

TRILL
Internet-Draft
Intended status: Standards Track
Expires: July 2, 2012

H. Zhai
F. Hu
ZTE Corporation
Dec 30, 2011

**Extending the Virtual Router Redundancy Protocol for TRILL campus
draft-hu-trill-rbridge-vrrp-02.txt**

Abstract

TRILL can be implemented in data center networks, which requires high reliability and stability. Whenever the egress RBridges or links break down, the TRILL rerouting time depends on the IS-IS topology convergence time, which may do not meet data center service requirements in terms of resiliency. VRRP provides a redundancy mechanism to avoid single point of failure and fast switching over. This draft proposes to extend VRRP protocol to TRILL in data center networks.

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 July 2, 2012.

Copyright Notice

Copyright (c) 2011 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 Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	3
2.	Terminology	3
3.	Application Scenario	4
4.	TRILL VRRP Frames	6
4.1.	TRILL-VRRP Multicast Address	7
4.2.	Source RBridge MAC Address	7
4.3.	L2-TRILL-VRRP Ethertype	7
4.4.	Frame Check Sequence (FCS)	8
5.	TRILL VRRP Payload Format	8
5.1.	Version	8
5.2.	Type	9
5.3.	Virtual RB ID	9
5.4.	Priority	9
5.5.	Count Nicknames	9
5.6.	Rsvd	9
5.7.	Maximum Advertisement Interval (Max Adver Int)	9
5.8.	Checksum	10
5.9.	Virtual System ID	10
5.10.	Nickname(s)	10
6.	VRRP Protocol State Machine	10
7.	IS-IS Adjacency	11
8.	Security Considerations	11
9.	Acknowledgements	11
10.	References	11
10.1.	Normative references	11
10.2.	Informative References	11
	Authors' Addresses	12

1. Introduction

TRILL (Transparent Interconnection of Lots of Links) provides optimal pair-wise data forwarding without configuration, safe forwarding even during periods of temporary loops, and support for multi-pathing of both unicast and multicast traffic[RFC6325]. TRILL is designed to replace STP (Spanning Tree Protocol) for data center networks. The IS-IS routing protocol is used as a control plane protocol to discover the topology and advertise link states. When the topology changes, IS-IS LSPs flood the link state information to other adjacencies. The topology convergence time can go up to about dozens of seconds for particular data center topologies, which may do not meet data center service requirements.

VRRP (Virtual Router Redundancy Protocol) specifies an election protocol that dynamically assigns responsibility for a virtual router to one of the VRRP routers on a LAN in IP network. There is one master equipment and one or several backup equipments in a VRRP group. The VRRP group looks like one equipment from the host side. This document is to extend VRRP to support nickname namespace and apply the VRRP protocol in TRILL network. VRRP is used to solve the single point failure of edge equipment, and fast the switching over time by avoiding topology changed in TRILL campus network. There is a VRRP group including one master BRB(Border RBridge) and one or several backup BRBs in the TRILL edge network. When the master BRB is failed, one of the backup BRB takes the role of master in the VRRP group. The master BRB floods the virtual nickname in TRILL campus network. The other R Bridges doesn't feel the change of master and the ISIS topology doesn't change.

The remainder of this document is organized as follows: Section 3 describes the VRRP application scenarios. There are two application scenarios introduced in the document. [Section 4](#) specifics the TRILL VRRP frame structure and encapsulation. The TRILL VRRP frame is encapsulated as the payload of Ethernet Frame. There is new type "L2-TRILL-VRRP" Ethertype to identify the TRILL VRRP frame. [Section 5](#) specifics the VRRP frame fields in details. [Section 6](#) describes the VRRP protocol state machine. [Section 7](#) describes IS-IS adjacency when deployed VRRP protocol in the R Bridges.

2. Terminology

Border RBridge: Abbr.BRB, a device locates the border of TRILL campus and runs TRILL protocol, BRB is used to communicate with other TRILL campus

VRRP RBridge: an RBridge running the Virtual Router Redundancy

Protocol. It may participate in one or more VRRP groups.

Virtual RBridge: An abstract object managed by VRRP that acts as a default RBridge for devices on a shared LAN. It consists of a Virtual System Identifier and a set of associated nickname (s) across a common LAN. A VRRP RBridge may backup one or more virtual RBridges.

Nickname Owner: The VRRP RBridge that has the virtual RBridge's nickname as one of its nickname addresses. This is the RBridge that, when up, will respond to packets addressed to one of these nickname addresses for ICMP pings, TCP connections, etc.

Virtual RBridge master: The VRRP RBridge that is assuming the responsibility of forwarding packets sent to the nickname associated with the virtual RBridge, and answering ARP requests for these nickname. Note that if the nickname owner is available, then it will always become the Master.

Virtual RBridge backup: The set of VRRP RBridge available to assume forwarding responsibility for a virtual RBridge should the current Master fail.

3. Application Scenario

Figure 1 is a data center application scenario. BRB is the edge of data center and the exit RBridge for the VLAN. In order to improve the reliability, BRB2 is the backup of BRB1 for the VLAN. If BRB1 is broken, RB1 and RB2 recalculate the route, and BRB2 becomes the exit RBridge for the VLAN. RB1 and RB2 should flush the new LSP in the network. It takes more than several seconds to switch data traffic from BRB1 to BRB2, which is not satisfied to the current data center video traffic. In addition, if the physical connection between RT1 and BRB1 is broken, RB1 can not feel the failure.

This document proposes to apply VRRP for ensure fast BRB switching upon failure. The VRRP mechanism can guarantee a sub-second (less than 50ms) switching time to ensure video data [VRRPv3]. Master BRB and backup BRBs are configured as belonging to a same VRRP group with the same virtual system ID and virtual nickname. The master BRB of the group floods the virtual nickname to adjacencies. If the Master BRB becomes unavailable, then the highest priority Backup BRB will be elected as Master after a short delay, providing a controlled transition of the virtual RBridge responsibility with minimal service interruption, and the Master BRB elected floods Link State Packets (LSPs) and is responsible for data forwarding in TRILL campus network, and the content of LSPs and the IS-IS link state topology

doesn't change.

In addition, two VRRP groupes can be configured in the BRBs, one group is used to down link, and the other is used to up link. The two VRRP groupes can be linkage: Once a VRRP group switches over the master, the other VRRP group will switch over the master BRB too. The physical connection between RT1 and RBR1 is broken, the up VRRP group switches the master to BRB2, and notifies the down link VRRP group to switch BRB2 too.

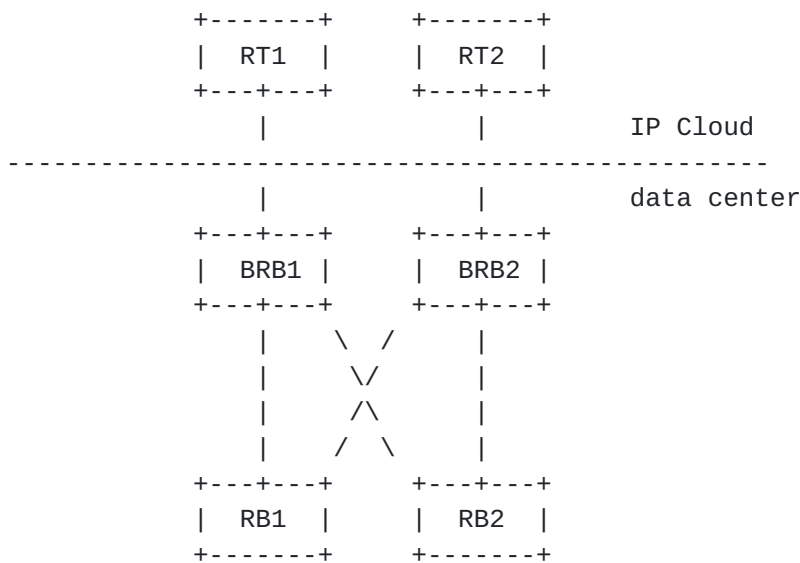


Figure 1

Figure 2 is multi-level TRILL deployment scenario. It is recommended that there is only one BRB in the border of two levels. The BRB becomes the bottleneck of TRILL network in case of failure, and is very easy to create such a single point of failure. Extension of VRRP can improve the reliability of BRB and avoiding the single point of failure.

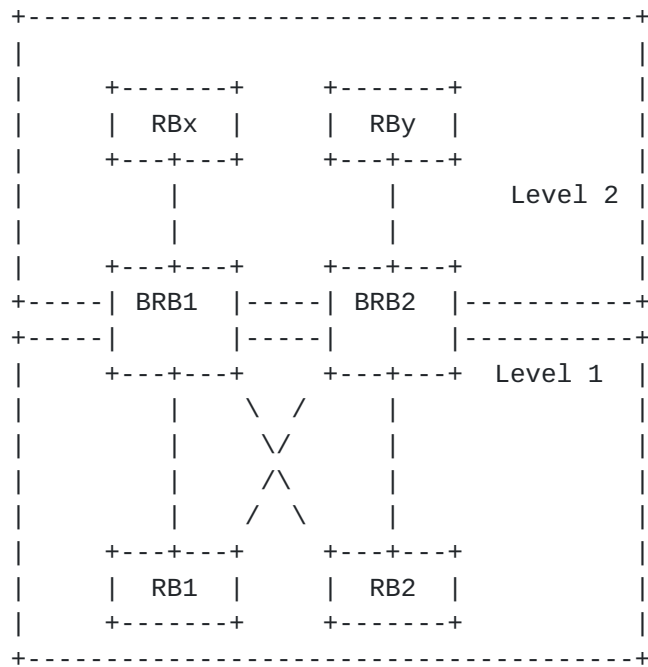


Figure 2

4. TRILL VRRP Frames

By multicasting periodically a TRILL VRRP frame, a master RBridge announces its existence and functionality to the backup RBridge(s) in a VRRP group. If none TRILL VRRP frame is received in a certain time, backup RBridge(s) will consider the master unavailable and trigger a new master RBridge election process.

A TRILL VRRP frame on an 802.3 link is structured as figure 3. All such frames are Ethertype encoded. The RBridge port out which such a frame is sent will strip the outer VLAN tag if configured to do so.

VRRP Frame Structure

Outer Ethernet Header:

```

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     TRILL-VRRP Multicast Address                                     |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|   TRILL-VRRP continued   |   Source RBridge MAC Address   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|   Source RBridge MAC Address continued   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|   Ethertype = C-Tag [802.1Q] | Outer.VLAN Tag Information   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|   L2-TRILL-VRRP Ethertype   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

VRRP for TRILL Payload:

```

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|                                     TRILL VRRP Payload                                     |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

Frame Check Sequence:

```

+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+
|   FCS (Frame Check Sequence)   |
+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+---+

```

Figure 3

4.1. TRILL-VRRP Multicast Address

The TRILL-VRRP multicast address is an IP-derived multicast MAC address. The IP address is:

224.0.0.18

The IP-derived multicast address is a link local scope multicast address. R Bridges MUST NOT forwards a frame with this destination address to another link.

4.2. Source RBridge MAC Address

It is a MAC address of RBridge port out which this TRILL VRRP frame is sent

4.3. L2-TRILL-VRRP Ethertype

It is used to indicate that the payload in the frame is a TRILL VRRP packet

4.4. Frame Check Sequence (FCS)

Each Ethernet frame has a single Frame Check Sequence (FCS) that is computed to cover the entire frame, for detecting frame corruption due to bit errors on a link. Thus, when a frame is encapsulated, the original FCS is not included but is discarded. Any received frame for which the FCS check fails SHOULD be discarded (this may not be possible in the case of cut through forwarding).

Although the FCS is normally calculated just before transmission, it is desirable, when practical, for an FCS to accompany a frame within an RBridge after receipt.

5. TRILL VRRP Payload Format

The format of TRILL VRRP payload is structured as figure 5.

VRRP Payload Format

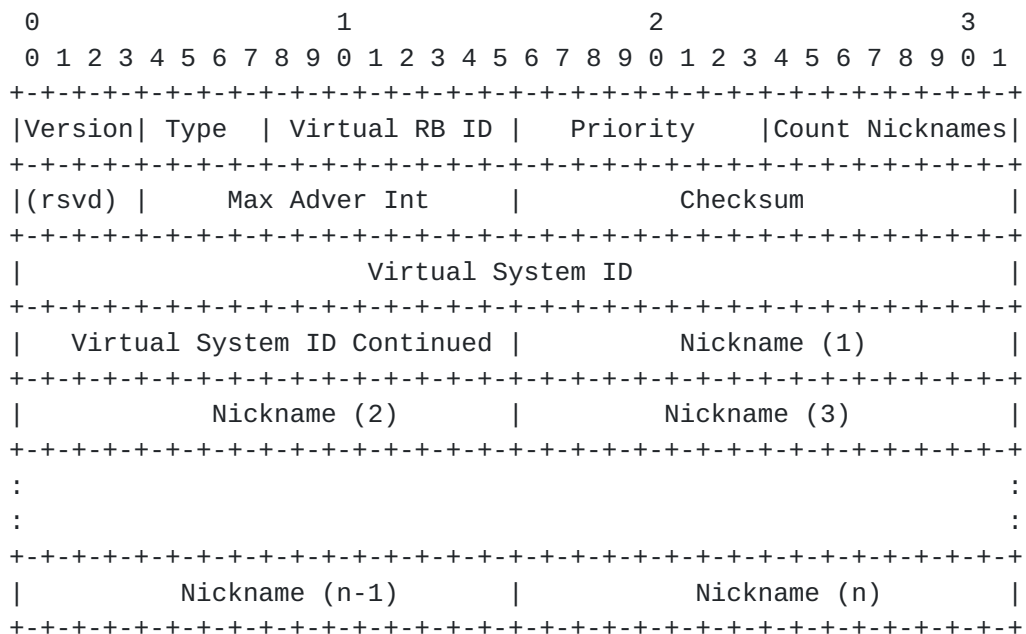


Figure 4

5.1. Version

The version field specifies the TRILL VRRP protocol version of this packet. This document defines version 1.

5.2. Type

The type field specifies the type of this TRILL VRRP packet. The only packet type defined in this version of the protocol is:

1 ADVERTISEMENT

A packet with unknown type MUST be discarded.

5.3. Virtual RB ID

The Virtual RBridge Identifier (VRBID) field identifies the virtual RBridge this packet is reporting status for. It is a configurable item in the range 1-255 (decimal). There is no default.

5.4. Priority

The priority field specifies the sending TRILL VRRP RBridge's priority for the virtual RBridge. Higher values equal higher priority. This field is an 8-bit unsigned integer field.

The priority value for the TRILL VRRP RBridge that owns the nicknames associated with the virtual nickname MUST be 255 (decimal).

TRILL VRRP RBridges backing up a virtual RBridge MUST use priority values between 1-254 (decimal) and the default priority value is 100(decimal).

The priority value zero (0) has special meaning, indicating that the current Master has stopped participating in TRILL VRRP. This is used to trigger backup RBridges to quickly transition to Master without having to wait for the current Master to time out.

5.5. Count Nicknames

The number of nicknames contained in this TRILL VRRP advertisement.

5.6. Rsvd

This field MUST be set to zero on transmission and ignored on reception.

5.7. Maximum Advertisement Interval (Max Adver Int)

The Maximum Advertisement Interval is a 12-bit field that indicates the time interval (in centiseconds) between ADVERTISEMENTS. The default is 100 centiseconds (1 second).

5.8. Checksum

The checksum field is used to detect data corruption in the TRILL VRRP message.

The checksum is the 16-bit one's complement of the one's complement sum of the entire TRILL VRRP message starting with the version field. For computing the checksum, the checksum field is set to zero. See [RFC1071](#) for more detail [CKSM].

5.9. Virtual System ID

The virtual system id is a 48-bit field that indicates the system id of the virtual RBridge this packet is reporting status for.

All the RBridges in a virtual RBridge MUST be configured with the same virtual system id. When a TRILL VRRP packet with different virtual system id from local virtual system id is received, the packet MUST be discarded. This field is used for troubleshooting misconfigured RBridges.

5.10. Nickname(s)

One or more nicknames are associated with the virtual RBridge. The number of nicknames included is specified in the "Count Nicknames" field. These fields are used for troubleshooting misconfigured RBridges.

6. VRRP Protocol State Machine

The VRRP protocol state machine is not changed. There are three states: Initialize, backup and master. Initialize state is to wait for a startup event; backup state is to monitor the availability and state of the master RBridge.

The master BRB election is according to the priority value. When the RBridge is elected as a virtual RBridge master, it floods LSP with virtual nickname to its adjacencies. If the RBridge is the nickname owner, it becomes the virtual nickname master automatically, and floods LSPs with owner nickname. Backup RBridge monitors and receives the VRRP packet from master. If backup RBridge has already enabled IS-IS protocol, it should flood LSP to withdraw its nickname LSA. Otherwise the backup RBridge should not flood LSP to its neighbors. The Backup RBridge exchanges hello packet with its neighbor, and receives LSPs from its adjacencies except from the master RBridge. Moreover, it never advertises local LSA, which is advertised by master RBridge.

7. IS-IS Adjacency

Master RBridge should setup and maintain all the adjacencies with the other RBridges except the backup RBridge. The Backup RBridge receives the other RBridges hello packets and IS-IS packets (such as LSP, CSNP, PSNP) besides master RBridge, but should not send any hello and IS-IS packets (LSP, CSNP, PSNP) to other RBridges. The backup RBridge can be detect, 2-way, and report states [[RFC6326](#)].

8. Security Considerations

9. Acknowledgements

The authors would like to gratefully acknowledge many people who have contributed discussion and ideas to the making of this proposal.

10. References

10.1. Normative references

- [RFC1071] Braden, R., Borman, D., Partridge, C., and W. Plummer, "Computing the Internet checksum", [RFC 1071](#), September 1988.
- [RFC1195] Callon, R., "Use of OSI IS-IS for routing in TCP/IP and dual environments", [RFC 1195](#), December 1990.
- [RFC3768] Hinden, R., "Virtual Router Redundancy Protocol (VRRP)", [RFC 3768](#), April 2004.
- [RFC5798] Nadas, S., "Virtual Router Redundancy Protocol (VRRP) Version 3 for IPv4 and IPv6", [RFC 5798](#), March 2010.
- [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.

10.2. Informative References

Authors' Addresses

Hongjun Zhai
ZTE Corporation
68 Zijinghua Road
Nanjing 200012
China

Phone: +86-25-52877345
Email: zhai.hongjun@zte.com.cn

Fangwei Hu
ZTE Corporation
889 Bibo Road
Shanghai 201203
China

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

