

TRILL Working Group
INTERNET-DRAFT
Updates: [6325](#)
Intended status: Proposed Standard

Donald Eastlake
Mingui Zhang
Huawei
Radia Perlman
Intel Labs
Vishwas Manral
Hewlett-Packard
Ayan Banerjee
Cumulus
July 16, 2012

Expires: January 15, 2013

Multiple Topology TRILL
<[draft-eastlake-trill-rbridge-multi-topo-03.txt](#)>

Abstract

The IETF TRILL (TRansparent Interconnection of Lots of Links) protocol is a solution for least cost transparent frame routing in multi-hop networks with arbitrary topologies and link technologies, using IS-IS (Intermediate System to Intermediate System) link-state routing and encapsulation with a hop count. IS-IS supports multiple topology routing. This document specifies a TRILL data plane encoding and procedures to make use of such routing.

Status of This Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of [BCP 78](#) and [BCP 79](#).

Distribution of this document is unlimited. Comments should be sent to the TRILL working group mailing list.

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

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

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

INTERNET-DRAFT

TRILL: Multi-Topology

Table of Contents

1. Introduction.....	3
1.1 Terminology.....	3
2. TRILL Multi-Topology.....	4
2.1 Nicknames and Topology Abbreviation.....	4
2.2 Nickname Allocation.....	5
2.3 SPF and Distribution Tree Calculation.....	6
2.3.1 Base Topology SPF.....	6
2.3.2 Non-Zero Topology SPF.....	6
2.3.3 Distribution Tree Calculation.....	6
3. Processing Multi-Topology Frames.....	8
3.1 Ingress Processing.....	8
3.2 Transit Processing.....	8
3.3 Egress Processing.....	9
3.4 Address Learning and Asymmetric Topologies.....	9
5. IS-IS Extensions.....	10
6. IANA Considerations.....	11
7. Security Considerations.....	11
8. Acknowledgements.....	11
9. References.....	12
9.1 Normative References.....	12
9.2 Informative References.....	12

INTERNET-DRAFT

TRILL: Multi-Topology

1. Introduction

The IETF has standardized RBridges (Routing Bridges), devices that implement the TRILL (TRansparent Interconnection of Lots of Links) protocol [[RFC6325](#)], a solution for least cost transparent frame routing in multi-hop networks with arbitrary topologies, using IS-IS (Intermediate System to Intermediate System) link-state routing [IS-IS] [[RFC1195](#)] [[RFC6326bis](#)] and encapsulation with a hop count.

TRILL supports VLANs (Virtual Local Area Networks) but the segregation provided by VLANs in TRILL is logical, not physical. While maintaining logical separation, the base TRILL standard shares inter-RBridge links across all VLANs and by default interconnects all end stations that are in the same VLAN and have connectivity to any RBridge in the campus.

IS-IS multi-topology [[RFC5120](#)] can provide physical separation of sub-sets of TRILL Data frames, assuming that the RBridges in a campus can be trusted. This can be used for a variety of purposes including such things as confining a sub-set traffic to an island within an RBridge campus, quality of service traffic engineering, excluding some frames from links that are particularly exposed to interception, and providing significant protection against some covert channels [[RFC4949](#)].

Familiarity with the TRILL base protocol standard [[RFC6325](#)], TRILL use of IS-IS [[RFC6326bis](#)], and IS-IS multi-topology [[RFC5120](#)] is assumed in this document.

[1.1](#) Terminology

The terminology and acronyms of [\[RFC6325\]](#) are used in this document with the additions listed below.

Highest Priority RBridge - The RBridge in the campus that is the highest priority to be a base topology tree root.

MT - Multi-Topology

TAS - Topology Abbreviation Size

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 Multi-Topology

The essence of TRILL multi-topology (MT) is that (a) when TRILL Data frames are ingressed or created they are assigned to a topology, (b) links can be labeled with one or more topologies and different cost, etc., for different topologies, and (c) a TRILL Data frame is only allowed on a link labeled with the frame's topology. In addition, there is a base topology: all links are considered to be in the base topology, and, by default, TRILL Data frames are considered to be base topology frames.

With minor exceptions, it is important that all transit R Bridges believe a TRILL Data frame is in the same topology to avoid persistent routing loops. This document specifies how to encode this information into the egress nickname in a TRILL Data frame. (Note: It is believed that this technique is supportable by most if not all TRILL fast path silicon implementations.)

Implementation of MT has a significant impact on shortest path and distribution tree computation. In general, it multiplies the level of effort by the number of different topologies in which the average R Bridge participates. Using the technique in this document, it also reduces the number of available nicknames, generally dividing it

by the number of topologies rounded up to the next power of two. For these reasons, the number of overlapping topologies in use should be minimized.

[2.1](#) Nicknames and Topology Abbreviation

The TRILL base protocol standard [[RFC6325](#)] specifies 16-bit nicknames as abbreviations for 7-bytes IS-IS IDs. In MT TRILL the nickname is subdivided into two fields. The least significant TAS (Topology Abbreviation Size) number of bits indicate the topology constraint of the nickname while the most significant (16 - TAS) bits continue to act as an abbreviation for an RBridge System ID.

The TAS value for an MT RBridge campus is dictated by the highest priority RBridge. To prevent transient disruption, other RBridges that might become the highest priority RBridge SHOULD be configured with the same TAS value. The value of TAS can vary from zero to seven. Zero indicates that only the base topology can be used.

In IS-IS, topologies have a 12-bit ID which we refer to as an absolute topology number. Topology zero is the base topology. All links are considered to be part of the base topology. Absolute topology numbers are mapped into abbreviations by a table dictated by the highest priority RBridge. To prevent transient disruption, other RBridges that might become the highest priority RBridge should be

configured with the same TAS topology abbreviation table. In the opening remarks to this [Section 2](#), "topology" should be read as referring to topology abbreviation.

Multiple absolute topology numbers MAY be mapped into the same abbreviation. This may or may not result in the merger of the mapped absolute topologies depending on whether their use is disjoint or overlapping. Absolute topology zero and topology abbreviation zero always refers to the base topology.

For example, assume that absolute topologies Tp1 and Tp2 exist and are both mapped into the same specific abbreviation. If all RBridges in the campus reachable by links labeled Tp1 or Tp2 are connected through such links, then, since MT TRILL forwarding is based on the topology abbreviation, Tp1 and Tp2 are necessarily merged to the

extent both topologies are in use. On the other hand, such RBridges could be divided into disjoint islands with no connection between the islands via links allowing Tp1 or Tp2. In that case, depending on how frames are topologically classified, each such island could independently be exclusively Tp1, exclusively Tp2, merged Tp1 and Tp2, or unused.

[2.2](#) Nickname Allocation

RBridges indicate in their LSP whether they support MT by the presence of the MT TLV [[RFC5120](#)]. If all RBridges in a campus support MT, then RBridges can be configured for and contend for nicknames as provided in [[RFC6325](#)], but only for nicknames that have TAS number of low order zero bits. If there are any RBridges in the campus that do not support MT, then MT RBridges must hold all nicknames from $(k * 2^{TAS})$ through $((k+1) * 2^{TAS} - 1)$ in order to, in effect, hold nickname $(k * 2^{TAS})$.

RBridges not supporting MT are unaware of the subdivision of the TRILL nickname into subfields. Thus, they may hold arbitrary 16-bit allowed quantities as nicknames. MT aware RBridges know that the bottom TAS bits of any nicknames held by such MT unaware RBridges are not topologically significant.

The nickname allocation described above and in [[RFC6325](#)] runs based on the nickname priority values appearing in Router Capabilities TLVs (TLV #242), which is what MT unaware RBridges will use. The Nicknames sub-TLV is allowed in the MT Router Capabilities TLV (TLV #144) for the sole purpose of permitting nicknames to have different priorities to be root in different topologies; in particular, the Nicknames sub-TLV field giving the priority to hold that nickname is ignored for Nicknames sub-TLVs in the MT Router Capabilities TLV.

[2.3](#) SPF and Distribution Tree Calculation

This section discusses MT SPF and distribution tree calculation.

[2.3.1](#) Base Topology SPF

Since MT TRILL cannot impose any changes on MT unaware RBridges, those RBridges will perform their SPF and distribution tree calculations as specified in [\[RFC6325\]](#). For compatibility, MT aware RBridge MUST perform their base topology SPF and distribution tree calculations in the same way. In particular, the base topology consists of those links listed in Extended IS Reachability TLVs (TLV #22). For backward compatibility, MT aware RBridges use only links listed in TLV #22 for the base topology, which SHOULD be all links, even if there exist MT-ISN TLVs (TLV #222) listing topology zero or if one or more non-zero absolute topologies are being mapped into the base topology abbreviation zero.

[2.3.2](#) Non-Zero Topology SPF

For MT aware RBridges, least cost (SPF) paths are also calculated per topology abbreviation for the non-zero abbreviations labeling any link connected to that RBridge. When paths are being calculated for a topology abbreviation, only links labeled with absolute topologies that map to that abbreviation are considered.

For topologies other than the base topology, link costs from the MT-ISN TLV (TLV #222 [\[RFC5120\]](#)) are used; however, such costs are reported per absolute topology, not per topology abbreviation. This is resolved for non-zero abbreviations by using, in SPF calculations, the lowest cost reported for the link for any absolute topology corresponding the topology abbreviation for which the calculation is being performed. Non-base topologies do not necessarily span the campus and there may be RBridges that are unreachable in such topologies. Frames destined for unreachable RBridges are discarded.

[2.3.3](#) Distribution Tree Calculation

Distribution trees are also calculated per topology abbreviation by MT aware RBridges. At least one distribution tree is always built as described in [\[RFC6325\]](#) for the base topology using the tree root nickname priorities advertised in the Router Capabilities TLV. That tree will span the campus. Care should be taken that the highest priority RBridge is configured to not request too many distribution

trees be calculated in the base topology. There may be silicon limits to how many distribution trees can be efficiently handled and too many base topology trees could starve other topologies that should have a distribution tree.

For non-zero topology abbreviations, a distribution tree will be composed only of links that are in at least one of the absolute topologies that map to that abbreviation and the tree may not span the campus. Such distribution trees might not span the campus and there might be multiple disjoint distribution trees for a particular topology abbreviation.

The calculation of additional distribution trees for non-zero topology abbreviations is accomplished using the same method of building from root as described in [[RFC6325](#)] but using link costs as described in the preceding section on TRILL non-zero topology SPF calculations. The number of such trees may be constrained by the capabilities of R Bridges in the campus as advertised in the Trees sub-TLV in the Router Capabilities TLV. However, that limit is of the number of trees actually including that R Bridge and would not include, for example, a distribution tree for some topology present only in a remote island of R Bridges.

After the base topology trees are calculated, trees for non-zero topologies are calculated, one per topology abbreviation, starting with the highest topology abbreviation in use and working down to the lowest non-zero topology. If additional trees are available and requested, they are distributed to the topology abbreviations. For the details see Appendix ... [[Since it is important that all R Bridges agree on the calculation of distribution trees, I think this is going to need some pseudo code in an appendix.]]

INTERNET-DRAFT

TRILL: Multi-Topology

[3. Processing Multi-Topology Frames](#)

This section specifies ingress, transit, and egress processing of MT TRILL Data frames, how asymmetric topologies can occur, and address learning.

[3.1 Ingress Processing](#)

On ingressing or creating a frame, an RBridge assigns it to an absolute topology ID. The method by which this 12-bit topology ID is assigned is beyond the scope of this document; however, different frames with the same source MAC address, VLAN, and egress RBridge SHOULD be assigned to the same topology. In TRILL, topology does not provide another level of VLAN but identifies a subset of traffic.

The resulting absolute ID is mapped to a topology abbreviation. That abbreviation is used as the bottom TAS bits in the ingress nickname field of the resulting TRILL Data frame.

For a unicast native frame the VLAN and the destination MAC address are used to look up the egress nickname field.

If the destination is unknown, or the native frame is multicast or broadcast, the ingress RBridge selects a distribution tree for that topology abbreviation that includes the ingress RBridge. If more than one tree is available, the method of choice is beyond the scope of this document, but by default, it should pick the tree whose root is least cost from the ingress RBridge.

[3.2 Transit Processing](#)

No change is required in transit frame processing assuming that SPF and distribution tree calculations have been performed as specified in [Section 2.3](#).

The next hop RBridge for TRILL unicast frames will automatically be restricted to the appropriate topology. The same applies to the zero or more next hops for the tree that a TRILL multi-destination frame

is being distributed on. There is no change in the Reverse Path Forwarding Check.

[3.3](#) Egress Processing

There is no change in egress processing.

[3.4](#) Address Learning and Asymmetric Topologies

There is no change in address learning. This can result in asymmetric topology use.

For example assume end stations ESa and ESb in the same VLAN-x attached to RB1 and RB2 respectively want to communicate. Generally, the initial frame from RB1 will be classified in topology abbreviation T1 and sent on a T1 distribution tree which should be pruned for VLAN-x. If that tree includes RB2, the frame will be delivered to RB2 that will egress it to ESb or, if it has not yet learned which of its links ESb is attached to, RB2 will egress the frame to all of its links in VLAN-x for which it is appointed forwarder [[RFC6439](#)]. In addition, RB2 will learn that ESa in VLAN-x is attached to RB1 but the nickname it learns from RB1 will have T1 encoded in it so that RB1 also learns, that ESa is in T1. When ESb send a return frame to ESa, RB2 will classify that frame as in topology abbreviation T2, encode that in the ingress nickname of the TRILL header, and send the frame to RB1 in T1 because the egress nickname RB2 will use was that learned from the receipt of the above frame from RB1 in T1. At this point, ESa and ESb can continue communicating using TRILL unicast frames in each direction with frames from ESa to ESb in T1 while those from ESb to ESa will be in T2.

If it is desired that T1 equal T2, then RB1 and RB2 must be configured to classify the incoming native frames as in the same topology. Note that, if many topologies are in use but there are

fewer distribution trees, the initial ESa frame in the example above might have been distributed in a T3 distribution tree. This will not affect RB2 learning which will learn from the T1 topology encoded into the ingress nickname in the encapsulated frame's TRILL Header.

[5](#). IS-IS Extensions

For the extensions to the TRILL use of IS-IS to make use multi-topology, see [[trill-mt](#)].

These include the addition of a topology mapping sub-TLV. This sub-TLV is across all topologies and occurs in the Router Capabilities TLV. It specifies the TAS for the campus and provides a mapping from absolute topology IDs to topology abbreviations. Absolute topology zero is always mapped to topology abbreviation zero. No entry is needed for this base topology mapping and any other mapping for topology zero is ignored. Multiple absolute topologies may be mapped to the same abbreviation. If there are mappings of the same absolute topology to multiple abbreviations, the largest abbreviation, considered as an unsigned integer dominates, and mappings of that absolute topology to smaller abbreviations are ignored.

INTERNET-DRAFT

TRILL: Multi-Topology

[6.](#) IANA Considerations

IANA considerations for this draft are in [[trill-mt](#)].

[7.](#) Security Considerations

See [[RFC6325](#)] for general RBridge Security Considerations.

As with any communications system, end-to-end encryption and authentication should be considered for particularly sensitive data.

More TBD

[8.](#) Acknowledgements

The comments and contributions of the following are gratefully acknowledged:

Meenakshi Kaushik and Dinesh Dutt.

[9.](#) References

The following sections list normative and informative references for this document.

[9.1](#) Normative References

[IS-IS] - International Organization for Standardization,

"Intermediate system to Intermediate system intra-domain routing information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode Network Service (ISO 8473)", ISO/IEC 10589:2002, Second Edition, Nov 2002.

[RFC1195] - Callon, R., "Use of OSI IS-IS for routing in TCP/IP and dual environments", [RFC 1195](#), December 1990.

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

[RFC5120] - Przygienda, T., Shen, N., and N. Sheth, "M-ISIS: Multi Topology (MT) Routing in Intermediate System to Intermediate Systems (IS-ISs)", [RFC 5120](#), February 2008.

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

[RFC6439] - Perlman, R., Eastlake, D., Li, Y., Banerjee, A., and F. Hu, "Routing Bridges (RBridges): Appointed Forwarders", [RFC 6439](#), November 2011.

[RFC6326bis] - Eastlake, D., A. Banerjee, D. Dutt, R. Perlman, A. Ghanwani, "TRILL Use of IS-IS", [draft-eastlake-isis-rfc6326bis](#), work in progress.

[trill-mt] - Manral, V., D. Eastlake, M. Zhang, A. Banerjee, "Multi Topology Routing Extensions for Transparent Interconnection of Lots of Links (TRILL)", [draft-manral-isis-trill-multi-topo](#), work in progress.

[9.2](#) Informative References

[RFC4949] - Shirey, R., "Internet Security Glossary, Version 2", FYI 36, [RFC 4949](#), August 2007.

Donald Eastlake 3rd
Huawei R&D USA
155 Beaver Street
Milford, MA 01757 USA

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

Mingui Zhang
Huawei Technologies Co., Ltd
HuaWei Building, No.3 Xinxu Rd., Shang-Di
Information Industry Base, Hai-Dian District,
Beijing, 100085 P.R. China

Email: zhangmingui@huawei.com

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

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

Vishwas Manral
Hewlett-Packard Co.
19111 Pruneridge Ave.
Cupertino, CA 95014 USA

Phone: 1-408-447-0000
Email: vishwas.manral@hp.com

Ayan Banerjee
Cumulus Networks
1089 West Evelyn Avenue
Sunnyvale, CA 94086 USA

Email: ayabaner@gmail.com

Copyright, Disclaimer, and Additional IPR Provisions

Copyright (c) 2012 IETF Trust and the persons identified as the document authors. All rights reserved.

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

