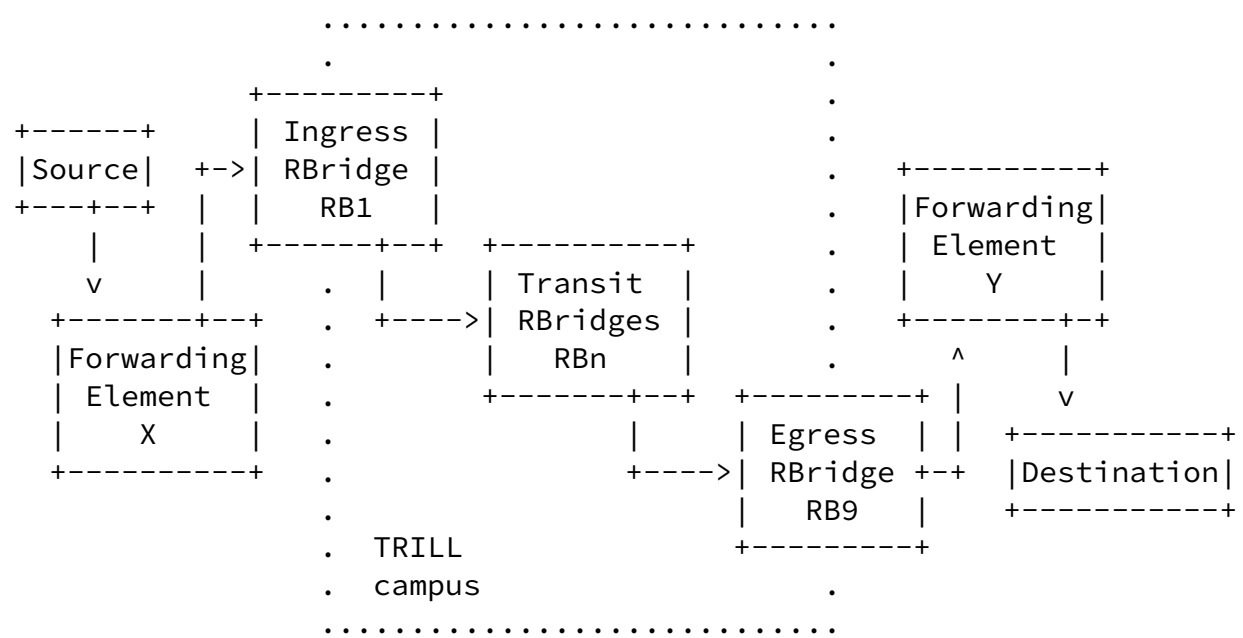
Explicit congestion notification (ECN) allows a forwarding element to notify downstream devices, including the destination, of the onset of congestion without having to drop packets. This document extends this capability to TRILL switches, including integration with IP ECN.

Table of Contents

1. Introduction

Explicit congestion notification (ECN [RFC3168]) allows a forwarding element, such as a router, to notify downstream devices, including the destination, of the onset of congestion without having to drop packets. Instead, the forwarding element can explicitly mark a proportion of packets in a two-bit ECN field. For example, a two-bit field in IP headers is available for ECN marking.

The transit of user data through a TRILL campus is similar to transport through a tunnel with the ingress and egress RBridges equivalent to the ends of the tunnel. Thus, existing ECN tunneling recommendatons, particularly [RFC6040], apply.



In the figure above, if ECN is implemented and assuming IP traffic, RB1 is effectivley a tunnel entrance and RB9 a tunnel exit. Traffic from Source to RB1 might or might not get marked as having experienced congestion in forwarding elements, such as X, before being encapsulated at ingress RB1. Any such ECN marking is encapsulated with a TRILL Header and provision is made in the TRILL Header extension Flags Word for ECN marking by the RBridges through which this traffic passes.

Any ECN marking in the traffic at the ingress is copied out to the TRILL Header Flags Word. At RB9, the TRILL egress, any ECN markings in the TRILL Header Flags Word and in the encapsulated traffic are combined so that subsequent forwarding elements, such as Y and the Destination, can see if congestion was experienced at any previous point in the path from Source if the forwarding elements are ECN capable and the Source marked packets as ECT (ECN Capabile Transport).

1.1 Conventions used in this document

The terminology and acronyms defined in [RFC6325] 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].

Acronyms:

CE - Congestion Experienced

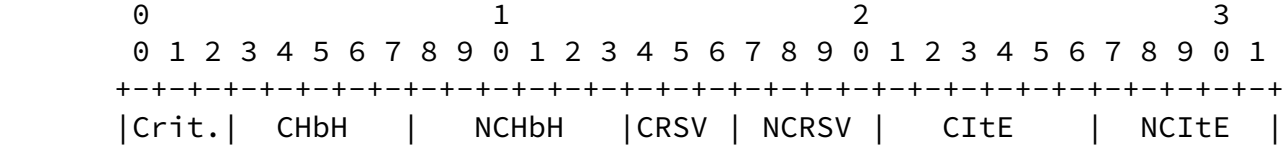
ECN - Explicit Congestion Notification

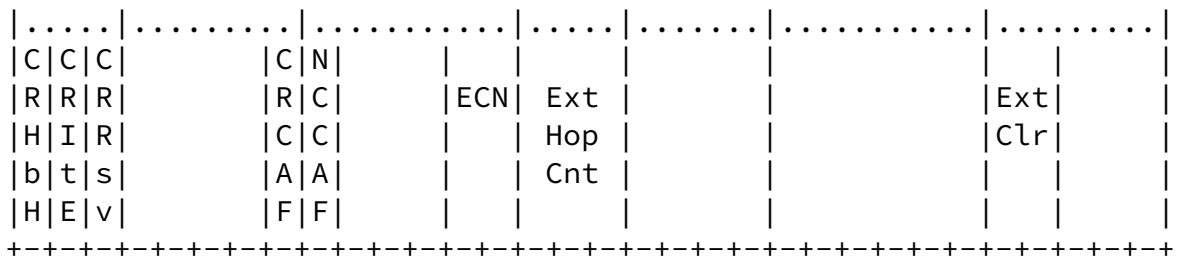
ECT - ECN Capabile Transport

2. The ECN Specific Extended Header Flags

RBridges MAY implement ECN (Explicit Congestion Notification) [RFC3168] through a two-bit field in the TRILL Header extension Flags Word [RFC7780]. If implemented, it SHOULD be enabled by default but can be disable on a per RBridge basis by configuration.

This field is show below as "ECN" and consists of bits 12 and 13 which are in the range reserved for non-critical hop-by-hop bits. See [RFC7780] and [RFC7179] for the meaining of the other bits.





The following table is modified from [RFC3168] and shows the meaning of bit values in TRILL Header extended flags 12 and 13. These are also the meanings of bits 6 and 7 of the DS field in the IPv4 and IPv6 heders as defined in [RFC3168]:

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 1. ECN Field Bit Combinations

3. ECN Support

An RBridge that has ECN support as specified herein advertises this through bit TBD in the Extended RBridge Capabilities APPsub-TLV [RFC7782] (see Section 4.2). On encapsulation, transit, and decapsulation it behaves as described in the subsections below, which correspond to the recommended provisions of [RFC6040].

3.1 Ingress ECN Support

Behavior at the ingress depends on whether the egress RBridge supports ECN. If it does, then the behavior is as follows (called "normal mode" in [RFC6040]):

- When encapsulating an IP frame that is ECN enabled (non-zero ECN field), the ingress RBridge MUST create a flags word as part of the TRILL Header, setting the F flag, and copy the two ECN bits from the IP header into flag word bits 12 and 13.

- When encapsulating a frame for a non-IP protocol, where that protocol has a means of indicating ECN that is understood by the ingress RBridge, it MAY add a flags word to the TRILL Header with the ECN bits set from the encapsulated native frame.

If the egress RBridge does not support ECN, the behavior is as follows (called "compatibility mode" in [RFC6040]):

- A TRILL Header Flags Word need not be created unless there is some reason other than ECN to do so.

- If a Flags Word is created, the ECN bits are set to zero (the Non-ECT value).

3.2 Transit ECN Support

When forwarding a TRILL Data packet encountering congestion at an RBridge, if the TRILL Header flags word is present, bits 12 and 13 are updated in the usual ECN manner [RFC3168]. An RBridge detects congestion either by monitoring its own queue depths or from participation in a link-specific protocol.

If, for reasons other than ECN, conditions at a transit RBridge require the insertion of a TRLL Header Flags Word into a TRILL Data packet, this implies that the egress RBridge is not ECN capable -- if it was, the Flags Word would have been included in the TRILL Data packet at the ingress. Thus, when a transit RBridge creates such a

Flags Word, it sets bits 12 and 13 to zero.

3.3 Egress ECN Support

Egress RBridge support of ECN is determined by looking at the Extended Capabilities APPsub-TLV that RBridge advertises. If bit TBD is zero, or the APPsub-TLV is absent, that RBridge does not support ECN. If the APPsub-TLV is present and bit TBD is one, then it does support ECN. If there are inconsistent APPsub-TLVs, the egress RBridge is assumed to support ECN if any of those APPsub-TLVs indicate that it does.

If the egress RBridge does not support ECN, it will ignore bits 12 and 13 of any Flags Word that is present, because it does not contain any special ECN logic.

If the egress RBridge supports ECN, it does the following:

- When decapsulating an IP frame, the RBridge MUST set the outgoing native IP frame ECN field to the code point at the intersection of the values for that field in the encapsulated IP frame (row) and the TRILL Header flags word ECN field (column) in Table 2 below or drop the frame in the case where the TRILL header indicates congestion experienced but the encapsulated native IP frame indicates a not ECN-capable transport. (Such frame dropping is necessary because IP transport that is not ECN-capable requires dropped frames to sense congestion.)

- When decapsulating a non-IP protocol frame with a means of indicating ECN that is understood by the RBridge, it MAY set the ECN information in the decapsulated native frame by combining that information in the TRILL Header flags word and the encapsulated non-IP native frame as specified in Table 2.

Table 2 below (adapted from [RFC6040]) shows how, at the egress, to combine the ECN information in the extended TRILL Header ECN field with the ECN information in an encapsulated frame to produce the ECN information to be carried in the resulting native frame.

Inner Native Header	Arriving TRILL Header Flag Word ECN Field			
	Not-ECT	ECT(0)	ECT(1)	CE
Not-ECT	Not-ECT	Not-ECT(*)	Not-ECT(*)	<drop>(*)
ECT(0)	ECT(0)	ECT(0)	ECT(1)	CE
ECT(1)	ECT(1)	ECT(1)(*)	ECT(1)	CE
CE	CE	CE	CE(*)	CE

Table 2: Egress ECN Behavior

An asterisk in the above table indicates a probably erroneous condition that SHOULD be logged.

4. IANA Considerations

This section summarizes IANA actions required.

4.1 Flags Word Bits

IANA is requested to assign bits 12 and 13 in the TRILL Header Flags Word for ECN and update the TRILL Extended Header Flags registry by replacing the line for bits 9-13 with the following"

Bits	Purpose	Reference
9-11	available non-critical hop-by-hop flags	
12-13	ECN (Explicit Congestion Notification)	[this document]

4.2 Extended RBridge Capability Bit

IANA is requested to assign bit TBD in the Extended RBridge Capabilities to indicate ECN support. The Extended RBridge Capabilities registry on the TRILL Parameters page is updated by adding the folloing line and updating any "Unassigned" line that is affected.

Bit	Mnemonic	Description	Reference
TBD	ECN	ECN Support	[this document]

5. Security Considerations

TBD

For ECN tunneling security considerations, see [RFC6040].

For general TRILL protocol security considerations, see [RFC6325].

6. Acknowledgements

This document was prepared with basic NROFF. All macros used were defined in the source file.

Normative References

[RFC2119] - Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <http://www.rfc-editor.org/info/rfc2119>.

[RFC3168] - Ramakrishnan, K., Floyd, S., and D. Black, "The Addition of Explicit Congestion Notification (ECN) to IP", RFC 3168, DOI 10.17487/RFC3168, September 2001, <http://www.rfc-editor.org/info/rfc3168>.

[RFC6040] - Briscoe, B., "Tunnelling of Explicit Congestion Notification", RFC 6040, DOI 10.17487/RFC6040, November 2010, <http://www.rfc-editor.org/info/rfc6040>.

[RFC6325] - Perlman, R., Eastlake 3rd, D., Dutt, D., Gai, S., and A. Ghanwani, "Routing Bridges (RBridges): Base Protocol Specification", RFC 6325, DOI 10.17487/RFC6325, July 2011,

<http://www.rfc-editor.org/info/rfc6325>.

[RFC7179] - Eastlake 3rd, D., Ghanwani, A., Manral, V., Li, Y., and C. Bestler, "Transparent Interconnection of Lots of Links (TRILL): Header Extension", RFC 7179, DOI 10.17487/RFC7179, May 2014, <http://www.rfc-editor.org/info/rfc7179>.

[RFC7780] - Eastlake 3rd, D., Zhang, M., Perlman, R., Banerjee, A., Ghanwani, A., and S. Gupta, "Transparent Interconnection of Lots of Links (TRILL): Clarifications, Corrections, and Updates", RFC 7780, DOI 10.17487/RFC7780, February 2016, <http://www.rfc-editor.org/info/rfc7780>.

[RFC7782] - Zhang, M., Perlman, R., Zhai, H., Durrani, M., and S. Gupta, "Transparent Interconnection of Lots of Links (TRILL) Active-Active Edge Using Multiple MAC Attachments", RFC 7782, DOI 10.17487/RFC7782, February 2016, <http://www.rfc-editor.org/info/rfc7782>.

Informative References

[none]

Authors' Addresses

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

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

Bob Briscoe (editor)
Simula Research Lab

Email: ietf@bobbriscoe.net
URI: http://bobbriscoe.net/