Transparent Interconnection of Lots of Links Working Group
Internet-Draft
Updates: 6325 (if approved)
Intended status: Standards Track
Expires: April 25, 2013

Q. Wu
W. Hao
Huawei

October 22, 2012

LSP extension for Tree Distribution Optimization across sites
draft-wu-trill-lsp-ext-tree-distr-opt-01

Abstract

This document specifies an extension to LSP for the Rbridge in one site to advertise Global VLAN scope and associated link attribute to all the Rbridges both in the site of that Border Rbridge and the other adjacent sites in the same campus. With this extension, RBridges can prune the distribution tree of multi-destination frames according to the scope of the VLAN and link attribute defined in this document.

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). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at http://datatracker.ietf.org/drafts/current/.

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."

This Internet-Draft will expire on April 25, 2013.

Table of Contents

1. Introduction

Large datacenters are often multi-site in nature and may contain a large number of Rbridges in each site. A trill Campus network may also be designed to be multilevel can be divided in to multiple IS-IS [IS-IS][RFC1195]L1 Areas interconnected by L2 backbone area. Routing between Rbridges within a IS-IS L1 area/ site is known as "Level 1 routing". Routing between IS-IS L1 areas or sites is known as "Level 2 routing". The IS-IS L1 area supports Level 1 routing and consists of Rbridges within the site and link between Rbridges within the site. The L2 backbone area supports Level 2 routing and consists of Border Rbridges and links between the Border Rbridges. Border Rbridges may participate in one or more L1 areas as Level-1 Rbridges inside each site, in addition to their role as Level 2 Rbridge across sites.

In Trill campus network, RBridges use distribution trees to forward multi-destination frames. In case of one Trill campus network having multiple sites, the traffic associated with some distributed trees may travel between sites while the traffic associated with other distributed trees may be limited to only one site and not allowed to go across other sites. The traffic spanning across sites is also referred to as the traffic with global scope. In order to support scaling and performance of large TRILL networks in the real deployments, it is desirable to forward most of Multi-destination Trill traffic within the site and reduce the traffic that is required to span across sites within the entire TRILL campus. According to The TRILL base protocol, each distribution tree SHOULD be pruned per VLAN. When it is inevitable to construct trees that have a scope across sites throughout the TRILL campus, it is necessary to treat traffic tagged with VLAN differently based on VLAN scope and distinct the link between Rbridges in one site and link between two Border Rbridge in two sites to support large scale multi-tenants application.

This document specifies an extension to LSP for the Rbridge in one site to advertise Global VLAN scope and associated link attribute to all the Rbridges both in the site of that Border Rbridge and the other adjacent sites in the same campus. With this extension, RBridges can prune the distribution tree of multi-destination frames according to the scope of the VLAN and link attribute defined in this document.

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 [RFC2119].

3. Motivations

Distinguishing global vlan from local vlan is to increase the number of tenants by not breaking the VLAN tag size limits. E.g. one campus being divided into n sites, without distinction between global vlan and local vlan, at most support 4K tenants. However, if we distinguish global vlan from local vlan, suppose each site support only local vlan. Then each site support 4K tenants, the total number of tenants supported by one campus can be increased to 4n*K.Suppose some sites support local vlan, some sites support both local vlan and global vlan, the total number of tenants supported by one campus (4K,4n*K).

4. TLV and Sub-TLV Extensions to IS-IS for Inter-site Distribution Tree

This section describes data formats and code points for the TLVs and sub-TLVs added to IS-IS defined by this specification to support the multi-level TRILL or re-used from that already contained in the standard IS-IS extensions defined in [RFC6326].

4.1. Global-VLANs Sub-TLV for the Router Capability TLV

The optional Global-VLANs sub-TLV specifies the VLANs that have global scope and enable Construction of global multi-destination trees among different sites. It has the following format:

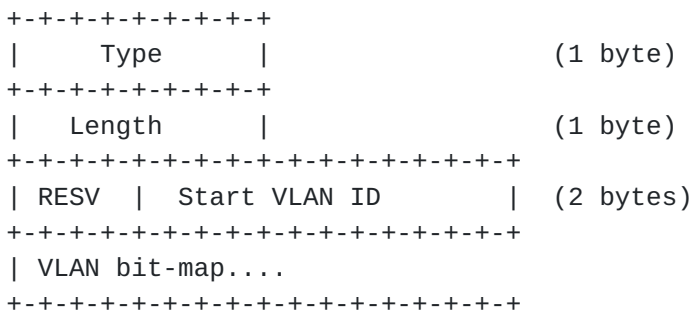

Figure 1: Report Block Structure

4.1.1. Definition of Fields in Sub-TLV

Type: 8bits

Router Capability sub-TLV type, set to TBD (GLOBAL-VLANs).

Length: 8bits

Variable, minimum 3.

RESV: 4bits

4 reserved bits that MUST be sent as zero and ignored on receipt.

Start VLAN ID:12bits

The 12-bit VLAN ID that is represented by the high order bit of the first byte of the VLAN bit-map.

VLAN bit-map:

The highest order bit indicates the VLAN equal to the start VLAN ID, the next highest bit indicates the VLAN equal to start VLAN ID + 1, continuing to the end of the VLAN bit-map field.

4.2. Link-Attributes Sub-TLV extension for extended IS reachability TLV

The link-attribute sub-TLV is carried within the TLV 22 and has a format identical to the sub-TLV format used by the Traffic Engineering Extensions for IS-IS ([RFC3784]): 1 octet of sub-type, 1 octet of length of the value field of the sub-TLV followed by the value field -- in this case, a 16 bit flags field.

The Link-attribute sub-type is 19 and the link-attribute has a length of 2 octets.

This sub-TLV is OPTIONAL and MUST appear at most once for a single IS neighbor. If a received Link State Packet (LSP) contains more than one Link-Attribute Sub-TLV, an implementation SHOULD decide to consider only the first encountered instance. The following bit is defined:

Public Link Type For TRILL(0x03) When set, this indicates that the link is public link for TRILL sites interconnection.

5. Use of TLV and Sub-TLV for Tree Distribution Optimization across sites

When the TRILL campus is divided into multiple sites, each site may have one or more Border Rbridges used to interconnect other remaining sites and form the Level 2 IS-IS Trill network. Such Level2 IS-IS Trill network can be used to construct global multi-destination tree spanning across various sites.

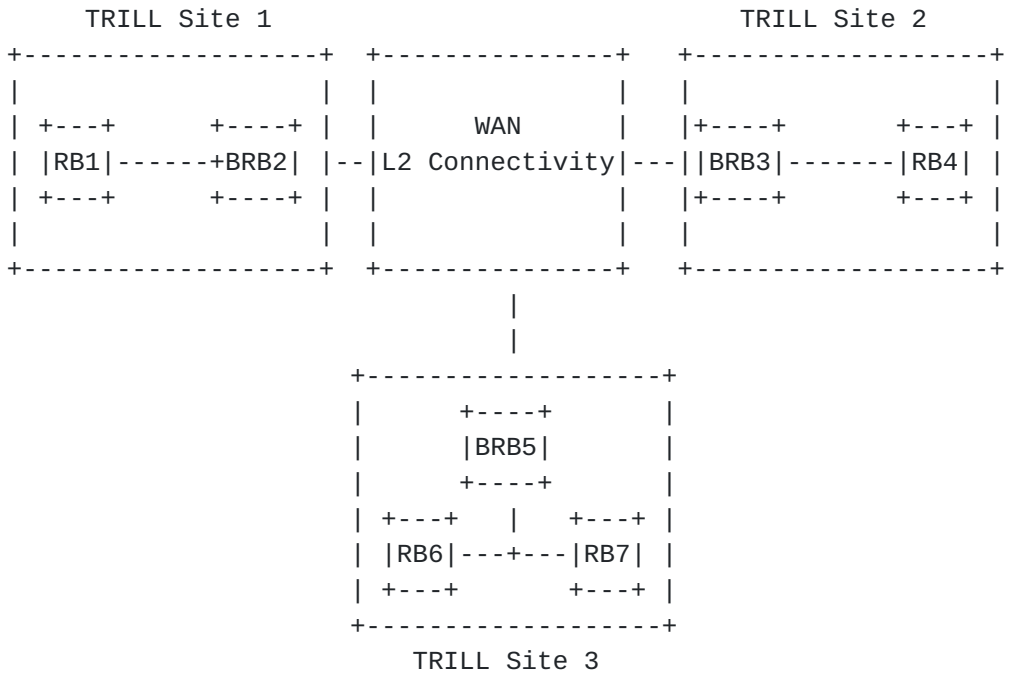

Figure 1: Example of multiple sites within one Trill Campus

In order to support scaling and performance of large TRILL networks in the real deployments, firstly, not all the links between the level 2 Rbridges need to be used to Construct global multi- destination trees. If the link between the level 2 Rbridges is allowed to construct global multi-destination trees, we can set this link attribute into "public interface for global tree construction". In this document, we reuse Link Attribute sub-TLV for the extended IS reachability TLV and allocate a new bit value inside link Attribute Sub-TLV to support indication of "public link for global tree Construction". The Border Rbridge in one site need to advertise this link attribute Sub-TLV to all the neighboring Border Rbridges in other neighboring sites and then this sub-TLV will be further forwarded to all the Rbridges in the site of each neighboring Border Rbridge. RBridges in each site can prune the distribution tree of multi-destination frames according to such link attribute.

Secondly, not all traffic should have global scope and need to span

across sites. Since each distribution tree SHOULD be pruned per VLAN according to [RFC6325], we can specify a set of Global VLANs to identify the traffic that has global scope. In this document, we define one new sub-TLV for the Router Capability TLV, i.e., Global-VLANs Sub- TLV. This Sub-TLV can be used by Rbriges in one site to determine whether Construction of global multi-destination trees across sites is allowed. In order to achieve this, the tree root or highest priority RBridge in one site configured to know a number of appropriate VLANs as Global VLANs and announce such information to the nearest border Rbridge; Then such Border Rbridge in this site need to advertise Global VLAN Sub-TLV to all the neighboring Border Rbridges in other neighboring sites and then this sub-TLV will be further forwarded to all the Rbridges in the site of each neighboring Border Rbridge. When Global VLAN and link attribute Sub-TLV described above has been distributed to all the corresponding Rbridges in the downstream of the tree root or highest priority RBridge, RBridges can prune the distribution tree of multi-destination frames according to the scope of the VLAN and link attribute defined in this document, eliminating branches that own link type mismatching with Distribution Tree scope identified by VLAN. If the distribution tree is local tree and has branches including a link with link attribute is set to public link for global tree construction, those branches should be eliminated. If the distribution tree is global tree and has branches containing a link with link attribute not set to public link for global tree construction, those branches also should be eliminated.

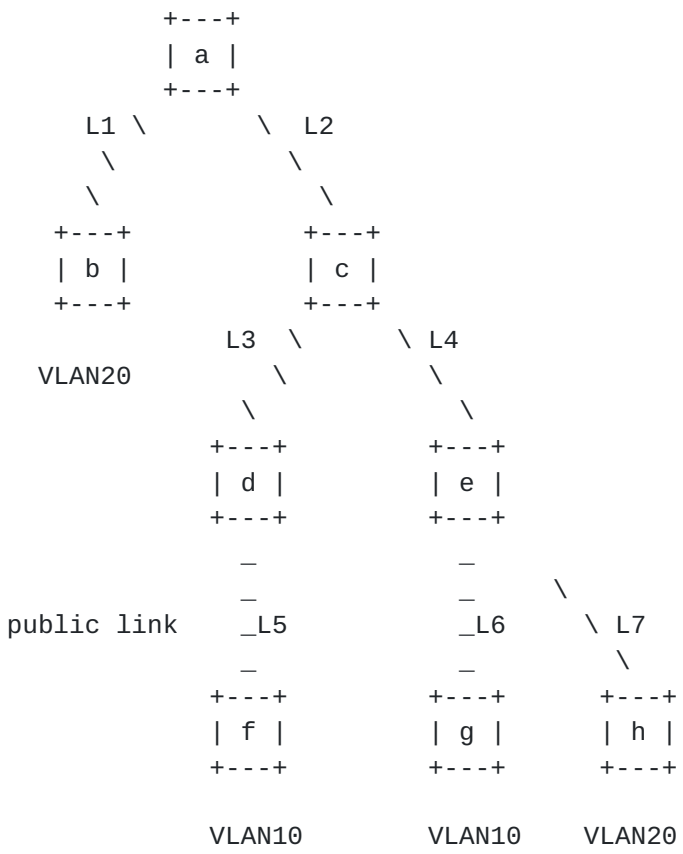

Figure 2: Distribution Tree

Take distribution tree in Figure 2 as an example, Rbridge a is root node. Rbridge f,g are leaf nodes that have end station on VLAN 10 while Rbridge b,h are another two leaf nodes and that have end station on VLAN 20. The link between Rbridge d and f is public link used across sites while the other links in the figure 2 are links owned by one single site. Assume VLAN 10 are local VLAN and VLAN 20 are Global VLAN, after distribution tree pruning is done, Rbrige c should eliminate branch that has Rridge d and f since distribution tree is pruned based on local VLAN 10 and Link 5 in that branch is public link, which mismatch with each other.

6. Unicast Forwarding Consideration

In unicast forwarding, the MAC forwarding table for a Trill Border Rbridge is usually learned through the data plane, i.e.,MAC address is learnt from received Broadcast,Unknown, Unicast,Multicast packet through distribution tree. For end stations on the local vlan, the broadcast scope is limited to one local site, the Border Rbridge only learns MAC address of locally attached end station and the forwarding path between end stations within one site can be built for unicast. For end stations on global VLAN, end stations between two sites are within the same layer 2 broadcast domain, the Border Rbridge can learn MAC address of end stations across sites and the forward path between two sites can be built as well for unicast. Therefore unicast forwarding between sites can be controlled through LSP extension we defined in this document.

If the Border Rbridge is statically configured with unicast forwarding table and the nickname of the destination Rbridge is specified as one Rbridge's nickname in other sites, the unicast packet must be forced to forward to the other sites. In this case, the Border Rbridge in other sites performs security check to the received packet. If the VLAN associated with the received packet is local VLAN and the packet is ingressed from public link across site, the packet should be discarded. If the VLAN associated with the received packet is Global VLAN, the packet should be allowed to ingress from public link across sites.

7. IANA Considerations

IANA is requested to assign a new codepoint for the Global-VLANs Sub-TLV defined in this document and carried within TLV 242.

IANA has created a registry for bit values inside the link-attributes sub-TLV called "link-attribute bit values for sub-TLV 19 of TLV 22".

This document instructs IANA to add a new bit value in the link-attribute bit values for sub-TLV 19 of TLV 22 registry as follows:

Value	Name	Reference
0x3	Public Link Type between sites	[This document]

Further values are to be allocated by the Standards Action process defined in [RFC2434], with Early Allocation (defined in [RFC4020]) permitted.

8. Security Considerations

The security considerations documented in [RFC4971][RFC5305] are applicable for the Sub-TLV extension defined in this document.

9. References

9.1. Normative References

[IS-IS] "Intermediate System to Intermediate System Intra-Domain Routing Exchange Protocol for use in Conjunction with the Protocol for Providing the Connectionless-mode Network Service (ISO 8473)", ISO/IEC 10589:2002 Second Edition, 2002.

[RFC1195] Ohta, M., "Use of OSI IS-IS for routing in TCP/IP and dual environments", December 1990.

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

[RFC2434] Narten, T. and H. Alvestrand, "Guidelines for Writing an IANA Considerations Section in RFCs", RFC 2434, October 1998.

[RFC3784] Smit, H., "Intermediate System to Intermediate System (IS-IS) Extensions for Traffic Engineering (TE)", June 2004.

[RFC4020] Kompella, K. and A. Zinin, "Early IANA Allocation of Standards Track Code Points", RFC 4020, February 2005.

[RFC4971] Vasseur, J., "Intermediate System to Intermediate System (IS-IS) Extensions for Advertising Router Information", July 2007.

[RFC5029] Vasseur, J. and S. Previdi, "SDP: Session Description Protocol", September 2007.

[RFC5305] Li, T. and H. Smit, "S-IS Extensions for Traffic Engineering", RFC 5305, October 2008.

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

[RFC6326] Eastlake , D., Banerjee, A., Dutt, D., Perlman, R., and A. Ghanwani, "Transparent Interconnection of Lots of Links (TRILL) Use of IS-IS", RFC 6326, July 2011.

9.2. Informative References

[TRILL-ML]
Perlman , R., Eastlake, D., Ghanwani, A., and H. Zhai,
"RBridges: Multilevel TRILL",
ID draft-perlman-trill-rbridge-multilevel-03,
October 2011.

Appendix A. Change Logs

A.1. draft-wu-trill-lsp-ext-tree-distr-opt-01

The following are the major changes to previous version draft-wu-trill-lsp-ext-tree-distr-opt-00:

- Add one new section to discuss Unicast Forwarding.
- Add one new section to clarify the motivation to write this draft.
- Some other editorial changes.

Authors' Addresses

Qin Wu
Huawei
101 Software Avenue, Yuhua District
Nanjing, Jiangsu 210012
China

Email: bill.wu@huawei.com

Weiguo Hao
Huawei
101 Software Avenue, Yuhua District
Nanjing, Jiangsu 210012
China

Email: haoweiguo@huawei.com