

TRILL WG
Internet-Draft
Intended status: Standards Track
Expires: January 31, 2013

Radia. Perlman
Intel Labs
Fangwei. Hu
ZTE Corporation
Donald. Eastlake 3rd
Huawei technology
July 30, 2012

TRILL Cloudlet
draft-perlman-trill-cloudlet-00

Abstract

This draft addresses the problems of nickname exhaustion, size of endnode learning table in access RBs, and the size of the core TRILL network. It does this by creating an invisible level of hierarchy at the edge that we refer to as a "cloudlet".

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 January 31, 2013.

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

TRILL Cloudlet

July 2012

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

Table of Contents

1.	Overview	3
2.	Terminology	3
3.	Smart Endnode	3
4.	Multi-homing	5
5.	Building a Cloudlet	5
6.	Security Considerations	6
7.	Acknowledgements	6
8.	IANA Considerations	6
9.	Normative References	6
	Authors' Addresses	7

[1.](#) Overview

The IETF TRILL (Transparent Interconnection of Lots of Links) protocol implemented by devices called RBridges (Routing Bridges, [\[RFC6325\]](#)), 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([\[RFC1195\]](#)) ([\[RFC6165\]](#)) ([\[RFC6326bis\]](#)) link state routing and encapsulating traffic using a header that includes a hop count. Devices that implement TRILL are called "RBridges" (Routing Bridges) or TRILL Switches.

This draft addresses the problems of nickname exhaustion, size of endnode learning table in access RBs, and the size of the core TRILL network. It does this by creating an invisible level of hierarchy at the edge that we refer to as a "cloudlet".

A cloudlet looks to the core like a collection of endnodes attached to a core access RB. The cloudlet can be a mixture of endnodes, hypervisors, switches, and simple RBs. A cloudlet will not consume nicknames, nor introduce LSPs into the core. In this draft we will build the concept from the simplest case; a cloudlet consisting of a single endnode, and expand the idea in subsequent sections.

[2.](#) Terminology

This document uses the acronyms defined in([\[RFC6325\]](#)) and the following phrase:

Smart end node: An end node that can do TRILL encapsulation/decapsulation.

Simple RBridge: An edge RBridge that assign its nickname to the smart end node attached. It should response the querying from its attached

smart end node.

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 ([\[RFC2119\]](#)).

[3.](#) Smart Endnode

Suppose endnode E is attached to RBridge R. E signals to R that E wishes to be a "smart endnode". Let's take the figure 1 as the example.

E-----R

figure1

E, will do TRILL encapsulation/decapsulation and endnode learning of endnodes it is in communication with, freeing its attached RB, R, from learning the addresses of nodes corresponding with E. When E encapsulates, E uses R's nickname as "ingress" in the TRILL header.

Although this draft is explaining the concept rather than exact details of packet formats, a logical way for E to signal R that E wishes to act as a smart endnode is by having E issue a TRILL-Hello, with a flag indicating that E wants to be a "smart endnode", and perhaps a flag indicating that E wishes to receive ESADI. A logical way for R to respond is by including E in its neighbor list in R's TRILL-Hello, with a flag indicating that R hears E's Hello, but acknowledges that E is a smart endnode rather than a true RB neighbor. R also includes R's nickname in its TRILL-Hello.

The endnode E does not issue LSPs, nor does it receive LSPs. It does the following:

- o It keeps an endnode table of (MAC, nickname) of end nodes with which the smart end node is communicating. Entries in this table are populated the same way that an edge RBridge populates the entries in its table:
 - * learning from (source, ingress) on packets it decapsulates

- * from ESADI([TRILL-ESADI](#)), TRILL ESADI is an end station system distribution protocol, which can be used to distribute the (MAC,nickname) entry.
- * by querying a directory,Smart end node E can get the (MAC, nickname) entry by querying a directory. The RBridges which the smart end nodes are attached act as directory server. Both the push and pull model are supported in this document([Directory](#)). In push model, the RBridge pushes down the MAC and egress nickname mapping for the hosts which might communicate with hosts attached to the RBridge. In pull model,smart end node sends a pull request to the RBridge if it has no the entry of (MAC, nickname) of the destination end node.
- * by having some entries configured

The Simple RBridge R does the following:

- o It marks the port to smart end node E as being "leave encapsulated"
- o For attached access ports, simple RBridge R keeps, for each such port, a set of MAC addresses of end nodes on those ports. For each of those MAC addresses, R keeps a new flag indicating whether that MAC address is a "smart endnode".
- o When receiving a packet with R's nickname as egress, if the destination MAC address in the enclosed packet is listed as "smart endnode", R leaves the packet encapsulated when forwarding to E.

[4.](#) Multi-homing

Now suppose E is attached to the TRILL campus in two places; to RBridges R1 and R2.

There are two ways for this to work:

- (1) E can choose either R1 or R2's nickname, when encapsulating a frame, whether the encapsulated frame is sent via R1 or R2.

Most likely, E would always choose the same one, unless that one was unreachable, and then E would switch to the other.

- (2) R1 and R2 might indicate, in their Hello, another nickname that attached end nodes may use if they are multihomed to R1 and R2, separate from R1 and R2's nicknames (which they would also list in their Hello). This would be useful if there were many end nodes multihomed to the same set of RBridges. This would be analogous to a pseudonode nickname; return traffic would go via the shortest path from the source to the endnode, whether it is R1 or R2. If E loses connectivity to R2, then E would revert to using R1's nickname. This does use a nickname, but hopefully would be shared by many end nodes multihomed to the same set of RBridges.

(When end node E loses connectivity to one of the RBridges, how to detect the connectionless and how to switch to another RBridge will be defined later in this document.)

[5.](#) Building a Cloudlet

Another level of functionality might be to build reasonably large cloudlets, with multiple hops of ordinary spanning tree bridges, and/or "simple RBridges". A "simple RBridge" is one that only is aware of the topology of the cloudlet. In other words, the cloudlet

operates like a lower level of routing. Inside the TRILL core, forwarding is based on nicknames. Inside the cloudlet, forwarding is based on MAC addresses of the end nodes in the cloudlet, although the packet being forwarded may be encapsulated with a TRILL header. If all end nodes in the cloudlet are smart end nodes, then packets inside the cloudlet would always be encapsulated. If there is a "normal" endnode N, in the cloudlet, then N would issue unencapsulated packets, and the first simple RBridge R1, would encapsulate it, and R1 would maintain the endnode cache entries for end nodes communicating with N. Likewise, when R1 is forwarding to N, R1 would decapsulate the packet.

RBridges within the cloudlet have to know whether the packet belongs in the cloudlet or the TRILL campus. This is done based on the "egress nickname" in the encapsulated packet. If the egress nickname

is the nickname to be used by the cloudlet, then the packet is forwarded only within the cloudlet. If the egress nickname is not one used by the cloudlet, the packet is forwarded to the "true RBridge" that attaches the cloudlet to the TRILL campus. If the cloudlet is multihomed to R1 and R2, say, and the "ingress nickname" indicates R1's nickname, then the cloudlet forwards the encapsulated packet towards R1. If the cloudlet is multihomed and is using a pseudonode nickname, then the cloudlet forwards to whichever of R1 or R2 is closer.

6. Security Considerations

For general TRILL Security Considerations, see([\[RFC6325\]](#)).

7. Acknowledgements

TBD

8. IANA Considerations

9. Normative References

[Directory]

Linda, D., Eastlake, D., Perlman, R., and I. Gashinsky, "TRILL Edge Directory Assistance Framework", *raft-ietf-trill-directory-framework-00* (work in process), July 2012.

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

Perlman, et al.

Expires January 31, 2013

[Page 6]

Internet-Draft

TRILL Cloudlet

July 2012

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

[RFC6165] Banerjee, A. and D. Ward, "Extensions to IS-IS for Layer-2 Systems", [RFC 6165](#), April 2011.

[RFC6325] Perlman, R., Eastlake, D., Dutt, D., Gai, S., and A. Ghanwani, "Routing Bridges (RBridges): Base Protocol

Specification", [RFC 6325](#), July 2011.

[RFC6326bis]

Eastlake, D., Banerjee, A., Dutt, D., Ghanwani, A., and R. Perlman, "Transparent Interconnection of Lots of Links (TRILL) Use of IS-IS", [draft-eastlake-isis-rfc6326bis-07.txt](#), work in process, March 2012.

[TRILL-ESADI]

Zhai, H., Hu, F., Perlman, R., and D. Eastlake, "TRILL (Transparent Interconnection of Lots of Links): The ESADI (End Station Address Distribution Information) Protocol", [draft-ietf-trill-esadi-00.txt](#) (work in process), June 2012.

Authors' Addresses

Radia Perlman
Intel Labs
2200 Mission College Blvd.
Santa Clara, CA 95054-1549
USA

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

Fangwei Hu
ZTE Corporation
No.889 Bibo Rd
Shanghai, 201203
China

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

Huawei technology
155 Beaver Street
Milford, MA 01757
USA

Phone: +1-508-634-2066
Email: d3e3e3@gmail.com