

TRILL Working Group
Internet Draft
Intended status: Standard Track

Tissa Senevirathne
Les Ginsberg
CISCO
Sam Aldrin
HuaWei
Ayan Banerjee
CISCO

May 28, 2013

Expires: November 2013

Default Nickname Based Approach for Multilevel TRILL
draft-tissa-trill-multilevel-02.txt

Status of this Memo

This Internet-Draft is submitted 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/ietf/lid-abstracts.txt>

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

This Internet-Draft will expire on May 28, 2009.

Copyright Notice

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

Internet-Draft

Multilevel TRILL

May 2013

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.

Abstract

Multilevel TRILL allows the interconnection of multiple TRILL networks to form a larger TRILL network without proportionally increasing the size of the IS-IS LSP DB. In this document, an approach based on default route concept is presented. Also, presented in the document is a novel method of constructing multi-destination trees using partial nickname space. Methods presented in this document are compatible with the [RFC6325](#) specified data plane operations.

Table of Contents

| | | |
|------------------------|--|--------------------|
| 1. | Introduction..... | 3 |
| 2. | Conventions used in this document..... | 4 |
| 3. | Solution Overview..... | 4 |
| 4. | Operational Overview..... | 5 |
| 4.1. | Unicast Forwarding..... | 5 |
| 4.2. | IS-IS Protocol changes for unicast forwarding..... | 5 |
| 4.3. | MAC Address Learning..... | 5 |
| 4.4. | Multicast..... | 6 |
| 4.5. | Life of Multicast frame..... | 9 |
| 5. | Area Affinity sub-TLV..... | 10 |
| 6. | Nickname acquisition and conflict resolution..... | 11 |
| 6.1. | Nickname Management sub-TLV..... | 13 |
| 7. | Further optimizations..... | 14 |
| 7.1. | Leaking of TRILL IS-IS sub-TLV within areas..... | 14 |
| 7.2. | Identification of Global Trees..... | 15 |
| 7.2.1. | Global Tree capability sub-TLV..... | 17 |
| 7.2.2. | Global Tree proposal sub-TLV..... | 17 |
| 7.2.3. | Global Tree Identifier sub-TLV..... | 18 |
| 7.3. | Announcing Group Addresses..... | 18 |

| | |
|--|--------------------|
| 8. Architecture Elements of Multi-level Multicast framework..... | 21 |
| 8.1. Bootstrap RBridge..... | 21 |
| 8.2. Rendezvous Point (RP)..... | 22 |
| 8.3. Default Affinity sub-TLV..... | 22 |

| | |
|---|--------------------|
| 8.4. Area Affinity sub-TLV..... | 22 |
| 8.5. TRILL BSR Protocol..... | 22 |
| 9. Security Considerations..... | 22 |
| 10. IANA Considerations..... | 23 |
| 11. References..... | 23 |
| 11.1. Normative References..... | 23 |
| 11.2. Informative References..... | 23 |
| 12. Acknowledgments..... | 24 |

[1.](#) Introduction

The TRILL Base Protocol Specification [[RFC6325](#)] provides a method for forwarding Layer 2 data frames over multiple active links, thereby optimizing network bandwidth and resiliency. TRILL requires native Layer 2 frames to be encapsulated with the TRILL header. TRILL devices (RBridges) are identified with a 16 bit identifier called a nickname. The TRILL header contains egress and ingress RBridge nicknames. Intermediate RBridges performs forwarding based on the egress nickname. TRILL utilize the IS-IS protocol to distribute RBridge nicknames.

TRILL Base Protocol Specification [[RFC6325](#)] specifies a tree based paradigm to forward broadcast and multicast traffic as well as unknown unicast traffic.

Traditional Layer 2 devices perform forwarding based on MAC addresses. As a result, in theory, all of the devices in the network are required to learn all of the MAC addresses in the Layer 2 domain. This leads to very large forwarding table sizes in the devices which limits the size of the layer 2 domain. Forwarding within the TRILL network occurs not based on MAC addresses but based on the RBridge nicknames. Hence, TRILL based networks have significant potential to be the core forwarding plane of very large datacenters.

Large datacenters are often multisite in nature and contain a large number of RBridges. TRILL is designed to be a single IS-IS area. As the size of the TRILL network grows, the size of the IS-IS LSP

database grows, leading to network convergence delays and increased volatility during transient conditions such as link flaps.

As mentioned above TRILL utilizes a tree based forwarding paradigm for multi-destination traffic. In large TRILL networks this may lead to sub-optimal forwarding. Additionally, entire network wide multicasting may lead to network bandwidth inefficiencies and have a negative impact on performance.

In order to support scaling and performance of large TRILL networks, it is important to have methods to:

1. Limit the size of the IS-IS LSP database
2. Optimize multicast forwarding
3. Limit the scope of flooding and broadcast traffic

In this document we propose methods that allow implementing multi-level TRILL without any data plane changes as well as meet the above design goals.

Also presented in this document is a novel method of constructing multi-destination trees using partial nickname space.

[2.](#) Conventions used in this document

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](#)].

In this document, these words will appear with that interpretation only when in ALL CAPS. Lower case uses of these words are not to be interpreted as carrying [RFC-2119](#) significance.

[3.](#) Solution Overview

Herein we present a high level view of the proposed solution; differing the details of the solution to subsequent sections.

The TRILL campus is divided in to multiple IS-IS L1 Areas interconnected by L2 backbone area. The backbone area may be interconnected by an IS-IS K2 area or by some other means. For example, one does not preclude a configuration based approach for the L2

backbone.

Area border RBridges indicate their ability to reach other areas by setting the Attach bit in IS-IS Link State PDU.

RBridges forward frames destined to RBridges in the area using the exact match of nicknames. Frames destined to RBridges outside of the L1 area are forwarded using the default route advertised by the area border RBridges.

Campus wide Multi-destination trees are partitioned in to two parts. A backbone tree rooted on the L2 backbone and local trees rooted within each L1 area. For campus wide trees the local trees and the

backbone tree have the same nickname. This avoids the need for egress RBridge nickname translation at the border RBridges).

One of the border RBridges performs connecting its local tree to the corresponding backbone tree. The Affinity sub-TLV and Area Affinity sub-TLV information facilitate constructing proper RPF states.

[4. Operational Overview](#)

[4.1. Unicast Forwarding](#)

The TRILL campus is divided in to multiple IS-IS L1 Areas with a Layer 2 backbone. (Figure 1).

The "Attached" bit in IS-IS PDU is used to indicate the advertising IS connected to other areas. In this document we propose to leverage the "Attached" bit to identify the border RBridges and forward traffic for all un-resolved nicknames to the border RBridges.

Border RBridges possess the complete nickname space of the TRILL campus. Utilizing this information, ingress area border RBridge forward TRILL frames to the egress area border RBridge via the L2 backbone.

Egress area border RBridges, forward the frame normally to the intended destination in the given L1 Area.

[4.2. IS-IS Protocol changes for unicast forwarding](#)

Support for non zero IS-IS areas for TRILL.

No new TLV or sub-TLV required.

Border RBridges advertise LSP with the "Attached" bit set

L1 Area RBridges are required to advertise TRILL related sub-TLV defined in [[RFC 6326](#)] with the router capability bit "S" set. The capability bit "D" MUST be set to zero (0). This allows leaking L1 PDU to the L2 backbone area but not to other L1 areas [[RFC4971](#)].

[4.3](#). MAC Address Learning

Egress RBridge learn remote MAC addresses against the actual nickname of the ingress RBRidge.

As an example: Let's Assume MAC address A is attached to RB1 and MAC address B is attached to RB7.

RB1 receives a frame destined to MAC B. RB1 does not know the location of MAC B. However, local policy such as, port or VLAN indicates that the frame is of global scope. RB1 transmits the frame on global tree "t" as an unknown unicast frame.

RB7 receives the frame and learn MAC A is associated with RB1. RB7 programs its forwarding tables such that MAC-A is associated with RB1. It also forwards the frame to MAC-B.

MAC-B responds. RB7 has the association of MAC-A to RB1. Hence, RB7 forwards the frame to RB1 as a unicast frame, with egress RBridge nickname as RB1 and ingress RBridge nickname as RB7.

The frame follows the normal uicast forwarding process explained above and arrives at RB1. RB1, learns the MAC-B association to RB7 and updates its forwarding tables accordingly. Also, RB1 forwards the frame to MAC-A.

[4.4](#). Multicast

Multicast forwarding has two parts: 1. Construction of the multicast trees 2. RPF (Reverse Path Forwarding) check.

In most of the real world deployments, not all of the traffic is required or desired to span across the entire TRILL campus. The majority of the traffic tends to have a local scope and some subset of traffic to have a global scope.

The scope of global traffic may be identified either through VLAN or via finegrain label that spans across the entire TRILL campus.

In this document we propose to classify TRILL multi-destination trees into two types:

1. Local trees (trees that have a scope within the local area)
2. Global trees (trees that have a scope throughout the TRILL campus)

Multi-destination traffic of local scope is forwarded using Local trees. Multi-destination traffic of global scope is forwarded using global trees.

Construction of global multi-destination trees and performing RPF check for such trees requires knowledge of all of the RBridges in the entire TRILL campus. In a large TRILL campus, construction of such global trees that need information of all RBridges may not only lack scalability but also may run in to instabilities during network changes. Additionally, when the TRILL campus is divided into

multiple IS-IS L1 areas, RBridges within an L1 areas do not possess reachability information for other areas. Thus, constructing such global trees may not be possible.

In this document we propose an [[RFC6325](#)] compatible approach of building multicast trees to address the issues mentioned above.

In the proposed method, global trees (campus wide trees) are partitioned in to two instances; a backbone tree instance and set of local tree instance per each area.

Backbone tree - An instance of the tree rooted in the L2 backbone. The Backbone tree is represented by the same nickname as the global tree.

Local tree - An instance of a tree per area, per campus wide tree, rooted within each area. Each instance of the tree in each area is represented by the same nickname as the global tree. (This is

important to avoid the need for tree translation at the border R Bridges)

L2 LSPs advertise the backbone tree.

L1 LSPs advertise the corresponding local-tree within each area.

One of the L1-L2 area border R Bridges in an given area is assigned the role of Rendezvous Point (RP) for the specific local tree (more details are presented in [section 8](#).).

Each RP functions as the plumb between the global tree and local tree. Both the trees have exactly the same nickname.

RP An RP advertises affinity to the rest of the campus for the specified tree by using the Affinity sub-TLV [trillcmt]. In the Affinity TLV associated nickname MUST be specified as zero to indicate the Affinity is related to default route. We refer to this usage of Affinity TLV as the default Affinity sub-TLV in the rest of the document.

Each R Bridge in the local L1 area builds its multi-destination SPF tree as specified in [[RFC6325](#)] and [trillcmt]. RPF checks are performed as specified in [[RFC6325](#)] and [trillcmt]. Additionally, RPF checks for default nicknames (i.e. all unknown nicknames to the local L1 area) are performed per the association specified by the default Affinity sub-TLV.

Additionally, each RP on behalf of the local Area it is representing for multi-cast tree Tx, advertises Area Affinity-TLV towards the L2 backbone area. The Area Affinity TLV, include the L1 Area ID of the associated area. The Area Affinity TLV, notifies R Bridges in the L2 area to enable the RPF check to accept nicknames in the associated L1 area from the announcing RP. The Area Affinity TLV allows greater scaling of the IS-IS LSP DB. If Affinity TLV contains all of the nicknames the IS-IS PDU size increases. Use of the Area Affinity sub-TLV to summarize the entire area in a single sub-TLV, limits the size of the LSP DB as well as PDU size. (Please see below for the structure of the Area Affinity sub-TLV).

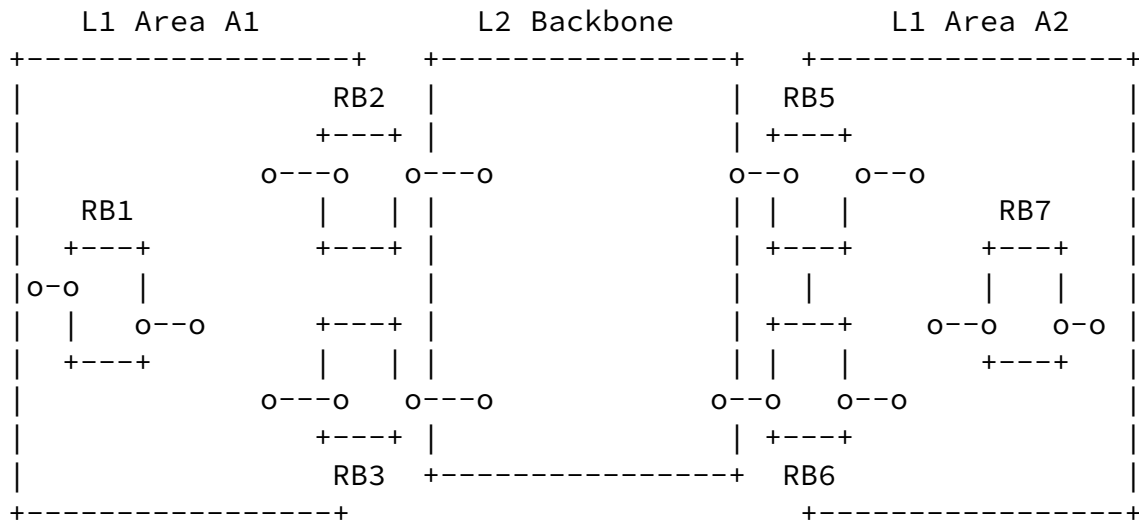
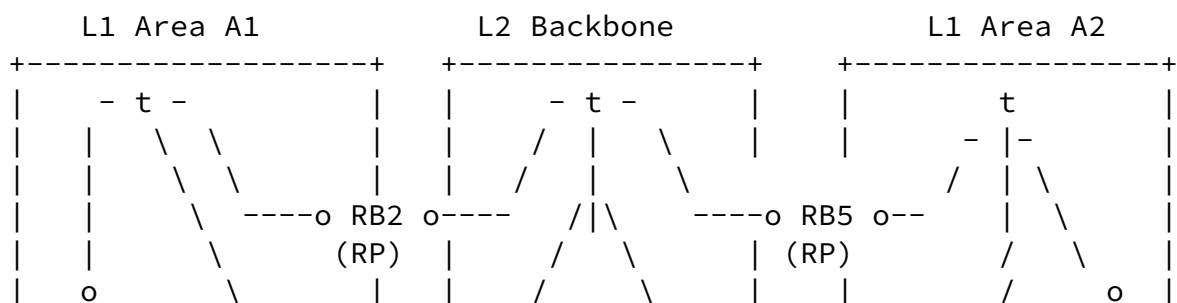


Figure 1 Sample Topology



contains the RPF check to accept frames from others areas on tree "t" on its L2 area interface. Hence, RB5, accept the multicast frame from tree "t" and forward along the local tree "t" and its local ports.

The multicast packet traverses along the local tree "t" in Area A2 and arrive at RB7. RB7 contains RPF check installed based on the Default Affinity TLV for tree "t", to receive multicast traffic for tree "t" that arrives from the RP facing interface. RB7 accepts the multicast frame arriving on local tree "t" with ingress RBridge nickname RB1 and performs applicable forwarding as specified in [\[RFC6325\]](#).

RB6 is an Area border RBridge for Area A2, but not RP for the area. RB6 receives the multicast frame through the local tree "t". Non RP area border RBridges for RPF and multi-destination forwarding purposes function like a L1 area RBridge. RB6, honors the default Affinity TLV received from RB5 for local multicast tree "t". Hence, it installs RPF check to accept all nicknames (default nickname) for tree "t" from the L1 Area interface pointing towards RB5.

RB6 accepts the incoming multicast frame along tree "t" with the ingress RBridge nickname RB1 and performs applicable forwarding to locally attached ports. RB6, which is a non RP for tree "t" for area A2, MUST not forward the multicast frame to the global tree "t".

[5.](#) Area Affinity sub-TLV

Area Affinity sub-TLV is a sub-TLV under IS-IS Router capability TLV and contains the following structure.

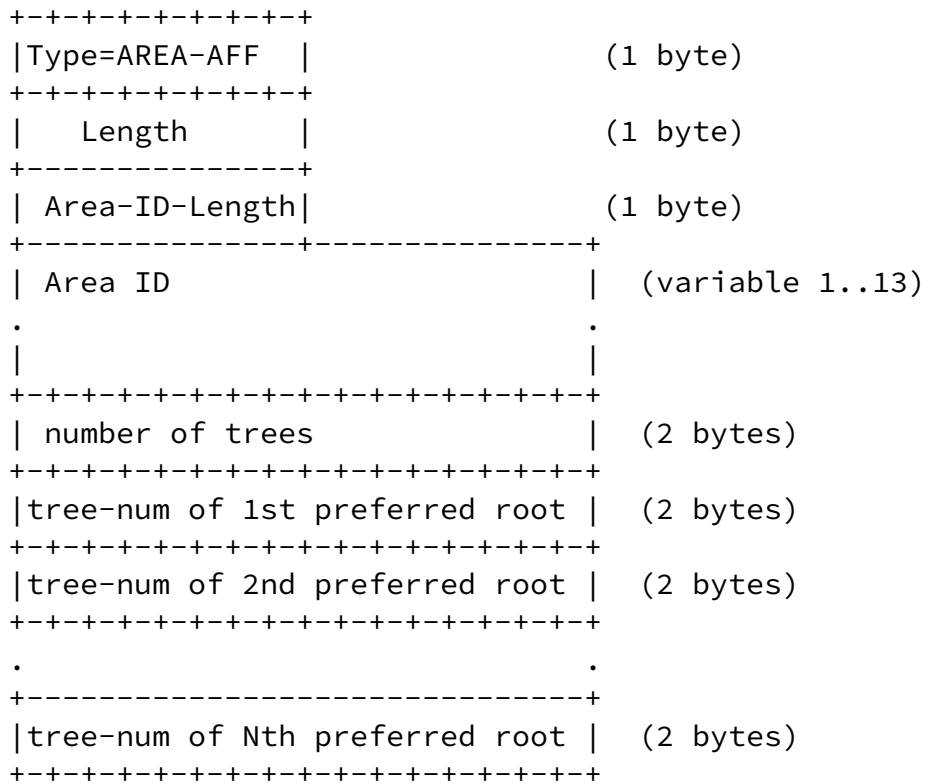


Figure 3 Area Affinity TLV structure

- o Type: AREA-AFF, (TBD).
- o Length: variable. Length is 1 + length of Area ID + 2*n, where n is the number listed tree numbers.
- o Area-ID : (variable length) Is the IS-IS Area ID. Length can be 1-13 bytes long.
- o Number of trees : number of trees for which association (affinity) being announced.
- o Tree-num of preferred root: The tree Number of the multicast tree.

Area-affinity conflicts MUST be resolved using methods specified in [trillcmt].

6. Nickname acquisition and conflict resolution

In the proposed method, nicknames of RBridges in remote L1 areas are not advertised into the local L1 area by area border RBridges. However, L2 backbone RBridges and L1-L2 border RBridges contain the nicknames used by all the RBridges in the campus. We propose to

Internet-Draft

Multilevel TRILL

May 2013

introduce a new IS-IS sub-TLV under Router capability TLV to distribute available nickname space. New IS-IS sub TLV is Nickname Management sub-TLV.

Nickname Management sub-TLV is announced by the area border RBridges in to the local L1 area with Router capability bits D set to one and S set to one. Nickname Management sub-TLV is announced in to L2 area by the area border RBridges with S and D bit clear. These settings ensure Nickname Management TLV is confined to the local L1 area L2 area and does not leak to other L1 areas.

Nickname Management sub-TLV instance announced in to the local area, contains two sets of ranges. Local nickname ranges and dynamic nickname ranges. Local nickname ranges are one or more sets of ranges that network administrator has configured on the border RBridges. Multiple local nickname ranges allow network administrator to configure multiple sets of non contiguous nickname ranges.

L1 area RBridges SHOULD, first, select a nickname or nicknames from the local ranges.

If entire local nickname space has exhausted (i.e. taken up by other RBridges), then L1 are RBridges SHOULD select a nickname or nicknames from the dynamic ranges.

It is recommended that RBridges use different nickname priorities to differentiate nickname acquired by different methods. Nickname priorities are assigned based on the acquisition method such that configured nicknames have highest priority, followed by nicknames derived from the local range, followed by nicknames derived from the dynamic range.

Dynamic ranges are derived by the area border RBridges based on the local nickname ranges of its and other areas. As an example let's assume area A1 has local nickname range of 100-200, A2 has a local nickname range of 201-300. Then the dynamic range is from 1-99 and 301 to 65471 (0xFFBF). Nickname values 0 and 0xFFC0 to 0xFFFF are reserved and MUST not be included in the dynamic nickname ranges.

Nickname Management sub-TLV instance announced in to the L2 backbone area, contains only the local nickname ranges. Local nickname ranges of each area allow other areas to derive applicable dynamic ranges to announce in to the corresponding L1 area.

[6.1.](#) Nickname Management sub-TLV

```

+---+---+---+---+---+
|Type=NICK-MGMT|          (1 byte)
+---+---+---+---+---+
|  Length      |          (1 byte)
+-----+
| Area-ID-Length|          (1 byte)
+-----+-----+
| Area ID      |          (variable 1..13)
.
|
+---+---+---+---+---+---+---+---+---+---+
| NL           |          (1 byte)
+---+---+---+---+---+---+---+---+---+---+
|starting nickname for l-range 1| (2 bytes)
+---+---+---+---+---+---+---+---+---+---+
|end nickname for l-range 1   | (2 bytes)
+---+---+---+---+---+---+---+---+---+---+
.
+---+---+---+---+---+---+---+---+---+---+
|starting nickname for l-range n| (2 bytes)
+---+---+---+---+---+---+---+---+---+---+
|end nickname for l-range n   | (2 bytes)
+---+---+---+---+---+---+---+---+---+---+
| ND           |          (1 byte)
+---+---+---+---+---+---+---+---+---+---+
|starting nickname for d-range 1| (2 bytes)
+---+---+---+---+---+---+---+---+---+---+
|end nickname for d-range 1   | (2 bytes)
+---+---+---+---+---+---+---+---+---+---+
.
+---+---+---+---+---+---+---+---+---+---+
|starting nickname for d-range n| (2 bytes)
+---+---+---+---+---+---+---+---+---+---+
|end nickname for d-range n   | (2 bytes)

```

+-----+

Figure 4 Nickname Management sub-TLV structure

- o Type: NICK-MGMT, (TBD).

- o Length: variable. Length is $1 + \text{length of Area ID} + 1 + 2 \times 2 \times \text{NL} + 1 + 2 \times 2 \times \text{ND}$, where NL is the number local ranges and ND is number of dynamic ranges.
- o Area-ID : (variable length) Is the IS-IS Area ID. Length can be 1-13 bytes long.
- o NL : number of local nickname ranges.
- o ND : number of dynamic ranges.
- o Starting nickname for l-range n: starting nickname of the local range n.
- o End nickname for l-range n: End nickname of the local range n.
- o Starting nickname for d-range n: starting nickname of the dynamic range n.
- o End nickname for d-range n: End nickname of the dynamic range n.

NOTE: Multiple instances of the nickname management sub-TLV MAY be included. Nickname management sub-TLV has usage in non multi-level deployments as well. Nickname management sub-TLV allows administrator to control nickname acquisition by RBridges.

7. Further optimizations

[RFC6325](#) allows multiple instance of nickname sub-TLV. If more specific forwarding is required, for some critical RBridges, such nicknames MAY be advertised in a separate Router capability TLV, with S bit set. So it leaks to all L1 areas.

[7.1.](#) Leaking of TRILL IS-IS sub-TLV within areas

At the boundary nodes, the following information needs to be leaked from the Level-1 database to the Level-2 database. The nickname TLVs that are learned from all nodes within the same site needs to be redistributed to the Level-2 database. All such redistributed nickname TLVs will have the root priority set to 0. Note, that all border area nodes will announce this TLV into the Level-2 database. As a result, all Level-2 nodes will be able to see the reachability of all other nodes. This enables unicast traffic flow. With respect to multicast, redistributed nicknames are not to be used in the root election for global trees. The roots of the global trees will be from the set of nicknames that are in the Level-2 database. Once the

roots have been identified, these nicknames need to be leaked back to the Level-1 areas. Calculation of multi-destination trees are presented in [section 7.2](#).

[7.2.](#) Identification of Global Trees

It is expected only a sub-set of traffic requires a global reach (inter Area). Majority of the traffic will be of local scope (intra area). Scope of the traffic can be identified either based on VLAN or fine-grain label. Traffic of global scope is forwarded using global trees. The trees of global scope may be identified:

1. By means of configuration
2. By means of multi-destination tree announcements sub-TLV.

Point number 2 above requires further clarifications. We propose to introduce 3 new sub-TLV under IS-IS router capability TLV to advertise global multi-destination trees. These new sub-TLV are listed below. Format of the new sub-TLVs are presented in [section 7.2.1](#) to 7.2.3.

- o The Global Tree capability sub-TLV
- o The Global Tree proposal sub-TLV
- o The Global Tree identifier sub-TLV

Each RBridge announces two sets of capabilities; global tree capabilities and local tree capabilities. It announces, maximum number of global distributions trees it can compute. RBridges

utilize Global Tree sub-TLV for the purpose of announcing its global tree capability. RBridges announce local tree capabilities using The Tree sub-TLV [[RFC6326](#)]. Global tree space is disjoint from the local tree space and MUST not have any effect on each other. Please refer to [[RFC6325](#)] for the process of how local trees are derived. In this section we present how the total number of global trees needed is calculated and identification of nickname of each tree root.

Each area border RBridge, using the global tree sub-TLV received from RBridges in its local area, derives the number of trees the area can support. The number of global trees "s", a local area can support is the fewest number of global trees that an RBridge in local area can support. Area border RBridges advertise in to the backbone, using the global Tree capability sub-TLV, the number of trees "s" that the given area can calculate. Global Tree capability sub-TLV announced in to the L2 backbone are advertised with "S" and "D" flags of Router Capability TLV set to 0. This prevent them leaking to L1 areas.

Rbridges in the L2 backbone calculate the number of global trees the campus can calculate. This number "i" is the fewest number of global trees among all areas.

Each RBridge in the L2 backbone using Global Tree proposal sub-TLV advertise the number of trees it want other RBridges in the L2 backbone to calculate. RBridges in the L2 backbone identifies number of global trees they need to calculate based on the number of trees "k" advertised by the highest priority RBridge in the L2 backbone.

The L2 area RBridge with the highest priority advertises set of nicknames for the global tree roots. These tree roots are selected based on tree root priority announced by L2 backbone RBridges. Global Tree Identifier sub-TLV is used for the purpose.

BSR RBridge of each L1 area advertises nicknames of global tree roots using Global Tree Identifier sub-TLV in to the corresponding local area. BSR also advertises the number of Global trees k the local area needs to calculate using Global tree proposal sub-TLV.

RBridges in the local area contain only nicknames of the local area. Global Tree Identifier sub-TLV announced by the BSR contains nicknames that are unknown to the local L1 area RBridges. How do

local RBridges calculate it SPF ?

Global Tree Identifier sub_TLV contain nickname of the global trees ordered in ascending order. Default Affinity TLV announced by RP RBridge contains the tree-id(s) that it is an RP. Tree-id k in the Default affinity TLV corresponds to nickname k in the Global Tree Identifier. RBridges in the local L1 area calculate its SPF tree assuming the tree-k is rooted at the RP RBridge announcing the Default affinity sub-TLV for that tree.

Global Tree proposal sub-TLV and Global Tree Identifier sub-TLV MUST only be advertised by BSR. Sub-TLV from highest priority RBridge is chosen, in the event of multiple RBridges advertising conflicting sub-TLVs.

[7.2.1.](#) Global Tree capability sub-TLV

```
+---+---+---+---+
|Type = GL-TREES|          (1 byte)
+---+---+---+---+
|  Length      |          (1 byte)
+---+---+---+---+---+---+---+---+---+---+
| Maximum trees able to compute | (2 byte)
+---+---+---+---+---+---+---+---+---+---+
```

Figure 5 Global Tree Capability sub-TLV

Type: Router Capability sub-TLV type, TBD

Length: 2.

Maximum trees able to compute: An unsigned 16-bit integer indicates maximum number of global trees the announcing RBridge able to

compute.

[7.2.2.](#) Global Tree proposal sub-TLV

```
+-----+
|Type = GL-TR-PR|          (1 byte)
+-----+
|  Length      |          (1 byte)
+-----+-----+
| Maximum trees to compute | (2 byte)
+-----+-----+
```

Figure 6 Global Tree Proposal sub-TLV

Type: Router Capability sub-TLV type, TBD (GL-TR-PR)

Length: 2.

Maximum trees to compute: An unsigned 16-bit integer indicates maximum number of global trees the announcing RBridge wants area RBRidges to calculate.

[7.2.3.](#) Global Tree Identifier sub-TLV

```
+-----+
|Type=GLTR-RT-IDs|        (1 byte)
+-----+
|  Length      |          (1 byte)
+-----+-----+
|Starting Tree Number      | (2 bytes)
+-----+-----+
|  Nickname (K-th root)    | (2 bytes)
+-----+-----+
|  Nickname (K+1 - th root) | (2 bytes)
+-----+-----+
```

```

|  Nickname (...)  |
+---+---+---+---+

```

Figure 7 Global Tree Identifier sub-TLV

Type: Router Capability sub-TLV type, set to TBD (GLTR-RT-IDs).

Length: $2 + 2 \times n$, where n is the number of nicknames listed.

Starting Tree Number: This identifies the starting tree number of the nicknames that are trees for the domain. This is set to 1 for the sub-TLV containing the first list. Other Tree-Identifiers sub-TLVs will have the number of the starting list they contain. In the event a tree identifier can be computed from two such sub-TLVs and they are different, then it is assumed that this is a transient condition that will get cleared. During this transient time, such a tree SHOULD NOT be computed unless such computation is indicated by all relevant sub-TLVs present.

Nickname: The nickname at which a distribution tree is rooted.

[7.3. Announcing Group Addresses](#)

Group Address announcements facilitate optimization of multicast forwarding. [[RFC6326](#)] and [[rfc6326bis](#)], define series of sub-TLV to announce various flavors of Group addresses. These sub-TLVs are encapsulated in Group Address TLV (142). We propose to define new

set of sub-TLV under Group Address TLV to carry Group Address announcements applicable to Global trees. IS-IS Group Address TLV (142) does not have flags to control the scope of the TLV. Hence, explicit, sub-TLV definitions are required to indentify group announcements that have global scope.

New sub-TLV numbers are required for the following.

- o Group MAC Address Sub-TLV
- o Group IPv4 Address Sub-TLV
- o Group IPv6 Address Sub-TLV
- o Group Labeled MAC Address Sub-TLV
- o Group Labeled IPv4 Address Sub-TLV
- o Group Labeled IPv6 Address Sub-TLV

Above sub-TLVs as defined in [[RFC6326](#)] and [[rfc6326bis](#)] applies to all trees within the TRILL campus. New sub-TLV definitions include flexibility to define the applicable multicast trees. This flexibility allows applications to further optimize multicast pruning per multicast tree basis.

New group address sub-TLV will be named as below and they have common TLV header as defined in Figure 8.

- o Group MAC Address-multicast tree Sub-TLV
- o Group IPv4 Address-multicast tree Sub-TLV
- o Group IPv6 Address-multicast tree Sub-TLV
- o Group Labeled MAC Address-multicast tree Sub-TLV
- o Group Labeled IPv4 Address-multicast tree Sub-TLV
- o Group Labeled IPv6 Address-multicast tree Sub-TLV

```

+-+--+--+--+--+--+
|Type=G-ADDR-TR |
+-+--+--+--+--+--+

```

(1 byte)

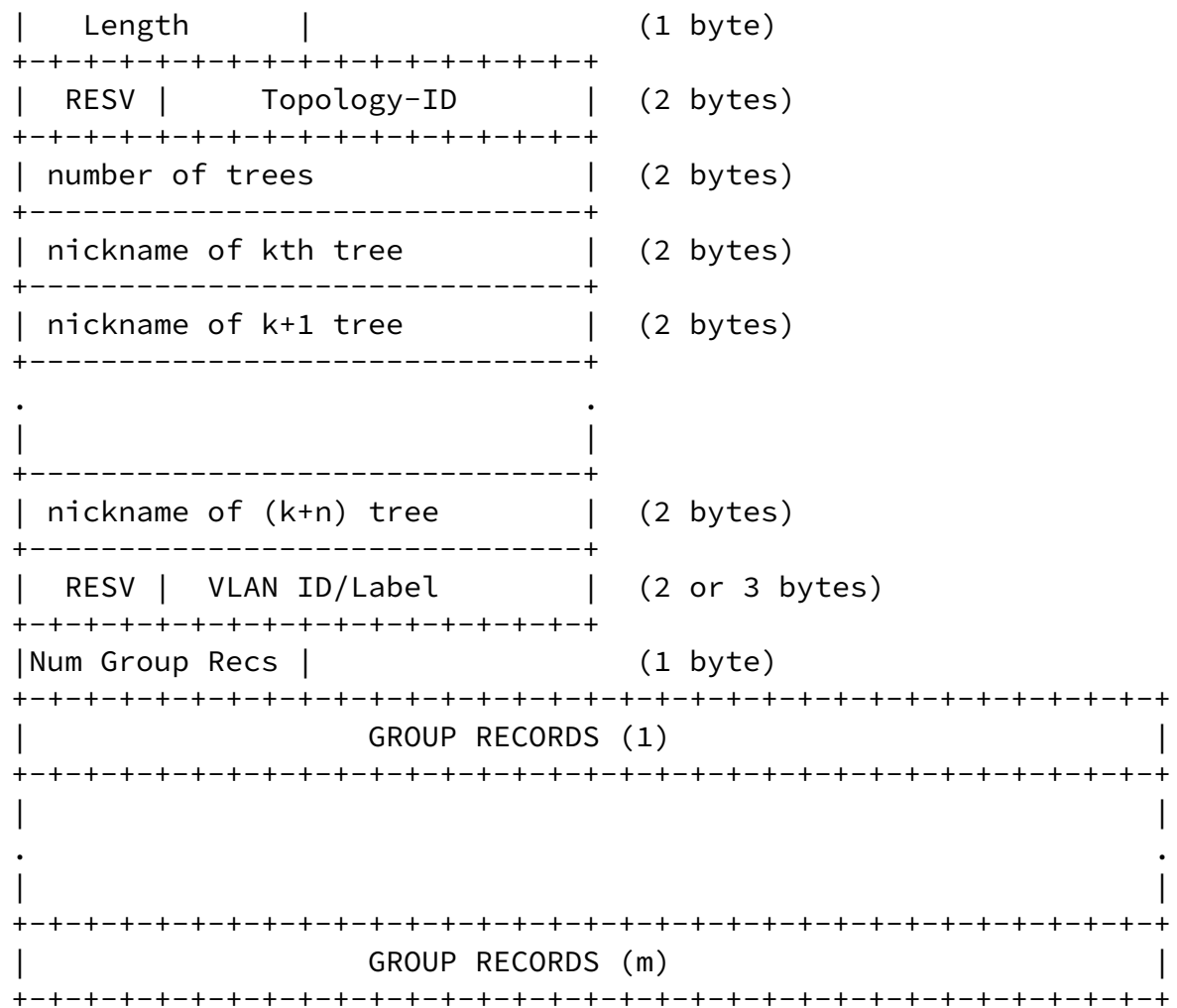


Figure 8 Common structure of Group-Address-multicast tree sub-TLVs

- o Type: G-ADDR-TR, (TBD). Defines sub-TLV for
 - o Group MAC Address-multicast tree Sub-TLV
 - o Group IPv4 Address-multicast tree Sub-TLV
 - o Group IPv6 Address-multicast tree Sub-TLV
 - o Group Labeled MAC Address-multicast tree Sub-TLV
- o Length: Applicable length of the sub-TLV in bytes. Excludes length of G-ADDR-TR and Length fields.
- o Topology ID: 2 byte identifier of the topology instance id.

- o Number of trees: Number of tree nicknames included in the sub-TLV. The nicknames included in the sub-TLV MUST be sorted in ascending order.
- o Nickname of kth tree: 2 byte nickname of the kth tree to which this Group address-tree announcement applies.
- o VLAN ID/Label: Either 4bit reserved followed by 12bit VLAN ID or 24bit finegrain label. Please see [[rfc6326bis](#)] for details.
- o Num Grp Recs: Number of Group Records to follow.
- o Group Record: Contents of the Group Records depend on the G-ADDR-TR definition and identical to their counterparts defined in [[rfc6326bis](#)].

8. Architecture Elements of Multi-level Multicast framework

- o Bootstrap RBridge
- o Rendezvous Point (RP)
- o Default Affinity sub-TLV
- o Area Affinity sub-TLV
- o RP Election Protocol

Five main elements of the multi-level multicast framework are listed above. Functional overviews of the elements are discussed below. Details, such as state machines, PDU format, etc, will be presented in later versions of this document.

8.1. Bootstrap RBridge

Each Area has one more area border RBridges between the Area and the L2 backbone area. For each area one of its area border RBridges is elected as the Bootstrap RBridge.

Border RBridges communicate with each other using the TRILL BSR protocol. Each border RBridge has a configured priority to be a Bootstrap RBridge with system-ID as the tie breaker. The RBridge with the highest priority become the bootstrap RBridge for the area.

Bootstrap RBridge selects the Rendezvous Point (RP) RBridges and assign each RP a set of trees for which RP will function as the gateway between the local and global multicast trees. To avoid loops and/or packet duplication the set of trees MUST only be allocated to one and only one RP.

[8.2.](#) Rendezvous Point (RP)

Rendezvous Point (RP) RBridge performs the function of gateway (or acts as a point of plumbing) between the local multicast tree and the corresponding global multicast tree rooted in the L2 backbone area.

Bootstrap RBridge designates one of the border RBridges (including itself) as the RP for a set of (mutually exclusive) trees.

Each border RBridge, using TRILL BSR protocol, announces its desire to be an RP. The desire to be an RP is a configurable option and enabled by default to be an RP.

[8.3.](#) Default Affinity sub-TLV

Default Affinity sub-TLV is announced by each RP to inform the RBridges in the L1 Area the association of the default nickname to a set of trees through the announcing RP [trillcmt].

RBridges in the L1 area, based on the Default Affinity sub-TLV installs the RPF check for default nickname for the specified tree "t" on an interface facing towards the announcing RP.

[8.4.](#) Area Affinity sub-TLV

The Area Affinity sub-TLV announced by each RP to inform the RBridges in L2 backbone Area the association of local L1 Area nicknames to a set of trees through the announcing RP (Figure 3).

RBridges in the L2 backbone area or has interfaces to the L2 backbone area, based on the Area Affinity sub-TLV, installs the RPF check for the nicknames in the indicated L1 Area, for the specified tree "t", on an interface facing towards the announcing RP.

[8.5.](#) TRILL BSR Protocol

9. Security Considerations

TBD

Senevirathne

Expires November 28, 2013

[Page 22]

Internet-Draft

Multilevel TRILL

May 2013

10. IANA Considerations

IANA is requested to add the Area Affinity sub-TLV, Nickname Management sub-TLV, Global Tree capability sub-TLV, Global tree proposal sub-TLV, Global tree Identifier sub-TLV as sub-TLVs under Router capability TLV.

11. References

11.1. Normative References

- [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", [BCP 14](#), [RFC 2119](#), March 1997.
- [RFC2234] Crocker, D. and Overell, P.(Editors), "Augmented BNF for Syntax Specifications: ABNF", [RFC 2234](#), Internet Mail Consortium and Demon Internet Ltd., November 1997.
- [RFC4971] Vasseur, JP, et.al, "Intermediate System to Intermediate System (IS-IS) Extensions for Advertising Router Information ", [RFC 4971](#), July 2007.
- [RFC6325] Perlma, R., et.al, "Routing Bridges (RBridges): Base Protocol Specification", [RFC 6325](#), July 2011.
- [trillcmt]Senevirathne, T., et.al, "Coordinated Multicast Trees (CMT)for TRILL", Work in Progress, January 2012.
- [RFC4971] Vasseur, JP, et.al, "Intermediate System to Intermediate System (IS-IS) Extensions for Advertising Router Information", [RFC 4971](#), July 2007.

11.2. Informative References

- [RFC6326] Eastlake, D, et.al, "Transparent Interconnection of Lots of Links (TRILL) Use of IS-IS", [RFC 6326](#), July 2011.
- [rfc6326bis] Eastlake, D, et.al, "Transparent Interconnection of

Lots of Links (TRILL) Use of IS-IS", Work in Progress,
[draft-eastlake-isis-rfc6326bis-04.txt](#), January 2012.

[trillml] Perlman, R., et.al, "RBridges: Multilevel TRILL", Work in Progress, [draft-perlman-trill-rbridge-multilevel-03.txt](#), October 2011.

Senevirathne

Expires November 28, 2013

[Page 23]

Internet-Draft

Multilevel TRILL

May 2013

[12](#). Acknowledgments

We wish to thank Leonard Tracy, Dinesh Dutt and Ashok Ganesan for reviewing and providing constructive feedback.

This document was prepared using 2-Word-v2.0.template.dot.

Authors' Addresses

Tissa Senevirathne
CISCO Systems
375 East Tasman Drive
San Jose CA 95134

Phone: 408-853-2291
Email: tsenevir@cisco.com

Les Ginsberg
CISCO Systems
510 McCarthy Blvd.
Milpitas CA 95035

Phone: 408-527-7729
Email: ginsberg@cisco.com

Sam Aldrin
Huawei Technologies
2330 Central Express Way
Santa Clara, CA 95051

Email: aldrin.ietf@gmail.com

Ayan Banerjee

CISCO Systems
425 East Tasman Drive
San Jose CA 95134

Phone: 408-527-0539
Email: ayabaner@cisco.com