

TRILL Working Group  
INTERNET-DRAFT  
Intended Status: Standard Track

Y. Li  
D. Eastlake  
W. Hao  
H. Chen  
Huawei Technologies  
R. Perlman  
EMC  
N. Nimmu  
Broadcom  
S. Chatterjee  
Cisco  
S. Rajagopalan  
IBM  
April 21, 2015

Expires: October 23, 2015

**TRILL: Data Label based Tree Selection for Multi-destination Data  
draft-ietf-trill-tree-selection-00**

Abstract

TRILL uses distribution trees to deliver multi-destination frames. Multiple trees can be used by an ingress RBridge for flows regardless of the VLAN, Fine Grained Label (FGL), and/or multicast group of the flow. Different ingress RBridges may choose different distribution trees for TRILL Data packets in the same VLAN, FGL, and/or multicast group. To avoid unnecessary link utilization, distribution trees should be pruned based on VLAN and/or FGL and/or multicast destination address. If any VLAN, FGL, or multicast group can be sent on any tree, for typical fast path hardware, the amount of pruning information is multiplied by the number of tree; however, there is a limited capacity for such pruning information.

This document specifies an optional facility to restrict the TRILL Data packets sent on particular distribution trees by VLAN, FGL, and/or multicast group thus reducing the total amount of pruning information so that it can more easily be accommodated by fast path hardware.

Status of this Memo

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

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>

## Copyright and License Notice

Copyright (c) 2015 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.

## Table of Contents

<a href="#">1. Introduction</a>	<a href="#">4</a>
<a href="#">1.1. Background Description</a>	<a href="#">4</a>
<a href="#">1.2. Motivations</a>	<a href="#">5</a>
<a href="#">2. Terminology Used in This Document</a>	<a href="#">7</a>
<a href="#">3. Data Label based Tree Selection</a>	<a href="#">8</a>
<a href="#">3.1 Overview</a>	<a href="#">8</a>
<a href="#">3.2. Sub-TLVs for the Router Capability TLV</a>	<a href="#">9</a>
<a href="#">3.2.1. The Tree and VLANs APPsub-TLV</a>	<a href="#">9</a>
<a href="#">3.2.2. The Tree and VLANs Used APPsub-TLV</a>	<a href="#">10</a>
<a href="#">3.2.3. The Tree and FGLs APPsub-TLV</a>	<a href="#">11</a>
<a href="#">3.2.4. The Tree and FGLs Used APPsub-TLV</a>	<a href="#">12</a>
<a href="#">3.3. Detailed Processing</a>	<a href="#">12</a>
<a href="#">3.4. Failure Handling</a>	<a href="#">13</a>
<a href="#">3.5. Multicast Extensions</a>	<a href="#">14</a>



<a href="#">4.</a>	Backward Compatibility . . . . .	<a href="#">14</a>
<a href="#">5.</a>	Security Considerations . . . . .	<a href="#">16</a>
<a href="#">6.</a>	IANA Considerations . . . . .	<a href="#">16</a>
<a href="#">7.</a>	References . . . . .	<a href="#">16</a>
<a href="#">7.1</a>	Normative References . . . . .	<a href="#">16</a>
<a href="#">7.2</a>	Informative References . . . . .	<a href="#">17</a>
<a href="#">8.</a>	Acknowledgments . . . . .	<a href="#">17</a>
	Authors' Addresses . . . . .	<a href="#">17</a>

## **1. Introduction**

### **1.1. Background Description**

One or more distribution trees, identified by their root nickname, are used to distribute multi-destination data in a TRILL campus [RFC6325]. The RBridge having the highest tree root priority announces the total number of trees that should be computed for the campus. It may also specify the ordered list of trees that RBridges need to compute using the Tree Identifiers (TREE-RT-IDs) sub-TLV [RFC7176]. Every RBridge can specify the trees it will use in the Trees Used Identifiers (TREE-USE-IDs) sub-TLV and the VLANs or fine grained labels (FGLs [RFC7172]) it is interested in are specified in Interested VLANs and/or Interested Labels sub-TLVs [RFC7176]. It is suggested that, by default, the ingress RBridge use the distribution tree whose root is the closest [RFC6325]. Trees Used Identifiers sub-TLVs are used to build the RPF Check table that is used for reverse path forwarding check; Interested VLANs and Interested Labels sub-TLVs are used for distribution tree pruning and the multi-destination forwarding table with pruning info is built based on that. Each distribution tree SHOULD be pruned per VLAN/FGL, eliminating branches that have no potential receivers downstream [RFC6325]. Further pruning based on Layer 2 or Layer 3 multicast address is also possible.

Defaults are provided but it is implementation dependent how many trees to calculate, where the tree roots are located, and which tree(s) are to be used by an ingress RBridge. With the increasing demand to use TRILL in data center networks, there are some features we can explore for multi-destination frames in the data center use case. In order to achieve non-blocking data forwarding, a fat tree structure is often used. Figure 1 shows a typical fat tree structure based data center network. RB1 and RB2 are aggregation switches and RB11 to RB14 are access switches. It is a common practice to configure the tree roots to be at the aggregation switches for more efficient traffic transportation. All the ingress RBridges that are access switches have the same distance to all the tree roots.



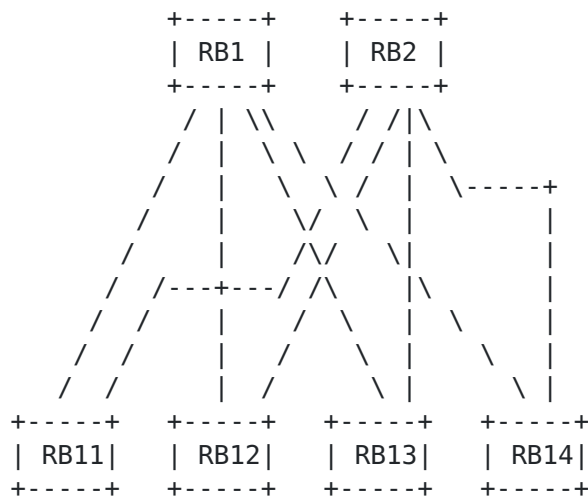


Figure 1. Fat Tree Structure based TRILL network

## 1.2. Motivations

In the structure of figure 1, if we choose to put the tree roots at RB1 and RB2, the ingress RBridge (e.g. RB11) would find more than one closest tree root (i.e. RB1 & RB2). An ingress RBridge has two options to select the tree root for multi-destination frames: choose one and only one as distribution tree root or use ECMP-like algorithm to balance the traffic among the multiple trees whose roots are at the same distance.

- For the former, a single tree used by each ingress RBridge, can have the obvious problem of inefficient link usage. For example, if RB11 chooses the tree1 that is rooted at RB1 as the distribution tree, the link between RB11 and RB2 will never be used for multi-destination frames ingressed by RB11.

- For the latter, ECMP based tree selection results in a linear increase in multicast forwarding table size with the number of trees as explained in the next paragraph.

A multicast forwarding table at an RBridge is normally used to map the key of (tree nickname + VLAN) to an index to a list of ports for multicast packet replication. The key used for mapping is simply the tree nickname when the RBridge does not prune the tree and the key could be (tree nickname + VLAN + Layer 2 or 3 multicast address) when the RBridge was programmed by control plane with Layer 2 or 3 multicast pruning information.

For any RBridge RBn, for each VLAN x, if RBn is in a distribution tree t for VLAN x, there will be an entry of (t, x, port list) in the





multicast forwarding table on RBn. Typically each entry contains a distinct combination of (tree nickname, VLAN) as the lookup key. If there are  $n$  such trees and  $m$  such VLANs, the multicast forwarding table size on RBn is  $n*m$  entries. If fine-grained label is used [[RFC7172](#)] and/or finer pruning is used (for example, VLAN + multicast group address is used for pruning), the value of  $m$  increases. In the larger scale data center, more trees would be necessary for better load balancing purpose and it results in the increasing of value  $n$ . In either case, the number of table entries  $n*m$  will increase dramatically.

The left table in Figure 2 shows an example of the multicast forwarding table on RB11 in the Figure 1 topology with 2 distribution trees in a campus using typical fast path hardware. The number of entries is approximately  $2 * 4K$  in this case. If 4 distribution trees are used in a TRILL campus and RBn has 4K VLANs with downstream receivers, it consumes 16K table entries. TRILL multicast forwarding tables have a limited size in hardware implementation. The table entries are a precious resource. In some implementations, the table is shared with Layer 3 IP multicast for a total of 16K or 8K table entries. Therefore we want to reduce the table size consumed as much as possible and at the same time maintain the load balancing among trees.

In cases where blocks of consecutive VLANs or FGLs can be assigned to a tree, it would be very helpful in compressing the multicast forwarding table if entries could have a Data Label value and mask and the fast path hardware could do longest prefix matching. But few if any fast path implementations provide such logic.

A straightforward way to alleviate the limited table entries problem is not to prune the distribution tree. However this can only be used in the restricted scenarios for the following reasons:

- Not pruning unnecessarily wastes bandwidth for multi-destination packets. There is broadcast traffic in each VLAN, like ARP and unknown unicast. In addition, if there is a lot of Layer 3 multicast traffic in some VLAN, no pruning may result in the worse consequence of Layer 3 user data unnecessarily flooded over the campus. The volume could be huge if certain applications like IPTV are supported. Finer pruning like pruning based on multicast group may be desirable in this case.
- Not pruning is only useful at pure transit nodes. Edge nodes always need to maintain the multicast forwarding table with the key of (tree nickname + VLAN) since the edge node needs to decide whether and how to replicate the frame to local access ports based on VLAN. It is very likely that edge nodes are relatively low scale switches with



the smaller shared table size, say 4K, available.

- Security concerns. VLAN based traffic isolation is a basic requirement in some scenarios. No pruning may result in the unnecessary leakage of the traffic. Misbehaved RBridges may take advantage of this.

In addition to the multicast table size concern, some silicon does not currently support hashing-based tree nickname selection at the ingress RBridge. VLAN based tree selection is used instead. The control plane of the ingress RBridge maps the incoming VLAN  $x$  to a tree nickname  $t$ . Then the data plane will always use tree  $t$  for VLAN  $x$  multi-destination frames. Though an ingress RBridge may choose multiple trees to be used for load sharing, it can use one and only one tree for each VLAN. If we make sure all ingress RBridges campus-wide send VLAN  $x$  multi-destination packets only using tree  $t$ , then there would be no need to store the multicast table entry with the key of (tree-other-than- $t$ ,  $x$ ) on any RBridge.

This document describes the TRILL control plane support for a VLAN based tree selection mechanism to reduce the multicast forwarding table size. It is compatible with the silicon implementation mentioned in the previous paragraph. Here VLAN based tree selection is a general term which also includes finer granularity case such as VLAN + Layer 2 or 3 multicast or FGL group based selection.

## **2. Terminology Used in This Document**

This document uses the terminology from [[RFC6325](#)] and [[RFC7172](#)], some of which is repeated below for convenience, along with some additional terms listed below:

campus: Name for a TRILL network, like "bridged LAN" is a name for a bridged network. It does not have any academic implication.

Data Label: VLAN or FGL.

ECMP: Equal Cost Multi-Path [[RFC6325](#)].

FGL: Finge Grainge Lable [[RFC7172](#)].

IPTV: "Television" (video) over IP.

RBridge: An alternative name for a TRILL switch.

TRILL: Transparent Interconnection of Lots of Links (or Tunneled Routing in the Link Layer).



TRILL switch: A device implementing the TRILL protocol. Sometimes called an RBridge.

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 [RFC-2119](#) [[RFC2119](#)].

### **3. Data Label based Tree Selection**

Data Label based tree selection can be used as a complementary distribution tree selection mechanism, especially when the multicast forwarding table size is a concern.

#### **3.1 Overview**

The tree root with the highest priority announces the tree nicknames and the Data Labels allowed on each tree. Such tree to Data Label correspondence announcements can be based on static configuration or some predefined algorithm beyond the scope of this document. An ingress RBridge selects the tree-VLAN correspondence it wishes to use from the list announced by the highest priority tree root. It SHOULD NOT transmit VLAN x frame on tree y if the highest priority tree root does not say VLAN x is allowed on tree y.

If we make sure one VLAN is allowed on one and only one tree, we can keep the number of multicast forwarding table entries on any RBridge fixed at 4K maximum (or up to 16M in case of fine grained label). Take Figure 1 as example, two trees rooted at RB1 and RB2 respectively. The highest priority tree root appoints the tree1 to carry VLAN 1-2000 and tree2 to carry VLAN 2001-4095. With such announcement by the highest priority tree root, every RBridge which understands the announcement will not send VLAN 2001-4095 traffic on tree1 and not send VLAN 1-2000 traffic on tree2. Then no RBridge would need to store the entries for tree1/VLAN2001-4095 or tree2/VLAN1-2000. Figure 2 shows the multicast forwarding table on an RBridge before and after we perform the VLAN based tree selection. The number of entries is reduced by a factor f, f being the number of trees used in the campus. In this example, it is reduced from 2\*4095 to 4095. This affects both transit nodes and edge nodes. Data plane encoding does not change.



+-----+-----+-----+	+-----+-----+-----+
tree nickname  VLAN  port list	tree nickname  VLAN  port list
+-----+-----+-----+	+-----+-----+-----+
tree 1   1	tree 1   1
+-----+-----+-----+	+-----+-----+-----+
tree 1   2	tree 1   2
+-----+-----+-----+	+-----+-----+-----+
tree 1   ...	tree 1   ...
+-----+-----+-----+	+-----+-----+-----+
tree 1   ...	tree 1   1999
+-----+-----+-----+	+-----+-----+-----+
tree 1   ...	tree 1   2000
+-----+-----+-----+	+-----+-----+-----+
tree 1   4094	tree 2   2001
+-----+-----+-----+	+-----+-----+-----+
tree 1   4095	tree 2   2002
+-----+-----+-----+	+-----+-----+-----+
tree 2   1	tree 2   ...
+-----+-----+-----+	+-----+-----+-----+
tree 2   2	tree 2   4094
+-----+-----+-----+	+-----+-----+-----+
tree 2   ...	tree 2   4095
+-----+-----+-----+	+-----+-----+-----+
tree 2   ...	
+-----+-----+-----+	
tree 2   ...	
+-----+-----+-----+	
tree 2   4094	
+-----+-----+-----+	
tree 2   4095	
+-----+-----+-----+	

Figure 2. Multicast forwarding table before (left) &amp; after (right)

### **3.2. Sub-TLVs for the Router Capability TLV**

Four new APPsub-TLVs that can be carried in E-L1FS FS-LSPs [[rfc7180bis](#)] are defined below. They can be considered analogous to finer granularity versions of the Tree Identifiers Sub-TLV and the Trees Used Identifiers Sub-TLV in [[RFC7176](#)].

#### **3.2.1. The Tree and VLANs APPsub-TLV**

The Tree and VLANs (TREE-VLANs) APPsub-TLV is used to announce the VLANs allowed on each tree by the RBridge that has the highest





priority to be a tree root. Multiple instances of this sub-TLV may be carried. The same tree nicknames may occur in the multiple Tree-VLAN RECORDs within the same or across multiple sub-TLVs. The sub-TLV format is as follows:

```

                                1 1 1 1 1 1
      0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
+---+---+---+---+---+---+---+---+---+
|   Type = tbd1                               |   (2 bytes)
+---+---+---+---+---+---+---+---+---+
|   Length                                   |   (2 bytes)
+---+---+---+---+---+---+---+---+---+...+---+
|   Tree-VLAN RECORD (1)                     |   (6 bytes)
+---+---+---+---+---+---+---+---+---+...+---+
|   .....                                   |
+---+---+---+---+---+---+---+---+---+...+---+
|   Tree-VLAN RECORD (N)                     |   (6 bytes)
+---+---+---+---+---+---+---+---+---+...+---+

```

where each Tree-VLAN RECORD is of the form:

```

+---+---+---+---+---+---+---+---+---+
|           Nickname                           |   (2 bytes)
+---+---+---+---+---+---+---+---+---+
| RESV |           Start.VLAN                   |   (2 bytes)
+---+---+---+---+---+---+---+---+---+
| RESV |           End.VLAN                     |   (2 bytes)
+---+---+---+---+---+---+---+---+---+

```

- o Type: TRILL GENINFO APPsub-TLV type, set to tbd1 (TREE-VLANs).
- o Length: 6\*n bytes, where there are n Tree-VLAN RECORDs. Thus the value of Length can be used to determine n. If Length is not a multiple of 6, the sub-TLV is corrupt and MUST be ignored.
- o Nickname: The nickname identifying the distribution tree by its root.
- o RESV: 4 bits that MUST be sent as zero and ignored on receipt.
- o Start.VLAN, End.VLAN: These fields are the VLAN IDs of the allowed VLAN range on the tree, inclusive. To specify a single VLAN, the VLAN's ID appears as both the start and end VLAN. If End.VLAN is less than Start.VLAN the Tree-VLAN RECORD MUST be ignored.

### 3.2.2. The Tree and VLANs Used APPsub-TLV

This APPsub-TLV has the same structure as the Tree and VLANs APPsub-TLV (TREE-VLANs) specified in [Section 3.2.1](#). The only difference is



that its APPsub-TLV type is set to tbd2 (TREE-VLAN-USE), and the Tree-VLAN RECORDs listed are those the originating RBridge allows.

### 3.2.3. The Tree and FGLs APPsub-TLV

The Tree and FGLs (TREE-FGLs) APPsub-TLV is used to announce the FGLs allowed on each tree by the RBridge that has the highest priority to be a tree root. Multiple instances of this APPsub-TLV may be carried. The same tree nicknames may occur in the multiple Tree-FGL RECORDs within the same or across multiple APPsub-TLVs. Its format is as follows:

```

                                1 1 1 1 1 1
      0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5
+---+---+---+---+---+---+---+---+---+
|   Type = tbd3                               |   (2 bytes)
+---+---+---+---+---+---+---+---+---+
|   Length                                   |   (2 bytes)
+---+---+---+---+---+---+---+---+---+
|   Tree-FGL RECORD (1)                     |   (8 bytes)
+---+---+---+---+---+---+---+---+---+
|   .....                                   |
+---+---+---+---+---+---+---+---+---+
|   Tree-FGL RECORD (N)                     |   (8 bytes)
+---+---+---+---+---+---+---+---+---+

```

where each Tree-VLAN RECORD is of the form:

```

+---+---+---+---+---+---+---+---+---+
|           Nickname                         |   (2 bytes)
+---+---+---+---+---+---+---+---+---+
|           Start.FGL                       |   (3 bytes)
+---+---+---+---+---+---+---+---+---+
|           End.FGL                         |   (3 bytes)
+---+---+---+---+---+---+---+---+---+

```

- o Type: TRILL GENINFO APPsub-TLV type, set to tbd3 (TREE-FGLs).
- o Length: 8\*n bytes, where there are n Tree-FGL RECORDs. Thus the value of Length can be used to determine n. If Length is not a multiple of 8, the sub-TLV is corrupt and MUST be ignored.
- o Nickname: The nickname identifying the distribution tree by its root.
- o RESV: 4 bits that MUST be sent as zero and ignored on receipt.
- o Start.FGL, End.FGL: These fields are the FGL IDs of the allowed FGL range on the tree, inclusive. To specify a single FGL, the FGL's



ID appears as both the start and end FGL. If End.FGL is less than Start.FGL the Tree-FGL RECORD MUST be ignored.

#### **3.2.4. The Tree and FGLs Used APPsub-TLV**

This APPsub-TLV has the same structure as the Tree and FGLs APPsub-TLV (TREE-FGLs) specified in [Section 3.2.3](#). The only difference is that its APPsub-TLV type is set to tbd4 (TREE-FGL-USE), and the Tree-FGL RECORDs listed are those the originating RBridge allows.

### **3.3. Detailed Processing**

The highest priority tree root RBridge MUST include all the necessary tree related APPsub-TLVs defined in [[RFC7176](#)] as usual in its E-L1FS FS-LSP and MAY include the Tree and VLANs Sub-TLV (TREE-VLANs) and or Tree and FGLs Sub-TLV (TREE-FGLs) in its E-L1FS FS-LSP [[rfc7180bis](#)]. In this way it MAY indicate that each VLAN and/or FGL is only allowed on one or some other number of trees less than the number of trees being calculated in the campus in order to save table space in the fast path forwarding hardware.

An ingress RBridge that understands the TREE-VLANs APPsub-TLV SHOULD select the tree-VLAN correspondences it wishes to use and put them in TREE-VLAN-USE APPsub-TLVs. If there were multiple tree nicknames announced in TREE-VLANs Sub-TLV for a VLAN x, ingress RBridge must choose one of them if it supports this feature. For example, the ingress RBridge may choose the closest (minimum cost) root from them. How to make such choice is out of the scope of this document. It may be desirable to have some fixed algorithm to make sure all ingress RBs choose the same tree for VLAN x in this case. Any single Data Label that the ingress RBridge is interested in should be related to one and only one tree ID in TREE-VLAN-USE to minimize the multicast forwarding table size on other RBridges but as long as the Data Label is related to less than all the trees being calculated, it will reduce the burden on the forwarding table size.

When an ingress RBridge tries to encapsulate a multi-destination frame for Data Label x, it SHOULD use the tree nickname that it selected previously in TREE-VLAN-USE or TREE-FGL-USE for Data Label x.

If RBridge RBn does not perform pruning, it builds the multicast forwarding table exactly same as that in [[RFC6325](#)].

If RBn prunes the distribution tree based on VLANs, RBn uses the information received in TREE-VLAN-USE APPsub-TLVs to mark the set of VLANs reachable downstream for each adjacency and for each related tree. If RBn prunes the distribution tree based on FGLs, RBn uses the



information received in TRILL-FGL-USE APPsub-TLVs to mark the set of FLGs reachable downstream for each adjacency and for each related tree.

Logically, an ingress RBridge that does not support VLAN based tree selection is equivalent to the one that supports it and announces all the combination pair of tree-id-used and interested-vlan as TREE-VLAN-USE and correspondingly for FGL.

### **3.4. Failure Handling**

Failure of a tree root that is not the highest priority: It is the responsibility of the highest priority tree root to inform other R Bridges of any change in the allowed tree-VLAN correspondence. When the highest priority tree root learns the root of tree t fails, it should re-assign the VLANs allowed on tree t to other trees or to a tree replacing the failed one.

Failure of the highest priority tree root: It is RECOMMENDED that the second highest priority tree root be pre-configured with the proper knowledge of the tree-VLAN correspondence allowed when the highest priority tree root fails. The information announced by the second priority tree root would be stored by all R Bridges but would not take effect unless the R Bridge noticed the failure of the highest priority tree root. When the highest priority tree root fails, the former second priority tree root will become the highest priority tree root of the campus. When an R Bridge notices the failure of the original highest priority tree root, it can immediately use the stored information announced by the original second priority tree root. It is recommended that the tree-VLAN correspondence information be pre-configured on the second highest priority tree root to be the same as that on the highest priority tree root for the trees other than the highest priority tree itself. This can minimize the change of multicast forwarding table in case of the highest priority tree root failure. For a large campus, it may make sense to pre-configure this information in a similar way on the third, fourth, or even lower priority tree root R Bridges.

In some transient conditions or in case of misbehavior by the highest priority tree root, an ingress R Bridge may encounter the following scenarios:

- No tree has been announced to allow VLAN x frames
- An ingress R Bridge is supposed to transmit VLAN x frames on tree t, but root of tree t is no longer reachable.

For the second case, an ingress R Bridge may choose another reachable





tree root which allows VLAN x according to the highest priority tree root announcement. If there is no such tree available, then it is same as the first case above. Then the ingress RBridge should be 'downgraded' to a conventional BRIDGE with behavior as specified in [RFC6325]. A timer should be set to allow the temporary transient stage to complete before the change of responsive tree or 'downgrade' takes effect. The value of timer should at least be set to the LSP flooding time of the campus.

### 3.5. Multicast Extensions

Data Label based tree selection is easily extended to (Data Label + Layer 2 or 3 multicast group) based tree selection. We can appoint multicast group 1 in VLAN 10 to tree1 and appoint group 2 in VLAN 10 to tree2 for better load sharing. One additional APPsub-TLV is specified as follows:

```

+---+---+---+---+---+---+---+---+---+
|  Type = tbd5                               | (2 byte)
+---+---+---+---+---+---+---+---+---+
|  Length                                     | (2 byte)
+---+---+---+---+---+---+---+---+---+
|      Tree Nickname                         | (2 bytes)
+---+---+---+---+---+---+---+---+---+
|  Group Sub-Sub-TLVs                       | (variable)
+---+---+---+---+---+---+---+---+....

```

- o Type: TRILL GENINFO APPsub-TLV type, set to tbd5 (TREE-GROUPS).
- o Length: 2 + the length of the Group Sub-Sub TLVs included
- o Nickname: The nickname identifying the distribution tree by its root.
- o RESV: 4 bits that MUST be sent as zero and ignored on receipt.
- o Group Sub-Sub-TLVs: Zero or more of the TLV structure that are allowed as sub-TLVs of the GADDR TLV [RFC7176]. Each such TLV structure specifies a multicast group and either a VLAN or FGL. Although these TLV structure are considered sub-TLVs when they appear inside a GADDR TLV, they are technically sub-sub-TLVs when they appear inside the TREE-GROUPS APPsub-TLV.

## 4. Backward Compatibility

RBridges MUST include the TREE-USE-IDs and INT-VLAN sub-TLVs in their



LSPs when required by [RFC6325](#) whether or not they supports the new TREE-VLAN-USE or TREE-FGL-USE sub-TLVs specified by this draft.

RBridges that understand the new TREE-VLAN-USE sub-TLV sent from another RBridge RBn should use it to build the multicast forwarding table and ignore the TREE-USE-IDs and INT-VLAN sub-TLVs sent from the same RBridge. TREE-USE-IDs and INT-VLAN sub-TLVs are still useful for some purposes other than building multicast forwarding table, for example RPF table building, spanning tree root notification, etc. If the RBridge does not receive TREE-VLAN-USE sub-TLV from RBn, it uses the conventional way described in [RFC6325](#) to build the multicast forwarding table.

For example, there are two distribution trees, tree1 and tree2 in the campus. RB1 and RB2 are RBridges that use the new APPsub-TLVs described in this document. RB3 is an old RBridge that is compatible with [RFC6325](#). Assume RB2 is interested in VLANs 10 and 11 and RB3 is interested in VLANs 100 and 101. Hence RB1 receives ((tree1, VLAN10), (tree2, VLAN11)) as TREE-VLAN-USE sub-TLV and (tree1, tree2) as TREE-USE-IDs sub-TLV from RB2 on port x. And RB1 receives (tree1) as TREE-USE-IDs sub-TLV and no TREE-VLAN-USE sub-TLV from RB3 on port y. RB2 and RB3 announce their interested VLANs in INT-VLAN sub-TLV as usual. Then RB1 will build the entry of (tree1, VLAN10, port x) and (tree2, VLAN11, port x) based on RB2's LSP and mechanism specified in this document. RB1 also builds entry of (tree1, VLAN100, port y), (tree1, VLAN101, port y), (tree2, VLAN100, port y), (tree2, VLAN101, port y) based on RB3's LSP in conventional way. The multicast forwarding table on RB1 with merged entry would be like the following.

+-----+-----+-----+
tree nickname  VLAN  port list
+-----+-----+-----+
tree 1   10   x
+-----+-----+-----+
tree 1   100   y
+-----+-----+-----+
tree 1   101   y
+-----+-----+-----+
tree 2   11   x
+-----+-----+-----+
tree 2   100   y
+-----+-----+-----+
tree 2   101   y
+-----+-----+-----+

It is expected that the table is not as small as the one where every RBridge supports the new TREE-VLAN-USE sub-TLVs. The worst case in a



hybrid campus is the number of entries equal to the number in current practice which does not support VLAN based tree selection. Such an extreme case happens when the interested VLAN set from the new R Bridges is a subset of the interested VLAN set from the old R Bridges.

VLAN based tree selection is compatible with the current practice. Its effectiveness increases with more R Bridge supporting this feature in the TRILL campus.

## 5. Security Considerations

This document does not change the general R Bridge security considerations of the TRILL base protocol. The APPsub-TLVs specified can be secured using the IS-IS authentication feature [[RFC5310](#)]. See [Section 6 of \[RFC6325\]](#) for general TRILL security considerations.

## 6. IANA Considerations

IANA is requested to assign five new TRILL APPsub-TLV type codes as specified in [Section 3](#) and update the TRILL Parameters registry as shown below.

Type	Name	Reference
----	----	-----
tbd1	TREE-VLANs	[this document]
tbd2	TREE-VLAN-USE	[this document]
tbd3	TREE-FGLs	[this document]
tbd4	TREE-FGL-USE	[this document]
tbd5	TREE-GROUPs	[this document]

## 7. References

### 7.1 Normative References

- [RFC6325] Perlman, R., et.al. "R Bridge: Base Protocol Specification", [RFC 6325](#), July 2011.
- [RFC6439] Eastlake, D. et.al., "R Bridge: Appointed Forwarder", [RFC 6439](#), November 2011.
- [RFC7172] Eastlake 3rd, D., Zhang, M., Agarwal, P., Perlman, R., and D. Dutt, "Transparent Interconnection of Lots of Links (TRILL): Fine-Grained Labeling", [RFC 7172](#), May 2014,

<<http://www.rfc-editor.org/info/rfc7172>>.

[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, <<http://www.rfc-editor.org/info/rfc7176>>.

[rfc7180bis] Eastlake 3rd, D. et. Al. [draft-eastlake-trill-rfc7180bis](#), work in progress.

## 7.2 Informative References

[RFC5310] - Bhatia, M., Manral, V., Li, T., Atkinson, R., White, R., and M. Fanto, "IS-IS Generic Cryptographic Authentication", [RFC 5310](#), February 2009, <<http://www.rfc-editor.org/info/rfc5310>>.

## 8. Acknowledgments

Authors wish to thank David M. Bond, Liangliang Ma, Rakesh Kumar R for the valuable comments (names in alphabet order).

### Authors' Addresses

Yizhou Li  
Huawei Technologies  
101 Software Avenue,  
Nanjing 210012  
China

Phone: +86-25-56624629  
Email: [liyizhou@huawei.com](mailto:liyizhou@huawei.com)

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

Phone: +1-508-333-2270  
Email: [d3e3e3@gmail.com](mailto:d3e3e3@gmail.com)

Weiguo Hao  
Huawei Technologies  
101 Software Avenue,  
Nanjing 210012



China

Phone: +86-25-56623144

Email: haoweiguo@huawei.com

Hao Chen  
Huawei Technologies  
101 Software Avenue,  
Nanjing 210012  
China

Email: philips.chenhao@huawei.com

Radia Perlman  
EMC  
2010 256th Avenue NE, #200  
Bellevue, WA 98007  
USA

Email: Radia@alum.mit.edu

Naveen Nimmu  
Broadcom  
9th Floor, Building no 9, Raheja Mind space  
Hi-Tec City, Madhapur,  
Hyderabad - 500 081, INDIA

Phone: +1-408-218-8893

Email: naveen@broadcom.com

Somnath Chatterjee  
Cisco Systems,  
SEZ Unit, Cessna Business Park,  
Outer ring road,  
Bangalore - 560087  
India

Email: somnath.chatterjee01@gmail.com

Sunny Rajagopalan  
IBM

Email: sunny.rajagopalan@us.ibm.com



