IDR Working Group
W. Hao
D. Eastlake
Huawei
Internet Draft
Intended status: Standard Track
S. Hares
Hickory Hill Consulting
S.Gupta
IP Infusion
M. Durrani
Cisco
Y. Li
Huawei
Expires: February 2017
August 16, 2016

Distribution of TRILL Link-State using BGP
draft-ietf-idr-ls-trill-02.txt

Abstract

This draft describes a TRILL link state and MAC address reachability information distribution mechanism using a BGP LS extension. External components such as an SDN Controller can use the information for topology visibility, troubleshooting, network automation, etc.

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

BGP has been extended to distribute IGP link-state and traffic engineering information to some external components [I-D.ietf-idr-ls-distribution], such as the PCE and ALTO servers. The information can be used by these external components to compute a MPLS-TE path

across IGP areas, visualize and abstract network topology, and the like.

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).

The TRILL protocol has been deployed in many data center networks. Data center automation is a vital step to increase the speed and agility of business. An SDN controller as an external component normally can be used to provide centralized control and automation for the data center network. Making a holistic view of whole network topology available to the SDN controller is an important part for data center network automation and troubleshooting.

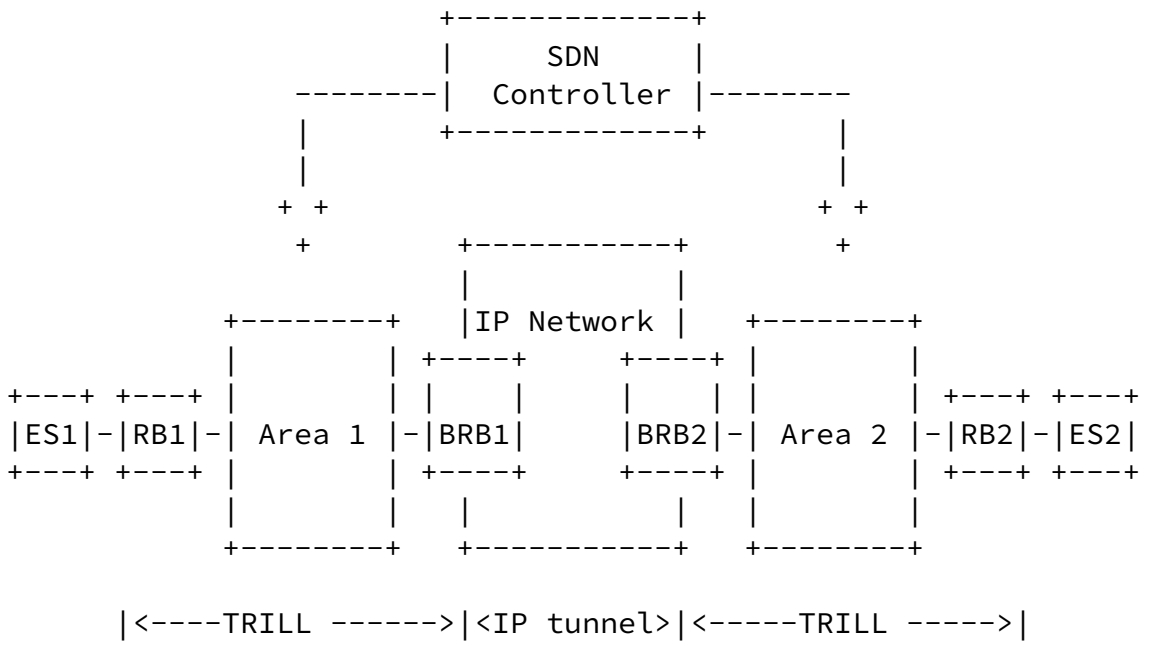


Figure 1: TRILL interconnection

In Data Center interconnection scenario illustrated in figure 1, a single SDN Controller or network management system (NMS) can be used for end-to-end network management. End-to-end topology visibility on

the SDN controller or NMS is very useful for whole network automation and troubleshooting. BGP LS can be used by the external SDN controller to collect multiple TRILL domain's link-state.

BGP LS also can be used for MAC address reachability information synchronization across multiple TRILL domains. The transported MAC reachability information and the like is for telemetry purposes and for use by SDN controller(s) where the coordination or protocol between the SDN controllers is out of scope.

This document describes the detailed BGP LS extension mechanisms for TRILL link state and MAC address reachability information distribution.

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 [RFC2119].

BGP - Border Gateway Protocol

BGP-LS - BGP Link-State

Data label - VLAN or FGL (Fine Grained Label [RFC7172])

IS - Intermediate System (for this document, all relevant intermediate systems are RBridges)

NLRI - Network Layer Reachability Information

SDN - Software Defined Networking

RBridge - A device implementing the TRILL protocol

TRILL - Transparent Interconnection of Lots of Links

3. Carrying TRILL Link-State Information in BGP

In [I-D.ietf-idr-ls-distribution], four NLRI types are defined as follows: Node NLRI, Link NLRI, IPv4 Topology Prefix NLRI and IPv6

Topology Prefix NLRI. For TRILL link-state distribution, the Node NLRI and Link NLRI are extended to carry layer 3 gateway role and link MTU information. TRILL specific attributes are carried using opaque Node Attribute TLVs, such as nickname, distribution tree number and identifiers, interested VLANs/Fine Grained Label, and multicast group address, etc.

To differentiate TRILL protocol from layer 3 IGP protocol, a new TRILL Protocol-ID is defined.

Protocol-ID	NLRI information source protocol
1	IS-IS Level 1
2	IS-IS Level 2
3	OSPFv2
4	Direct
5	Static configuration
6	OSPFv3
TBD	TRILL

Table 1: Protocol Identifiers

ESADI (End Station Address Distribution Information) protocol [RFC7357] is a per data label control plane MAC learning solution. MAC address reachability information is carried in ESADI packets. Compared with data plane MAC learning solution, ESADI protocol has security and fast update advantage that are pointed out in [RFC7357].

For an RBridge that is announcing participation in ESADI, the RBridge can distribute MAC address reachability information to external components using BGP. A new NLRI type of ''MAC Reachability NLRI'' is requested for the MAC address reachability distribution.

Type	NLRI Type
1	Node NLRI

2	Link NLRI
3	IPv4 Topology Prefix NLRI
4	IPv6 Topology Prefix NLRI
TBD	MAC Reachability NLRI

Table 2: NLRI Types

The MAC Reachability NLRI uses the format as shown in the following figure.

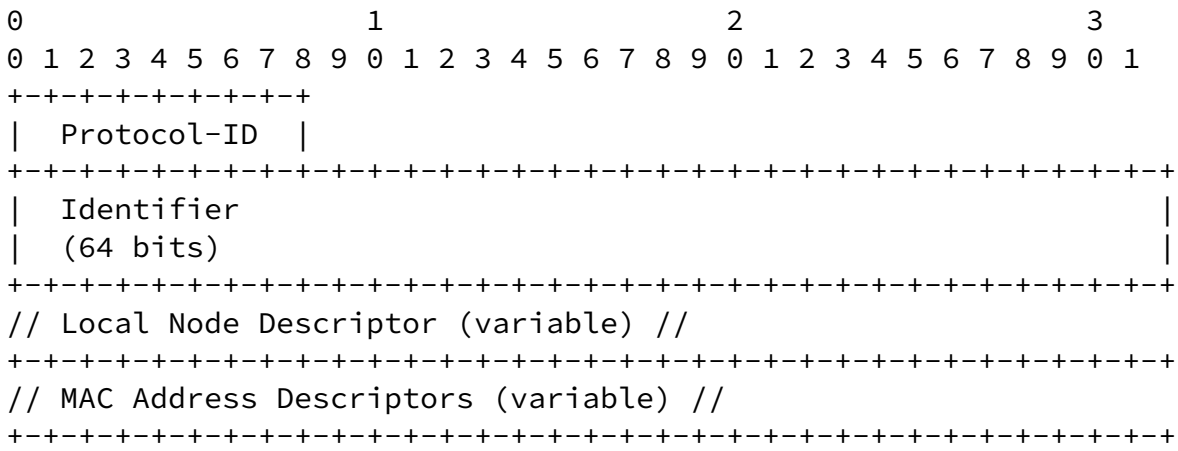
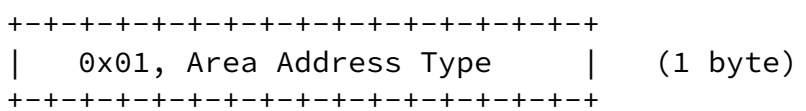


Figure 2: The MAC Reachability NLRI format

3.1. Node Descriptors

The Node Descriptor Sub-TLV types include Autonomous System and BGP-LS Identifier, IS-IS Area-ID and IGP Router-ID. TRILL uses a fixed zero Area Address as specified in [RFC6325], Section 4.2.3. This is encoded in a 4-byte Area Address TLV (TLV #1) as follows:



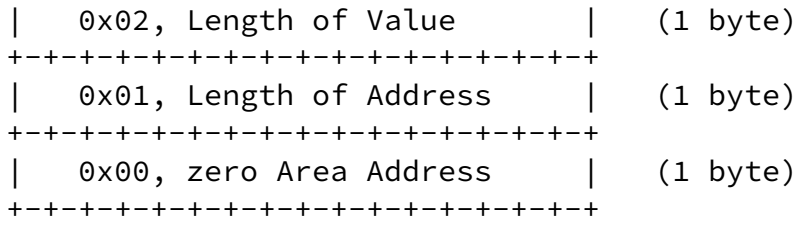


Figure 3: Area Address TLV

3.1.1. IGP Router-ID

Similar to layer 3 IS-IS, TRILL protocol uses 7-octet "IS-IS ID" as the identity of an RBridge or a pseudonode, IGP Router ID sub-TLV in Node Descriptor TLVs contains the 7-octet "IS-IS ID". In TRILL network, each RBridge has a unique 48-bit (6-octet) IS-IS System ID. This ID may be derived from any of the RBridge's unique MAC addresses or configured. A pseudonode is assigned a 7-octet ID by the DRB (Designated RBridge) that created it, the DRB is similar to the "Designated Intermediate System" (DIS) corresponding to a LAN.

3.2. MAC Address Descriptors

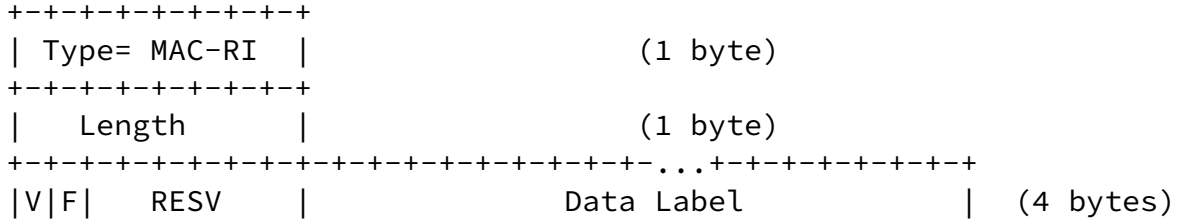
The ''MAC Address Descriptor'' field is a set of Type/Length/Value (TLV) triplets. ''MAC Address Descriptor'' TLVs uniquely identify an

MAC address reachable by a Node. The following attributes TLVs are defined:

TLV Code Point	Description	Length	Value defined in:
1	MAC-Reachability	variable	section 3.2.1

Table 3: MAC Address Descriptor TLVs

3.2.1. MAC-Reachability TLV



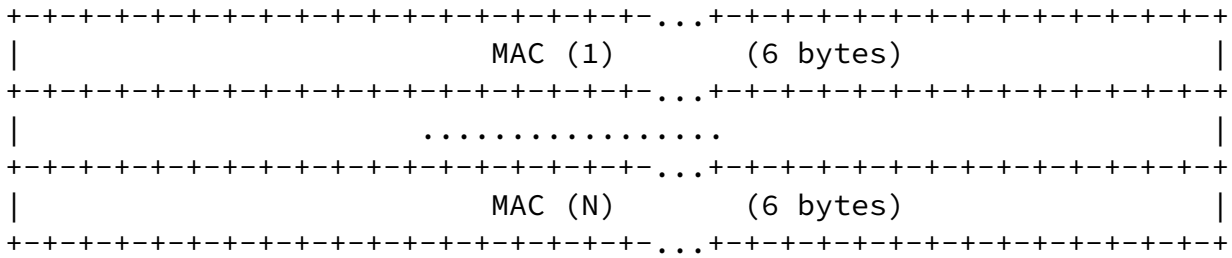


Figure 4: MAC-Reachability TLV format

Length is 4 plus a multiple of 6.

The bits of 'V' and 'F' are used to identify Data Label type and are defined as follows:

Bit	Description
'V'	VLAN
'F'	Fine Grained Label

Table 4: Data Label Type Bits Definitions

Notes: If BGP LS is used for NVO3 network MAC address distribution between external SDN Controller and NVE, Data Label can be used to represent 24 bits VN ID.

3.3. The BGP-LS Attribute

3.3.1. Node Attribute TLVs

3.3.1.1. Node Flag Bits TLV

A new Node Flag bit is added as follows:

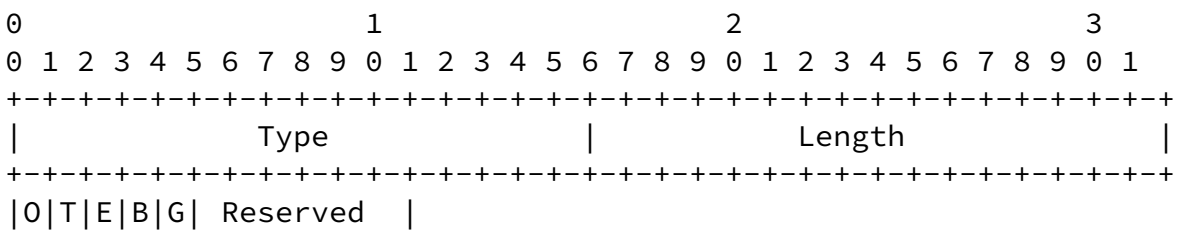


Figure 5: Node Flag Bits TLV format

The new bit and remaining reserved bits are defined as follows:

Bit	Description	Reference
'G'	Layer 3 Gateway Bit	[RFC7176]
Reserved	Reserved for future use	

Table 5: Node Flag Bits Definitions

3.3.1.2. Opaque Node Attribute TLV

The Opaque Node Attribute TLV is used as the envelope to transparently carry TRILL specific information. In [RFC7176], there are the following Sub-TLVs in the Router Capability and MT-Capability TLVs and the Group Address (GADDR) TLV that need to be carried. Future possible TRILL TLVs/Sub-TLVs extension also can be carried using the Opaque Node Attribute TLV.

Descriptions	IS-IS TLV/Sub-TLV
TRILL-VER	22/13
NICKNAME	22/6
TREES	22/7
TREE-RT-IDs	22/8
TREE-USE-IDs	22/9
INT-VLAN	22/10
VLAN-GROUP	22/14
INT-LABEL	22/15

RBCHANNELS	22/16
AFFINITY	22/17
LABEL-GROUP	22/18
GMAC-ADDR	142/1
GIP-ADDR	142/2
GIPV6-ADDR	142/3
GLMAC-ADDR	142/4
GLIP-ADDR	142/5
GLIPV6-ADDR	142/6

Table 6: TRILL TLVs/Sub-TLVs

3.3.2. Link Attribute TLVs

Link attribute TLVs are TLVs that may be encoded in the BGP-LS attribute with a link NLRI. Besides the TLVs that has been defined in [I-D.ietf-idr-ls-distribution] section 3.3.2 table 9, the following 'Link Attribute' TLV is provided for TRILL.

TLV Code Point	Description	IS-IS TLV /Sub-TLV	Defined in:
TBD	Link MTU	22/28	[RFC7176]/2.4

Table 7: Link Attribute TLVs

4. Operational Considerations

This document does not require any MIB or Yang model to configure operational parameters.

Any implementation of the protocol in this specification (i.e. that distributes TRILL Link-State information using BGP), MUST do the

malformed attribute checks below, and if it detects a malformed attribute, it should use the 'Attribute Discard' action per [I-D.ietf.idr-error-handling] section 2.

An implementation MUST perform the following expanded [BGP-LS]

syntactic check for determining if the message is malformed:

- Does the sum of all TLVs found in the BGP LS attribute correspond to the BGP LS path attribute length ?

- Does the sum of all TLVs found in the BGP MP_REACH_NLRI attribute correspond to the BGP MP_REACH_NLRI length ?

- Does the sum of all TLVs found in the BGP MP_UNREACH_NLRI attribute correspond to the BGP MP_UNREACH_NLRI length ?

- Does the sum of all TLVs found in a Node-, Link, prefix (IPv4 or IPv6) NLRI attribute correspond to the Node-, Link- or Prefix Descriptors 'Total NLRI Length' field ?

- Does any fixed length TLV correspond to the TLV Length field in this document ?

- Does the sum of MAC reachability TLVs equal the length of the field?

In addition, the following checks need to be made for the fields specific to the BGP LS for TRILL:

PROTOCOL ID is TRILL

NLRI types are valid per table 2

MAC Reachability NLRI has correct format including:

- Identifier (64 bits),

- local node descriptor with AREA address TLV has the form found in figure 2

opaque TLV support the range of ISIS-TLV/SUB-TLV shown in table 3, and link TLVs support the range in figure 8.

5. Security Considerations

Procedures and protocol extensions defined in this document do not affect the BGP security model. See [RFC6952] for details.

6. IANA Considerations

For all of the following assignments, [this document] is the reference.

IANA is requested to assign one Protocol-ID for "TRILL" from the BGP-LS registry of Protocol-IDs.

IANA is requested to assign one NLRI Type for "MAC Reachability" from the BGP-LS registry of NLRI Types.

IANA is requested to assign one Node Flag bit for "Layer 3 Gateway" from the BGP-LS registry of BGP-LS Attribute TLVs.

IANA is requested to assign one new TLV type for "Link MTU" from the BGP-LS registry of BGP-LS Attribute TLVs.

7. References

7.1. Normative References

[1] [I-D.ietf-idr-ls-distribution] Gredler, H., Medved, J., Previdi, S., Farrel, A., and S.Ray, "North-Bound Distribution of Link-State and TE Information using BGP", draft-ietf-idr-ls-distribution-10(work in progress), January 2015.

[2] [I-D.ietf.idr-error-handling] Enke, C., John, S., Pradosh, M., Keyur,P., "Revised Error Handling for BGP UPDATE Messages", draft-ietf-idr-error-handling-19(work in progress), April 2015.

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

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

[5] [RFC7172] Eastlake 3rd, D., Zhang, M., Agarwal, P., Perlman, R., and D. Dutt, "Transparent Interconnection of Lots of Links (TRILL): Fine-Grained Labeling", RFC 7172, DOI 10.17487/RFC7172, May 2014, <http://www.rfc-editor.org/info/rfc7172>.

[6] [RFC7176] Eastlake, D., Senevirathne, T., Ghanwani, A., Dutt, D., Banerjee, A.," Transparent Interconnection of Lots of Links (TRILL) Use of IS-IS'', May 2014.

[7] [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>.

7.2. Informative References

8. Acknowledgments

Authors like to thank Ross Callon, Andrew Qu, Jie Dong, Mingui Zhang, Qin Wu, Shunwan Zhuang, Zitao Wang, Lili Wang for their valuable inputs.

Authors' Addresses

Weiguo Hao
Huawei Technologies
101 Software Avenue,
Nanjing 210012
China

Phone: +86-25-56623144
Email: haoweiguo@huawei.com

Donald E. Eastlake
Huawei Technologies
155 Beaver Street
Milford, MA 01757 USA

Phone: +1-508-333-2270
Email: d3e3e3@gmail.com

Susan K. Hares
Hickory Hill Consulting
7453 Hickory Hill
Saline, MI 48176 USA

Email: shares@ndzh.com

Sujay Gupta
IP Infusion
Email: sujay.gupta@ipinfusion.com

Muhammad Durrani
Cisco
Phone: +1-408-527-6921
Email: mdurrani@cisco.com

Yizhou Li
Huawei Technologies
101 Software Avenue,
Nanjing 210012, China

Phone: +86-25-56625375
Email: liyizhou@huawei.com