Internet Engineering Task Force (IETF)
Request for Comments: 7172
Updates: 6325
Category: Standards Track
ISSN: 2070-1721

D. Eastlake 3rd
M. Zhang
Huawei
P. Agarwal
Broadcom
R. Perlman
Intel Labs
D. Dutt
Cumulus Networks

May 2014

Transparent Interconnection of Lots of Links (TRILL): Fine-Grained Labeling

Abstract

The IETF has standardized Transparent Interconnection of Lots of Links (TRILL), a protocol for least-cost transparent frame routing in multi-hop networks with arbitrary topologies and link technologies, using link-state routing and a hop count. The TRILL base protocol standard supports the labeling of TRILL Data packets with up to 4K IDs. However, there are applications that require a larger number of labels providing configurable isolation of data. This document updates RFC 6325 by specifying optional extensions to the TRILL base protocol to safely accomplish this. These extensions, called fine-grained labeling, are primarily intended for use in large data centers, that is, those with more than 4K users requiring configurable data isolation from each other.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 5741.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc7172.

Table of Contents

1. Introduction 4
1.1. Terminology 5
1.2. Contributors 5
2. Fine-Grained Labeling 5
2.1. Goals 6
2.2. Base Protocol TRILL Data Labeling 7
2.3. Fine-Grained Labeling (FGL) 7
2.4. Reasons for VL and FGL Coexistence 9
3. VL versus FGL Label Differences 10
4. FGL Processing 11
4.1. Ingress Processing 11
4.1.1. Multi-Destination FGL Ingress 11
4.2. Transit Processing 12
4.2.1. Unicast Transit Processing 12
4.2.2. Multi-Destination Transit Processing 12
4.3. Egress Processing 13
4.4. Appointed Forwarders and the DRB 14
4.5. Distribution Tree Construction 14
4.6. Address Learning 15
4.7. ESADI Extension 15
5. FGL TRILL Interaction with VL TRILL 15
5.1. FGL and VL Mixed Campus 15
5.2. FGL and VL Mixed Links 17
5.3. Summary of FGL-Safe Requirements 18
6. IS-IS Extensions 19
7. Comparison with Goals 19
8. Allocation Considerations 20
8.1. IEEE Allocation Considerations 20
8.2. IANA Considerations 20
9. Security Considerations 20
Appendix A. Serial Unicast 22
Appendix B. Mixed Campus Characteristics 23
B.1. Mixed Campus with High Cost Adjacencies 23
B.2. Mixed Campus with Data Blocked Adjacencies 24
Acknowledgements 25
References 25
Normative References 25
Informative References 26

1. Introduction

The IETF has standardized the Transparent Interconnection of Lots of Links (TRILL) protocol [RFC6325], which provides a solution for least-cost transparent routing in multi-hop networks with arbitrary topologies and link technologies, using [IS-IS] [RFC6165] [RFC7176] link-state routing and a hop count. TRILL switches are sometimes called RBridges (Routing Bridges).

The TRILL base protocol standard supports the labeling of TRILL Data packets with up to 4K IDs. However, there are applications that require a larger number of labels of data for configurable isolation based on different tenants, service instances, or the like. This document updates [RFC6325] by specifying optional extensions to the TRILL base protocol to safely accomplish this. These extensions, called fine-grained labeling, are primarily intended for use in large data centers, that is, those with more than 4K users requiring configurable data isolation from each other.

This document describes a format for allowing a data label of 24 bits, known as a "fine-grained label", or FGL. It also describes coexistence and migration from current RBridges, known as "VL" (for "VLAN Labeled") RBridges, to TRILL switches that can support FGL ("Fine-Grained Labeled") packets. Because various VL implementations might handle FGL packets incorrectly, FGL packets cannot be introduced until either all VL RBridges are upgraded to what we will call "FGL-safe", which means that they will not "do anything bad" with FGL packets, or all FGL RBridges take special precautions on any port by which they are connected to a VL RBridge. FGL-safe requirements are summarized in Section 5.3.

It is hoped that many RBridges can become FGL-safe through a software upgrade. VL RBridges and FGL-safe RBridges can coexist without any disruption to service, as long as no FGL packets are introduced.

If all RBridges are upgraded to FGL-safe, FGL traffic can be successfully handled by the campus without any topology restrictions. The existence of FGL traffic is known to all FGL RBridges because some RBridge (say, RB3) that might source or sink FGL traffic will advertise interest in one or more fine-grained labels in its contribution to the link state (its LSP). If any VL RBridges remain at the point when any RBridge announces that it might source or sink FGL traffic, the adjacent FGL-safe RBridges MUST ensure that no FGL packets are forwarded to their VL RBridge neighbor(s). The details are specified in Section 5.1 below.

1.1. Terminology

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

DEI - Drop Eligibility Indicator [802.1Q].

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

FGL-edge - An FGL TRILL switch advertising interest in an FGL label.

FGL link - A link where all of the attached TRILL switches are FGL.

FGL-safe - A TRILL switch that can safely be given an FGL data packet, as summarized in Section 5.3.

RBridge - Alternative name for a TRILL switch.

TRILL switch - Alternative name for an RBridge.

VL - VLAN Labeling or VLAN Labeled or VLAN Label.

VL link - A link where any one or more of the attached RBridges are VL.

VL RBridge - A TRILL switch that supports VL but is not FGL-safe.

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].

1.2. Contributors

Thanks for the contributions of the following:

Tissa Senevirathne and Jon Hudson

2. Fine-Grained Labeling

The essence of Fine-Grained Labeling (FGL) is that (a) when frames are ingressed or created they may incorporate a data label from a set consisting of significantly more than 4K labels, (b) TRILL switch ports can be labeled with a set of such fine-grained data labels,

and (c) an FGL TRILL Data packet cannot be egressed through a TRILL switch port unless its fine-grained label (FGL) matches one of the data labels of the port.

Section 2.1 lists FGL goals. Section 2.2 briefly outlines the more coarse TRILL base protocol standard [RFC6325] data labeling. Section 2.3 outlines FGL for TRILL Data packets. Section 2.4 discusses VL and FGL coexistence.

2.1. Goals

There are several goals that would be desirable for FGL 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 labeling supported by [RFC6325] provides for only 2**12 - 2 valid identifiers or labels (VLANs). A substantially larger number is required.

2. Silicon

   Fine-grained labeling (FGL) should, to the extent practical, use existing features, processing, and fields that are already supported in many fast path silicon implementations that support the TRILL base protocol.

3. Base RBridge Interoperation

   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 if RBridges not implementing the FGL features can exchange VL TRILL Data packets with FGL TRILL switches.

4. Alternate Priority

   Under some circumstances, it would be desirable for traffic from an attached non-TRILL network to be handled, while transiting a TRILL network, with a different priority from the priority of the original native frames. This could be accomplished by the ingress TRILL switch assigning a different priority to the FGL TRILL Data packet resulting from ingressing the native frames. The original priority should be restored on egress.

2.2. Base Protocol TRILL Data Labeling

This section provides a brief review of the [RFC6325] TRILL Data packet VL Labeling and changes the description of the TRILL Header by moving the point at which the TRILL Header ends. This change in description does not involve any change in the bits on the wire or in the behavior of VL TRILL switches.

VL TRILL Data packets have the structure shown below:

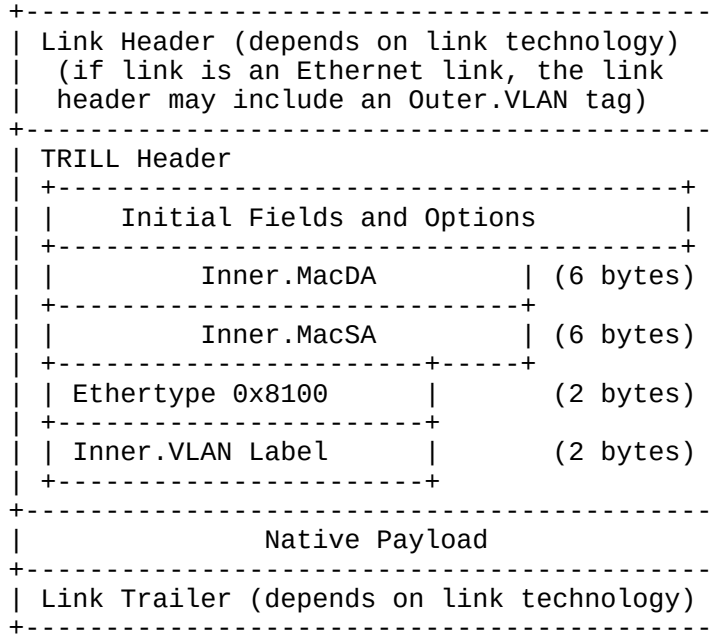


Figure 1: TRILL Data with VL

In the base protocol as specified in [RFC6325], the 0x8100 value is always present and is followed by the Inner.VLAN field, which includes the 12-bit VL.

2.3. Fine-Grained Labeling (FGL)

FGL expands the variety of data labels available under the TRILL protocol to include a fine-grained label (FGL) with a 12-bit high order part and a 12-bit low order part. In this document, FGLs are denoted as "(X.Y)", where X is the high order part and Y is the low order part of the FGL.

FGL TRILL Data packets have the structure shown below.

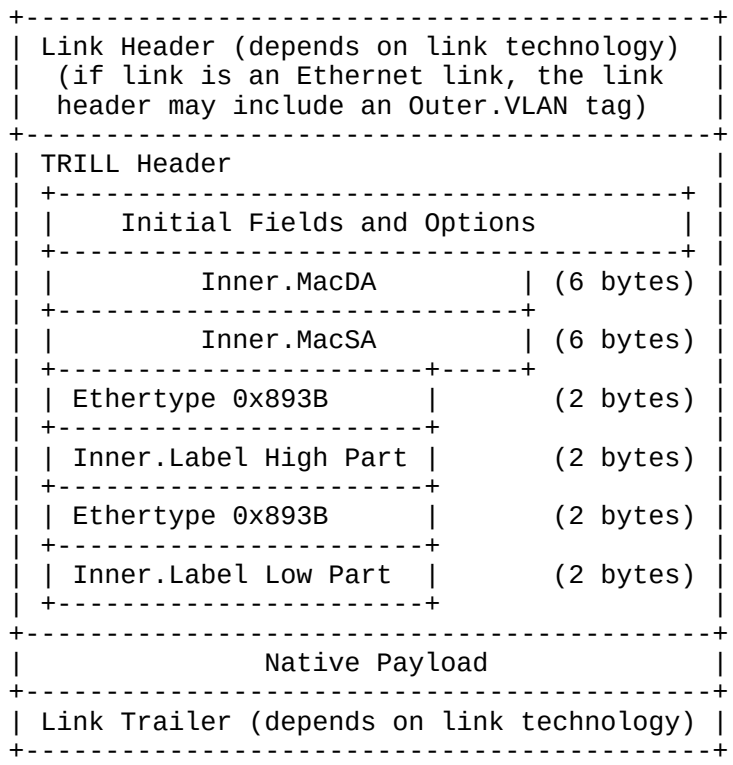


Figure 2: TRILL Data with FGL

For FGL packets, the inner Media Access Control (MAC) address fields are followed by the FGL information using 0x893B. There MUST be two occurrences of 0x893B, as shown. Should a TRILL switch processing an FGL TRILL Data packet notice that the second occurrence is actually some other value, it MUST discard the packet. (A TRILL switch transiting a TRILL Data packet is not required to examine any fields past the initial fixed fields and options, although it may do so to support Equal-Cost Multi-Path (ECMP) or distribution tree pruning.)

The two bytes following each 0x893B have, in their low order 12 bits, fine-grained label information. The upper 4 bits of those two bytes are used for a 3-bit priority field and one Drop Eligibility Indicator (DEI) bit as shown below.

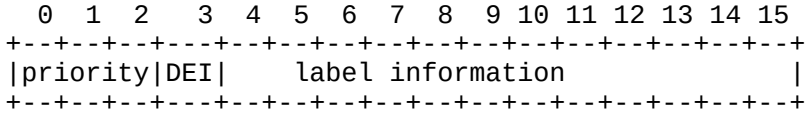


Figure 3: FGL Part Data Structure

The priority field of the Inner.Label High Part is the priority used for frame transport across the TRILL campus from ingress to egress. The label bits in the Inner.Label High Part are the high order part of the FGL, and those bits in the Inner.Label Low Part are the low order part of the FGL. The priority field of the Inner.Label Low Part is remembered from the data frame as ingressed and is restored on egress.

The appropriate FGL value for an ingressed or locally originated native frame is determined by the ingress TRILL switch port as specified in Section 4.1.

2.4. Reasons for VL and FGL Coexistence

For several reasons, as listed below, it is desirable for FGL TRILL switches to be able to handle both FGL and VL TRILL Data packets.

- Continued support of VL packets means that, by taking the precautions specified herein, in many cases such arrangements as VL TRILL switches easily exchanging VL packets through a core of FGL TRILL switches are possible.

- Due to the way TRILL works, it may be desirable to have a maintenance VLAN or FGL [RFC7174] in which all TRILL switches in the campus indicate interest. It will be simpler to use the same type of label for all TRILL switches for this purpose. That implies using VL if there might be any VL TRILL switches in the campus.

- If a campus is being upgraded from VL to FGL, continued support of VL allows long-term support of edges labeled as VL.

3. VL versus FGL Label Differences

There are differences between the semantics across a TRILL campus for TRILL Data packets that are data labeled with VL and FGL.

With VL, data label IDs have the same meaning throughout the campus and are from the same label space as the C-VLAN IDs used on Ethernet links to end stations.

The larger FGL data label space is a different space from the VL data label space. For ports configured for FGL, the C-VLAN on an ingressed native frame is stripped and mapped to the FGL data label space with a potentially different mapping for each port. A similar FGL-to-C-VLAN mapping occurs per port on egress. Thus, for ports configured for FGL, the native frame C-VLAN on one link corresponding to an FGL can be different from the native frame C-VLAN corresponding to that same FGL on a different link elsewhere in the campus or even a different link attached to the same TRILL switch. The FGL label space is flat and does not hierarchically encode any particular number of native frame C-VLAN bits or the like. FGLs appear only inside TRILL Data packets after the inner MAC addresses.

It is the responsibility of the network manager to properly configure the TRILL switches in the campus to obtain the desired mappings. Such configuration is expected to be automatic in many cases, based on configuration databases and orchestration systems.

With FGL TRILL switches, many things remain the same because an FGL can appear only as the Inner.Label inside a TRILL Data packet. As such, only TRILL-aware devices will see a fine-grained label. The Outer.VLAN that may appear on native frames and that may appear on TRILL Data packets if they are on an Ethernet link can only be a C-VLAN tag. Thus, ports of FGL TRILL switches, up through the usual VLAN and priority processing, act as they do for VL TRILL switches: TRILL switch ports provide a C-VLAN ID for an incoming frame and accept a C-VLAN ID for a frame being queued for output. Appointed Forwarders [RFC6439] on a link are still appointed for a C-VLAN. The Designated VLAN for an Ethernet link is still a C-VLAN.

FGL TRILL switches have capabilities that are a superset of those for VL TRILL switches. FGL TRILL switch ports can be configured for FGL or VL, with VL being the default. As with a base protocol [RFC6325] TRILL switch, an unconfigured FGL TRILL switch port reports an untagged frame it receives as being in VLAN 1.

4. FGL Processing

This section specifies ingress, transit, egress, and other processing details for FGL TRILL switches. A transit or egress FGL TRILL switch determines that a TRILL Data packet is FGL by detecting that the Inner.MacSA is followed by 0x893B.

4.1. Ingress Processing

FGL-edge TRILL switch ports are configurable to ingress native frames as FGL. Any port not so configured performs the previously specified [RFC6325] VL ingress processing on native frames resulting in a VL TRILL Data packet. (There is no change in Appointed Forwarder logic (see Section 4.4).) An FGL-safe TRILL switch may have only VL ports, in which case it is not required to support the capabilities for FGL ingress described in this section.

FGL-edge TRILL switches support configurable per-port mapping from the C-VLAN of a native frame, as reported by the ingress port, to an FGL. FGL TRILL switches MAY support other methods to determine the FGL of an incoming native frame, such as methods based on the protocol of the native frame or based on local knowledge.

The FGL ingress process MUST copy the priority and DEI (Drop Eligibility Indicator) associated with an ingressed native frame to the upper 4 bits of the Inner.Label Low Order part. It SHOULD also associate a possibly different mapped priority and DEI with an ingressed frame, but a TRILL switch might not be able to do so because of implementation limitations. The mapped priority is placed in the Inner.Label High Part. If such mapping is not supported, then the original priority and DEI MUST be placed in the Inner.Label High Part.

4.1.1. Multi-Destination FGL Ingress

If a native frame that has a broadcast, multicast, or unknown MAC destination address is FGL ingressed, it MUST be handled in one of the following two ways. The choice of which method to use can vary from frame to frame, at the choice of the ingress TRILL switch.

1. Ingress as a TRILL multi-destination data packet (TRILL Header M bit = 1) on a distribution tree rooted at a nickname held by an FGL RBridge or by the pseudonode of an FGL link. FGL TRILL Data packets MUST NOT be sent on a tree rooted at a nickname held by a VL TRILL switch or by the pseudonode of a VL link.

2. Serially TRILL unicast the ingressed frame to the relevant egress TRILL switches by using a known unicast TRILL Header (M bit = 0). An FGL ingress TRILL switch SHOULD unicast a multi-destination TRILL Data packet if there is only one relevant egress FGL TRILL switch. The relevant egress TRILL switches are determined by starting with those announcing interest in the frame's (X.Y) label. That set SHOULD be further filtered based on multicast listener and multicast router attachment LSP announcements if the native frame was a multicast frame.

Using a TRILL unicast header for a multi-destination frame when it has only one actual destination RBridge almost always improves traffic spreading and decreases latency as discussed in Appendix A. How to decide whether to use a distribution tree or serial unicast for a multi-destination TRILL Data packet that has more than one destination TRILL switch is beyond the scope of this document.

4.2. Transit Processing

Any FGL TRILL switch MUST be capable of TRILL Data packet transit processing. Such processing is fairly straightforward as described in Section 4.2.1 for known unicast TRILL Data packets and in Section 4.2.2 for multi-destination TRILL Data packets.

4.2.1. Unicast Transit Processing

There is very little change in TRILL Data packet unicast transit processing. A transit TRILL switch forwards any unicast TRILL Data packet to the next hop towards the egress TRILL switch as specified in the TRILL Header. All transit TRILL switches MUST take the priority and DEI used to forward a packet from the Inner.VLAN label or the FGL Inner.Label High Part. These bits are in the same place in the packet.

An FGL TRILL switch MUST properly distinguish flows if it provides ECMP for unicast FGL TRILL Data packets.

4.2.2. Multi-Destination Transit Processing

Multi-destination TRILL Data packets are forwarded on a distribution tree selected by the ingress TRILL switch, except that an FGL ingress TRILL switch MAY TRILL unicast such a frame to all relevant egress TRILL switches, all as described in Section 4.1. The distribution trees do not distinguish between FGL and VL multi-destination packets, except in pruning behavior if they provide pruning. There is no change in the Reverse Path Forwarding Check.

An FGL TRILL switch (say, RB1) having an FGL multi-destination frame for label (X.Y) to forward on a distribution tree SHOULD prune that tree based on whether there are any TRILL switches on a tree branch that are advertising connectivity to label (X.Y). In addition, RB1 SHOULD prune multicast frames based on reported multicast listener and multicast router attachment in (X.Y).

Pruning is an optimization. If a transit TRILL switch does less pruning than it could, there may be greater link utilization than strictly necessary but the campus will still operate correctly. A transit TRILL switch MAY prune based on an arbitrary subset of the bits in the FGL label, for example, only the High Part or only the Low Part of the label.

4.3. Egress Processing

Egress processing is generally the reverse of ingress progressing described in Section 4.1. An FGL-safe TRILL switch may have only VL ports, in which case it is not required to support the capabilities for FGL egress described in this section.

An FGL-edge TRILL switch MUST be able to convert, in a configurable fashion, from the FGL in an FGL TRILL Data packet it is egressing to the C-VLAN ID for the resulting native frame with different mappings on a per-port basis. The priority and DEI of the egressed native frame are taken from the Inner.Label Low Order Part. A port MAY be configured to strip output VLAN tagging.

It is the responsibility of the network manager to properly configure the TRILL switches in the campus to obtain the desired mappings.

FGL egress is similar to VL egress, as follows:

1. If the Inner.MacDA is All-Egress-RBridges, special processing applies, based on the payload Ethertype (for example, End-Station Address Distribution Information (ESADI) [RFC6325] or RBridge Channel [RFC7178]), and if the payload Ethertype is unknown, the packet is discarded. If the Inner.MacDA is not All-Egress-RBridges, then either item 2 or item 3 below applies, as appropriate.

2. A known unicast FGL TRILL Data packet (TRILL Header M bit = 0) with a unicast Inner.MacDA is egressed to the FGL port or ports matching its FGL and Inner.MacDA. If there are no such ports, it is flooded out of all FGL ports that have its FGL, except any ports for which the TRILL switch has knowledge that the frame’s Inner.MacDA cannot be present on the link out of that port.

3. A multi-destination FGL TRILL Data packet is decapsulated and flooded out of all ports that have its FGL, subject to multicast pruning. The same processing applies to a unicast FGL TRILL Data packet with a broadcast or multicast Inner.MacDA that might be received due to serial unicast.

An FGL TRILL switch MUST NOT egress an FGL packet with label (X.Y) to any port not configured with that FGL, even if the port is configured to egress VL packets in VLAN X.

FGL TRILL switches MUST accept multi-destination TRILL Data packets that are sent to them as TRILL unicast packets (packets with the TRILL Header M bit set to 0). They locally egress such packets, if appropriate, but MUST NOT forward them (other than egressing them as native frames on their local links).

4.4. Appointed Forwarders and the DRB

There is no change in adjacency [RFC7177], DRB (Designated RBridge) election, or Appointed Forwarder logic [RFC6439] on a link, regardless of whether some or all the ports on the link are for FGL TRILL switches, with one exception: implementations SHOULD provide that their default priority for a VL RBridge port to be the DRB is less than their default priority for an FGL RBridge to be the DRB. This will assure that, in the unconfigured case, an FGL RBridge will be elected DRB when using that implementation.

4.5. Distribution Tree Construction

All distribution trees are calculated as provided for in the TRILL base protocol standard [RFC6325] as updated by [RFC7180], with the exception that the default tree root priority for a nickname held by an FGL TRILL switch or an FGL link pseudonode is 0x9000. As a result, they will be chosen in preference to VL nicknames in the absence of configuration. If distribution tree roots are configured, there MUST be at least one tree rooted at a nickname held by an FGL TRILL switch or by an FGL link pseudonode. If distribution tree roots are misconfigured so there would not be such a tree, then the highest priority FGL nickname to be a tree root is used to construct an additional tree, regardless of configuration. (VL TRILL switches will not know about this additional distribution tree but, through the use of Step (A) or (B) in Section 5.1, no VL TRILL switch should ever receive a multi-destination TRILL Data packet using this additional tree.)

4.6. Address Learning

An FGL TRILL switch learns addresses from the data plane on ports configured for FGL based on the fine-grained label rather than the native frame's VLAN. Addresses learned from ingressed native frames on FGL ports are logically represented by { MAC address, FGL, port, confidence, timer }, while remote addresses learned from egressing FGL packets are logically represented by { MAC address, FGL, remote TRILL switch nickname, confidence, timer }.

4.7. ESADI Extension

The TRILL ESADI (End-Station Address Distribution Information) protocol is specified in [RFC6325] as optionally transmitting MAC address connection information through TRILL Data packets between participating TRILL switches over the virtual link provided by the TRILL multi-destination packet distribution mechanism. In [RFC6325], the VL to which an ESADI packet applies is indicated only by the Inner.VLAN label, and no indication of that VL is allowed within the ESADI payload.

ESADI is extended to support FGL by providing for the indication of the FGL to which an ESADI packet applies only in the Inner.Label of that packet, and no indication of that FGL is allowed within the ESADI payload.

5. FGL TRILL Interaction with VL TRILL

This section discusses mixing FGL-safe and VL TRILL switches in a campus. It does not apply if the campus is entirely FGL-safe or if there are no FGL-edges. Section 5.1 specifies what behaviors are needed to render such mixed campuses safe. See also Appendix B for a discussion of campus characteristics when these behaviors are in use. Section 5.2 gives details of link-local mixed behavior.

It is best, if possible, for VL TRILL switches to be upgraded to FGL-safe before introducing FGL-edges (and therefore FGL data packets).

5.1. FGL and VL Mixed Campus

By definition, it is not possible for VL TRILL switches to safely handle FGL traffic, even if the VL TRILL switch is only acting in the transit capacity. If a TRILL switch can safely transit FGL TRILL Data packets, then it qualifies as FGL-safe but will still be assumed to be VL until it advertises in its LSP that it is FGL-safe.

VL frames are required to have 0x8100 at the beginning of the data label, where FGL frames have 0x893B. VL TRILL switches conformant to [RFC6325] should discard frames with this new value after the inner MAC addresses. However, if they do not discard such frames, they could be confused and egress them into the wrong VLAN (see Section 9 below) or persistently reorder them due to miscomputing flows for ECMP, or they could improperly prune their distribution if they are multi-destination so that they would fail to reach some intended destinations. Such difficulties are avoided by taking all practical steps to minimize the chance of a VL TRILL switch handling an FGL TRILL Data packet. These steps are specified below.

FGL-safe switches will report their FGL capability in LSPs. Thus, FGL-safe TRILL switches (and any management system with access to the link-state database) will be able to detect the existence of TRILL switches in the campus that do not support FGL.

Once a TRILL switch advertises an FGL-edge, any FGL-safe TRILL switch (RB1 in this discussion) that observes, on one of its ports, a VL RBridge on the link out of that port, MUST take Step (A) or (B) below for that port and also take Step (C) further below. ("Observes" means that it has an adjacency to the VL TRILL switch that is in any state other than Down [RFC7177] and holds an LSP fragment zero for it, showing that it is not FGL-safe.) Finally, for there to be full FGL connectivity, the campus topology must be such that all FGL TRILL switches are reachable from all other FGL TRILL switches without going through a VL TRILL switch.

(A) If RB1 can discard any FGL TRILL Data packet that would be output through a port where it observes a VL RBridge, while allowing the output of VL TRILL Data packets through that port, then

   A1. RB1 MUST so discard all FGL TRILL Data output packets that would otherwise be output through the port, and

   A2. For all adjacencies out of that port (even adjacencies to other FGL RBridges or a pseudonode) in the Report state [RFC7177], RB1 MUST report that adjacency cost as 2**23 greater than it would have otherwise reported, but not more than 2**24 - 2 (the highest link cost still usable in least-cost path calculations and distribution tree construction). This assures that if any path through FGL-safe TRILL switches exists, such a path will be computed.

(B) If RB1 cannot discard any FGL TRILL Data packet that would be output through a port where it observes a VL RBridge while allowing VL TRILL Data packets, then RB1 MUST, for all adjacencies out of that port (even adjacencies to other FGL-safe

RBridges or a pseudonode) in the Report state [RFC7177], report the adjacency cost as 2**24 - 1. As specified in IS-IS [RFC5305], that cost will stop the adjacency from being used in least-cost path calculations, including distribution tree construction (see Section 2.1 of [RFC7180]) but will still leave it visible in the topology and usable, for example, by any traffic engineered path mechanism.

(C) The roots for all distribution trees used for FGL TRILL Data packets must be nicknames held by an FGL-safe TRILL switch or by a pseudonode representing an FGL link. As provided in Section 4.5, there will always be such a distribution tree.

Using the increased adjacency cost specified in part A2 of Step (A) above, VL links will be avoided unless no other path is available for typical data center link speeds using the default link cost determination method specified in Item 1 of Section 4.2.4.4 of [RFC6325]. However, if links have low speed (such as about 100 megabits/second or less) or some non-default method is used for determining link costs, then link costs MUST be adjusted such that no adjacency between FGL-safe TRILL switches has a cost greater than 200,000.

To summarize, for a mixed TRILL campus to be safe once FGL-edges are introduced, it is essential that the steps above be followed by FGL-safe RBridges, to ensure that paths between such RBridges do not include VL RBridges, and to ensure that FGL packets are never forwarded to VL RBridges. That is, all FGL-safe switches MUST do Step (A) or (B) for any port out of which they observe a VL RBridge neighbor. Also, for full FGL connectivity, all FGL-safe TRILL switches MUST do Step (C) and be connected in a single FGL contiguous area.

5.2. FGL and VL Mixed Links

The usual DRB election operates on a link with mixed FGL and VL ports. If an FGL TRILL switch port is a DRB, it can handle all native traffic. It MUST appoint only other FGL TRILL switch ports as Appointed Forwarder for any VLANs that are to be mapped to FGL.

For VLANs that are not being mapped to FGL, if Step (A) is being followed (see Section 5.1), it can appoint either a VL or FGL TRILL switch for a VLAN on the link to be handled by a VL. If Step (B) is being followed, an FGL DRB MUST only appoint FGL Appointed Forwarders, so that all end stations will get service to the FGL campus. If a VL RBridge is a DRB, it will not understand that FGL TRILL switch ports are different. To the extent that Step (B) is in effect and a VL DRB handles native frames or appoints other VL TRILL

switch ports on a link to handle native frames for one or more VLANs, the end stations sending and receiving those native frames may be isolated from the FGL campus. When a VL DRB happens to appoint an FGL port as Appointed Forwarder for one or more VLANs, the end stations sending and receiving native frames in those VLANs will get service to the FGL campus.

5.3. Summary of FGL-Safe Requirements

The list below summarizes the requirements for a TRILL switch to be FGL-safe.

1. For both unicast and multi-destination data, RB1 MUST NOT forward an FGL packet to a VL neighbor RB2. This is accomplished as specified in Section 5.1.

2. For both unicast and multi-destination data, RB1 MUST NOT egress a packet onto a link that does not belong in that FGL.

3. For unicast data, RB1 must forward the FGL packet properly to the egress nickname in the TRILL Header. This means that it MUST NOT delete the packet because of not having the expected VLAN tag, it MUST NOT insert a VLAN tag, and it MUST NOT misclassify a flow so as to persistently misorder packets, because the TRILL fields are now 4 bytes longer than in VL TRILL packets.

4. For multi-destination data, RB1 must forward the packet properly along the specified tree. This means that RB1 MUST NOT falsely prune the packet. RB1 is allowed not to prune at all, but it MUST NOT prevent an FGL packet from reaching all the links with that FGL by incorrectly refusing to forward the FGL packet along a branch in the tree.

5. RB1 must advertise, in its LSP, that it is FGL-safe.

Point 1 above, for a TRILL switch to correctly support ECMP, and point 2, for a TRILL switch to correctly prune distribution trees, require that the TRILL switch properly recognize and distinguish between the two Ethertypes that can occur immediately after the Inner.MacSA in a TRILL Data packet.

6. IS-IS Extensions

Extensions related to TRILL's use of IS-IS are required to support FGL and must include the following:

1. A method for a TRILL switch to announce itself in its LSP as FGL-safe (see Section 8.2).

2. A sub-TLV analogous to the Interested VLANs and Spanning Tree Roots sub-TLV of the Router Capabilities TLV but indicating FGLs rather than VLs. This is called the Interested Labels and Spanning Tree Roots (INT-LABEL) sub-TLV in [RFC7176].

3. Sub-TLVs analogous to the GMAC-ADDR sub-TLV of the Group Address TLV that specifies an FGL rather than a VL. These are called the GLMAC-ADDR, GLIP-ADDR, and GLIPV6-ADDR sub-TLVs in [RFC7176].

7. Comparison with Goals

Comparing TRILL FGL, as specified in this document, with the goals given in Section 2.1, we find the following:

1. Fine-Grained: FGL provides 2**24 labels, vastly more than the 4094 (4K) VLAN labels supported in TRILL as specified in [RFC6325].

2. Silicon: Existing TRILL fast path silicon chips can 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 native-frame-to-FGL mapping and the encoding of the FGL as specified herein, as well as the inverse decoding and mapping. Some existing silicon chips can perform only one of these operations on a frame in one pass through the fast path; however, other existing chips are believed to be able to perform both operations on the same frame in one pass through their fast path. It is also believed that most FGL TRILL switches will be capable of having their ports configured to discard FGL packets. Such a capability makes interoperation with VL TRILL switches practical using Step (A) as opposed to Step (B) (see Section 5.1).

3. Base RBridge Interoperation: As described in Section 3, FGL is not generally compatible with TRILL switches conformant to the base specification [RFC6325]. In particular, a VL TRILL switch cannot be an FGL TRILL switch because there is a risk that it would mishandle FGL packets. However, a contiguous set of VL TRILL switches can exchange VL frames, regardless of the presence of FGL TRILL switches in the campus. The provisions of Section 5 support reasonable interoperation and migration scenarios.

4. Alternate Priority: The encoding specified in Section 2.3 and the ingress/egress processing specified in Section 4 provide for a new priority and DEI in the Inner.Label High Part and a place to preserve the original user priority and DEI in the Low Part so that it can be restored on egress.

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 0x893B for TRILL FGL.

8.2. IANA Considerations

IANA has allocated capability flag 1 in the TRILL-VER sub-TLV capability flags [RFC7176] to indicate that a TRILL switch is FGL-safe.

9. Security Considerations

See [RFC6325] for general TRILL security considerations.

As with any communications system, end-to-end encryption and authentication should be considered for sensitive data. In this case, that would be encryption and authentication extending from a source end station and carried through the TRILL campus to a destination end station.

Confusion between a packet with VL X and a packet with FGL (X.Y) or confusion due to a malformed frame is a potential problem if an FGL TRILL switch did not properly check for the occurrence of 0x8100 or 0x893B immediately after the Inner.MacSA (see Sections 2.2 and 2.3) and handle the frame appropriately.

[RFC6325] requires that the Ethertype immediately after the Inner.MacSA be 0x8100. A VL TRILL switch that did not discard a packet with some other value there could cause problems. If it received a TRILL Data packet with FGL (X.Y) or with junk after the Inner.MacSA that included X where a VLAN ID would appear, then:

1. It could egress the packet to an end station in VLAN X. If the packet was a well-formed FGL frame, the payload of such an egressed native frame would appear to begin with Ethertype 0x893B, which would likely be discarded by an end station. In any case,

such an egress would almost certainly be a violation of security policy requiring the configurable separation of differently labeled data.

2. If the packet was multi-destination and the TRILL switch pruned the distribution tree, it would incorrectly prune it on the basis of VLAN X. For an FGL packet, this would probably lead to the multi-destination data packet not being delivered to all of its intended recipients.

Possible problems with an FGL TRILL switch that (a) received a TRILL Data packet with junk after the Inner.MacSA that included X where a VLAN ID would appear and (b) did not check the Ethertype immediately after the Inner.MacSA would be that it could improperly egress the packet in VLAN X, violating security policy. If the packet was multi-destination and was improperly forwarded, it should be discarded by properly implemented TRILL switches downstream in the distribution tree and never egressed, but the propagation of the packet would still waste bandwidth.

To avoid these problems, all TRILL switches MUST check the Ethertype immediately after the Inner.MacSA and, if it is a value they do not know how to handle, either discard the frame or make no decisions based on any data after that Ethertype. In addition, care must be taken to avoid FGL packets being sent to or through VL TRILL switches that will discard them if the VL TRILL switch is properly implemented or mishandle them if it is not properly implemented. This is accomplished as specified in Section 5.1.

Appendix A. Serial Unicast

This informational appendix discusses the advantages and disadvantages of using serial unicast instead of a distribution tree for multi-destination TRILL Data packets. See Sections 4.1 and 4.3. This document requires that FGL TRILL switches accept serial unicast, but there is no requirement that they be able to send serial unicast.

Consider a large TRILL campus with hundreds of TRILL switches in which, say, 300 end stations are in some particular FGL data label.

At one extreme, if all 300 end stations were on links attached to a single TRILL switch, then no other TRILL switch would be advertising interest in that FGL. As a result, it is likely that because of pruning a multi-destination (say, broadcast) frame from one such end station would not be sent to any another TRILL switch, even if put on a distribution tree.

At the other extreme, assume that the 300 end stations are attached, one each, to 300 different TRILL switches; in that case, you are almost certainly better off using a distribution tree because if you tried to serially unicast you would have to output 300 copies, probably including multiple copies through the same port, and would cause much higher link utilization.

Now assume that these 300 end stations are connected to exactly two TRILL switches, say, 200 to one and 100 to the other. Using unicast TRILL Data packets between these two TRILL switches is best because the frames will follow least-cost paths, possibly with such traffic spread over a number of least-cost paths with equal cost. On the other hand, if distribution trees were used, each frame would be constrained to the tree used for that frame and would likely follow a higher cost route and only a single path would be available per tree. Thus, this document says that unicast SHOULD be used if there are exactly two TRILL switches involved.

The decision of whether to use a distribution tree or serial unicast if the end stations are connected to more than two TRILL switches is more complex. Which would be better would depend on many factors, including network topology and application data patterns. How to make this decision in such cases is beyond the scope of this document.

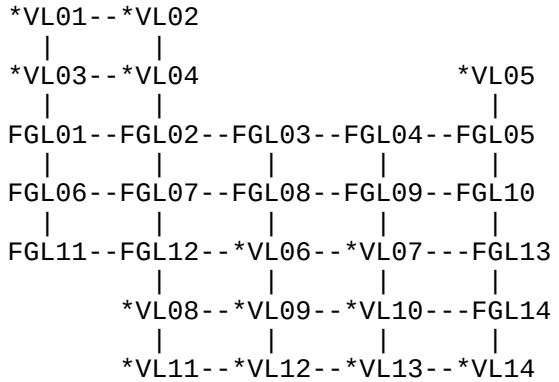
Appendix B. Mixed Campus Characteristics

This informational appendix describes the characteristics of a TRILL campus with mixed FGL-safe and VL TRILL switches for two cases: Appendix B.1 discusses the case where all FGL adjacencies with VL are handled by Step (A) in Section 5.1, and Appendix B.2 discusses the case where all FGL adjacencies with VL are handled by Step (B) in Section 5.1.

B.1. Mixed Campus with High Cost Adjacencies

If the FGL TRILL switches use Step (A) in Section 5.1, then VL and FGL TRILL switches will be able to interoperate for VL traffic. Least-cost paths will avoid any FGL -> VL TRILL switch hops unless no other reasonable path is available. In conjunction with Section 4.5, there will be at least one distribution tree rooted at a nickname held by an FGL TRILL switch or the pseudonode for an FGL link. Furthermore, if the FGL TRILL switches in the campus form a single contiguous island, this distribution tree will have a fully connected sub-tree covering that island. Thus, any FGL TRILL Data packets sent on this tree will be able to reach any other FGL TRILL switch without attempting to go through any VL TRILL switches. (Such an attempt would cause the FGL packet to be discarded as specified in part A1 of Step (A).)

If supported, Step (A) is particularly effective in a campus with an FGL TRILL switch core and VL TRILL switches in one or more islands around that core. For example, consider the campus below. This campus has an FGL core consisting of FGL01 to FGL14 and three VL islands consisting of VL01 to VL04, VL05, and VL06 to VL14.



Assuming that the FGL TRILL switches in this campus all implement Step (A), then end stations connected through a VL port can be connected anywhere in the campus to VL or FGL TRILL switches and, if in the same VLAN, will communicate. End stations connected through an FGL port on FGL TRILL switches will communicate if their local VLANs are mapped to the same FGL.

Due to the high cost of FGL-to-VL adjacencies used in path computations, VL TRILL switches are avoided on paths between FGL TRILL switches. For example, even if the speed and default adjacency cost of all the connections shown above were the same, traffic from FGL12 to FGL13 would follow the 5-hop path FGL12 - FGL07 - FGL08 - FGL09 - FGL10 - FGL13 rather than the 3-hop path FGL12 - VL09 - VL10 - FGL14.

B.2. Mixed Campus with Data Blocked Adjacencies

If the FGL TRILL switches use Step (B) in Section 5.1, then least-cost and distribution tree TRILL Data communication between VL and FGL TRILL switches is blocked, although TRILL IS-IS communication is normal. This data blocking, although implemented only by FGL TRILL switches, has relatively symmetric effects. The following paragraphs assume that such data blocking between VL and FGL is in effect throughout the campus.

A campus of mostly FGL TRILL switches implementing Step (B) with a few isolated VL TRILL switches scattered throughout will work well in terms of connectivity for end stations attached to those FGL switches, except that they will be unable to communicate with any end stations for which a VL switch is appointed forwarder. The VL TRILL switches will be isolated and will only be able to route TRILL Data to the extent that they happen to be contiguously connected to other VL TRILL switches. Distribution trees computed by the FGL switches will not include any VL switches (see Section 2.1 of [RFC7180]).

A campus of mostly VL TRILL switches with a few isolated FGL TRILL switches scattered throughout will also work reasonably well as described immediately above but with all occurrences of "FGL" and "VL" swapped.

However, a campus so badly misconfigured that it consists of a randomly intermingled mixture of VL and FGL TRILL switches using Step (B) is likely to offer very poor data service, due to many links being blocked for data.

Acknowledgements

The comments and suggestions of the following, listed in alphabetic order, are gratefully acknowledged:

Stewart Bryant, Spencer Dawkins, Adrian Farrel, Anoop Ghanwani, Sujay Gupta, Weiguo Hao, Phanidhar Koganti, Yizhou Li, Vishwas Manral, Rajeev Manur, Thomas Narten, Gayle Nobel, Erik Nordmark, Pete Resnick, Olen Stokes, Sean Turner, Ilya Varlashkin, and Xuxiaohu.

References

Normative References

[802.1Q] IEEE 802.1, "IEEE Standard for Local and metropolitan area networks--Media Access Control (MAC) Bridges and Virtual Bridged Local Area Networks", IEEE Std 802.1Q-2011, August 2011.

[IS-IS] ISO/IEC 10589:2002, Second Edition, "Information technology -- Telecommunications and information exchange between systems -- Intermediate System to Intermediate System intra-domain routeing information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode network service (ISO 8473)", 2002.

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

[RFC5305] Li, T. and H. Smit, "IS-IS Extensions for Traffic Engineering", RFC 5305, October 2008.

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

[RFC7176] Eastlake 3rd, D., Senevirathne, T., Ghanwani, A., Dutt, D., and A. Banerjee, "Transparent Interconnection of Lots of Links (TRILL) Use of IS-IS", RFC 7176, May 2014.

[RFC7177] Eastlake 3rd, D., Perlman, R., Ghanwani, A., Yang, H., and V. Manral, "Transparent Interconnection of Lots of Links (TRILL): Adjacency", RFC 7177, May 2014.

[RFC7180] Eastlake 3rd, D., Zhang, M., Ghanwani, A., Manral, V., and A. Banerjee, "Transparent Interconnection of Lots of Links (TRILL): Clarifications, Corrections, and Updates" RFC 7180, May 2014.

Informative References

[RFC6165] Banerjee, A. and D. Ward, "Extensions to IS-IS for Layer-2 Systems", RFC 6165, April 2011.

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

[RFC7174] Salam, S., Senevirathne, T., Aldrin, S., and D. Eastlake 3rd, "Transparent Interconnection of Lots of Links (TRILL) Operations, Administration, and Maintenance (OAM) Framework", RFC 7174, May 2014.

[RFC7178] Eastlake 3rd, D., Manral, V., Li, Y., Aldrin, S., and D. Ward, "Transparent Interconnection of Lots of Links (TRILL): RBridge Channel Support", RFC 7178, May 2014.

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

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

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

Dinesh G. Dutt
Cumulus Networks
1089 West Evelyn Avenue
Sunnyvale, CA 94086
USA

EMail: ddutt.ietf@hobbesdutt.com