

November 19, 2007

Issues and Approaches to Scaling RBridge Deployments
draft-gray-rbridge-scaling-01.txt

Status of this Memo

By submitting this Internet-Draft, each author represents that any applicable patent or other IPR claims of which he or she is aware have been or will be disclosed, and any of which he or she becomes aware will be disclosed, in accordance with [Section 6 of 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/1id-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 May 19, 2008.

Abstract

RBridges are link layer (L2) devices that use routing protocols as a control plane. They combine several of the benefits of the link layer with network layer routing benefits. RBridges may use existing link state routing (without requiring configuration) to improve RBridge to RBridge aggregate throughput. RBridges also

provide support for IP multicast and IP address resolution optimizations. They are intended to be applicable to similar L2 network sizes as conventional bridges and are intended to be backward compatible with those bridges as both ingress/egress and transit. They also support VLANs (although this generally requires configuration) and otherwise attempt to retain as much 'plug and play' as is already available in existing bridges. There has been a lot of discussion within the TRILL working group about the potential for scaling issues when using IS-IS in combination with (possibly as many as 4K) VLANs. This document is intended to provide information on potential scaling issues and the possible solutions that may be applied in deploying RBridges.

Table of Contents

1	Introduction.....	4
1.1	Terminology.....	4
1.2	Routing Protocol Operation.....	5
1.3	Other Bridging and Ethernet Protocol Operations.....	5
2	Scaling Issues With IS-IS in Combination with VLANs.....	6
2.1	Peering Complexity.....	6
2.2	Designated RBridge Election and Load Splitting.....	6
2.3	Messaging Complexity and Message Length.....	7
3	Security Considerations.....	7
4	IANA Considerations.....	7
5	Acknowledgments.....	7
6	References.....	8
6.1	Normative References.....	8
6.2	Informative References.....	8
	Author's Addresses.....	8
	Intellectual Property Statement.....	9
	Disclaimer of Validity.....	9
	Copyright Statement.....	9
	Acknowledgment.....	10

[1](#) Introduction

This document describes issues with, and potential solutions for, scaling deployments of RBridge standard implementations in combination with standard VLANs.

[1.1](#) Terminology

The following terminology is used, as described in this section, throughout this document.

- o IS-IS: Intermediate System to Intermediate System routing protocol. See [\[2\]](#) for further information on IS-IS.
- o LAN: Local Area Network, is a computer network covering a small geographic area, like a home, office, or group of buildings, e.g., as based on IEEE 802.3 technology, see also IEEE 802.1Q-2005, [Section 3.11](#) [\[3\]](#).
- o Spanning Tree Protocol (STP): an Ethernet (802.1D) protocol for establishing and maintaining a single spanning tree among all the bridges on a local Ethernet segment. Also, Rapid Spanning Tree Protocol (RSTP). In this document, STP and RSTP are considered to be the same.
- o SPF: Shortest Path First - an algorithm name associated with routing, used to determine a shortest path graph traversal.
- o TRILL: Transparent Interconnect over Lots of Links - the working group and working name for the problem domain to be addressed in this document.
- o VLAN: Virtual Local Area Network, see IEEE 802.1Q-2005 [\[3\]](#).
- o Adjacent RBridges: RBridges that communicate directly with each other without relay through other RBridges.
- o Designated RBridge (DR): the RBridge that is elected to handle ingress and egress traffic to a particular Ethernet link having shared access among multiple RBridges; that RBridge is such a link's "Designated RBridge". The Designated RBridge is determined by an election process among those RBridges having shared access via a single LAN.

- o Edge RBridge (edge of a TRILL Campus): describes RBridges that serve to ingress frames into the TRILL Campus and egress frames from the TRILL Campus. L2 frames transiting an TRILL Campus enter, and leave, it via an edge RBridge.
- o Peer RBridge: The term "Peer RBridge", or (where usage is not ambiguous) the term "Peer", are used in the RBridge context to refer to any of the RBridges that make up a TRILL campus.
- o RBridge: a logical device as described in this document, which incorporate both routing and bridging features, thus allowing for the achievement of TRILL Architecture goals. A single RBridge device which can cooperate with other RBridge devices to create a TRILL Campus.
- o TRILL Campus: this term, or the term "Campus" (where usage is not ambiguous) is used in the RBridge context to refer to the set of cooperating RBridges and TRILL Links that connect them to each other.
- o TRILL Link: this term, or the term "Link" (where its usage is not ambiguous) is used in the RBridge context to refer to the Layer 2 connection that exists either between RBridges, or between an RBridge and Ethernet end stations.

1.2 Routing Protocol Operation

The details of routing protocol operation are as specified for IS-IS.

1.3 Other Bridging and Ethernet Protocol Operations

Perhaps the most severe scaling issue associated with RBridge specific behaviors is that relating to interactions with VLANs and - in particular with 802.1Q bridges.

The simplest solution would be to effectively minimize - to the point of disallowing - VLAN configuration, particularly between RBridges.

Unfortunately, this approach cannot be relied on except in the case of a point to point connection between two RBridges. This is because it is relatively easy to show a reasonable network topology involving multiple VLAN (802.1Q) bridges in which using

a single VLAN for control purposes will hide a bridge failure resulting in an undetected partition in the VLAN network.

Even in the point to point case, it is essential that VLAN state information is shared across the point to point link for all VLANs (at least those for which the two associated RBridges are configured to participate in). Several proposals have been discussed and it is very likely that one approach will be to use a compressed vector representation such as has been defined for Multiple VLAN Registration Protocol (MVRP).

2 Scaling Issues With IS-IS in Combination with VLANs

Related to the issue above is the complexity associated with IS-IS peering on a per-VLAN basis. Peer state information needs to be maintained using a refresh-based messaging mechanism. IS-IS peering between IS-IS adjacent peers for possibly as many as 4,094 VLANs effectively multiplies that traffic by the number of VLANs possibly resulting in easily over-whelming slower control plane processing devices.

However, the use of even a compressed vector scheme - such as has been suggested - adds to message size and processing complexity and does nothing much to reducing complexity of VLAN state information, as mentioned in the section above, nor to reducing the complexity associated with forwarding table size.

2.1. Peering Complexity

The details of peering complexity are yet to be determined. Ideally, this section will contain not only analysis of this information, but possibly simulation results - based on a number of scenarios - as well.

2.2. Designated RBridge Election and Load Splitting

One of the possible complications that can result in the need to have visibility into VLAN information relates to the need (on any LAN link where VLAN traffic may flow along different paths or be delivered to different local end stations) to allow for multiple DRB elections - using VLAN IDs as a differentiator. This component of the complexity issues clearly only applies on non point-to-point TRILL Links.

One approach suggested in dealing with this was to use VLAN ID ranges. However, it is trivial to construct a network where this necessarily results in sub-optimal to worst-case DRB selections

for at least some portion of the network. At least one such scenario has been discussed on the mailing list to date.

With minimal additional complexity, the proposal to use (compressed) vector representation(s) - possibly including directly re-using the representation used by MVRP - offers a better approach at marginally higher cost.

However there is one flaw with either of these approaches in that they both propose to use a reduction in messages by combining VLAN specific messages for multiple VLANs and propagating these messages over a reduced subset of the affected VLANs. In fact, in most of the proposals to date a single VLAN is proposed as a control VLAN - even in the case where multiple 802.1Q bridges are known to be present - and this can lead to undetected partitioning of the network.

This could conceivably be detected if all of the 802.1Q bridges also supported use of the MVRP VLAN state vector information, but there is some indication that MVRP may not be commonly deployed in 802.1Q bridges.

2.3. Messaging Complexity and Message Length

The details of messaging complexity are yet to be determined. It is clear that there is a trade-off involved here between many small messages (for the per-VLAN case) verses aa much smaller number of much larger (and/or more complicated) messages.

Ideally, this section will contain not only analysis of this information, but possibly simulation results - based on the same set of scenarios used for peering complexity - as well.

3 Security Considerations

TBD.

4 IANA Considerations

TBD. Should be none.

5 Acknowledgments

TBD.

[6](#) References

[6.1](#) Normative References

TBD.

[6.2](#) Informative References

- [1] 802.1D-2004 IEEE Standard for Local and Metropolitan Area Networks: Media Access Control (MAC) Bridges
- [2] Callon, R., "Use of OSI IS-IS for Routing in TCP/IP and Dual Environments", [RFC 1195](#), December, 1990.
- [3] 802.1Q-2005 IEEE Standard for Local and Metropolitan Area Networks: Virtual Bridged Local Area Networks (Incorporates IEEE Std 802.1Q-1998, IEEE Std 802.1uT-2001, IEEE Std 802.1vT-2001, and IEEE 802.1sT-2002)

Author's Addresses

Editor:

Eric Gray
Ericsson
900 Chelmsford Street
Lowell, MA, 01851
Phone: +1 (978) 275-7470
EMail: Eric.Gray@Ericsson.com

Intellectual Property Statement

The IETF 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 this 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. Information on the procedures with respect to rights in RFC documents can be found in [BCP 78](#) and [BCP 79](#).

Copies of IPR 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 this standard. Please address the information to the IETF at ietf-ipr@ietf.org.

Disclaimer of Validity

This document and the information contained herein 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 HEREIN WILL NOT INFRINGE ANY RIGHTS OR ANY IMPLIED WARRANTIES OF MERCHANTABILITY OR FITNESS FOR A PARTICULAR PURPOSE.

Copyright Statement

Copyright (C) The IETF Trust (2007).

This document is subject to the rights, licenses and restrictions contained in [BCP 78](#), and except as set forth therein, the authors retain all their rights.

Acknowledgment

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