TRILL Working Group
INTERNET-DRAFT
Intended status: Proposed Standard
Updates: 6325
Donald Eastlake
Mingui Zhang
Huawei
Puneet Agarwal
Broadcom
Dinesh Dutt
Cisco
Radia Perlman
Intel Labs
Expires: April 27, 2012
October 28, 2011

RBridges: Fine-Grained Labeling
<draft-eastlake-trill-rbridge-fine-labeling-02.txt>

Abstract

The IETF has standardized RBridges (Routing Bridges), devices that implement the TRILL (TRansparent Interconnection of Lots of Links) protocol, a standard for least cost transparent frame routing in multi-hop networks with arbitrary topologies, using link-state routing and encapsulation with a hop count.

The TRILL base protocol standard supports up to 4K VLAN IDs (Virtual Local Area Network IDentifiers). However, there are applications that require more fine-grained labeling of data and end stations. This document updates RFC 6325 by specifying extensions to the TRILL protocol to accomplish this.

Status of This Memo

http://www.ietf.org/shadow.html

Table of Contents

1. Introduction

The IETF has standardized RBridges (Routing Bridges), devices that implement the TRILL (TRansparent Interconnection of Lots of Links) protocol [RFC6325]. RBridges provide a solution for least cost transparent frame routing in multi-hop networks with arbitrary topologies, using [IS-IS] [RFC6326bis] link-state routing and encapsulation with a hop count addressing the problems outlined in [RFC5556].

The TRILL base protocol standard supports labeling with up to 4K VLAN IDs (Virtual Local Area Network IDentifiers). However, there are applications that require more fine-grained labeling of data and end stations. This document updates [RFC6325] by specifying extensions to the TRILL protocol to accomplish this.

Familiarity with [RFC6325] and [RFC6326bis] is assumed in this document.

1.1 Terminology

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

VL - VLAN Labeling or VLAN Labeled or VLAN Label

VL RBridge - An RBridge that support VL and does not support FGL

FGL - Fine-Grained Labeling or Fine-Grained Labeled or Fine-Grained Label

FGL RBridge - An RBridge that support both FGL and VL

Edge RBridge - An RBridge announcing VL or FGL connectivity in its link state

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. Fine-Grained Labeling

The essence of Fine-Grained Labeling (FGL) is that (a) when TRILL Data frames are ingressed or created they may incorporate a label from a set of significantly more than 4K labels, (b) RBridge ports can be labeled with a set of such labels, and (c) an FGL TRILL Data frame cannot be egressed through an RBridge port unless its FGL matches one of the labels of the port.

Section 2.1 lists fine-grained labeling requirements. Section 2.2 briefly outlines TRILL Data VLAN Labeling (VL) in the TRILL base protocol standard [RFC6325]. And Section 2.3 then outlines a method of Fine-Grained Labeling (FGL) of TRILL Data frames.

2.1 Requirements

There are several requirements that should be met by FGL in TRILL. They are briefly described in the list below in approximate order by priority with the most important first.

1. Fine-Grained

   Some networks have a large number of entities that need configurable isolation, whether those entities are independent customers, applications, or branches of a single endeavor or some combination of these or other entities. The VLAN labeling supported by [RFC6325] provides for only (2**12 - 2) valid identifiers or labels. A substantially larger number is required.

2. Silicon Considerations

   Fine-grained labeling (FGL) should, to the extent practical, use existing features, processing, and fields that are already supported in at least some of the existing TRILL fast path silicon implementations.

3. Base RBridge Compatibility

   To support some incremental conversion scenarios, it is desirable that not all RBridges in a campus using FGL be required to be FGL aware. That is, it is desirable that RBridges not implementing the FGL feature and performing at least the transit forwarding function can usefully process TRILL Data frames that incorporate FGL.

4. Alternate Priority

It would be desirable for an ingress RBridge to be able to assign a different priority to an FGL TRILL Data frame for its ingress-to-egress propagation from the priority of the original native frame. The original priority should be restored on egress.

2.2 Existing TRILL VLAN Labeling

This section provides a brief review of existing TRILL Data frame VLAN Labeling (VL) and changes the description of VL from that appearing in [RFC6325] by moving the end of the TRILL Header. This description change does not involve any change in the bits on the wire or in the behavior of existing [RFC6325] RBridges.

Currently TRILL Data frames have the VL structure shown below:

```
+--------------------------------------------+
| Link Header (Depends on Link Technology)   |
+--------------------------------------------+
| TRILL Header                               |
| +--------------------------------------+   |
| | Initial Fields and Options, FGLflag=0 |  |
| +--------------------------------------+   |
| |         Inner.MacDA         | (6 bytes)  |
| +-----------------------------+            |
| |         Inner.MacSA         | (6 bytes)  |
| +-----------------------------+            |
| |C-VLAN EtherType (0x8100)|     (2 bytes)  |
| +-------------------------+                |
| | Inner.VLAN Label        |     (2 bytes)  |
| +-------------------------+                |
+--------------------------------------------+
|                Native Payload              |
+--------------------------------------------+
| Link Trailer (Depends on Link Technology)  |
+--------------------------------------------+
```

The FGLflag bit is the formerly reserved TRILL Header bit immediately adjacent to the V field that [RFC6325] specifies as always set to zero and ignored on receipt.

The C-VLAN EtherType is always present and is followed by the Inner.VLAN field including the 12-bit VLAN ID field.

2.3 Fine-Grained Labeling (FGL)

FGL expands the 12-bit VLAN label available under the TRILL base protocol standard to a 24-bit fine-grained label. In this document, FGLs are usually denoted as "(X.Y)" where X is the high order 12 bits and Y is the low order 12 bits of the FGL. The FGL information appears in the TRILL Header as shown below.

```
            +--------------------------------------------+
            | Link Header (Depends on Link Technology)   |
            +--------------------------------------------+
            | TRILL Header                               |
            | +----------------------------------------+ |
            | | Initial Fields and Options, FGLflag=1 | |
            | +----------------------------------------+ |
            | |         Inner.MacDA         | (6 bytes) |
            | +-----------------------------+           |
            | |         Inner.MacSA         | (6 bytes) |
            | +-----------------------------+           |
            | |C-VLAN EtherType (0x8100)|     (2 bytes) |
            | +-------------------------+               |
            | | Inner.Label First Part  |     (2 bytes) |
            | +-------------------------+               |
            | |EX-TAG EtherType (0xTBD) |     (2 bytes) |
            | +-------------------------+               |
            | | Inner.Label Second Part |     (2 bytes) |
            | +-------------------------+               |
            +--------------------------------------------+
            |               Native Payload               |
            +--------------------------------------------+
            | Link Trailer (Depends on Link Technology) |
            +--------------------------------------------+
```

The FGLflag bit is the formerly reserved TRILL Header bit immediately adjacent to the V field that [RFC6325] specifies as always set to zero and ignored on receipt. This document updates [RFC6325] by providing that that bit is set to one if and only if the TRILL Data frame is FGL.

The fixed format area of the TRILL Header with the Inner.Label and EtherType fields 0x8100 and 0xTBD is mandatory for FGL frames. It is designed to be backward compatibility with [RFC6325] conformant RBridges although such RBridges will only be aware of the most significant 12-bits of the FGL.

The two byte following the EX-TAG EtherType have, in their low order 12 bits, a low order extension to the VLAN ID in the preceding VLAN tag and together they constitute the fine-grained label. The upper 4 bits of those two bytes are used for a 3-bit priority field and one unused bit.

The priority field of the initial C-VLAN is the priority used for frame transport from ingress to egress.

The appropriate FGL value for an ingressed native frame is determined by the input RBridge port as specified in Section 4.1. Ports of RBridges supporting FGL also have capabilities to transmit frames being forwarded or egressed as untagged or Outer.VLAN tagged as specified in Section 4.3.

3. Campus Wide VL versus FGL Semantic Differences

There are significant differences between the semantics across a campus for VLAN labels (VLs) and fine-grained labels (FGLs).

With VL, VLAN label IDs have the same meaning throughout the campus. In addition, the TRILL Header label in the Inner.VLAN is from the same label space as the VLAN IDs used on Ethernet links in the campus.

With TRILL FGL, many things remain the same. Ports of FGL RBridges act as they do for VL RBridges: Ethernet links still have C-VLAN labeling on them and RBridge ports provide a VLAN ID for an incoming frame and accept a VLAN ID for a frame being queued for output. Appointed Forwarders [RFCaf] on a link are still appointed for a C-VLAN. The Designated VLAN for a link is a C-VLAN. However, the 24-bit FGL label space is a different flat space from the 12-bit C-VLAN space. For ports configured for FGL, the C-VLAN on an ingressed native frame is mapped to the FGL space with a potentially different mapping for each port. A similar FGL to C-VLAN mapping occurs on egress. Thus, for ports configured for FGL, the C-VLAN corresponding to an FGL on one link can be different from the C-VLAN corresponding to that same FGL on a different link elsewhere in the campus or even a different link attached to the same RBridge. The FGL label space is flat and does not hierarchically encode any particular number of C-VLAN bits or the like.

FGL RBridge ports can be configured for FGL or VL with VL being the default. As with a base protocol [RFC6325] RBridge, by default an FGL RBridge port reports an untagged frame it receives as being in VLAN 1.

4. Coexistence with VL RBridges

VLAN Labeling (VL) RBridges will operate properly as transit RBridges. Transit RBridges look at the Inner.VLAN ID field only for the filtering of multi-destination frames. If an RBridge does not perform filtering, or filters on only some of the fields in the packet, the only consequence is that multi-destination frames will use more bandwidth than necessary. VL RBridges would only look at the high order 12 bits of the FGL, which are in the position where a VL RBridge would expect to find a VLAN ID. Thus they will not be able to prune as effectively as transit FGL RBridges could because they will ignore the lower 12 bits of the FGL.

It would be more serious if a VL edge RBridge, RB1, unaware of FGL, forwarded an FGL frame with FGL (X.Y) onto a link through an RB1 port configured as VL VLAN-X. RB1 would strip the TRILL Header only through the Inner.Label First Part, which it thinks is an Inner.VLAN Label, and forward the packet with the Inner.Label Second Part and preceding 0xTBD field still present. This might cause other problems on the link. It would also be problematic if a malicious end station could forge an apparent FGL label (X.Y) frame by including extra fields in native frames ingressed by a VL edge RBridge. Therefore, it is highly desirable for all the edge RBridges to be FGL RBridges.

FGL RBridges will report the FGL capability in LSPs, so FGL RBridges (and any management system with access to the link state database) will be able to detect the existence of VL edge RBridges.

It might be useful, in a particular campus with mixed VL and FGL RBridges, to have some end station VLANs accessible via VL edge RBridges. This is supported by reserving some number of VLANs (say the first k), to be VL-addressable. These VLANs will be specified with a single Inner.VLAN label, whether or not the edge RBridges attached to these VLANs are FGL-capable. When VL-specifiable VLANs are used in a FGL campus, and where there are VL edge RBridges advertising connectivity to those VLANs, the upper 12 bits in an FGL MUST NOT be equal to the value of any VL-specifiable VLAN.

If this rule is violated, the network misconfiguration is detected by the FGL RBridges that will then refuse in ingress to or egress from label (X.Y) while VLAN X connectivity is being advertised by a VL edge RBridge.

5. Fine-Grained Labeling Details

This section specifies ingress, transit, egress, and other processing of TRILL Data frames with regard to Fine-Grained Labels (FLGs). A transit or egress FGL RBridge detects FGL TRILL Data frames by noticing that the FGLflag in the TRILL Header is set.

5.1 Ingress Processing

An FGL RBridge may be configured, on one or more ports, to FGL ingress native frames. There is no change in VL ingress processing, which is the default unless a port has been configured for FGL, and no change in Appointed Forwarder logic (see Section 5.4).

FGL RBridges MUST support configurable per port mapping from the C-VLAN ID associated with a native frame to a 24-bit fine-grained label. FGL RBridges MAY support other methods to determine the FGL ID of an incoming native frame, such as based on the protocol of the native frame. If the resulting label (X.Y) is such that VLAN X connectivity is being advertised by a VL edge RBridge in the campus, the ingressed frame MUST be dropped.

The FGL ingress process MUST place the priority associated with an ingressed native frame in upper 3 bits of the second Inner.Label part. It SHOULD also associate a possibly different mapped priority with an ingressed frame. The mapped priority is placed in the Inner.Label First Part. If such mapping is not supported then the original priority is also placed in the Inner.Label First Part.

An FGL ingress RBridge MAY serially TRILL unicast a multi-destination TRILL Data frame to the relevant egress RBridges, if those egress RBridges are all FGL RBridges, after encapsulating it as a TRILL known unicast data frame (M=0) and SHOULD so unicast such a multi-destination TRILL Data frame if there is only one relevant egress FGL RBridge. The relevant egress RBridges are determined by starting with those announcing connectivity to the frame's (X.Y) label. That set SHOULD be further filtered based on multicast listener and router connectivity if the native frame was a multicast frame.

Use of S-tags is beyond the scope of this document but is an obvious extension.

5.2 Transit Processing

TRILL Data frame transit processing is fairly straightforward as described in Section 5.2.1 for known unicast TRILL Data frames and in Section 5.2.2 for multi-destination TRILL Data frames.

5.2.1 Unicast Transit Processing

There is almost no change in TRILL Data frame unicast transit processing. A transit RBridge forwards any unicast TRILL Data frame to the next hop towards the egress RBridge as specified in the TRILL Header. Just as transit RBridges conformant to the TRILL base protocol standard [RFC6325] do not examine the Inner.VLAN of unicast TRILL Data frames, transit FGL RBridges do not examine the 24-bit FGL of unicast TRILL Data frames.

All transit RBridges, whether VL or FGL, MUST take the priority used to forward a frame from the Inner.VLAN label or the FGL Inner.Label First Part. These bits are in the same relative position for VL and FGL frames so VL RBridges will do this automatically even though they do not fully understand FGL frames.

5.2.2 Multi-Destination Transit Processing

All multi-destination TRILL Data frames are forwarded on a distribution tree selected by the ingress RBridge. The distribution trees for FGL and VL multi-destination frames are the same and are calculated as provided for in the TRILL base protocol standard [RFC6325]. There is no change in the Reverse Path Forwarding Check.

An FGL RBridge, say RB1, having an FGL multi-destination frame for label (X.Y) to forward on a distribution tree, SHOULD prune based on whether there are any edge RBridges on the tree branch that are connected to label (X.Y). In addition, RB1 SHOULD prune multicast frames based on reported multicast listener and multicast router attachment in (X.Y). Finally, a transit FGL RBridge MAY drop any multi-destination frame for label (X.Y) if some VL RBridge is advertising connectivity to VLAN X. "MAY" is chosen in this case to minimize the checking burden on transit RBridges.

To ensure that a transit VL RBridge does not falsely filter traffic for FGL label (X.Y), an FGL edge RBridge attached to FGL label (X.Y) MUST report connection to VLAN X, as if X were a VLAN label, in addition to reporting connectivity to label (X.Y). Because of this, VL transit RBridges can safely apply pruning to all TRILL Data frames, both VL and FGL, based on the reported VLAN-X connectivity of all downstream RBridges.

To ensure that a transit VL RBridge does not falsely prune traffic for FGL label (X.Y) base on multicast filtering, an FGL edge RBridge attached to label (X.Y) MUST also report for VLAN X either (1) that it is attached to both IPv4 and IPv6 multicast routers or (2) its merged FGL label (X.Y) multicast listener and router connectivity for all Y.

5.3 Egress Processing

Egress processing is generally the reverse of ingress progressing described in Section 5.1.

If any VL RBridge in the campus is announcing connectivity to VLAN-X, an FGL RBridge MUST NOT egress a frame with FGL label (X.Y) but must drop such a frame.

An FGL RBridge MUST be able to configurably convert the 24-bit fine grained label in an FGL TRILL Data frame it is egressing to a 12-bit C-VLAN ID for the resulting native frame on a per port basis. A port MAY be configured to strip such tagging. It is the responsibility of the network manager to properly configure the RBridges and ports in the campus to obtain the desired mappings.

An FGL RBridge egresses FGL frames with the above tag conversion similarly to the egressing of VL frames, as follows:

1. A known unicast FGL frame is egressed to the FGL port matching its fine-grained label and Inner.MacDA. If there is no such port, it is flooded out all FGL ports with its fine-grained label unless the RBridge has knowledge that the frames Inner.MacDA cannot be out that port.

2. A multi-destination FGL frame is decapsulated and flooded out all ports with its fine-grained label, subject to multicast pruning.

FGL RBridges MUST accept multi-destination encapsulated frames that are sent to them as TRILL unicast frames, that is, frames with a multicast or broadcast Inner.MacDA and the TRILL Header M bit = 0. They locally egress such frames, if appropriate, but MUST NOT forward them (other than egressing them as native frames on their local links).

Use of S-tags is beyond the scope of this document but is an obvious extension.

5.4 Appointed Forwarders and the DRB

There is no change in Adjacency [RFC6327] or Appointed Forwarder logic [RFCaf] on a link regardless of whether some or all the ports on the link are for FGL RBridges. However, if it is intended for native frames on a link in some VLAN-X to be ingressed and egressed with FGL, the Appointed Forwarder for VLAN-X for that link obviously MUST be an FGL RBridge.

If there are FGL and VL RBridges connected to a link, it may be best if the priorities are configured so that the DRB is an FGL RBridge. However, there is no inherent difficulty in a VL DRB RBridge appointing an FGL RBridge connected to the link as Appointed Forwarder for whatever VLANs are appropriate.

5.5 Address Learning

An FGL RBridge learns addresses on FGL ports based on the fine-grained label rather than VLAN ID. Addresses learned from ingressed native frames are logically represented by { MAC address, fine-grained label, port, confidence, timer } while remote addresses learned from egressing FGL frames are logically represented by { MAC address, fine grained label, remote RBridge nickname, confidence, timer }.

6. IS-IS Extensions

Extensions to the TRILL use of IS-IS are required to support the following:

1. An method for an RBridge to announce itself in its LSP as supporting FGL.

2. A sub-TLV analogous to Interested VLANs and Spanning Tree Roots sub-TLV of the Router Capabilities TLV but indicating fine-grained labels rather than VLANs.

3. A sub-TLV analogous to the GMAC-ADDR sub-TLV of the Group Address TLV that specifies a fine-grained label rather than a VLAN.

See [RFC6326bis] and Section 8.2.

7. Comparison to Requirements

Comparing TRILL fine-grained labeling (FGL), as specified in this document, with the requirements given in Section 2.1, we find they are met as follows:

1. Fine-Grained: FGL provides approaching 2**24 labels, vastly more labels than the 4K VLAN IDs.

2. Silicon Considerations: Existing TRILL fast path silicon chips can, almost by definition, perform base TRILL Header insertion and removal to support ingress and egress. In addition, it is believed that most such silicon chips can also perform the 12-bit VLAN ID and port to fine-grained label mapping and the encoding of the fine-grained label as specified herein, as well as the inverse decoding and mapping. Some existing silicon can perform only one of these operations on a frame in the fast path and is thus not suitable to implement fast path TRILL FGL processing; however, other existing chips are believed to be able to perform both operations on the same frame in the fast path and are suitable for FGL implementation.

3. Base RBridge Compatibility: As described in Section 3, FGL is compatible with base specification RBridges [RFC6325] acting as transit RBridges and, as described in Section 5.4, there is no particular problem in mixing VL and FGL RBridge on the same link.

4. Alternate Priority: The encoding specified in Section 2.3 provides for a new priority in the two bytes with the high order 12 FGL bits and a place to preserve the original user priority, so it can be restored on egress, within the two bytes with the low order 12 FGL bits.

8. Allocation Considerations

Allocations by the IEEE Registration Authority and IANA are listed below.

8.1 IEEE Allocation Considerations

The IEEE Registration Authority has assigned EtherType TBD for EX-TAG.

8.2 IANA Considerations

IANA is requested to allocate capability bit TBD (0 recommended) in the TRILL-VER sub-TLV capability bits to indicate an RBridge is FGL-capable.

9. 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.

Confusion between a frame with VLAN-X labeling and FGL label (X.Y) is a potential problem: A TRILL Data frame with FGL label (X.Y) could be egressed to VLAN-X by a VL RBridge that is Appointed Forwarder for VLAN-X on one of its ports. This is solved by prohibiting FGL RBridges from ingressing to FGL labeling (X.Y) if the RBridge campus is misconfigured so that a VL edge RBridge is reporting connectivity to VLAN-X while label (X.Y) is in use.

10. Acknowledgements

The comments and contributions of the following are gratefully acknowledged:

Anoop Ghanwani, Sujay Gupta, Jon Hudson, Vishwas Manral, Erik Nordmark, and Ilya Varlashkin.

11. Normative References

[IS-IS] - ISO/IEC 10589:2002, Second Edition, "Intermediate System to Intermediate System Intra-Domain Routeing Exchange Protocol for use in Conjunction with the Protocol for Providing the Connectionless-mode Network Service (ISO 8473)", 2002.

[802.1Q] - IEEE 802.1, "IEEE Standard for Local and metropolitan area networks - Virtual Bridged Local Area Networks", IEEE Std 802.1Q-2011, May 2011.

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

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

[RFC6326bis] - Eastlake, D., Banerjee, A., Dutt, D., Perlman, R., and A. Ghanwani, "Transparent Interconnection of Lots of Links (TRILL) Use of IS-IS", draft-eastlake-isis-rfc6326bis-00.txt, work in progress.

12. Informative References

[RFC5556] - Touch, J. and R. Perlman, "Transparent Interconnection of Lots of Links (TRILL): Problem and Applicability Statement", RFC 5556, May 2009.

[RFC6327] - Eastlake 3rd, D., Perlman, R., Ghanwani, A., Dutt, D., and V. Manral, "Routing Bridges (RBridges): Adjacency", RFC 6327, July 2011

[RFCaf] - Perlman, R., D. Eastlake, A. Banerjee, H. Fangwei, "RBridges: Appointed Forwarders", draft-ietf-trill-rbridge-af, work in progress.

Authors' Addresses

Donald Eastlake 3rd
Huawei Technologies
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.156 Beiqing Rd.
Z-park ,Shi-Chuang-Ke-Ji-Shi-Fan-Yuan,Hai-Dian District,
Beijing 100095 P.R. China

Email: zhangmingui@huawei.com

Puneet Agarwal
Broadcom Corporation
3151 Zanker Road
San Jose, CA 95134 USA

Phone: +1-949-926-5000
Email: pagarwal@broadcom.com

Dinesh G. Dutt
Cisco Systems
170 Tasman Drive
San Jose, CA 95134-1706 USA

Phone: +1-408-527-0955
Email: ddutt@cisco.com

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

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