TRILL Working Group
INTERNET-DRAFT
Intended status: Proposed Standard
Updates: RFCtrill

Radia Perlman
Intel Labs
Donald Eastlake 3rd
Huawei
Ayan Banerjee
Cisco
Hu Fangwei
ZTE

Expires: September 28, 2011

March 29, 2011

RBridges: Appointed Forwarders
<draft-perlman-trill-rbridge-af-00.txt>

Abstract

The IETF TRILL protocol provides optimal pair-wise data forwarding without configuration, safe forwarding even during periods of temporary loops, and support for multipathing of both unicast and multicast traffic. TRILL accomplishes this by using IS-IS link state routing and by encapsulating traffic using a header that includes a hop count. Devices that implement TRILL are called RBridges.

TRILL supports multi-access LAN links that can have multiple end stations and RBridges attached. Where multiple RBridges are attached to a link, native traffic to and from end stations on that link is handled by a subset of those RBridges called Appointed Forwarders, with the intent that native traffic in each VLAN be handled by at most one RBridge. The purpose of this document is to improve the documentation of the Appointed Forwarder mechanism.

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 TRILL working group mailing list <rbridge@postel.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

Acknowledgements

The authors of [RFCtrill] and [RFCadj] and those listed in the Acknowledgements section of [RFCtrill] and [RFCadj] are hereby acknowledged by reference.

Table of Contents

1. Introduction

The IETF TRILL protocol [RFCtrill] provides optimal pair-wise data frame forwarding without configuration, safe forwarding even during periods of temporary loops, and support for multipathing of both unicast and multicast traffic. TRILL accomplishes this by using [IS-IS] [RFC1195] link state routing and encapsulating traffic using a header that includes a hop count. The design supports VLANs and optimization of the distribution of multi-destination frames based on VLANs and IP derived multicast groups. Devices that implement TRILL are called RBridges.

Section 2 of [RFCadj] explains the environment for which the TRILL protocol is designed and the differences between that environment and the typical Layer 3 routing environment.

TRILL supports multi-access LAN links that can have multiple end stations and RBridges attached. Where multiple RBridges are attached to a link, native traffic to and from end stations on that link is handled by a subset of those RBridges called Appointed Forwarders, with the intent that native traffic in each VLAN be handled by at most one RBridge. An RBridge can be Appointed Forwarder for many VLANs.

The purpose of this document is to improve the documentation of the Appointed Forwarder mechanism. It includes reference implementation details. Alternative implementations that interoperate on the wire are permitted.

The Appointed Forwarder mechanism is irrelevant to any link on which end station service is not offered. This includes links configured as point-to-point and any link with all RBridge ports on that link configured as trunk ports. (In TRILL, configuration of a port as a "trunk port" just means that no end station service will be provided. It does not imply that all VLANs are enabled on that port.)

The Appointed Forwarder mechanism has no affect on the formation of adjacencies, the election of the DRB for a link, MTU matching, or pseudonode formation. Those topics are covered in [RFCadj]. Furthermore, Appointed Forwarder status has no effect on the handling of TRILL Data frames. It only affects the handling of native frames.

For other aspects of the TRILL base protocol see [RFCtrill] and [RFCadj]. Familiarity with [RFCtrill] and [RFCadj] is assumed in this document. In case of conflict between this document and [RFCtrill], this document prevails.

1.1 Terminology and Acronyms

This document uses the acronyms defined in [RFCtrill].

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 [RFC2119].

2. Appointed Forwarders and Their Appointment

The Appointed Forwarder on a link for VLAN-x is the RBridge that ingresses native frames from the link and egresses native frames to the link in VLAN-x. By default, the DRB on a link is in charge of native traffic for all VLANs on the link. The DRB may, if it wishes, act as Appointed Forwarder for any VLAN and it may appoint other RBridges that have ports on the link as Appointed Forwarder for one or more VLANs.

It is important that there not be two Appointed Forwarders on a link that are ingressing and egressing native frames for the same VLAN at the same time. Should this occur, it could form a loop where frames are not protected by a TRILL Hop Count for part of the loop. While TRILL tries to avoid such a situation, for loop safety there is also an "inhibition" mechanism (see Section 3) that can cause an RBridge that is Appointed Forwarder to not ingress or egress native frames.

As discussed in Section 5, an RBridge may have multiple ports on a link. As discussed in [RFCadj], if there are multiple ports with the same MAC address on a link, all but one will be suspended. While the cases of multiple ports on a link for one RBridge and multiple ports with the same MAC address are fully accommodated, multiple ports on a link for one RBridge is expected to be a rare condition and duplicate MAC addresses are not recommended by either TRILL or IEEE 802.1 standards.

Appointed Forwarder status has no affect on the handling of TRILL Data frames. It only affects the handling of native frames.

There are two mechanisms by which an RBridge can be appointed or un-appointed as Appointed Forwarder, as a result of DRB elections [RFCadj] as discussed in Section 2.1, or as a result of action by the DRB as discussed in Section 2.2.

2.1 Appointment Effects of DRB Elections

When an RBridge believes that it has become the DRB on a link, by default it can act as Appointed Forwarder for any VLANs on that link that it chooses as long as that VLAN is enabled on its port, or at least one of its ports if it has more than one port on the link.

When an RBridge believes that it has lost DRB status on a link or it observes a change in the RBridge that is DRB for the link without itself becoming DRB, it loses all Appointed Forwarder status.

In the rare corner case where an RBridge has more than one port on a link, one of which was previously the DRB election winner but that

port has just lost the DRB election to a different port of the same RBridge (possibly due to management configuration of port priorities), there is no change in which RBridge is DRB. Therefore neither of the above points applies and there is no change in Appointed Forwarder status.

2.2 Appointment and Removal by the DRB

The DRB may appoint other RBridges on the link through inclusion of one or more Appointed Forwarders sub-TLVs [RFCtisis] in a TRILL Hello it sends on the Designated VLAN out the port that won the DRB election. When the DRB sends any appointments in a TRILL Hello, it must send all appointments for that link in that Hello. Any previous appointment not included is implicitly revoked.

The DRB MUST NOT send any appointments on a link until its DRB inhibition timer (see Section 3) for that link has expired.

How the DRB decides what other RBridges on the link, if any, to appoint forwarder for which VLANs is beyond the scope of this document.

2.2.1 Processing Forwarder Appointments

When a non-DRB RBridge that can offer end station service on a link receives a TRILL Hello from the DRB on the Designated VLAN sent from the port that won the DRB election, that RBridge examines the Hello for Appointed Forwarder sub-TLVs. If none are present, there is no change in the receiver's Appointed Forwarder status. If one or more Appointed Forwarder sub-TLVs are present and none of them appoints the receiving RBridge, then it loses all Appointed Forwarder status and is no longer Appointed Forwarder for any VLAN. If one or more Appointed Forwarder sub-TLVs are present that appoint the receiving RBridge they are handled as described below.

An appointment in an Appointed Forwarder sub-TLV is for a specific RBridge and a contiguous interval of VLAN IDs; however, it actually appoints that RBridge forwarder only for the VLAN(s) in that range that are enabled on one or more ports that RBridge has on the link. The RBridge becomes Appointed Forwarder for all such VLANs in all such Appointed Forwarder sub-TLVs in the Hello PDU and, if it was Appointed Forwarder for any additional VLANs, it loses Appointed Forwarder status for such additional VLANs.

Since the network manager normally controls what VLANs are enabled on RBridge ports, they can appoint an RBridge forwarder for an arbitrary

set of VLANs by enabling only those VLANs on the relevant port (or ports) and then having the DRB send an appointment that appears to appoint the target RBridge forwarder for all VLANs. However, to facilitate inter-RBridge communication, the Designated VLAN for a link SHOULD be enabled on RBridge ports on that link and it may not be desired to appoint the RBridge forwarder for the Designated VLAN. Thus, in the general case, it would require two appointments, although it would still only require one appointment if the Designated VLAN were an extreme low or high value, that is, VLAN 1 or VLAN 0xFFE.

For example, assume the DRB wants RB2 to be Appointed Forwarder for all even numbered VLANs and the Designated VLAN for the link is VLAN 101. The network manager could cause all even numbered VLANs plus VLAN 101 to be enabled on the relevant port of RB2 and then, with the desired effect, cause the DRB to send appointments to RB2 appointing it forwarder for all VLANs from 1 through 100 and from 102 through 4,095.

2.2.2 Frequency of Appointments

It is not necessary for the DRB to include the forwarder appointments in every TRILL Hello that it sends on the designated VLAN. For loop safety, every RBridge is required to indicate, in every TRILL Hello it sends in VLAN-x to a link, whether it is acting as the Appointed Forwarder for VLAN-x for that link (see item 4 in Section 3). And it is RECOMMENDED that the DRB have all VLANs for which end station service will be offered on the link, as well as the Designated VLAN, enabled. Thus the DRB will generally be informed by other RBridges on the link of the VLANs for which they believe they are Appointed Forwarder. If this matches the appointments the DRB wishes to make, it is not required to re-send its forwarder appointments; however, for robustness, especially in cases such as VLAN misconfigurations in a bridged LAN link, it is RECOMMENDED that the DRB send its forwarder appointments once per Holding Time.

2.2.3 Appointed Forwarders Limit

The mechanism of DRB forwarder appointment and the limited length of TRILL Hellos imposes a limit on the number of RBridges on a link that can be Appointed Forwarders. To obtain a conservative estimate, assume that no more than 1000 bytes are available in a TRILL Hello for such appointments. Assume it is desired to appoint various RBridges on a link forwarders for arbitrary non-intersecting sets of VLANs. Using the technique discussed above would generally require two appointments, or 12 bytes, per RBridge. With allowance for sub-

TLV and TLV overhead, appointments for 83 RBridges would fit in under 1000 bytes. Including the DRB, this implies a link with 84 or more RBridges attached. Links with more than a hand full of RBridges attached are expected to be very rare.

(If the Designated VLAN were an extreme low or high value, such as VLAN 1, which is the default and may be a common value in practice, only 6 bytes per RBridge would be required. This would permit twice as many different Appointed Forwarder RBridges than indicated by the general analysis above or, alternatively, take only half as much space to appoint the same number of Appointed Forwarders.)

Unnecessary changes in the DRB and unnecessary changes in Appointed Forwarders SHOULD NOT be made as they result in transient lack of end station service. Large numbers of Appointed Forwarders on a link (in excess of 65) are NOT RECOMMENDED due to the complexity of their establishment and maintenance.

3. The Inhibition Mechanism

An RBridge has, for every link on which it can offer end station service (that is every link for which it can act as an Appointed Forwarder), the following timers denominated in seconds:

a DRB inhibition timer,

a root change inhibition timer, and

up to 4,094 VLAN inhibition timers, one for each legal VLAN ID.

The DRB and root change inhibition timers MUST be implemented.

The loss of native traffic due to inhibition will be minimized by logically implementing a VLAN inhibition timer per each VLAN for which end station service will ever be offered on the link by the RBridge and this SHOULD be done. However, if implementation limitations make a full set of such timers impractical, the VLAN inhibition timers for more than one VLAN can, with care, be merged into one timer. In particular, an RBridge MUST NOT merge the VLAN inhibition timers for together for two VLANs if it is Appointer Forwarder for one and not for the other as this can lead to unnecessary and indefinitely prolonged inhibition. In the limit, there will be safe operations, albeit with more native frame loss than would otherwise be required, even if only two VLAN inhibition timers are provided, one for VLANs for which the RBridge is Appointed Forwarder and one for all other VLANs. Where a VLAN inhibition timer represents more than one VLAN, an update or test that would have be done to the timer for any of the VLANs is performed on the merged timer.

These timers are set as follows:

1. On booting or management reset, each port will have its own set of timers, as even if two or more are on the same link the RBridge will not have had a chance to learn that yet. All inhibition timers are set to expired except the DRB inhibition timer that, because each port will initially believe it is DRB, is set in accordance with item 2 below.

2. When an RBridge believes that it has become DRB on a link, it sets the DRB inhibition timer to the Holding Time of its port on that link (or the largest Holding Time of any of its ports on that link if it knows it has more than one).

3. When an RBridge believes that it has lost DRB status on a link, it sets the DRB inhibition timer to expired.

Note: In the rare corner case where one port of an RBridge was

the DRB election winner but later loses the DRB election to a different port of the same RBridge on that link (perhaps due to management configuration of port priority), neither 2 nor 3 above applies and the DRB timer is not changed.

4. When an RBridge RB1 receives a TRILL Hello on VLAN-x from some other RBridge RB2 and RB2 asserts in the Hello that it is Appointed Forwarder, then RB1 sets the VLAN-x inhibition timer for that link to the maximum of its existing value and the Holding Time in the received Hello. An RBridge MUST maintain VLAN inhibition timers for a link to which it connects if it can offer end station service on that link even if it is not currently Appointed Forwarder for any VLAN on that link.

5. When an RBridge RB1 enables VLAN-x on a port connecting to a link and VLAN-x was previously disabled on all of RB1's ports on that link, it sets the VLAN inhibition timer for VLAN-x for that link to its Holding Time for that port.

6. When an RBridge detects a spanning tree root bridge change on a port, it sets the root change inhibition timer for the link to an amount of time that defaults to 30 seconds and is configurable to any value from 30 down to zero seconds. This condition will not occur unless the RBridge is receiving BPDUs on the port from an attached bridged LAN. It is safe to configure this inhibition time to the settling time of an attached bridged LAN. For example, if it is known that Rapid Spanning Tree Protocol (RSTP [802.1Q-2005]) is running throughout the attached bridged LAN, it should be safe to configure this inhibition time to 4 seconds. Note that, while an RBridge could determine what version of spanning tree is running on the physical link between it and any directly connected bridge by examination of the BPDUs it receives, it could not tell if inter-bridge links beyond those directly connected bridges were running classic Spanning Tree Protocol (STP), which might require the root change inhibition timer to be set to 30 seconds for safety.

7. When an RBridge decides that one of its ports (or a set of its ports) P1 is on the same link as another of its ports (or set of its ports) P2, then the inhibition timers are merged to a single set of inhibition timers by using the maximum value of the corresponding timers.

8. When an RBridge decides that a set of its ports that it had been treating as being on the same link are no longer on the same link, those ports will necessarily be on two or more links (one link per port in the limit). This is handled by cloning a copy of the timers for each of the two or more links the RBridge has decided these ports connect to.

4. Inhibited Appointed Forwarder Behavior

An Appointed Forwarder for a link is inhibited for VLAN-x unless (1) its DRB inhibition timer for that link, (2) its root inhibition time for that link, and (3) its VLAN inhibition timer for that link for VLAN-x, are all three expired.

If a VLAN-x Appointed Forwarder for a link is inhibited and receives a TRILL Data frame whose encapsulated frame is in VLAN-x and would normally be egressed to that link, it decapsulates the native frame as usual. However, it does not output it to or queue it for that link although, if appropriate, it may output it to or queue it for other links.

If a VLAN-x Appointed Forwarder for a link is inhibited and receives a native frame in VLAN-x that would normally be ingressed from that link, the native frame is ignored.

An RBridge with an un-expired VLAN inhibition timer for VLAN-x is still required to indicate in TRILL Hellos it sends on VLAN-x whether it is Appointed Forwarder for that VLAN.

Inhibition has no effect on the receipt or forwarding of TRILL Data frames.

5. Multiple Ports on the Same Link

An RBridge may have multiple ports on the same link. Some of these ports may be suspended due to MAC address duplication as described in [RFCadj]. Suspended ports never ingress or egress native frames.

If an RBridge has one or more non-suspended ports on a link and those ports offer end station service, that is, those ports are not configured as point-to-point or trunk ports, then that RBridge is eligible to be an Appointed Forwarder. It can become Appointed Forwarder either by default because it is DRB, or by appointment by the DRB as described in Sections 2.1 and 2.2.

If an RBridge which is Appointed Forwarder for VLAN-x on a link has multiple non-suspended ports on that link, it may load share the task of ingressing and egressing VLAN-x native frames across those ports however it chooses, as long as there is no case in which a frame it egresses onto the link from one port can be ingressed on another port, creating a loop. If the RBridge is Appointed Forwarder for multiple VLANs, a straightforward thing to do would be to partition those VLANs among the ports it has on the link.

6. Security Considerations

This memo provides improved documentation of the TRILL Appointed Forwarder mechanism. It does not change the security considerations of the TRILL base protocol. See Section 6 of [RFCtrill].

7. IANA Considerations

This document requires no IANA actions. RFC Editor: Please delete this section before publication.

8. References

Normative and Informational references for this document are listed below.

8.1 Normative References

[IS-IS] - ISO/IEC 10589:2002, Second Edition, "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)", 2002.

[RFC1195] - Callon, R., "Use of OSI IS-IS for routing in TCP/IP and dual environments", RFC 1195, December 1990.

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

[RFCtisis] - Eastlake, D., A. Banerjee, D. Dutt, R. Perlman, A. Ghanwani, "TRILL Use of IS-IS", draft-ietf-isis-trill-05, in RFC Editor's queue.

[RFCtrill] - Perlman, R., D. Eastlake, D. Dutt, S. Gai, and A. Ghanwani, "RBridges: Base Protocol Specification", draft-ietf-trill-rbridge-protocol-16.txt, in RFC Editor's queue.

[RFCadj] - Eastlake, D., R. Perlman, A. Ghanwani, D. Dutt, V. Manral, "RBridges: Adjacency", draft-ietf-trill-adj, work in progress.

8.2 Informative References

None.

Authors' Addresses

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

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

Donald Eastlake 3rd
Huawei Technologies
155 Beaver Street
Milford, MA 01757 USA

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

Ayan Banerjee
Cisco Systems
170 West Tasman Drive
San Jose, CA 95134 USA

Email: ayabaner@cisco.com

Fangwei Hu
ZTE Corporation
889 Bibo Road
Shanghai 201203
China

Phone: +86-21-68896273
Email: hu.fangwei@zte.com.cn

Copyright and IPR Provisions