TRILL Working Group
INTERNET-DRAFT
Intended Status: Standard Track

W. Hao
Y. Li
Huawei
A. Qu
MediaTec
M. Durrani
Cisco
P. Sivamurugan
IP Infusion
L. Xia
Huawei

Expires: January 06, 2016
July 06, 2015

# TRILL Distributed Layer 3 Gateway
# draft-ietf-trill-irb-06.txt

Abstract

Currently the TRILL protocol provides optimal pair-wise data frame forwarding for layer 2 intra-subnet traffic but not for layer 3 inter-subnet traffic. A centralized gateway solution is typically used for layer 3 inter-subnet traffic forwarding but has the following issues:

1. Sub-optimum forwarding paths for inter-subnet traffic.

2. A centralized gateway may need to support a very large number of gateway interfaces in a data center, one per tenant per data label used by that tenant, to provide interconnect functionality for all the layer 2 virtual networks in entire TRILL network.

3. A traffic bottleneck at the gateway.

This document specifies an optional TRILL distributed gateway solution that resolves these centralized gateway issues.

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/1id-abstracts.html

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



Table of Contents

## 1. Introduction

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

Currently, TRILL provides optimal unicast forwarding for Layer 2 intra-subnet traffic but not for Layer 3 inter-subnet traffic, where subnet means different IP address prefix and typically a different Data Label (VLAN or FGL). In this document, an optional TRILL-based distributed Layer 3 gateway solution is specified to provide optimal unicast forwarding for Layer 3 inter-subnet traffic. With distributed gateway support an edge RBridge provides both routing based on Layer 2 identity (address and virtual network (VN, i.e. Data Label)) among end stations (ESs) that belong to same subnet and routing based on Layer 3 identity among ESs that belong to different subnets of the same routing domain. An edge RBridge supporting this feature needs to provide routing instances and Layer 3 gateway interfaces for local connected ESs. Such routing instances provide IP address isolation between tenants. In the TRILL distributed Layer 3 gateway solution, inter-subnet traffic can be fully spread over edge RBridges, so there is no single bottleneck.

### 1.1. Document Organization

This document is organized as follows: Section 3 gives a simplified example and more detailed problem statement. Section 4 gives the Layer 3 traffic forwarding model. Section 5 provides a distributed gateway solution overview. Section 6 gives a detailed distributed gateway solution example. And Section 7 describes the TRILL protocol extensions needed to support this distributed gateway solution.

## 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].

The terms and acronyms in [RFC6325] are used with the following additions:

ARP: Address Resolution Protocol [RFC826].

Data Label: VLAN or FGL [RFC7172].

DCN: Data Center Network.

ES: End Station. VM (Virtual Machine) or physical server, whose address is

either the destination or source of a data frame.

Gateway interface: Layer 3 virtual interface on gateway aka gateway interface) terminates layer 2 forwarding and forwards IP traffic to the destination as per IP forwarding rules. Incoming traffic from a physical port on a gateway will be distributed to its virtual gateway interface based on Data Label (VLAN or FGL).

L2: Layer 2.

L3: IP Layer 3.

ND: IPv6's Neighbor Discovery [RFC4861].

ToR: Top of Rack.

VN: Virtual Network. In a TRILL campus, each virtual network is identified by a unique 12-bit VLAN ID or 24-bit Fine Grained Label [RFC7172].

VRF: Virtual Routing and Forwarding. In IP-based computer networks, Virtual Routing and Forwarding (VRF) is a technology that allows multiple instances of a routing table to co-exist within the same router at the same time.

## 3. Simplified Example and Problem Statement

Section 3.1 gives a simplified example in a TRILL campus with and without a distributed layer 3 gateway using VLAN Data Labels. A more detailed description of the problem without a distributed layer 3 gateway is given in Section 3.2. The remainder of this document, particularly Section 5, describes the distributed gateway solution in more detail.

### 3.1. Distributed Gateway Simplified Example

Assuming a tenant has four subnets, each subnet corresponds to one VLAN indicating one individual layer 2 virtual network, say the VLANs are VLAN 10 to VLAN 13, the end stations in VLAN 10 and VLAN 11 are connected to RB1 and RB2, and the end stations(ESs) in VLAN 12 and VLAN 13 are connected to RB3 and RB4. TRILL makes all end stations in each VLAN appear to be on the same layer 2 link. Their layer 3 IP gateway also appears as an end station on each link. Each tenant end station finds the layer 3 gateway's MAC address by using ARP or ND to ask for the MAC address corresponding to its IP router.

For traffic within a subnet, that is IP traffic to another end station in the same VLAN attached to the TRILL campus, the end station just ARPs for the MAC address for the destination end station's IP. It then uses this MAC address for traffic to that destination and TRILL routes the ingressed TRILL data packets to the destination's edge RBridge based on the egress nickname for that destination MAC address and VLAN. This is the regular process as defined in TRILL base protocol [RFC6325].

In centralized layer 3 gateway solution, all traffic within that tenant between different VLANs must go through the centralized layer 3 gateway device, say Gateway 1, even if the traffic is between two end stations connected to the same edge RBridge, because only the layer 3 gateway can change the VLAN labeling of the traffic. Gateway 1 has four gateway interfaces for these four VLANs.

With the distributed layer 3 gateway, each edge RBridge acts as a default layer 3 gateway for local connecting ESs, it also has IP router capabilities to provide IP communication with other edge RBridges. Each edge RBridge only needs gateway interfaces for local connecting ESs, i.e., RB1 and RB2 have gateway interfaces for VLAN 10 and VLAN 11, RB3 and RB4 have gateway interfaces for VLAN 12 and VLAN 13. No RBridges should normally need to maintain gateway interfaces all VLANs, because each RBridge normally only supports a limited number of VNs. This will enhance the scalability of tenants number and subnets number per tenant.

When each end station ARPs for their layer 3 gateway, that is, their IP router, the edge RBridge to which it is connected will respond with that RBridge's 'gateway MAC'. When the end station later sends IP traffic to the layer 3 gateway, because the destination IP is outside of its subnet, the edge RBridge intercepts the IP packet because the destination MAC is its gateway MAC. That RBridge routes the IP packet using the routing instance associated with that tenant, handling it in one of three ways:

(1) If the destination IP is connected to the same edge RBridge, that RBridge can simply transmit the IP packet out the right edge port in the destination VLAN.

(2) If the destination IP is located in an outside network, the edge RBridge encapsulates it as a TRILL Data packet and sends it to the actual TRILL campus edge RBridge connecting to the outside network.

(3) if the destination in an end station connected to a different edge RBridge, the ingress RBridge uses TRILL encapsulation to route the IP packet to the correct egress RBridge, using that RBridge's gateway MAC and an Inner.VLAN identifying the tenant. Finally, the egress RBridge terminates the TRILL encapsulation and routes the IP packet to the destination end station based on the routing instance for that tenant.

Through the distributed layer 3 gateway solution, the inter-subnet traffic are fully dispersed and are transmitted along optimal pair-wise forwarding path, improving network efficiency.

## 3.2. Problem Statement

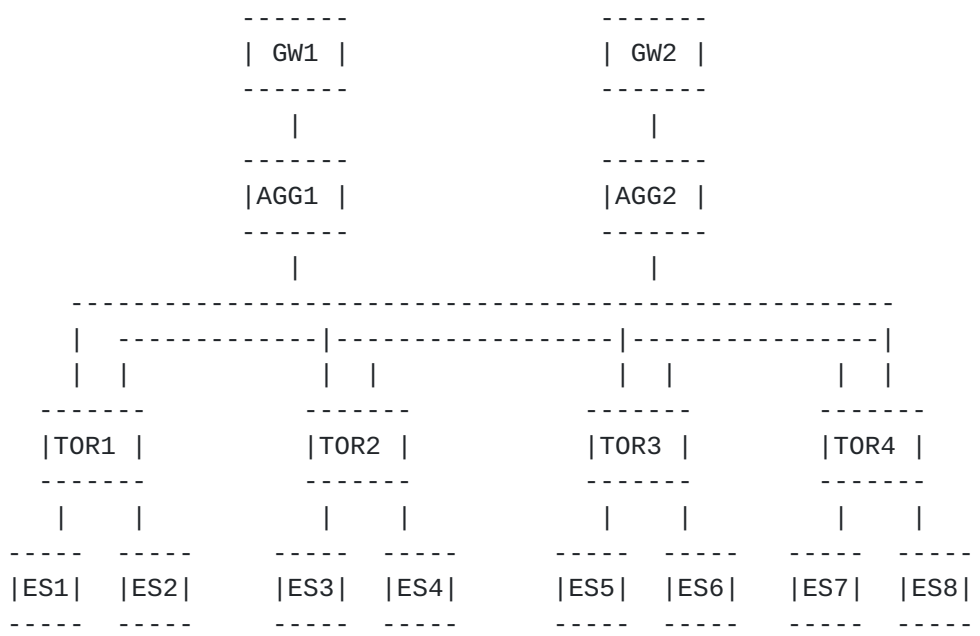



Figure 1. A Typical DC Network

Figure 1 depicts a TRILL Data Center Network (DCN) where edge RBridges are Top of Rack (ToR) switches. Centralized gateway GW1 and GW2 in

figure 1 provide the layer 3 packet forwarding for both north-south traffic and east-west inter-subnet traffic between ESs.

End stations in one IP subnet expect to send IP traffic for a different subnet to an IP router. In addition, there is normally a Data Label (VLAN or FGL) associated with each IP subnet but there is no facility in the base TRILL protocol [RFC6325] to change the Data Label for traffic between subnets. If two end stations of the same tenant are on two different subnets and need to communicate with each other, their packets are typically forwarded all the way to a centralized IP Layer 3 gateway to perform L3 routing and, if necessary, change the Data Label.

This is generally sub-optimal because the two end stations may be connected to the same ToR where L3 switching could have been performed locally. For example, in above Figure 1, assuming ES1 (10.1.1.2 ) and ES2 (20.1.1.2) belong to different subnets of same tenant, the unicast IP traffic between them has to go through a centralized gateway. It can't be locally router between them on TOR1. However, if an edge RBridge has the distributed gateway capabilities specified in this document, then it can still perform optimum L2 forwarding for intra-subnet traffic and, in addition, optimum L3 forwarding for inter-subnet traffic, thus delivering optimum forwarding for unicast packets in all important cases.

With Fine Grained Labeling [RFC7172], in theory up to 16 million Layer 2 VN can be supported in a TRILL campus. To support inter- subnet traffic, a very large number of Layer 3 gateway interfaces could be needed on a centralized gateway if each VN corresponds to a subnet and there are many tenant with many subnets per tenant. It is a big burden for the centralized gateway to support so many interfaces. In addition all inter-subnet traffic will go through the centralized gateway that may become the traffic bottleneck.

In summary, the centralized gateway has the following issues:

1. Sub-optimum forwarding paths for inter-subnet traffic due to the requirements to perform IP routing and possibly change Data Labels at a centralized gateway.

2. The centralized gateway may need to support a very large number of gateway interfaces, in a data center one per tenant per data label used by that tenant, to provide interconnect functionality for all the layer 2 virtual networks in entire TRILL network.

3. A traffic bottleneck at the centralized gateway.

A distributed gateway on edge RBridges addresses these issues.

## 4. Layer 3 Traffic Forwarding Model

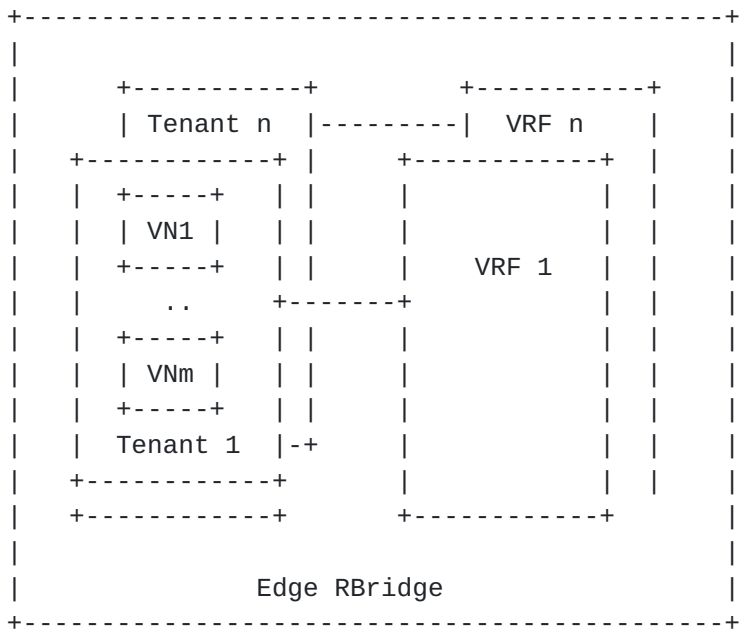



Figure 2. Edge RBridge Model as Distributed Gateway

In a data center network (DCN), each tenant has one or more Layer 2 virtual networks and, in normal cases, each tenant corresponds to one routing domain. Normally each Layer 2 virtual network uses a different Data Label and corresponds to one or more subnets.

Each Layer 2 virtual network in a TRILL campus is identified by a unique 12-bit VLAN ID or 24-bit Fine Grained Label [RFC7172]. Different routing domains may have overlapping address space but need distinct and separate routes. The end stations that belong to the same subnet communicate through L2 forwarding, end systems of the same tenant that belong to different subnets communicate through L3 routing.

Figure 2 depicts the model where there are n VRFs corresponding to n tenants with each tenant having up to m segments/subnets (virtual network).

## 5. Distributed Gateway Solution Overview

In the TRILL distributed gateway scenario, an edge RBridge must perform Layer 2 routing for the ESs that are on the same subnet and IP routing for the ESs that are on the different subnets of the same tenant.

As the IP address space in different routing domains can overlap, VRF instances need to be created on each edge RBridge to isolate the IP

forwarding process for different routing domains present on the edge RBridge. A globally unique tenant ID identifies each routing domain. The network operator should ensure the consistency of the tenant ID on each edge RBridge for each routing domain. If a routing domain spreads over multiple edge RBridges, routing information for the routing domain must be synchronized among these edge RBridges to ensure the reachability to all ESs in that routing domain. The Tenant ID is carried with the routing information to differentiate the routing domains.

From the data plane perspective, all edge RBridges are connected to each other via one or multiple TRILL hops, however they are always a single IP hop away. When an ingress RBridge receives inter-subnet traffic from a local ES whose destination MAC is the edge RBridge's gateway MAC, that RBridge will perform Ethernet header termination and look up in its IP routing table to route the traffic to the IP next hop. If the destination ES is connected to a remote edge RBridge, the remote RBridge will be the IP next hop for traffic forwarding. The ingress RBridge will perform TRILL encapsulation for such inter-subnet traffic and route it to the remote RBridge through the TRILL campus.

When that remote RBridge receives the traffic, it will decapsulate the packet and then lookup in the RBridge's IP forwarding table to route it to the destination ES. Through this method, TRILL with distributed gateways provides pair-wise data routing for inter-subnet traffic.

### 5.1. Local routing information

An ES can be locally connected to an edge RBridges through a layer 2 network or externally connected through a layer 3 IP network.

If the ES is connected to an edge RBridge through a Layer 2 network,then the edge RBridge must act as a Layer 3 Gateway for the ES. A gateway interface should be established on the edge RBridge for the connecting ES. Because the ESs in a subnet may be spread over multiple edge RBridges, each of these edge RBridges should establish its gateway interface for the subnet and these gateway interfaces on different edge RBridges share the same gateway MAC and gateway IP address.

Before an ES starts to send inter-subnet traffic, it should acquire its gateway's MAC through the ARP/ND process. Local connecting edge RBridges that are supporting this distributed gateway feature always respond with the gateway MAC address when receiving ARP/ND requests for the gateway IP. Through the ARP/ND process, the edge RBridge can learn the IP and MAC correspondence of a local ES connected to the edge RBridge by Layer 2 and then generate local IP routing entries for the ES in the corresponding routing domain.

An IP router looks to TRILL like an ES. If a router/ES is located in an external IP network, normally it provides access to one or more IP prefixs. The router/ES should run an IP routing protocol with the connecting TRILL edge Rbridge. The edge RBridge will learn the IP prefixes behind the router/ES through the IP routing protocol, then the RBridge will generate local IP routing entries in the corresponding routing domain.

## 5.2. Local routing information synchronization

When a routing instance is created on an edge RBridge, the tenant ID, tenant Label (VLAN or FGL), tenant gateway MAC, and their correspondence should be set and globally advertised (see Section 7.1).

When an ingress RBridge performs inter-subnet traffic TRILL encapsulation, the ingress RBridge uses the Label advertised by the egress RBridge as the inner VLAN or FGL and uses the tenant gateway MAC advertised by the egress RBridge as the Inner.MacDA. The egress RBridge relies on this tenant Data Label to find the local VRF instance for the IP forwarding process when receiving inter-subnet traffic from the TRILL campus. (The role of tenant Label is akin to an MPLS VPN Label in an MPLS IP/MPLS VPN network.) Tenant Data Labels are independently allocated on each edge RBridge for each routing domain, an edge RBridge can pick up an access Data Label in a routing domain to act as the inter-subnet Label, or the edge RBridge can use a different Label from any access Labels to act as tenant Label. It's implementation dependant and there is no restriction on this. The tenant gateway MAC differentiates inter-subnet Layer 3 traffic or intra-subnet Layer 2 traffic on the egress RBridge. Each tenant on a RBridge can use a different gateway MAC or same tenant gateway MAC for inter-subnet traffic purposes. This is also implementation dependant and there is no restriction on it.

When a local IP prefix is learned in a routing instance on an edge RBridge, the edge RBridge should advertise the IP prefix information for the routing instance to other edge RBridges to generate IP routing entries. If the ESs in a VN are spread over multiple RBridges, these RBridges should advertise each local connecting end station's IP address in the VN to other RBridges. If the ESs in a VN are only connected to one edge RBridge, that RBridge only needs to advertise the subnet corresponding to the VN to other RBridges. A globally unique tenant ID also should be carried to differentiate IP prefixes between different tenants, because the IP address space of different tenants can overlap (see Sections 7.3 and 7.4).

If a tenant is deleted on an edge Rbridge, the edge Rbridge should notify all other edge Rbridges to delete local IP prefixs, tenant Label and tenant gateway MAC. If there is a new tenant which is created and the

original's tenant label is assigned to the new tenant immediately, it may cause a security policy violation for the traffic in flight, because when the egress Rbridge receives traffic from the old tenant, it will forward it in the new tenant's routing instance and deliver it to wrong destination. So tenant Label MUST NOT be re-allocated until a reasonable amount of time has passed to allow any traffic in flight to be discarded.

When the ARP entry in an edge Rbridge for an ES times out, it will trigger an edge Rbridge LSP advertisement to other edge Rbridges with the corresponding IP routing entry deleted. If the ES is an IP router, the edge Rbridge also notifies other edge Rbridges that they must delete the routing entries corresponding to the IP prefixs accessible through that IP router. During the IP prefix deleting process, if there is traffic in flight, the traffic will be discarded at the egress Rbridge because there is no local IP routing entry to the destination.

If an edge Rbridge changes its tenant gateway MAC, it will trigger an edge Rbridge LSP advertisement to other edge Rbridges giving the new gateway MAC as Inner.MacDA for future traffic destined to the edge Rbridge. During the gateway MAC changing process, if there is traffic in flight using the old gateway MAC as Inner.MacDA, the traffic will be discarded or be forwarded as layer 2 intra-subnet traffic on the edge Rbridge. If the inter-subnet tenant Label is a unique Label which is different from any access Labels, when the edge Rbridge receives the traffic whose Inner.MacDA is different from local tenant gateway MAC, the traffic will be discarded. If the edge RBridge uses one of the access Labels as inter-subnet tenant Label, the traffic will be forwarded as layer 2 intra-subnet traffic unless special traffic filtering policy is enforced on the edge Rbridge.

If there are multiple nicknames owned by an edge RBridge, the edge RBridge also can specify one nickname as the egress nickname for inter-subnet traffic forwarding. A NickFlags APPsub-TLV with the SE-flag set can be used for this purpose. If the edge RBridge doesn't specify a nickname for this purpose, the ingress RBridge can use any one of the nicknames owned by the egress as the egress nickname for inter-subnet traffic forwarding.

TRILL E-L1FS FS-LSP [rfc7180bis] APPsub-TLVs can be used for IP routing information synchronization in each routing domain among edge RBridges. Based on the synchronized information from other edge RBridges, each edge RBridge generates remote IP routing entries in each routing domain.

Through this solution, the intra-subnet forwarding function and inter-subnet IP routing functions are integrated and network management and deployment will be simplified.

### 5.3. Active-active access

TRILL active-active service provides end stations with flow level load balance and resilience against link failures at the edge of TRILL campuses as described in [RFC7379].

If an ES is connected to two TRILL RBridges, say RB1 and RB2, in active-active mode, RB1 and RB2 can act as distributed layer 3 gateway for the ES. RB1 and RB2 each learn the ES's IP address through ARP/ND process and then they announce the IP address to the TRILL campus independently. The remote ingress RBridge will generate an IP routing entry corresponding with the IP address with two IP next hops of RB1 and RB2.

When the ingress RBridge receives inter-subnet traffic from a local access network, the ingress RBridge selects RB1 or RB2 as the IP next hop based on local load balancing algorithm, then the traffic will be transmitted to the selected next hop destination RB1 or RB2 through the TRILL campus.

### 5.4. Data traffic forwarding process

After a Layer 2 connected ES1 in VLAN-x acquires its gateway's MAC, it can start inter-subnet data traffic transmission to ES2 in VLAN-y.

When the edge RBridge attached to ES1 receives inter-subnet traffic from ES1, that RBridge performs Layer 2 header termination, then, using the local VRF corresponding to VLAN-x, it performs the IP routing process in that VRF.

If destination ES2 is attached to the same edge RBridge, the traffic will be locally forwarded to ES2 by that RBridge. Compared to the centralized gateway solution, the forwarding path is optimal and a traffic detour is avoided.

If ES2 is attached to a remote edge RBridge, the remote edge RBridge is IP next hop and the inter-subnet traffic is forwarded to the IP next hop through TRILL encapsulation. If there are multiple equal cost shortest path between ingress RBridge and egress RBridge, all these path can be used for inter-subnet traffic forwarding, so load spreading can be achieved for inter-subnet traffic.

When the remote RBridge receives the inter-subnet TRILL encapsulated traffic, the RBridge decapsulates the TRILL encapsulation and checks the Inner.MacDA, if that MAC address is the local gateway MAC corresponding to the inner Label (VLAN or FGL), the inner Label will be used to find the corresponding local VRF, then the IP routing process in

that VRF will be performed, and the traffic will be locally forwarded to the destination ES2.

In summary, this solution avoids traffic detours through a central gateway, both inter-subnet and intra-subnet traffic can be forwarded along pair-wise shortest paths, and network bandwidth is conserved.

## 6. Distributed Layer 3 Gateway Process Example

This section gives a detailed description of a distributed layer 3 gateway solution example.

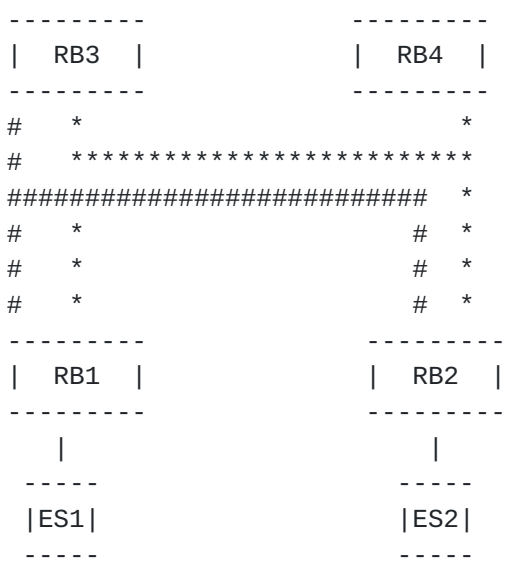



Figure 3. Distributed gateway scenario

In figure 3, RB1 and RB2 support the distribution gateway function, ES1 connects to RB1, ES2 connects to RB2. ES1 and ES2 belong to Tenant1, but are in different subnets.

The IP address, VLAN, and subnet information of ES1 and ES2 are as follows:

| ES | Tenant | IP Address | Subnet | VLAN |
|---|---|---|---|---|
| ES1 | Tenant1 | 10.1.1.2 | 10.1.1.1/32 | 10 |
| ES2 | Tenant1 | 20.1.1.2 | 20.1.1.1/32 | 20 |

Figure 4. ES information

The nickname, VRF, tenant VLAN, tenant gateway MAC for Tenant1 on RB1 and RB2 are as follows:

| RB | Nickname | Tenant | VRF | Tenant VLAN | Gateway MAC |
|---|---|---|---|---|---|
| RB1 | nick1 | Tenant1 | VRF1 | 100 | MAC1 |
| RB2 | nick2 | Tenant1 | VRF2 | 100 | MAC2 |

Figure 5. RBridge information

## 6.1. Control plane process

RB1 announces the following local routing information to the TRILL campus:

Tenant ID: 1

Tenant gateway MAC: MAC1

Tenant VLAN for Tenant1: VLAN 100.

IP prefix in Tenant1: 10.1.1.2/32.

RB2 announces the following local routing information to TRILL campus:

Tenant ID: 1

Tenant gateway MAC: MAC2

Tenant VLAN for Tenant1: VLAN 100.

IP prefix in Tenant1: 20.1.1.2/32.

Relying on the routing information from RB2, remote routing entries on RB1 are generated as follows:

| Prefix/Mask | Inner.MacDA | inner VLAN | egress nickname |
|---|---|---|---|
| 20.1.1.2/32 | MAC2 | 100 | nick2 |

Figure 6. Tenant 1 remote routing table on RB1

Similarly, relying on the routing information from RB1, remote routing entries on RB2 are generated as follows:

| Prefix/Mask | Inner.MacDA | inner VLAN | egress nickname |
|---|---|---|---|

```
|10.1.1.2/32|     MAC1     |   100     |    nick1      |
+-----------+--------------+-----------+---------------+
```

Figure 7. Tenant 1 remote routing table on RB1

## 6.2. Data plane process

Assuming ES1 sends unicast inter-subnet traffic to ES2, the traffic forwarding process is as follows:

1. ES1 sends unicast inter-subnet traffic to RB1 with RB1's gateway's MAC as the destination MAC.

2. Ingress RBridge (RB1) forwarding process:

RB1 checks the destination MAC, if the destination MAC equals the local gateway MAC, the gateway function will terminate the Layer 2 header and perform L3 routing.

RB1 looks up IP routing table information by destination IP and Tenant ID to get IP next hop information, which includes the egress RBridge's gateway MAC (MAC2), tenant VLAN (VLAN 100) and egress nickname (nick2). Using this information, RB1 will perform inner Ethernet header encapsulation and TRILL encapsulation. RB1 will use MAC2 as the Inner.MacDA, MAC1 (RB1's own gateway MAC) as the Inner.MacSA, VLAN 100 as the Inner.VLAN, nick2 as the egress nickname and nick1 as the ingress nickname.

RB1 looks up TRILL forwarding information by egress nickname and sends the traffic to the TRILL next hop as per [RFC6325]. The traffic will be sent to RB3 or RB4 as a result of load balancing.

Assuming the traffic is forwarded to RB3, the following occurs:

3. Transit RBridge (RB3) forwarding process:

RB3 looks up TRILL forwarding information by egress nickname and forwards the traffic to RB2 as per [RFC6325].

4. Egress RBridge forwarding process:

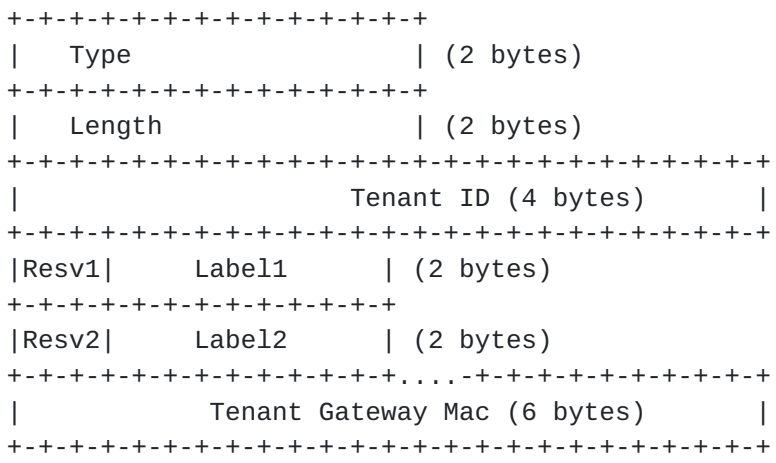
As the egress nickname is RB2's own nickname, RB2 performs TRILL decapsulation. Then it checks the Inner.MacDA and, because that MAC is equal to the local gateway MAC, performs inner Ethernet header termination. Using the inner VLAN, RB2 finds the local corresponding VRF and looks up the packets destination IP address in the VRF's IP routing table. The traffic is then be locally forwarded to ES2.

## 7. TRILL Protocol Extensions

If an edge RBridge RB1 participates in the distributed gateway function, it announces its tenant gateway MAC and tenant Data Label to the TRILL campus through the tenant Label and gateway MAC APPsub-TLV, it should announce its local IPv4 and IPv6 prefixs through the IPv4 Prefix APPsub-TLV and the IPv6 Prefix APPsub-TLV respectively. If RB1 has multiple nicknames, it can announce one nickname for distributed gateway use using Nickname Flags APPsub-TLV with "SE" Flag set to one.

The remote ingress RBridges belonging to the same routing domain use this information to generate IP routing entries in that routing domain. These RBridges use the nickname, tenant gateway MAC and tenant Label of RB1 to perform inter-subnet traffic TRILL encapsulation when they receive inter-subnet traffic from a local ES. The nickname is used as the egress nickname, the tenant gateway MAC is used as the Inner.MacDA, and the tenant Data Label is used as the Inner.Label. The following APPsub-TLVs MUST be included in a TRILL GENINFO TLV in E-L1FS FS-LSPs [rfc7180bis].

### 7.1. The tenant Label and gateway MAC APPsub-TLV




o Type: Set to TENANT-LABEL sub-TLV type (TBD1). Two bytes, because this APPsub-TLV appears in an extended TLV [RFC7356].

o Length: If Label1 field is used to represent a VLAN, the value of the length field is 12. If Label1 and Label2 field are used to represent an FGL, the value of the length field is 14.

o Tenant ID: This identifies a global tenant ID.

o Resv1: 4 bits that MUST be sent as zero and ignored on receipt.

o Label1: If the value of the length field is 12, it identifies a tenant VLAN ID, If the value of the length field is 14, it identifies the higher 12 bits of a tenant FGL.

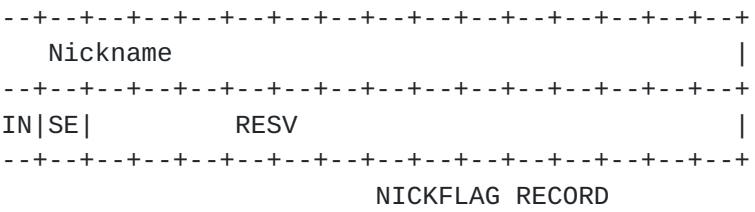
o Resv2: 4 bits that MUST be sent as zero and ignored on receipt. Only present if the length field is 14.

o Label2: This field has the lower 12 bits of tenant FGL. Only present if the length field is 14.

o Tenant Gateway MAC: This identifies the local gateway MAC corresponding to the tenant ID. The remote ingress RBridges use the Gateway MAC as Inner.MacDA. The advertising TRILL RBridge uses the gateway MAC to differentiate layer 2 intra-subnet traffic and layer 3 inter-subnet traffic in the egress direction.

## 7.2. "SE" Flag in NickFlags APPsub-TLV

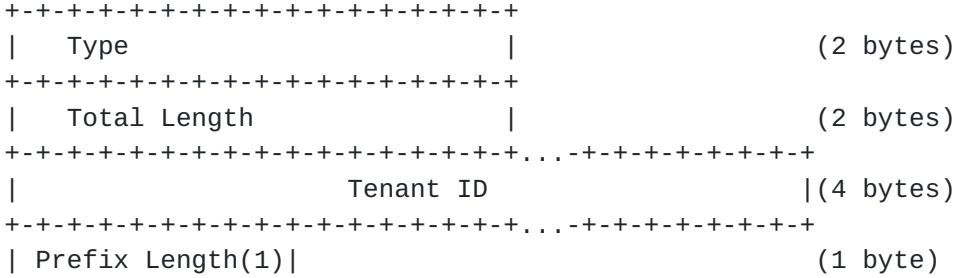
The NickFlags APPsub-TLV is specified in [rfc7180bis]. The SE Flag is assigned as follows:

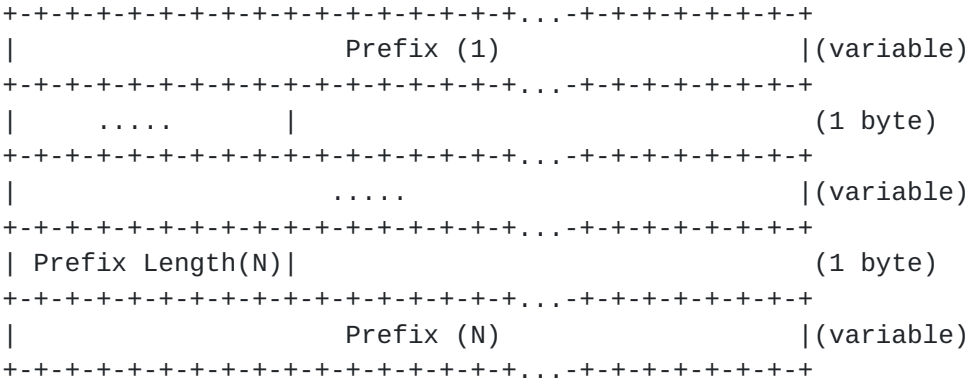



NICKFLAG RECORD

o SE. If the SE flag is one, it indicates that the advertising RBridge suggests the nickname should be used as the Inter-Subnet Egress nickname for inter-subnet traffic forwarding. If flag is zero, that nickname will not be used for that purpose.

## 7.3. The IPv4 Prefix APPsub-TLV






o Type: Set to IPV4-PREFIX sub-TLV type (TBD2). Two bytes, because this APPsub-TLV appears in an extended TLV [RFC7356].

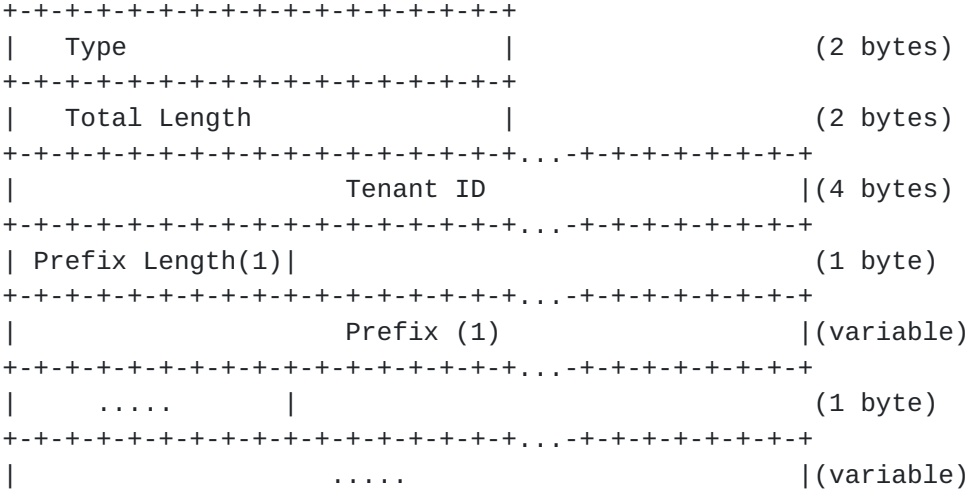
o Total Length: This 2-byte unsigned integer indicates the total length of the Tenant ID, the Prefix Length, and the Prefix fields in octets. A value of 0 indicates that no IPv4 prefix is being advertised.

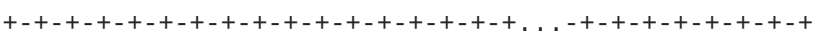
o Tenant ID: This identifies a global tenant ID.

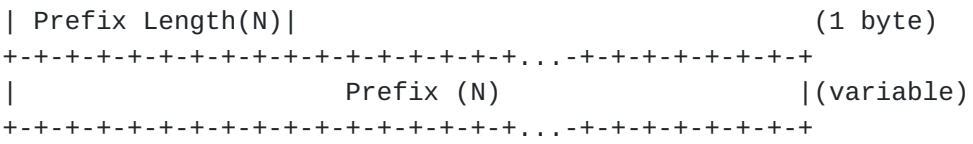
o Prefix Length: The Prefix Length field indicates the length in bits of the IPv4 address prefix. A length of zero indicates a prefix that matches all IPv4 addresses (with prefix, itself, of zero octets).

o Prefix: The Prefix field contains an IPv4 address prefix, followed by enough trailing bits to make the end of the field fall on an octet boundary. Note that the value of the trailing bits is irrelevant. For example, if the Prefix Length is 12, indicating 12 bits, then the Prefix is 2 octets and the low order 4 bits of the Prefix are irrelevant.

## 7.4. The IPv6 Prefix APPsub-TLV








o Type: Set to IPV6-PREFIX sub-TLV type (TBD3). Two bytes, because this APPsub-TLV appears in an extended TLV [RFC7356].

o Total Length: This 2-byte unsigned integer indicates the total length of the Tenant ID, the Prefix Length, and the Prefix fields in octets. A value of 0 indicates that no IPv6 prefix is being advertised.

o Tenant ID: This identifies a global tenant ID.

o Prefix Length: The Prefix Length field indicates the length in bits of the IPv6 address prefix. A length of zero indicates a prefix that matches all IPv6 addresses (with prefix, itself, of zero octets).

o Prefix: The Prefix field contains an IPv6 address prefix, followed by enough trailing bits to make the end of the field fall on an octet boundary. Note that the value of the trailing bits is irrelevant. For example, if the Prefix Length is 100, indicating 100 bits, then the Prefix is 13 octets and the low order 4 bits of the Prefix are irrelevant.

## 8. Security Considerations

Correct configuration of the edge RBridges participating is important to assure that data is not delivered to the wrong tenant, which would violate security constrains. IS-IS security [RFC5310] can be used to secure the information advertised by the edge RBridges in LSPs and FS-LSPs.

Particularly sensitive data should be encrypted end-to-end, that is, from the source end station to the destination end station.

For general TRILL Security Considerations, see [RFC6325].

## 9. IANA Considerations

IANA is requested to assign three APPsub-TLV type numbers less than 255 and update the "TRILL APPsub-TLV Types under IS-IS TLV 251 Application Identifier 1" registry as follows:

| Type | Name | References |
|---|---|---|

| | | |
|---|---|---|
| TBD1 | TENANT-GWMAC-LABEL | [this document] |
| TBD2 | IPV4-PREFIX | [this document] |
| TBD3 | IPV6-PREFIX | [this document] |

IANA is requested to assign a flag bit in the NickFlags APPsub-TLV as described in Section 7.2 and update the registry created by Section 11.2.3 of [rfc7180bis] as follows:

| Bit | Mnemonic | Description | Reference |
|---|---|---|---|
| 1 | SE | Inter-Subnet Egress | [this document] |

## 10. Normative References

[IS-IS] - ISO/IEC, "Intermediate system to Intermediate system routeing information exchange protocol for use in conjunction with the Protocol for providing the Connectionless-mode Network Service (ISO 8473)", ISO/IEC 10589:2002.

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

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

[RFC7172] - Eastlake, D., M. Zhang, P. Agarwal, R. Perlman, D. Dutt, "TRILL (Transparent Interconnection of Lots of Links): Fine-Grained Labeling", RFC7172, May 2014.

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

[RFC7356] - Ginsberg, L., Previdi, S., and Y. Yang, "IS-IS Flooding Scope Link State PDUs (LSPs)", RFC 7356, September 2014, <http://www.rfc-editor.org/info/rfc7356>.

[RFC7357] - Zhai, H., Hu, F., Perlman, R., Eastlake 3rd, D., and O. Stokes, "Transparent Interconnection of Lots of Links (TRILL): EndStation

Address Distribution Information (ESADI) Protocol", RFC 7357, September 2014, <http://www.rfc-editor.org/info/rfc7357>.

[rfc7180bis] - Eastlake, D., et al, "TRILL: Clarifications, Corrections, and Updates", draft-ietf-trill-rfc7180bis, work in progress.

## 11. Informative References

[RFC826] - Plummer, D., "Ethernet Address Resolution Protocol: Or Converting Network Protocol Addresses to 48.bit Ethernet Address for Transmission on Ethernet Hardware", STD 37, RFC 826, November 1982, <http://www.rfc-editor.org/info/rfc826>.

[RFC4861] - Narten, T., Nordmark, E., Simpson, W., and H. Soliman, "Neighbor Discovery for IP version 6 (IPv6)", RFC 4861, September 2007, <http://www.rfc-editor.org/info/rfc4861>.

[RFC5310] - Bhatia, M., Manral, V., Li, T., Atkinson, R., White, R., and M. Fanto, "IS-IS Generic Cryptographic Authentication", RFC 5310, February 2009.

[RFC7379] - Li, Y., Hao, W., Perlman, R., Hudson, J., and H. Zhai, "Problem Statement and Goals for Active-Active Connection at the Transparent Interconnection of Lots of Links (TRILL) Edge", RFC 7379, October 2014, <http://www.rfc-editor.org/info/rfc7379>.

## Acknowledgments

The authors wish to acknowledge the important contributions of Donald Eastlake, Gayle Noble, Guangrui Wu, Zhenbin Li, Zhibo Hu.

## Authors' Addresses

Weiguo Hao
Huawei Technologies
101 Software Avenue,
Nanjing 210012, China
Phone: +86-25-56623144
Email: haoweiguo@huawei.com

Yizhou Li
Huawei Technologies
101 Software Avenue,
Nanjing 210012, China
Phone: +86-25-56625375

Email: liyizhou@huawei.com

Andrew Qu
MediaTec
Email: laodulaodu@gmail.com

Muhammad Durrani
Cisco
Email: mdurrani@cisco.com

Ponkarthick Sivamurugan
Address: IP Infusion,
RMZ Centennial
Mahadevapura Post
Bangalore - 560048
Email: Ponkarthick.sivamurugan@ipinfusion.com

Liang Xia(Frank)
Huawei Technologies
101 Software Avenue,
Nanjing 210012, China
Phone: +86-25-56624539
Email: frank.xialiang@huawei.com