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: October 23, 2015 April 23, 2015

TRILL Distributed Layer 3 Gateway
draft-ietf-trill-irb-05.txt

Abstract

Currently 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. Huge number of gateway interfaces, millions in the extreme case, need to be supported on the centralized gateway.

3. Traffic bottleneck at the gateway.

An optional TRILL distributed gateway solution that resolves these centralized gateway issues is specified in this document.

Status of this Memo

Copyright Notice

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 traffic for each subnet uses 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 needs to provide routing instances and Layer 3 gateway interfaces for local connected ESs. The routing instances are for 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.

This document is organized as follows: Section 3 describes why a distributed gateway solution is beneficial. Section 4 gives the Layer 3 traffic forwarding model. Section 5 provides a distributed gateway solution overview. Section 6 gives a distributed gateway 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.

GW: Gateway.

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].

RD: Routing Domain.

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. Problem Statement

```
                  -------                 -------
                  | GW1 |                 | GW2 |
                  -------                 -------
                     |                       |
                  -------                 -------
                  |AGG1 |                 |AGG2 |
                  -------                 -------
                     |                       |
       -----------------------------------------------------
       |  -------------|------------------|----------------|
       |  |            |  |               |  |           |  |
     -------          -------          -------        -------
     |TOR1 |          |TOR2 |          |TOR3 |        |TOR4 |
     -------          -------          -------        -------
      |    |           |    |           |    |         |    |
   -----  -----     -----  -----     -----  -----   -----  -----
   |ES1|  |ES2|     |ES3|  |ES4|     |ES5|  |ES6|   |ES7|  |ES8|
   -----  -----     -----  -----     -----  -----   -----  -----
```

Figure 1 A typical DC network

Figure 1 depicts a Data Center Network (DCN) using TRILL 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 TRILL 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 forwarding 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 forwarded on TOR1. If an edge RBridge has distributed gateway capabilities, then it can perform optimum L2 forwarding for intra-subnet traffic and optimum

L3 forwarding for inter-subnet traffic, delivering optimum forwarding for unicast packets in all important cases.

When Fine Grained Labeling [RFC7172] is introduced, up to 16 million Layer 2 VN can be supported in a TRILL campus. To support inter-subnet traffic, up to 16 million Layer 3 gateway interfaces should be created on a centralized gateway if each VN corresponds to a subnet. It is a huge 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 the centralized gateway.

2. Huge number of gateway interfaces, millions in the extreme case, need to be supported on the centralized gateway.

3. Traffic bottleneck at the centralized gateway.

A distributed gateway on edge RBridges addresses these issues.

4. Layer 3 Traffic Forwarding Model

```
+---------------------------------------------+
|                                             |
|      +-----------+         +-----------+    |
|      | Tenant n  |---------|  VRF n    |    |
|   +------------+ |     +------------+  |    |
|   |  +-----+   | |     |            |  |    |
|   |  | VN1 |   | |     |            |  |    |
|   |  +-----+   | |     |    VRF 1   |  |    |
|   |     ..     +-------+            |  |    |
|   |  +-----+   | |     |            |  |    |
|   |  | VNm |   | |     |            |  |    |
|   |  +-----+   | |     |            |  |    |
|   |  Tenant 1  |-+     |            |  |    |
|   +------------+       |            |  |    |
|   +------------+       +------------+       |
|                                             |
|               Edge RBridge                  |
+---------------------------------------------+
```

Figure 2 Edge RBridge model as distributed GW

In a data center network (DCN), each tenant may include one or more Layer 2 virtual networks and, in normal cases, each tenant corresponds to one routing domain (RD). 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 among 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 should be synchronized among these edge RBridges to ensure the reachability to all ESs in that routing domain. The Tenant ID should be 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 GW for the ES. A gateway interface should be established on the edge RBridge for the connecting ES. Because the ESs in the same 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.

If an ES is located in an external IP network, the ES also can be connected to the TRILL campus through a TRILL edge RBridge. The TRILL edge RBridge runs a unified routing protocol with the external IP network for each routing domain. The edge RBridge learns the IP prefix corresponding to the ES through the IP routing protocol, then the RBridge generates 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. 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 there are multiple nicknames attached to 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 SE-flag can be used for this purpose. If the edge RBridge doesn't specify the nickname, the ingress RBridge can use any one of the nicknames as 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, RB1 and RB2 in active-active mode, RB1 and RB2 act as distributed layer 3 gateway for the ES, RB1 and RB2 will 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 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 process to ES2 in VLAN-y. When the local connecting edge RBridge receives inter-subnet traffic from ES1, the RBridge performs Layer 2 header termination, then, using the local VRF corresponding to VLAN-x, it performs the IP forwarding process in that VRF.

If destination ES2 is also attached to the ingress RBridge, the traffic will be locally forwarded to ES2 on the ingress 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 pair-wise 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 forwarding process in that VRF will be performed, and the traffic will be locally forwarded to the destination ES2.

In summary, through this solution, traffic detours to a central gateway are avoided, 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

```
---------             ---------
|  RB3  |             |  RB4  |
---------             ---------
#   *                         *
#   **************************
###########################  *
#   *                     #  *
#   *                     #  *
#   *                     #  *
---------              ---------
|  RB1  |              |  RB2  |
---------              ---------
   |                       |
 -----                   -----
 |ES1|                   |ES2|
 -----                   -----
```

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 forwarding process.

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 table 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 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. Relying on 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 should announce 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 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

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Type                    | (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Length                  | (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                     Tenant ID (4 bytes)         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Resv1|     Label1      | (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+
| Resv2|     Label2      | (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+....-+-+-+-+-+-+-+-+-+-+-+
|            Tenant Gateway Mac (6 bytes)         |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

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 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:

```
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|   Nickname                                    |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
|IN|SE|         RESV                            |
+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+--+
                         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

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Type                        |                   (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Total Length                |                   (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...-+-+-+-+-+-+-+-+-+
|                     Tenant ID                      |(4 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...-+-+-+-+-+-+-+-+-+
| Prefix Length(1)|                                  (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...-+-+-+-+-+-+-+-+-+
|                     Prefix (1)                     |(variable)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...-+-+-+-+-+-+-+-+-+
|     .....       |                                  (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...-+-+-+-+-+-+-+-+-+
|                    .....                           |(variable)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...-+-+-+-+-+-+-+-+-+
| Prefix Length(N)|                                  (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...-+-+-+-+-+-+-+-+-+
|                     Prefix (N)                     |(variable)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...-+-+-+-+-+-+-+-+-+
```

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

```
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Type                        |                  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Total Length                |                  (2 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...-+-+-+-+-+-+-+-+-+
|                     Tenant ID                      |(4 bytes)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...-+-+-+-+-+-+-+-+-+
| Prefix Length(1)|                                  (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...-+-+-+-+-+-+-+-+-+
|                     Prefix (1)                     |(variable)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...-+-+-+-+-+-+-+-+-+
|     .....       |                                  (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...-+-+-+-+-+-+-+-+-+
|                    .....                           |(variable)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...-+-+-+-+-+-+-+-+-+
| Prefix Length(N)|                                  (1 byte)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...-+-+-+-+-+-+-+-+-+
|                     Prefix (N)                     |(variable)
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+...-+-+-+-+-+-+-+-+-+
```

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): End Station 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