TRILL Routing Requirements in Support of RBridges
draft-ietf-trill-routing-reqs-02.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 July 31, 2007.

Abstract

RBridges provide the ability to have an entire domain, with multiple physical links, look to IP like a single subnet. The design allows for zero configuration of switches within an RBridge domain, optimal pair-wise routing, safe forwarding even during periods of temporary loops, and potential additional advantages as well. The design also supports VLANs, allows forwarding tables to be based on destinations within the RBridge domain (rather than station destinations, allowing internal forwarding tables to be smaller than in conventional bridge systems).
The intent is to re-uses one or more existing routing protocols to accomplish these goals.

This document lays out the background and specific requirements of potential routing protocols to be considered for use in this design. In particular, this document defines what is needed to accommodate this design using IS-IS (or OSPF) as a basis for RBridge protocols.

Table of Contents

1. Introduction......................................................3
   1.1. Terminology................................................4
   1.2. Specific TRILL Goals......................................5
2. General Requirements Potentially Affecting Routing............6
3. Link State Protocol Specific Requirements......................6
4. Potential Issues................................................7
   4.1. Interactions with Spanning Tree Forwarding.................7
   4.2. Computing Routes..........................................8
   4.3. RBridge Interactions with Routing........................8
5. Security Considerations.........................................8
6. Conclusions.....................................................9
7. IANA Considerations............................................9
8. Acknowledgments................................................9
9. References......................................................10
   9.1. Normative References.....................................10
   9.2. Informative References...................................10
10. Author’s Address(es)..........................................10
11. Intellectual Property Statement.............................10
12. Disclaimer of Validity........................................11
13. Copyright Statement...........................................11
14. Acknowledgment................................................11
1. Introduction

The current dominant approach to segregating network traffic relies on a hierarchical arrangement of bridges and routers. Hierarchy is further extended - both within routing protocols (such as IS-IS and OSPF) and between routing protocols (for example, between IGPs and BGP). At least part of the current network structure is based on a determined trade-off between limitations of IP routing and similar disadvantages of 802 bridging.

Bridging Limitations

For example, bridged networks consist of single broadcast/flooding domains. Ethernet/802 encapsulation (on which bridging is based) does not provide mechanisms for reducing the impact of looping data traffic that may result from a transient change in network topology and the existence of multiple paths.

The impact of looping traffic is far worse with flooded or broadcast traffic as this results in exponentially increasing traffic load. Consideration of the impacts of looping data lead to the use of STP/RSTP to establish a connected - loop free - tree by disabling forwarding on a subset of links that might create a loop. This has also the effect of eliminating redundant paths, reducing aggregate throughput and increasing latency through the network.

Because of the potential for severe impact from looping traffic, many (if not most) current bridge implementations stop forwarding of traffic frames following a topology change event and restart only after STP/RSTP is complete.

As a result, the process of eliminating potential loops in existing bridging deployments:

1) Results in inefficient use of inter-bridge connections and
2) May cause delays in forwarding traffic as a result of changes in the network topology.

The combined effect of broadcast/flooding traffic, and the use of spanning trees for loop avoidance, sets practical limits on bridged network size in the network hierarchy and results in inefficient bandwidth use of inter-bridge connections. Inefficient inter-bridge connection usage similarly limits the usefulness of bridging with high-speed (and consequently high cost) interfaces.
IP Routing Issues

For IP routed networks, any link (or subnet) must have at least one unique prefix. This means that a node that moves from one IP subnet to another must change its IP address. Also, nodes with multiple IP subnet attachments must have multiple IP addresses. In IP routed networks, there are frequent trade-off considerations between using smaller subnets (longer prefix length) to minimize wasted IP address space (as a result of unused addresses in the fixed address range defined by the prefix and length) and using larger subnets (shorter prefix length) to minimize the need for (changes in) configuration.

In any case - with current IP routing technology - subnets must be configured for each routed interface.

Proposed Solution

Using a hybrid of routing and bridging - an RBridge - we hope to gain the benefits of both technologies, in particular, gaining the bandwidth advantages of shortest path routing while retaining the simplified configuration associated with bridging.

1.1. Terminology

The following terms are used in this document in the way that they are defined in "TRILL RBridge Architecture" [TARCH]:

ARP (Address Resolution Protocol)
Bridge Learning
Broadcast Domain
Broadcast Traffic
Cooperating RBridges
Egress RBridge
Ethernet
Flooded Traffic
Flooding
Frame
IGP (Interior Gateway Protocol)
Ingress RBridge
Ingress RBridge Tree
IS-IS (Intermediate System to Intermediate System)
ND (Neighbor Discovery)
OSPF (Open Shortest Path First)
RBridge
SPF (Shortest Path First)
Station
STP/RSTP (Spanning Tree Protocol/Rapid Spanning Tree Protocol)
TRILL (TRansport Interconnect over Lots of Links)
Unknown Destination
VLAN (Virtual Local Area Network)
1.2. Specific TRILL Goals

(Near) Zero Configuration

It is a TRILL requirement that it must be possible to deploy RBridges in at least a nominal set of potential deployment scenarios without a need to perform any configuration at each RBridge. It is possible to meet this goal for a sub-set of all possible deployment scenarios by making realistic restrictions on deployment - such as restricting the deployment scenarios to exclude those involving a "trust model" that imposes a need for configuration of some form of "shared secret" or other configuration required to restrict access to "trusted" devices.

It is also conceivable that a minimal configuration may be required for deployment of an initial (set of) device(s) - with subsequently deployed devices deriving that configuration information during the process of - for example - peer discovery. This would constitute a mechanism for "near zero configuration".

Efficient Unicast Bandwidth Usage

For unicast, non-flooded traffic, RBridges are intended to merge the efficient bandwidth use of IP routing with the simplicity of Ethernet (or 802.1) bridging for networks possibly larger - and with greater forwarding capacity - than is the case with these networks presently.

The approach that we will use in accomplishing this is based on the idea of extending (one or more) link state routing protocol(s) to include distribution of Ethernet/802 reachability information between RBridges. In addition, there may be specific requirements imposed on the interactions between these extensions and Spanning Tree Protocol (STP) and between RBridges and (potentially co-located) IP routing instances.

Potentially More Efficient Multicast and Broadcast Usage

There are clear advantages to incorporating mechanisms for improved efficiency in forwarding (layer 2) multicast frames and - possibly in reducing the amount of broadcast traffic as well. To the extent that these efficiency improvements may be considered "optimizations" and may be defined orthogonally to the process of specifying basic RBridge functionality, the potential to include these optimizations is (highly) desirable, but not mandatory.

Examples of this type of optimization include use of any intrinsic multicast routing capabilities and optimizations of ARP/ND.
Backward Compatibility

RBridges must be fully compatible with current bridges, IPv4 and IPv6 routers and stations. They should be invisible to current IP routers (just as bridges are), and like routers, they terminate a bridged spanning tree, (i.e. - they do not forward BPDUs). Unlike Routers, RBridges do not terminate a broadcast domain.

2. General Requirements Potentially Affecting Routing

Candidate IGP Routing protocols - IS-IS or OSPF - must be evaluated for compatibility with the above goals.

For example, since IS-IS requires a unique System ID for each IS-IS instance (at least within a "scoped" deployment), a requirement for "(near) zero configuration" implies a need for mechanisms that allow auto-configuration and/or negotiation of these (scoped) unique IDs.

Similar requirement must apply for OSPF as well, if selected.

In addition, forwarding of protocol messages must be compatible with (or reasonably adaptable to) use of forwarding at layer 2, or there must be a means for deriving suitable higher layer addresses for the purpose of protocol exchanges - without imposing the need to manually configure higher-layer addresses.

3. Link State Protocol Specific Requirements

Assuming that link state routing protocols meet above requirements, running a link state protocol among RBridges is straightforward. It is the same as running a level 1 routing protocol in an area. This would be theoretically true for either IS-IS or OSPF, assuming that both of these meet the general requirements above.

From the perspective of simply extending existing routing protocols, IS-IS is a more appropriate choice than OSPF because it is easy in IS-IS to define new TLVs for use in carrying a new information type. This document, however, does not mandate a specific link-state, IGP, routing protocol. Instead, it sets forth the requirements that will apply to any link-state routing protocol that may be used.

For implementations providing co-located Router/RBridge function, it is necessary to have mechanisms for distinguishing any protocol interactions in Routing instances from protocol interactions in the co-located RBridge instance. Specific mechanisms we will use are very likely to be determined by the Link State Routing Protocol that
we select. Potential distinguishing mechanisms include use of a new well-known Ethernet/802 multicast address, higher-layer protocol ID or other - routing protocol specific - approaches.

The mechanism chosen should be consistent with the TRILL goals. If, for example, a routing protocol specific approach required use of a unique "area" identifier, the RBridge area identifier should be a constant, well-known, value for all RBridges, and would not be one that would ever appear as a real routing area identifier - in order to allow for a potential for configuration-free operation.

Information that RBridge link state information will carry includes:

- layer 2 addresses of nodes within the domain which have transmitted frames but have not transmitted ARP or ND replies
- layer 3, layer 2 addresses of IP nodes within the domain. For data compression, perhaps only the portion of the address following the domain-wide prefix need be carried. This will be more of an issue for IPv6 than for IPv4.
- VLANs directly connected to this RBridge

Station information need only be delivered to RBridges supporting the VLAN in which the station resides. So, for instance, if station E is discovered through a VLAN A frame, then E’s location need only be delivered to other RBridges that are attached to VLAN A links.

Given that RBridges must support delivery only to links within a VLAN (for multicast or unknown frames marked with the VLAN’s tag), this mechanism can be used to advertise station information solely to the RBridges "within" a VLAN (i.e. – having connectivity or configuration that associates them with a VLAN). Although a separate instance of the link state protocol could be run for this purpose, the topology is so restricted (just a single broadcast domain), that it may be preferable to define special case mechanisms whereby each Designated RBridge advertises attached stations, and receives explicit acknowledgments from other RBridges.

4. Potential Issues

4.1. Interactions with Spanning Tree Forwarding and Bridge Learning

Spanning tree forwarding applies within parts of the RBridge domain, where two or more RBridges are connected by a link that includes multiple 802.1 bridges.
In order to simplify the interactions between RBridges and bridges — in particular, relative to spanning tree forwarding — RBridges block BPDUs in spanning tree protocol with attached 802.1 bridges.

Hence, from the Link State Routing protocol perspective, the protocol will be able to treat spanning tree links as indistinguishable from any other Ethernet/802.1 link, in the same way that routing protocols do today.

However, support for multi-pathing is potentially problematic and is assumed — in this document — to be a non-goal. Multi-path forwarding has the potential to confound the bridge learning process.

4.2. Computing Routes

RBridges must calculate an L2 "route table" consisting of Next Hop information for each known L2 unicast destination address within a (possibly VLAN scoped) broadcast domain. This is computed using a routing protocol’s SPF algorithm and based on destination layer 2 address reachability advertisements.

4.3. RBridge Interactions with Routing

The fact that RBridges participate in flooding, and will have other significant differences in forwarding behavior, provides additional reasons to maintain separate routing instances if an RBridge and Router are co-located. Otherwise, interactions between routers and RBridges should be identical to interactions between routers and bridges.

One of the potential interactions to consider in evaluating specific routing protocols is how a specific routing protocol advertises on a shared media. For instance, if a "pseudo-node" is used as a target for such advertisements, how does this interact with the advertising requirements for RBridges?

We expect this to be routing and RBridge protocol specific, and a subject for further study.

5. Security Considerations

The goal is not to add additional security issues over what would be present with traditional bridges and routers. RBridge Interactions with Routers must be defined such that there is no "leaking" of info used in authentication and/or encryption between router and r-bridge instances.
As with routing schemes, authentication of RBridge messages would be a simple addition to protocol (and it could be accomplished the same way as it would be in existing routing protocol). However, any sort of authentication requires additional configuration, which might interfere with the requirement that RBridges need no configuration.

The essential requirement that RBridges do not require configuration provides a forceful argument that most RBridge components are likely to be physically separate (verses logically separate instances within a single physical device) from routers. However, implementers may choose to provide devices with both Routing and RBridge instancing capabilities.

Implementers should consider the differences in trust models implied in Routing and Bridging domains and apply appropriate trust boundary safeguards in addition to instance isolation in general.

6. Conclusions

Routing protocols must be evaluated using the criteria in sections 2 and 3 above, with a clear objective of satisfying the TRILL goals outlined in section 1.2. In addition, specific protocol solutions should use discussion in section 4 above in making a determination as to what approaches TRILL should use, for that (or those) routing protocols that is determined to be useful for RBridge implementation.

Because of the requirement to be able to extend the routing protocol to carry new information, and potentially support new types of peer negotiation, the selected routing protocol must include mechanisms to allow simple routing protocol extensions, new message formats and potentially new types of message exchanges.

For reasons stated in above sections, we believe it is clear that the IS-IS routing protocol may easily be adapted to satisfy TRILL routing protocol requirements.

7. IANA Considerations

There is no action required of IANA by this document.

8. Acknowledgments

Thanks and appreciation are due Radia Perlman and Erik Nordmark for their efforts in reviewing this document. Also appreciated are the review efforts of Joel Halpern and Russ White.
9. References

9.1. Normative References


9.2. Informative References

None.

10. Author’s Addresses

Eric Gray
Ericsson
900 Chelmsford Street,
Lowell, MA - 01851
Email: Eric.Gray@Ericsson.com

11. 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.
12. 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.

13. 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.

14. Acknowledgment

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