INTERNET-DRAFT
Intended status: Proposed Standard
Updates: 7752

D. Eastlake
W. Hao
Y. Li
Huawei
S. Gupta
IP Infusion
M. Durrani
Equinix

Expires: April 12, 2018
October 13, 2018

# Distribution of TRILL Link-State using BGP
<draft-ietf-idr-ls-trill-05.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, and the like. This document updates RFC 7752.

## Status of This Memo

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

Distribution of this document is unlimited. Comments should be sent to the authors or the IDR working group mailing list: idr@ietf.org.

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 [RFC7752], 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] [RFC7780] 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. Providing a holistic view of whole network topology to the SDN controller is an important part of data center network automation and troubleshooting.

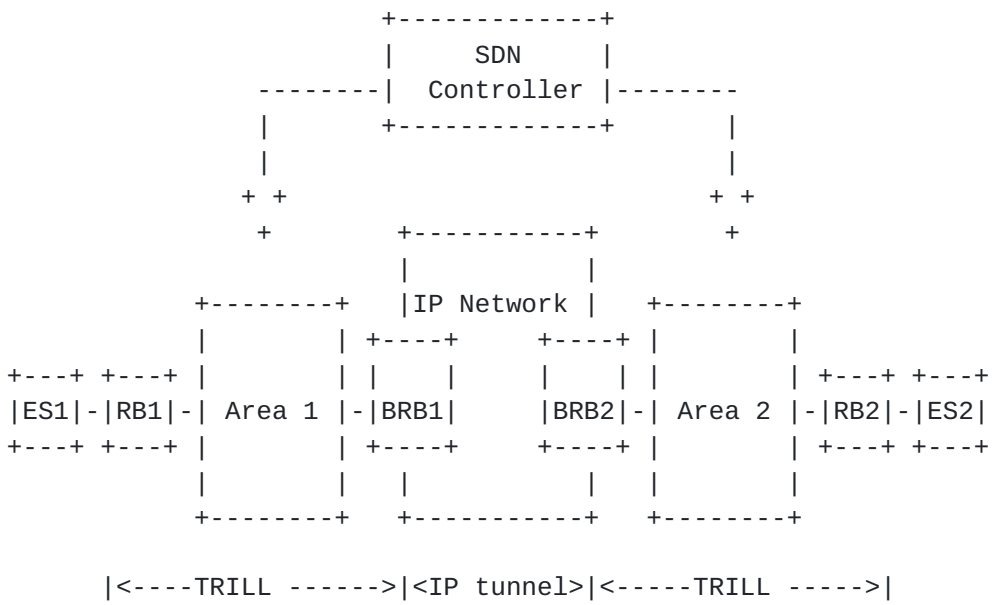


Figure 1: TRILL interconnection

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. This document updated [RFC7752] by creating a new IANA registry for BGP-LS Node Descriptor Flag Bits.

## 2. Conventions used in this document

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

BGP - Border Gateway Protocol

BGP-LS - BGP Link-State

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

IGP - Interior Gateway Protocol

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

LS - Link State

NLRI - Network Layer Reachability Information

SDN - Software Defined Networking

RBridge - A device implementing the TRILL protocol

TRILL - Transparent Interconnection of Lots of Links [RFC6325] [RFC7176] [RFC7780]

## 3. Carrying TRILL Link-State Information in BGP

In [RFC7752], several BGP-LS NLRI types are defined. 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. Examples of such attributes are nickname, distribution tree number and identifiers, interested VLANs/Fine Grained Label, and multicast group address, etc.

To differentiate the TRILL protocol from layer 3 IGP protocols, a new TRILL Protocol-ID = TBD1 is specified.

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 "MAC Reachability NLRI" NLRI type TBD2 is used for the MAC address reachability distribution.

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

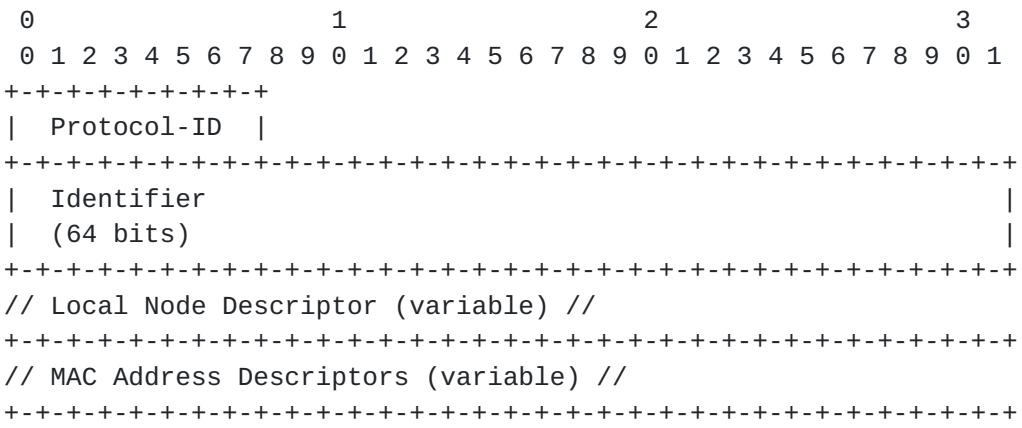


Figure 2: The MAC Reachability NLRI format

### 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:

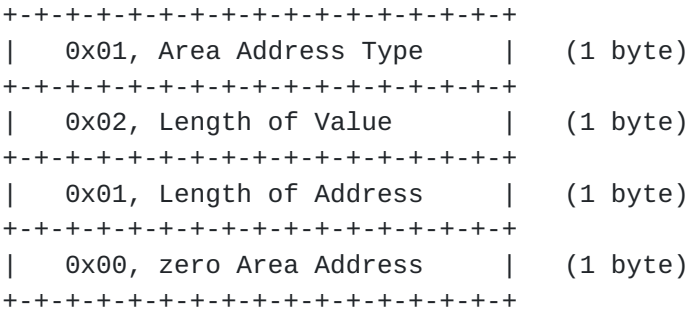


Figure 3: Area Address TLV

### 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 1: MAC Address Descriptor TLVs

### 3.2.1 MAC-Reachability TLV

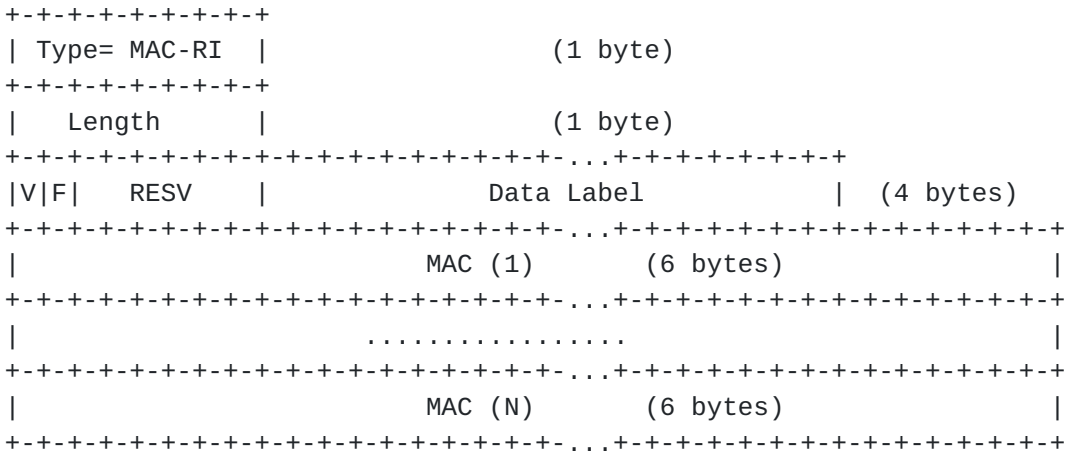


Figure 4: MAC-Reachability TLV format

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 2: Data Label Type Bits Definitions

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

## 3.3 The BGP-LS Attributes

### 3.3.1 Node Attribute TLVs

#### 3.3.1.1 Node Flag Bits TLV

A new Node Flag bit TBD4 is added to the Node Flag Bits TLV. This flag indicates that the node is a distributed Layer 3 gateway [RFC7956].

#### 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 most cases, this information is encoded as sub-TLVs within the IS-IS Router Capability and MT-Capability TLVs or as the Group Address (GADDR) TLV. Many of these are specified in [RFC7176] but additional sub-TLVs have been specified and may be specified in the future that also can be carried using the Opaque Node Attribute TLV.

### 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 [RFC7752] 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: |
|---|---|---|---|
| TBD3 | Link MTU | 22/28 | [RFC7176] Section 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 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 every 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

MAC Reachability NLRI has correct format including:

- Identifier (64 bits),
- local node descriptor with AREA address TLV has the form found in Figure 2

## 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 TBD1 for "TRILL" from the BGP-LS registry of Protocol-IDs.

IANA is requested to assign one NLRI Type TBD2 for "MAC Reachability" from the BGP-LS registry of NLRI Types.

IANA is requested to assign one new TLV type TBD3 for "Link MTU" from the BGP-LS registry of BGP-LS Attribute TLVs.

IANA is requested to create a registry for BGP-LS Node Descriptor Flag Bits and to assign one Node Flag bit TBF4 [bit 6 suggested] for "Layer 3 Gateway". This new registry is to be added to the IANA Border Gateway Protocol - Link State (BGP-LS) Parameters web page as follows:

Name: BGP-LS Node Descriptor Flag Bits
Registration Procedure: Expert Review
Reference: [RFC7752]
Note: These bits are in the payload of the Node Flag Bits TLV. The bit array is variable length so the maximum bit value assignable is very large; however, length of the TLV grows linearly with the highest numbered flag bit used and there may be practical limits on the length of the TLV.

| Bit | Letter | Description | Reference |
|---|---|---|---|
| 0 | O | Overload Bit | [ISO10589] |
| 1 | T | Attached Bit | [ISO10589] |
| 2 | E | External Bit | [RFC2328] |
| 3 | B | ABR Bit | [RFC2328] |
| 4 | R | Router Bit | [RFC5340] |
| 5 | V | V6 Bit | [RFC5340] |
| 6 | G | L3 Gateway Bit | [this document][RFC7956] |
| 7-up | - | Unassigned | |

Normative References

[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.

[IS-IS] - International Organization for Standardization, "Information technology -- Telecommunications and information exchange between systems -- Intermediate System to Intermediate System intra-domain routeing information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode network service (ISO 8473)", ISO/IEC 10589:2002, Second Edition, November 2002.

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

[RFC6325] - Perlman, R., Eastlake 3rd, D., Dutt, D., Gai, S., and A. Ghanwani, "Routing Bridges (RBridges): Base Protocol Specification", RFC 6325, DOI 10.17487/RFC6325, July 2011, <https://www.rfc-editor.org/info/rfc6325>.

[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>.

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

[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>.

[RFC7752] - Gredler, H., Ed., Medved, J., Previdi, S., Farrel, A., and S. Ray, "North-Bound Distribution of Link-State and Traffic Engineering (TE) Information Using BGP", RFC 7752, DOI 10.17487/RFC7752, March 2016, <https://www.rfc-editor.org/info/rfc7752>.

[RFC7780] - Eastlake 3rd, D., Zhang, M., Perlman, R., Banerjee, A., Ghanwani, A., and S. Gupta, "Transparent Interconnection of Lots of Links (TRILL): Clarifications, Corrections, and Updates", RFC 7780, DOI 10.17487/RFC7780, February 2016, <https://www.rfc-editor.org/info/rfc7780>.

[RFC7956] - Hao, W., Li, Y., Qu, A., Durrani, M., and P. Sivamurugan, "Transparent Interconnection of Lots of Links (TRILL) Distributed Layer 3 Gateway", RFC 7956, DOI 10.17487/RFC7956, September 2016, <https://www.rfc-editor.org/info/rfc7956>.

[RFC8174] - Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

## Informative References

[ISO10589] - SO, "Intermediate System to Intermediate System intra-domain routeing information exchange protocol for use in conjunction with the protocol for providing the connectionless-mode network service (ISO 8473)", International Standard 10589:2002, Second Edition, 2002.

[RFC2328] - Moy, J., "OSPF Version 2", STD 54, RFC 2328, DOI 10.17487/RFC2328, April 1998, <https://www.rfc-editor.org/info/rfc2328>.

[RFC5340] - Coltun, R., Ferguson, D., Moy, J., and A. Lindem, "OSPF for IPv6", RFC 5340, DOI 10.17487/RFC5340, July 2008, <https://www.rfc-editor.org/info/rfc5340>.

[RFC6952] - Jethanandani, M., Patel, K., and L. Zheng, "Analysis of BGP, LDP, PCEP, and MSDP Issues According to the Keying and Authentication for Routing Protocols (KARP) Design Guide", RFC 6952, DOI 10.17487/RFC6952, May 2013, <https://www.rfc-editor.org/info/rfc6952>

## Acknowledgments

Authors thank Susan Hares, John Scudder, 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
1424 Pro Shop Court
Davenport, FL 33896 USA

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

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

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

Sujay Gupta
IP Infusion

Email: sujay.gupta@ipinfusion.com

Muhammad Durrani
Equinix

Email: mdurrani@equinix.com

Copyright, Disclaimer, and Additional IPR Provisions