

TRILL Working Group  
Internet Draft  
Intended status: Standards Track

Yizhou Li  
Weiguo Hao  
Huawei Technologies

Expires: November 2012

July 09, 2012

Aware Spanning Tree Topology Change on R Bridges  
draft-yizhou-trill-tc-awareness-00.txt

## 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/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 January 9, 2009.

## Copyright Notice

Copyright (c) 2012 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

Internet-Draft

VLAN based Tree Selection

July 2012

the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Abstract

Table of Contents

<a href="#">1.</a>	<a href="#">Introduction .....</a>	<a href="#">2</a>
<a href="#">1.1.</a>	<a href="#">Motivations .....</a>	<a href="#">4</a>
<a href="#">2.</a>	<a href="#">Conventions used in this document.....</a>	<a href="#">5</a>
<a href="#">3.</a>	<a href="#">BPDU RBridge Channel.....</a>	<a href="#">5</a>
<a href="#">4.</a>	<a href="#">Operations .....</a>	<a href="#">6</a>
<a href="#">4.1.</a>	<a href="#">Sending BPDU using RBridge channel .....</a>	<a href="#">7</a>
<a href="#">4.2.</a>	<a href="#">Receiving BPDU in RBridge channel .....</a>	<a href="#">7</a>
<a href="#">4.3.</a>	<a href="#">Informing the remote site.....</a>	<a href="#">8</a>
<a href="#">5.</a>	<a href="#">Security Considerations.....</a>	<a href="#">9</a>
<a href="#">6.</a>	<a href="#">IANA Considerations .....</a>	<a href="#">9</a>
<a href="#">7.</a>	<a href="#">References .....</a>	<a href="#">10</a>
<a href="#">7.1.</a>	<a href="#">Normative References.....</a>	<a href="#">10</a>
<a href="#">7.2.</a>	<a href="#">Informative References.....</a>	<a href="#">10</a>
<a href="#">8.</a>	<a href="#">Acknowledgments .....</a>	<a href="#">10</a>

## [1.](#) Introduction

TRILL protocol [[RFC6325](#)] [[RFC6439](#)] described appointed forwarder mechanism for loop avoidance in the scenario shown by Figure 1. Only one of the RBridges is responsible for encapsulating/decapsulating a given VLAN data frame on a link. Local bridged LAN runs normal spanning tree protocol for loop avoidance. RBridge keeps track of the root bridge by listening to BPDUs received on the local port. This information is reported per VLAN by the RBridge in its LSP and is used to detect the root change. Root change will trigger the reset of the inhibition timer of the appointed forwarder. When an RBridge ceases to be appointed forwarder for a VLAN on a port, it sends topology change BPDUs to purge the MAC table on local bridged LAN switches. An RBridge never encapsulates or forwards any BPDU frame it receives [[RFC6325](#)].

[RFC6325] A.2 & A.3 presented the problems using the conventional approach shown in Figure 1. Native frames enter and leave a link via the link's appointed forwarder for the VLAN of the frame can cause congestion or suboptimal routing. Four methods was illustrated in

[RFC6325] to solve the problem,

1. Use RBridge instead of conventional bridge

2. Re-arrange network topology
3. Carefully select the different appointed forwarders for VLANs if end stations on local bridged LAN can be separated into multiple VLANs
4. Configure the RBridges to be like one STP tree root in local bridged LAN. The RBridge ports that are connected to the bridged LAN send spanning tree configuration BPDUs. Then the bridged LAN is forced into partitions. Figure 2 shows its network topology.

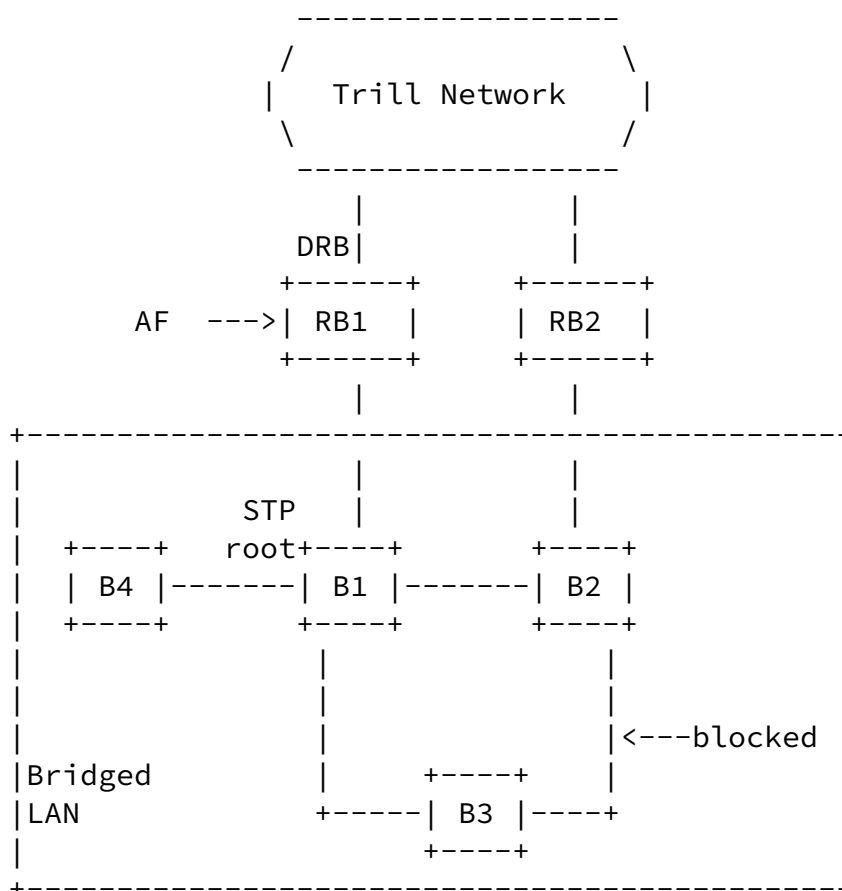


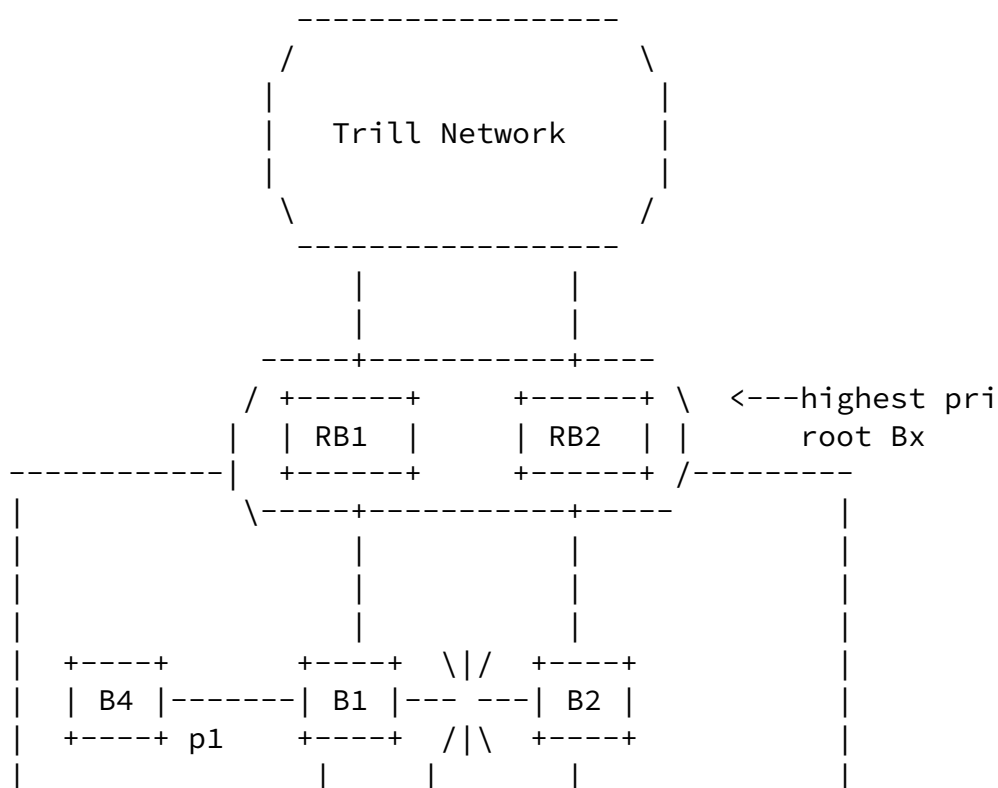
Figure 1 TRILL and bridged LAN topology

Method 1 and 2 highly depends on the network topology and equipment types and therefore have very limited applicability. Method 3 and 4 have broader applicability. Method 4 is more applicable than method 3 if all end stations in bridged LAN are on the same VLAN or intra VLAN load balancing is required to avoid per VLAN congestion and suboptimal routing. The traffic discontinuity was caused by inhibition timer setting in case of root change in method 3. Proper timeout value has to be carefully chosen for tradeoff between unnecessary traffic continuity and potential loop. Method 4

eliminates the requirement of setting inhibition timer in case of root change. Therefore method 4 is considered as a very common practice in real deployment.

### [1.1](#). Motivations

Bridged LAN may have topology change at any time. When RB1 & RB2 serve as one single STP tree root, it is required that RB1 and RB2 have to tunnel some BPDUs to help the bridged LAN convergence in certain circumstances. Figure 2 is used to show such motivation in the given topology.



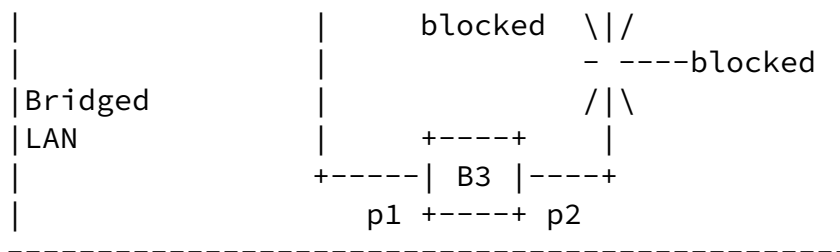


Figure 2 RBs function as STP tree root topology

RB1 & RB2 use the same bridge ID to emit spanning tree BPDUs as the highest priority root Bx. All bridges in LAN see RB1 and RB2 as a single tree root. Therefore B1-B2 and B2-B3 links are blocked for loop avoidance after running spanning tree protocol. RB1 and RB2 will not receive TRILL-Hello from each other. Bridged LAN is logically

partitioned into two parts. RB1 is DRB and AF for all VLANs in left partition and RB2 is DRB and AF in right partition.

If B1-B3 link fails for some reason, alternate port p2 on B3 will send topology change (TC) BPDU to B2. B2-B3 link will start forwarding frames. TC BPDU is then sent from B2 to RB2. As RB2 never forwards BPDU frame to TRILL campus, left partition has no way to know the topology change. Therefore B4 will not be able to correctly purge the MACs learnt from port p1 for end stations connected to B3. MAC table entry aging is the last resort in this case. In addition, a remote end station may keep sending traffic to an end station connected to B3 via RB1-B1 which causes frame loss. Therefore some mechanism must be used to purge the MACs learned both in the left partition of the bridged LAN and the remote Rbridges when topology changes. This draft proposes to use RBridge channel [[TRILLChannel](#)] to tunnel the TC BPDU to solve the issue.

## 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 [RFC-2119](#) [[RFC2119](#)].

This document uses the terminologies defined in [[RFC6325](#)] along with the following:

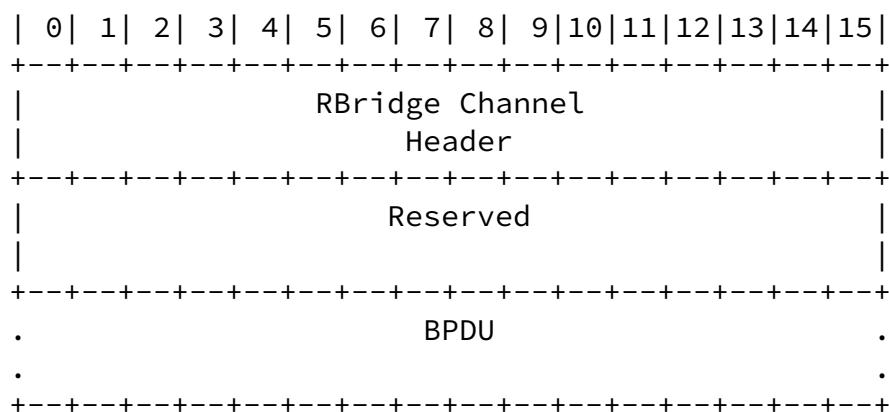
Root Bridge Group - A group of RBridges acting as a single tree root

in a spanning tree instance in local bridged LAN

### [3. BPDU RBridge Channel](#)

A new channel protocol is defined to carry BPDU.

Channel protocol code: TBD (BPDU)



### [4. Operations](#)

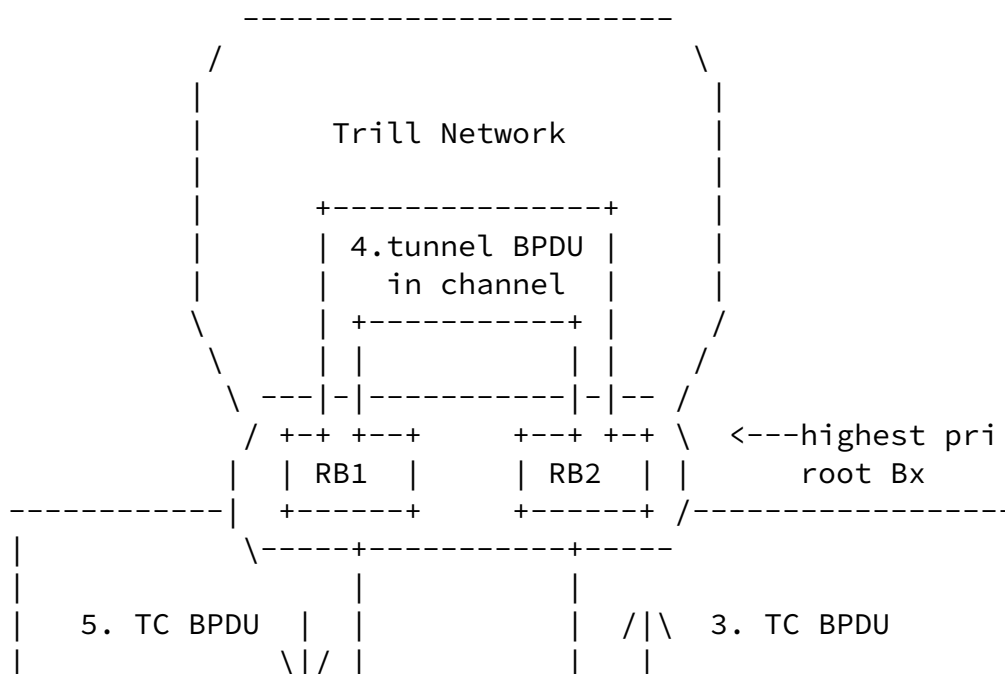
Figure 3 shows TC BPDU tunneled from RB2 to RB1 using RBridge Channel.

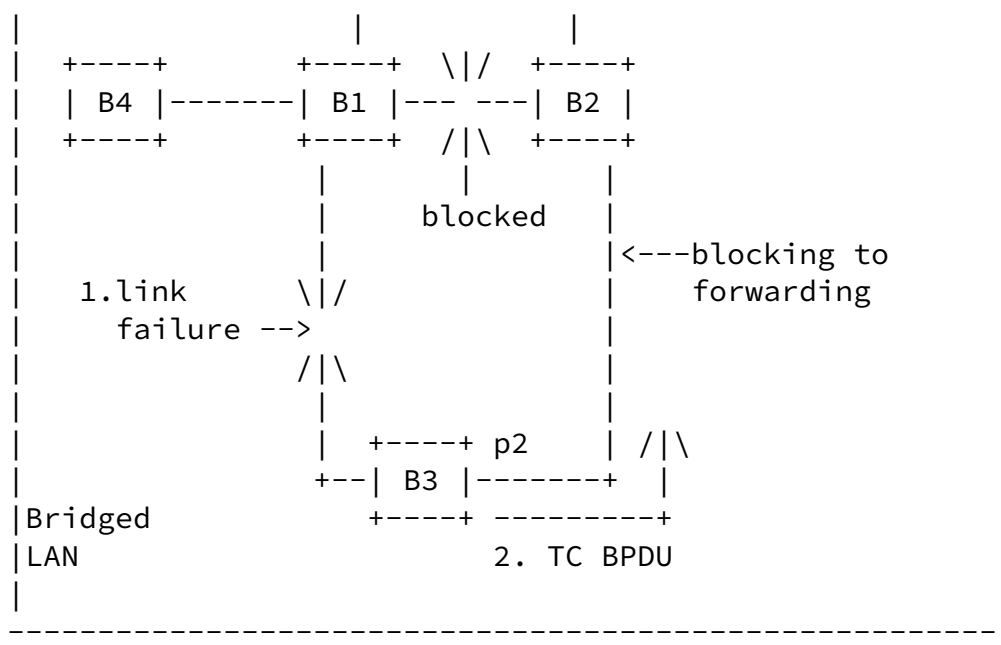
#### [4.1. Sending BPDU using RBridge channel](#)

In figure 3, when B1-B3 link fails, port p2 on B3 will start to send TC BPDU and go to forwarding state. RB2 receives TC BPDU from B2 sequentially. RB2 encapsulates the TC BPDU in RBridge channel and sends it to RB1.

Interested VLANs and Spanning Tree Roots Sub-TLV [RFC6326] carries spanning tree root bridge IDs seen for all ports for which the RBridge is the appointed forwarder for a VLAN. As RB1 and RB2 use the same bridge ID and that bridge ID is the spanning tree root, RB1 and RB2 are considered as in a root bridge group.

When RBridge receives TC BPDUs from an access port, it tunnels the frame to all the other R Bridges in the same root bridge group using RBridge channel protocol specified in [section 3](#). Normally the number of R Bridges in a root bridge group is limited, say 2 or 3; such tunneling is performed using TRILL unicast encapsulation. N members in a root bridge group results in N-1 unicast tunneled BPDU sent. In figure 3, RB2 knows RB1 is in the same root bridge group from LSP exchange; hence RB2 uses RB1's nickname as egress nickname and encapsulates the TC BPDU in RBridge channel.





#### 4.2. Receiving BPDU in RBridge channel

When an RBridge receives a TC BPDUs from RBridge channel, it determines the frame was sent from a RB in the same root bridge group. Then RBridge decapsulates the frame and sends the original TC BPDUs to

its local bridged LAN. TC BPDU will be flooded throughout in left partition to clear MAC table in bridges.

### 4.3. Informing the remote site

When local topology changes, the correspondence of end station and its attaching RBridge cached by remote RB may become invalid. The RBridges who is the appointed forwarder for the specified VLAN in remote sites should be informed to clear the stale correspondence table entry.

When traffic is bi-directional, the remote RBridge will receive the data frames from the newly attached RBridge of the local end station. The remote RBridge will update its MAC-Nickname correspondence table.





```

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|           ...           |      nickname n      |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|      nickname n      | Num of VLAN blocks|
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|      Start.VLAN      |           |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|      End.VLAN        |           |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|      Other Start/End VLAN list ...      |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

## 5. Security Considerations

This document does not change the general RBridge security considerations of the TRILL base protocol and TRILL RBridge Channel. See [Section 6 of \[RFC6325\]](#) and section 7 of [\[TRILLChannel\]](#).

Forged TC BPDU may trigger RBridge continuously sending tunneled BPDU and MAC purge. It may cause denial-of-service in TRILL campus. Similar as the traditional bridged LAN running spanning tree, it is suggested to monitor the receiving rate of TC BPDU on bridged LAN facing port of RBridges. If the receiving rate is beyond the threshold, RBridge should only process and tunnel the TC BPDU in the configured rate.

## 6. IANA Considerations

IANA is requested to allocate the new channel protocol codes as following.

Channel protocol code X1: BPDU

Channel protocol code X2: MAC purge

## 7. References

### 7.1. Normative References

- [RFC6325] Perlman, R., Eastlake 3rd, 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, "TRILL Use of IS-IS", [RFC 6326](#), July 2011.
- [6326bis] Eastlake, D. et.al., "'Transparent Interconnection of Lots of Links (TRILL) Use of IS-IS'", [draft-eastlake-isis-rfc6326bis-07.txt](#), Work in Progress, December 2011.
- [RFC6439] Eastlake, D. et.al., "'RBridge: Appointed Forwarder'", [RFC 6439](#), November 2011.
- [TRILLChannel] - Eastlake, D., V. Manral, Y. Li, S. Aldrin, D. Ward, "RBridges: RBridge Channel Support in TRILL", [draft-ietf-trill-rbridge-channel](#), work in progress.
- [RFC6327] Eastlake 3rd, D., Perlman, R., Ghanwani, A., Dutt, D., and V. Manral, "Routing Bridges (RBridges): Adjacency", [RFC 6327](#), July 2011
- [802.1D] "IEEE Standard for Local and metropolitan area networks /Media Access Control (MAC) Bridges", 802.1D-2004, 9 June 2004.

## [7.2](#). Informative References

- [RFC6165] Banerjee, A. and D. Ward, "Extensions to IS-IS for Layer-2 Systems", [RFC 6165](#), April 2011.
- [802.1Q-2011] "IEEE Standard for Local and metropolitan area networks /Virtual Bridged Local Area Networks", 802.1Q-2011, 31 Aug 2011.

## [8](#). Acknowledgments

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

## Authors' Addresses

Yizhou Li  
Huawei Technologies  
101 Software Avenue,  
Nanjing 210012  
China

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

Weiguo Hao  
Huawei Technologies  
101 Software Avenue,  
Nanjing 210012  
China

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