INTERNET-DRAFT
Intended Status: Proposed Standard

Mingui Zhang
Donald Eastlake
Huawei
Radia Perlman
EMC

Expires: September 10, 2015

March 9, 2015

Single Area Border RBridge Nickname for TRILL Multilevel
draft-zhang-trill-multilevel-single-nickname-00.txt

Abstract

A major issue in multilevel TRILL is how to manage RBridge nicknames. In this document, the area border RBridge uses a single nickname in both Level 1 and Level 2. RBridges in Level 2 must obtain unique nicknames but RBridges in different Level 1 areas may have the same nicknames.

Status of this Memo

This Internet-Draft is submitted to IETF 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/1id-abstracts.html

The list of Internet-Draft Shadow Directories can be accessed at http://www.ietf.org/shadow.html

Copyright and License Notice

Table of Contents

1. Introduction

TRILL multilevel techniques are designed to improve TRILL scalability issues. As described in [MultiL], there have been two proposed approaches. One approach, which is referred as the "unique nickname" approach, gives unique nicknames to all the TRILL switches in the multilevel campus, either by having the Level-1/Level-2 border TRILL switches advertise which nicknames are not available for assignment in the area, or by partitioning the 16-bit nickname into an "area" field and a "nickname inside the area" field. The other approach, which is referred as the "aggregated nickname" approach, involves assigning nicknames to the areas, and allowing nicknames to be reused in different areas, by having the border TRILL switches rewrite the nickname fields when entering or leaving an area.

The approach specified in this document is different from both "unique nickname" and "aggregated nickname" approach. In this document, the nickname of an area border RBridge is used in both Level 1 (L1) and Level 2 (L2). No additional nicknames are assigned

to the L1 areas. Each L1 area is denoted by the group of all nicknames of those border RBridges of the area. For this approach, nicknames in L2 MUST be unique but nicknames inside different L1 areas MAY be reused.

2. Acronyms and Terminology

2.1. Acronyms

Data Label: VLAN or FGL

IS-IS: Intermediate System to Intermediate System [ISIS]

2.2. Terminology

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].

Familiarity with [RFC6325] is assumed in this document.

3. Nickname Handling on Border RBridges

This section provides an illustrative example and description of the border learning border RBridge nicknames.

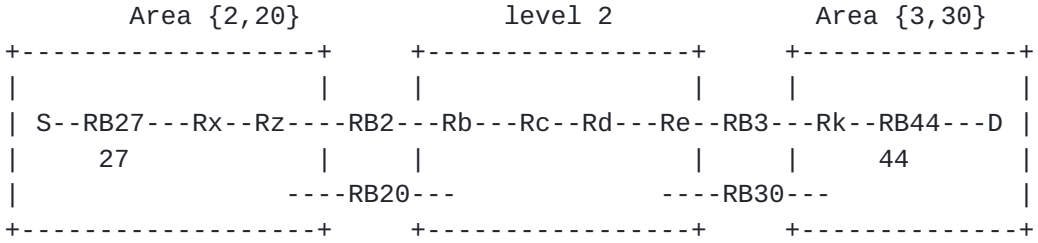


Figure 3.1: An example topology for TRILL multilevel

In Figure 3.1, RB2, RB20, RB3 and RB30 are area border TRILL switches (RBridges). Their nicknames are 2, 20, 3 and 30 respectively. Area border RBridges use the set of border nicknames to denote the L1 area that they are attached to. For example, RB2 and RB20 use nicknames {2,20} to denote the L1 area on the left.

A source S is attached to RB27 and a destination D is attached to RB44. RB27 has a nickname, say 27, and RB44 has a nickname, say 44 (and in fact, they could even have the same nickname, since the TRILL switch nickname will not be visible outside these Level 1 areas).

3.1. Actions on Unicast Packets

Let's say that S transmits a frame to destination D and let's say that D's location is learned by the relevant TRILL switches already. These relevant switches have learned the following:

1) RB27 has learned that D is connected to nickname 3.
2) RB3 has learned that D is attached to nickname 44.

The following sequence of events will occur:

- S transmits an Ethernet frame with source MAC = S and destination MAC = D.

- RB27 encapsulates with a TRILL header with ingress RBridge = 27, and egress RBridge = 3 producing a TRILL Data packet.

- RB2 and RB20 have announced in the Level 1 IS-IS instance in area {2,20}, that they are attached to all those area nicknames, including {3,30}. Therefore, IS-IS routes the packet to RB2 (or RB20, if RB20 on the least-cost route from RB27 to RB3).

- RB2, when transitioning the packet from Level 1 to Level 2, replaces the ingress TRILL switch nickname with its own nickname, so replaces 27 with 2. Within Level 2, the ingress RBridge field in the TRILL header will therefore be 2, and the egress RBridge field will be 3. Also RB2 learns that S is attached to nickname 27 in area {2,20} to accommodate return traffic. RB2 SHOULD synchronize with RB20 that MAC = S is attached to nickname 27.

- The packet is forwarded through Level 2, to RB3, which has advertised, in Level 2, its L2 nickname as 3.

- RB3, when forwarding into area {3,30}, replaces the egress nickname in the TRILL header with RB44's nickname (44). The ingress nickname MAY be replaced with an area nickname selected from {2,20}. See Section 4 for the detail of the selection method. Suppose nickname 2 is selected. So, within the destination area, the ingress nickname will be 2 and the egress nickname will be 44.

- RB44, when decapsulating, learns that S is attached to nickname 2, which is one of the area nicknames of the ingress.

3.2. Actions on Multi-Destination Packets

Now suppose that D's location has not been learned by RB27 and/or RB3. What will happen, as it would in TRILL today, is that RB27 will forward the packet as multi-destination, choosing a tree. As the multi-destination packet transitions into Level 2, RB2 replaces the ingress nickname with its own nickname for the area. If RB27 does not

know the location of D, the packet must be flooded, subject to possible pruning, in Level 2 and, subject to possible pruning, from Level 2 into every Level 1 area that it reaches on the Level 2 distribution tree. There may be multiple eligible border RBridges for this area to transit the multi-destination packets from Level 2 to a Level 1. It's important that these area border RBridges agree on an election method to determine who is the Designated Boarder RBridge (DBRB) for the transition, otherwise RBridges in this area will see packet duplication. It's RECOMMNEDED that the pseudorandom algorithm as defined in Section 5.3 of [RFC7357] is used as the election method.

Now suppose that RB27 has learned the location of D (attached to nickname 3), but RB3 does not know where D is. In that case, RB3 must turn the packet into a multi-destination packet within area {3,30}. In this case, care must be taken so that, another border TRILL switch in that area not forward the now multi-destination packet back into Level 2. Therefore, it would be desirable to have a marking, somehow, that indicates the scope of this packet's distribution to be "only this area" (see also Section 4 of [MultiL]).

4. Per-flow Load Balancing

When a packet from other areas arrives at an area border RBridge, this RBridge MAY select one area nickname of the ingress to replace the ingress nickname of the packet. The selection is simply based on a pseudorandom algorithm as defined in Section 5.3 of [RFC7357]. With the random ingress nickname replacement, the border RBridge actually achieves a per-flow load balance for returning traffic.

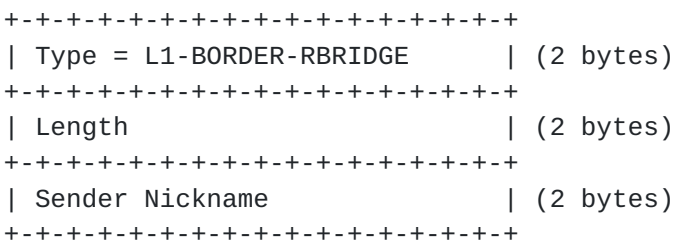
All area border RBridges in an L1 area MUST agree on the same pseudorandom algorithm. The source MAC address, ingress area nicknames, egress area nicknames and the Data Label are candidate factors of the input of this pseudorandom algorithm. Note that the value of the destination MAC address SHOULD be excluded from the input of this pseudorandom algorithm, otherwise the egress RBridge will see one source MAC address flip flopping among multiple ingress RBridges.

When a packet originated from an area arrives at the area border RBridge, this RBridge MAY select one area nickname of the egress to replace the egress nickname of the packet. By default, it SHOULD choose the egress area border RBridge with the least cost route to reach. The pseudorandom algorithm as defined in Section 5.3 of [RFC7357] may be used as well. In that case, however, the ingress area border RBridge may take the non-least-cost Level 2 route to forward the TRILL data packet to the egress area border RBridge.

5. Protocol Extensions for Discovery

5.1. Discovery of Border RBridges in L1

The following Level 1 Border RBridge APPsub-TLV will be included in an E-L1FS FS-LSP fragment zero [RFC7180bis] as an APPsub-TLV of the TRILL GENINFO-TLV. Through listening to this Appsub-TLV, an area border RBridge discovers all other area border RBridges in this area.



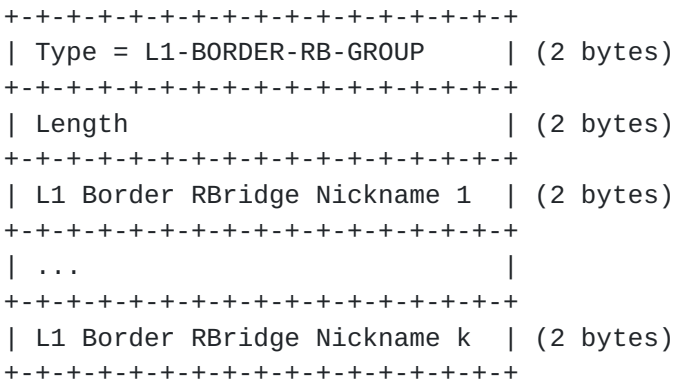
o Type: Level 1 Border RBridge (TRILL APPsub-TLV type tbd1)

o Length: 2

o Sender Nickname: The nickname the originating IS will use as the L1 Border RBridge nickname. This field is useful because the originating IS might own multiple nicknames.

5.2. Discovery of Border RBridge Sets in L2

The following APPsub-TLV will be included in an E-L2FS FS-LSP fragment zero [RFC7180bis] as an APPsub-TLV of the TRILL GENINFO-TLV. Through listening to this APPsub-TLV in L2, an area border RBridge discovers all groups of L1 border RBridges and each such group identifies an area.



o Type: Level 1 Border RBridge Group (TRILL APPsub-TLV type tbd2)

o Length: 2*k. If length is not a multiple of 2, the APPsub-TLV is

corrupt and MUST be ignored.

- o L1 Border RBridge Nickname: The nickname that an area border RBridge uses as the L1 Border RBridge nickname. The L1-BORDER-RB-GROUP TLV generated by an area border RBridge MUST include all L1 Border RBridge nicknames of the area. It's RECOMMENDED that these k nicknames are ordered in ascending order according to the 2-octet nickname considered as an unsigned integer.

6. E-L1FS/E-L2FS Backwards Compatibility

All Level 2 RBridges MUST support E-L2FS [RFC7356] [rfc7180bis]. The Extended TLVs defined in Section 5 are to be used in Extended Level 1/2 Flooding Scope (E-L1FS/E-L2FS) PDUs. Area border RBridges MUST support both E-L1FS and E-L2FS. RBridges that do not support either E-L1FS or E-L2FS cannot serve as area border RBridges but they can well appear in an L1 area acting as non-area-border RBridges.

7. Security Considerations

For general TRILL Security Considerations, see [RFC6325].

The newly defined TRILL APPsub-TLVs in Section 5 are transported in IS-IS PDUs whose authenticity can be enforced using regular IS-IS security mechanism [ISIS][RFC5310]. This document raises no new security issues for IS-IS.

8. IANA Considerations

8.1. TRILL APPsub-TLVs

IANA is requested to allocate two new types under the TRILL GENINFO TLV [RFC7357] for the TRILL APPsub-TLVs defined in Section 5. The following entries are added to the "TRILL APPsub-TLV Types under IS-IS TLV 251 Application Identifier 1" Registry on the TRILL Parameters IANA web page.

Type	Name	Reference
tbd1[256]	L1-BORDER-RBRIDGE	[This document]
tbd2[257]	L1-BORDER-RB-GROUP	[This document]

9. References

9.1. Normative References

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

[RFC6325] Perlman, R., Eastlake 3rd, D., Dutt, D., Gai, S., and A. Ghanwani, "Routing Bridges (RBridges): Base Protocol Specification", RFC 6325, July 2011.

[RFC7356] L. Ginsberg, S. Previdi, et al, "IS-IS Flooding Scope LSPs", RFC 7356, June 2014.

[RFC7357] Zhai, H., Hu, F., Perlman, R., Eastlake 3rd, D., and O. Stokes, "Transparent Interconnection of Lots of Links (TRILL): End Station Address Distribution Information (ESADI) Protocol", RFC 7357, September 2014.

9.2. Informative References

[ISIS] ISO, "Intermediate system to Intermediate system routeing information exchange protocol for use in conjunction with the Protocol for providing the Connectionless-mode Network Service (ISO 8473)", ISO/IEC 10589:2002.

[RFC5310] Bhatia, M., Manral, V., Li, T., Atkinson, R., White, R., and M. Fanto, "IS-IS Generic Cryptographic Authentication", RFC 5310, February 2009.

[RFC7180bis] D. Eastlake, M. Zhang, et al, "TRILL: Clarifications, Corrections, and Updates", draft-eastlake-trill-rfc7180bis, work in progress.

[MultiL] Perlman, R., Eastlake, D., et al, "Flexible Multilevel TRILL", draft-perlman-trill-rbridge-multilevel, work in progress.

Author's Addresses

Mingui Zhang
Huawei Technologies
No.156 Beiqing Rd. Haidian District,
Beijing 100095 P.R. China

EMail: zhangmingui@huawei.com

Donald E. Eastlake, 3rd
Huawei Technologies
155 Beaver Street
Milford, MA 01757 USA

Phone: +1-508-333-2270
EMail: d3e3e3@gmail.com

Radia Perlman
EMC
2010 256th Avenue NE, #200
Bellevue, WA 98007 USA

EMail: radia@alum.mit.edu