TRILL working group Internet Draft

Intended status: Standard Track

Expires: Sept 2013

L. Dunbar
D. Eastlake
Huawei
Radia Perlman
Intel
I. Gashinsky
Yahoo
February 22, 2013

Directory Assisted TRILL Encapsulation draft-dunbar-trill-directory-assisted-encap-03.txt

Status of this Memo

This Internet-Draft is submitted to IETF in full conformance with the provisions of \underline{BCP} 78 and \underline{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/ietf/lid-abstracts.txt

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

This Internet-Draft will expire on August 22, 2013.

Copyright Notice

Copyright (c) 2013 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP-78 and the IETF Trust's Legal Provisions Relating to IETF Documents
(http://trustee.ietf.org/license-info) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in

Section 4.e of the <u>Trust Legal Provisions</u> and are provided without warranty as described in the Simplified BSD License.

Abstract

This draft describes how data center network can benefit from non-RBridge nodes performing TRILL encapsulation with assistance from directory service.

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 0.

The term ''TRILL'' and ''RBridge'' are used interchangeably in this document. The term ''subnet'' and ''VLAN'' are also used interchangeably because it is very common to map one subnet to one VLAN.

Table of Contents

<u>1</u> . Introduction	2
<u>2</u> . Terminology	3
3. Directory Assistance to Non-RBridge	3
$\underline{\textbf{4}}.$ Source Nickname in Frames Encapsulated by Non-RBridge Nodes	6
5. Conclusion and Recommendation	6
6. Manageability Considerations	6
7. Security Considerations	6
8. IANA Considerations	6
9. Acknowledgments	6
<u>10</u> . References	7
Authors' Addresses	7
Intellectual Property Statement	8
Disclaimer of Validity	9

1. Introduction

This draft describes how data center network can benefit from non-RBridge nodes performing TRILL encapsulation with assistance from directory service.

[RBridge-directory] describes the framework for RBridge edge to get MAC&VLAN<->RBridgeEdge mapping from a directory service in data center environment instead of flooding unknown DAs across TRILL domain. When directory is used, any node, even non-RBridge node, can perform the TRILL encapsulation. This draft is to demonstrate the benefits of non-RBridge nodes performing TRILL encapsulation.

2. Terminology

AF Appointed Forwarder RBridge port

Bridge: IEEE 802.1Q compliant device. In this draft, Bridge is used

interchangeably with Layer 2 switch.

DA: Destination Address

DC: Data Center

End of Row switches in data center. Also known as EoR:

Aggregation switches in some data centers

FDB: Filtering Database for Bridge or Layer 2 switch

Application running on a physical server or a virtual Host:

machine. A host usually has at least one IP address and at

least one MAC address.

SA: Source Address

Top of Rack Switch in data center. It is also known as ToR:

access switches in some data centers.

VM: Virtual Machines

3. Directory Assistance to Non-RBridge

With directory assistance [RBridge-Directory], a non-RBridge can determine if a packet needs to be forwarded across the RBridge domain. Suppose the RBridge domain boundary starts at network switches (i.e. not virtual switches embedded on servers), a directory can assist Virtual Switches embedded on servers to encapsulate proper TRILL header by providing the information of the egress RBridge edge to which the target is attached. If a target is not attached to other RBridge edge nodes based on the directory [RBridge-Directory], the non-RBridge node can forward the data frames natively, i.e. not encapsulating any TRILL header.

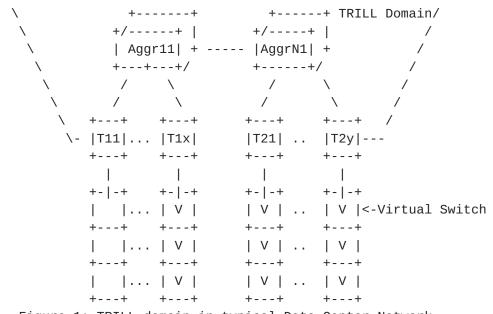


Figure 1: TRILL domain in typical Data Center Network

When a TRILL encapsulated data packet reaches the ingress RBridge, the ingress RBridge can simply forward the pre-encapsulated packet to the RBridge that is specified in the DA field of the TRILL header of the data frame. When the ingress RBridge receives a native Ethernet frame, it only forward the data frame to the directly attached bridged LAN.

Under this environment, the ingress RBridge doesn't need to flood the received Ethernet data frames to TRILL domain when the DA in the Ethernet data frames is unknown. Under this scheme, for an RBridge with multiple ports connected to a bridged LAN, data frames received from TRILL domain, decapsulated and forwarded to the bridged LAN via one port, and flooded back to the RBridge via another port, won't be encapsulated again and forwarded back TRILL domain.

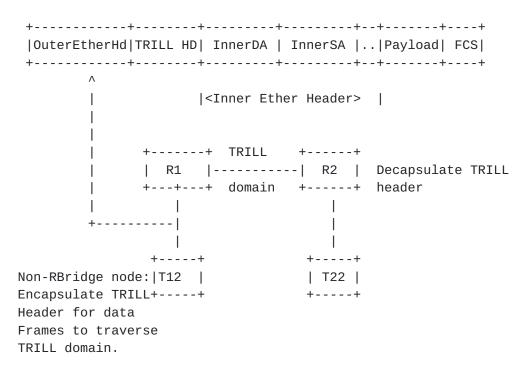
That means there is no need to worry about AF ports and all RBridge edge ports connected to one bridged LAN can receive and forward preencapsulated traffic, which greatly improves the overall network utilization.

Note: [RBridge] Section 4.6.2 Bullet 8 specifies that an RBridge port can be configured to accept TRILL encapsulated frames from a neighbor that is not an RBridge.

When data frames do not need to be sent across RBridge domain, they are switched by all nodes/ports per IEEE802.1Q and RBridge edge will not encapsulate and forward native Ethernet frames across RBridge domain.

When a pre-encapsulated TRILL frame arrives at an RBridge whose nickname matches with the destination nickname in the TRILL header, the processing is exactly same as normal, i.e. it decapsulates the native frame from the received TRILL frame and forwards the decapsulated Ethernet frame to the target attached to its edge ports.

We call a node that only performs the TRILL encapsulation but doesn't participate in RBridge's IS-IS routing a ''TRILL Encapsulating node'' or ''Simplified RBridge''. The TRILL Encapsulating Node gets the MAC&VLAN<->RBridgeEdge mapping table pushed down or pulled from directory servers [RBridge-directory]. Upon receiving a native Ethernet frame, the TRILL Encapsulating Node checks the MAC&VLAN<->RBridgeEdge mapping table, and perform the corresponding TRILL encapsulation if the entry is found in the mapping table. If the destination address and VLAN of the received Ethernet frame doesn't exist in the mapping table, the Ethernet frame is forwarded per IEEE802.1Q.



4. Source Nickname in Frames Encapsulated by Non-RBridge Nodes

The TRILL header includes a Source RBridge's Nickname (ingress) and Destination RBridge's Nickname (egress). When a TRILL header is added by a non-RBridge node, using the Ingress RBridge edge node's nickname in the source address field will make the ingress RBridge node receive TRILL frames with its own nickname in the frames' source address field, which can be confusing.

To avoid confusion of edge RBridges receiving TRILL encapsulated frames with their own nickname in the frames' source address field from neighboring non-RBridge nodes, a new nickname can be given to an RBridge edge node, e.g. Phantom Nickname, to represent all the TRILL Encapsulating Nodes attached to the RBridge edge node.

When the Phantom Nickname is used in the Source Address field of a TRILL frame, it is understood that the TRILL encapsulation is actually done by a non-RBridge node which is attached to an edge port of an RBridge Ingress node.

5. Conclusion and Recommendation

When directory service is available, nodes that are outside TRILL domain, i.e. don't participate in TRILL IS/IS routing protocol, become capable of encapsulating TRILL header for data frames destined for remote RBridges that is not on the same bridged LAN. The non-RBridge encapsulation approach is especially useful when there are many servers in a data center equipped with hypervisor-based virtual switches. It is relatively easy for virtual switches, which are usually software based, to get directory assistance and perform network address encapsulation.

Manageability Considerations

TBD.

Security Considerations

TBD.

8. IANA Considerations

TBD

9. Acknowledgments

This document was prepared using 2-Word-v2.0.template.dot.

10. References

[RBridge-Directory] Dunbar, et, al ''TRILL (Transparent Interconnection of Lots of Links) Edge Directory Assistance Framework'', < draft-ietf-trill-directory-framework-03>, March, 2013

[RBridges] Perlman, et, al ''RBridge: Base Protocol Specification'', <draft-ietf-trill-rbridge-protocol-16.txt>, March, 2010

[RBridges-AF] Perlman, et, al ''RBridges: Appointed Forwarders'', <<u>draft-ietf-trill-rbridge-af-02.txt</u>>, April 2011

[ARMD-Problem] Dunbar, et,al, ''Address Resolution for Large Data Center Problem Statement'', Oct 2010.

[ARP reduction] Shah, et. al., "ARP Broadcast Reduction for Large Data Centers", Oct 2010

Authors' Addresses

Linda Dunbar Huawei Technologies 1700 Alma Drive, Suite 500 Plano, TX 75075, USA

Phone: (972) 543 5849 Email: ldunbar@huawei.com Donald Eastlake Huawei Technologies 155 Beaver Street Milford, MA 01757 USA

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

Radia Perlman Intel Labs 2200 Mission College Blvd. Santa Clara, CA 95054-1549 USA

Phone: +1-408-765-8080 Email: Radia@alum.mit.edu

Igor Gashinsky Yahoo 45 West 18th Street 6th floor New York, NY 10011 Email: igor@yahoo-inc.com

Intellectual Property Statement

The IETF Trust takes no position regarding the validity or scope of any Intellectual Property Rights or other rights that might be claimed to pertain to the implementation or use of the technology described in any IETF Document or the extent to which any license under such rights might or might not be available; nor does it represent that it has made any independent effort to identify any such rights.

Copies of Intellectual Property disclosures made to the IETF Secretariat and any assurances of licenses to be made available, or the result of an attempt made to obtain a general license or permission for the use of such proprietary rights by implementers or users of this specification can be obtained from the IETF on-line IPR repository at http://www.ietf.org/ipr

The IETF invites any interested party to bring to its attention any copyrights, patents or patent applications, or other proprietary rights that may cover technology that may be required to implement any standard or specification contained in an IETF Document. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

All IETF Documents and the information contained therein are provided on an "AS IS" basis and THE CONTRIBUTOR, THE ORGANIZATION HE/SHE REPRESENTS OR IS SPONSORED BY (IF ANY), THE INTERNET SOCIETY, THE IETF TRUST AND THE INTERNET ENGINEERING TASK FORCE DISCLAIM ALL WARRANTIES, EXPRESS OR IMPLIED, INCLUDING BUT NOT LIMITED TO ANY WARRANTY THAT THE USE OF THE INFORMATION THEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Acknowledgment

Funding for the RFC Editor function is currently provided by the Internet Society.