TRILL

Weiguo Hao
Yizhou Li
Zhibo Hu
Huawei

Internet Draft
Intended status: Standards Track
Expires: August 2014

February 14, 2014

Frame Duplication Avoidance For TRILL Active-Active Access
draft-hao-trill-dup-avoidance-active-active-01.txt

Status of this Memo

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 August 14, 2009.

Copyright Notice

Abstract

The problems in active-active connection to TRILL campus was introduced in [TRILL-Active-PS]. Frame duplication was a well recognized issue in that scenario. This draft proposes a Designated Forwarder (DF) mechanism to solve the issue of frame duplications. The basic idea is to elect one RBridge per VLAN from an edge group to be responsible for egressing the multicast frames. An edge group is a group of edge RBridges that a CE is multi-homed to in active-active mode. TRILL protocol extension for DF election is described in this draft.

Table of Contents

1. Introduction 3
2. Conventions used in this document..... 5
3. DF election mechanism overview..... 5
4. DF election algorithm per VLAN..... 6
5. Link and node failure process..... 7
6. TRILL protocol extension..... 7
   6.1. The LAGID TLV 7
7. Security Considerations..... 8
8. IANA Considerations 8
9. References 8
   9.1. Normative References..... 8
   9.2. Informative References..... 8
10. Acknowledgments 8

1. Introduction

The IETF TRILL (Transparent Interconnection of Lots of Links) [RFC6325] protocol provides loop free and per hop based multipath data forwarding with minimum configuration. TRILL uses IS-IS [RFC6165] [RFC6326bis] as its control plane routing protocol and defines a TRILL specific header for user data.

Customer edge(CE) devices typically are multi-homed to several RBridges which form an edge group. All of the uplinks of CE is considered as an Multi-Chassis Link Aggregation (MC-LAG) bundle. An active-active flow-based load-sharing mechanism is implemented to achieve better load balancing and high reliability. A CE device can be a layer3 end system by itself or a bridge switch through which layer3 end systems are accessed to TRILL campus.

Draft [TRILL-Active-PS] lists basic problems which any active-active solutions should address:

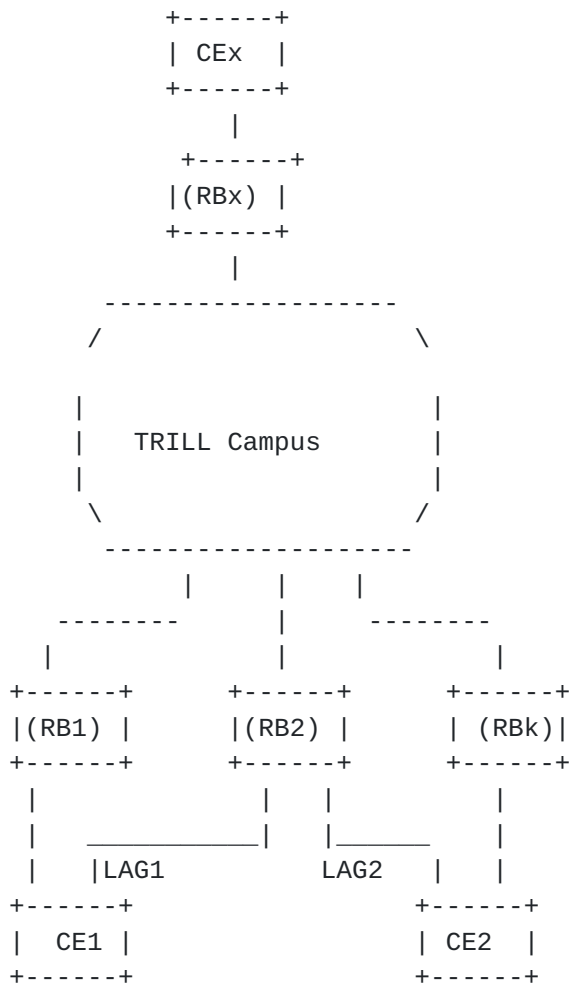


Figure 1 TRILL Active-Active Access Scenario

1. Frame duplications

2. Loop

3. Address flip-flop

4. Unsynchronized information among member RBridges

This document proposes a Designated Forwarder (DF) mechanism to solve the issue of frame duplications. Frame duplication may occur when a remote host sends multi-destination frame to a local CE which has active-active connection to the TRILL campus. The basic idea of DF is to elect one RBridge per VLAN from an edge group to be responsible for egressing the multicast traffic. An edge group is the group of edge RBridges that a CE is multi-homed to in active-

active mode. TRILL protocol extension can be easily extended to support DF election mechanism.

2. Conventions used in this document

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].

This document uses the same terminology as found in the active-active problem statement draft [TRILL-Active-PS]. Some of the terms defined in the that document have been repeated in this section for the convenience of the reader, along with additional terminology that is used by this document.

CE - customer equipment. Could be a bridge or end station.

FGL -Fine Grained Label.

Edge group - a group of edge RBs to which one CE is multiply Attached, a edge group corresponds to a MC-LAG. One RB can be in more than one edge group.

3. DF election mechanism overview

Each CE is multi-homed to multiple edge RBs in active-active mode through an MC-LAG. These edge RBs form one edge group. An edge group corresponds to an MC-LAG. A RB may join multiple edge groups to provide active-active access for multiple CE devices.

Each MC-LAG should be assigned a globally unique LAGID in network as identification. When an edge RB, say RB1, has any downlink port running in MC-LAG mode, the following DF election procedures are executed.

1. RB1 acquires the unique LAGID for that MC-LAG through manual or automatic provisioning.

2. RB1 then starts a timer (default value = 10 seconds) to allow the reception of LAGID TLV LSP from other RB nodes of same edge group. This timer value MUST be same across all RBs of same edge group.

3. RB1 announces self LAGID in TRILL LSP to TRILL campus before local timer expires.

4. All RBs that have same LAGID provisioned as in RB1's announcement discover each other and form an edge group.

5. DF election process starts when the timer expires.

6. Each RB in an edge group elects a DF using same algorithm which guarantees the same RB elected as DF per MC-LAG per VLAN. The RB that is elected as a DF for a given VLAN will unblocks multi-destination traffic in the egress direction towards the CE. All non-DF PEs continue to drop multi-destination traffic in the egress direction towards the CE.

All edge RBs, including DF and non-DF, can ingress the traffic to TRILL campus as usual. On the other hand, only the elected DF RBridge can egress the multi-destination frame from TRILL campus to local access network. It ensures no duplicated frame received by local CE when a remote RB sends multicast traffic.

4. DF election algorithm per VLAN

Assuming there are k RBs in an edge group, DF election algorithm per VLAN is as follows:

1. All RBs in an edge group are ordered and numbered from 0 to k-1 in ascending order according to the 7-octet IS-IS ID.

2. For VLAN ID m, choose RB whose number equals (m mod k) as DF.

FGL are primarily intended for >4K tenants use in large Data Centers to provide traffic isolation among each tenants. When a CE is multi-homed to TRILL campus through FGL in active-active mode, DF election should be considered per FGL. If there are some specific multicast groups receivers connecting to TRILL campus, DF election should be considered per multicast group.

If VLANs and multicast group addresses are statically configured on edge RBridge access port, the operators will ensure configuration consistency in an edge group. The algorithm always give consistent calculation result during DF election.

If VLANs is dynamically enabled through VLAN registration protocol (VRP) (GVRP or MVRP), VLANs enabled for each MC-LAG should be either provisioned on all the corresponding ports on RBs of the same edge group or synchronized among those RBs. Similarly, the multicast groups dynamically joined through IGMP or MLD protocol by a RBridge port should follow the same logic. The detailed synchronization mechanism is beyond the scope of this draft.

In a steady state, all VLANs and multicast group addresses are consistent among all edge RBs in an edge group. Therefore the DF election algorithm gives the consistency calculation result for each MC-LAG and each VLAN(or FGL/multicast group).

5. Link and node failure process

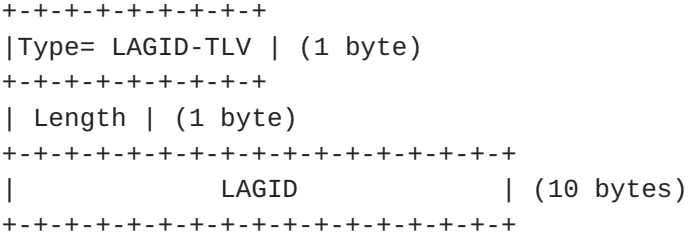
When an edge RB, say RB1, has all downlink ports in a particular MC-LAG failed, RB1 should announce a new LSP with the remaining alive LAGIDs. When other RBs in the same edge group as RB1 receive the LSP, these RBs think RB1 has left the edge group and will trigger the DF re-election process.

If RB1 suffers a node failure, after other RBs in the same edge group know this information because RB1 becomes unreachable in the core IS-IS link state data base (all links to RB1 will appear to be down), other RBs will re-run the DF election process.

6. TRILL protocol extension

The LAGID TLV is introduced to support global LAGID announcement. It is carried in an LSP PDU. Edge RBs rely on the LAGID TLV to automatically form edge group corresponding to each MC-LAG. DF election of the edge group is followed by such auto-discovery.

6.1. The LAGID TLV



o Type: TLV Type, TBD.

o Length: indicates the length of LAGID field, it is a fixed value of ten.

o LAGID: the global LAGID corresponding to a MC-LAG.

7. Security Considerations

This draft does not introduce any extra security risks. For general TRILL Security Considerations, see [RFC6325].

8. IANA Considerations

TBD

9. References

9.1. Normative References

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

[2] [RFC6325] Perlman, R., et.al. "RBridge: Base Protocol Specification", RFC 6325, July 2011.

[3] [RFC6326bis] Eastlake, D., Banerjee, A., Dutt, D., Perlman, R., and A. Ghanwani, "TRILL Use of IS-IS", draft-eastlake-isis-rfc6326bis, work in progress.

9.2. Informative References

[4] [TRILL-Active-PS] Li,Y., et.al., "Problems of Active-Active connection at the TRILL Edge", draft-yizhou-trill-active-active-connection-prob-02, Work in progress, July

2014.

10. Acknowledgments

The authors wish to acknowledge the important contributions of Donald Eastlake, Junlin Zhang.

Authors' Addresses

Weiguo Hao
Huawei Technologies
101 Software Avenue,
Nanjing 210012
China

Phone: +86-25-56623144
Email: haoweiguo@huawei.com

Yizhou Li
Huawei Technologies
101 Software Avenue,
Nanjing 210012
China

Phone: +86-25-56625375
Email: liyizhou@huawei.com

Zhibo Hu
Huawei Technologies
156 Beiqing Road,
Beijing 100095
China

Phone: +86-010-60613388
Email: huzhibo@huawei.com