L2VPN
Internet Draft
Intended status: Standards Track
Expires: December 2013

Weiguo Hao
Yizhou Li
Pei Xu
Huawei

June 14, 2013

Multi-homed network in EVPN
draft-hao-evpn-mhn-00.txt

Status of this Memo

This Internet-Draft will expire on December 14, 2013.

Copyright Notice

Abstract

To enhance the reliability, bridged network is normally multi-homed to an EVPN network, there are two categories of mechanisms to avoid the layer 2 traffic loop. The first category does not require the PEs participating in the control protocol of the bridged network, while the second category requires that. [EVPN] described one of the first category mechanisms called designated forwarder (DF) election to achieve loop avoidance and vlan-based load balancing. This draft mainly focuses on the second category of mechanisms which can achieve intra-vlan MAC-based load balancing. MAC-based VLAN balancing is more applicable than DF election mechanism if all end stations in bridged network are on the same VLAN which can cause traffic congestion in DF

link.

Table of Contents

1. Introduction

[EVPN] introduces a solution for multipoint L2VPN services. In EVPN networks, MAC learning between PEs is not via the data plane(different from what happens in traditional bridging network) but via the control plane using multi-protocol (MP) BGP.

To enhance the reliability, the PE nodes need offer multi-homed connectivity to a CE or access Network, i.e, both multi-homed device (MHD) as well as multi-homed network (MHN) scenarios in [EVPN-REQ] should be covered by E-VPN solution. In MHN scenario, the multi-homed Ethernet network would typically run a resiliency mechanism such as Multiple Spanning Tree Protocol [802.1Q] or Ethernet Ring Protection Switching [G.8032]. For example, EVPN can be used for Data Center(DC) interconnection to provide LAN extension for each DC site and each site is an MSTP networks. Normally each site should be multi-homed to multiple EVPN PEs to ensure the reliability.

As defined in [EVPN-REQ], the following solutions should be provided

for MHN scenario:

A solution MUST support multi-homed network connectivity with active/standby redundancy.

A solution MUST also support multi-homed network with active/active VLAN-based load balancing (i.e. disjoint VLAN sets active on disparate PEs).

A solution MAY support VLAN-based load balancing among PEs that are member of a redundancy group spanning multiple ASes.

A solution MAY support multi-homed network with active/active MAC-based load balancing (i.e. different MAC addresses on a VLAN are reachable via different PEs).

The former three requirements can be addressed through designated forwarder (DF) election mechanism as described in [EVPN], a brief review of DF election mechanism will be given in section 3.

This draft will mainly focus on a new mechanism to achieve active/active MAC-based load balancing to fulfil the fourth requirement. The details of the solution will be illustrated in section 4. Protocol extensions of EVPN for this mechanism will be given in section 5.

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 terminologies defined in [RFC6325] along with the following:

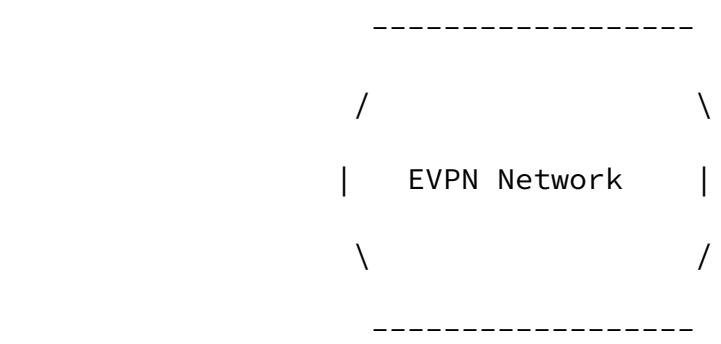
EVPN: Ethernet virtual private network.

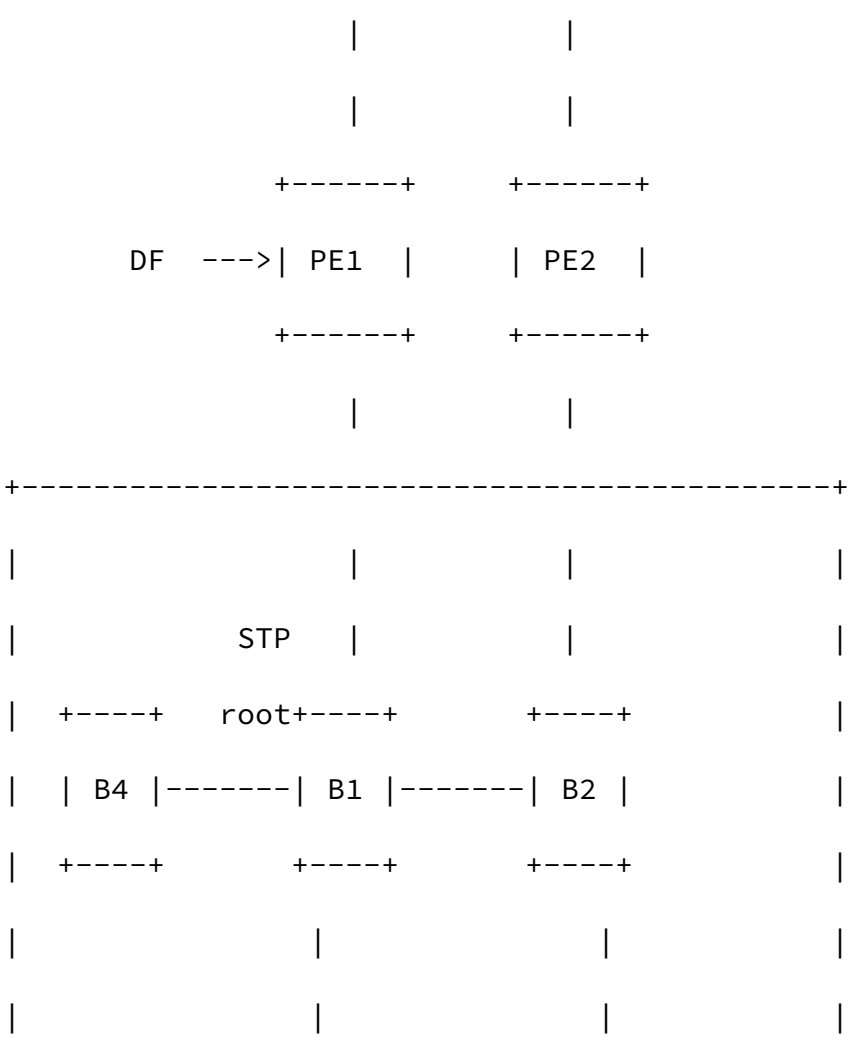
G.8032: Ethernet ring protection switching.

NVO3: Network virtualization over layer3.

STP: Spanning Tree Protocol.

3. Recap on Designated Forwarder (DF) election mechanism





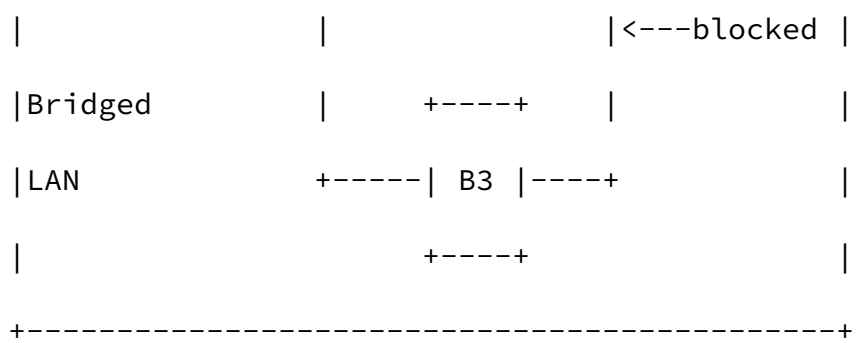


Figure 1 DF election mechanism

As described in [EVPN], designated forwarder(DF) mechanism is required for loop avoidance. Only one of the links between the switched bridged network and the PEs is active for a given Ethernet tag, as shown by Figure 1. This mechanism does not require the PEs to participate in the control protocol of the bridged network. Bridges in the local bridged network runs normal Multiple Spanning Tree Protocol [802.1Q] or Ethernet Ring Protection Switching [G.8032].

Through this method VLAN-based load balancing among PEs can be achieved. All end systems of one VLAN can access the EVPN network through only one PE.

In this case, the Ethernet A-D route per Ethernet segment MUST be advertised with the "Active-Standby" flag set to one. Only one PE is elected as DF for each EVI(E-VPN Instance). Only DF is responsible for sending multicast, broadcast and unknown unicast traffic, on a given Ethernet tag to the bridged network. In order to perform better traffic load-balancing within a given segment, multiple DFs per Ethernet segment can be elected and each PE is the DF for a disjoint set of EVIs. An EVI is an E-VPN routing and forwarding instance on a PE and consists of one or more broadcast domains which is identified by an Ethernet Tag which are assigned to the broadcast domains of a given E-VPN instance by the provider of that E-VPN. The information about an Ethernet Tag on a particular Ethernet segment is advertised using an "Ethernet Auto-Discovery route(Ethernet A-D route)". In the case of a multi-homed CE, this route MUST carry the "ESI Label Extended Community" to enable split horizon. Also, the route can be used for Designated Forwarder (DF) election and MAY be used to optimize the withdrawal of MAC addresses upon failure.

For fast convergence case, upon a failure in connectivity to the

attached segment, the PE withdraws the corresponding Ethernet A-D route. This triggers all PEs that receive the withdrawal to update their next-hop adjacencies for all MAC addresses associated with the Ethernet segment in question. If there is any other PE advertising an Ethernet A-D route for the same segment, the PE updates the next-hop adjacencies to point to this backup PE(s).

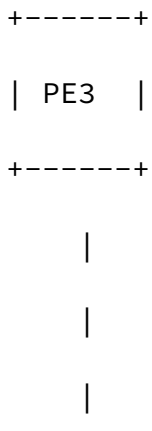
With DF mechanism, native frames enter and leave bridged network via the same designated forwarder for a given VLAN. It may cause congestion or suboptimal routing. PE and bridges should be carefully configured so that end stations on a remnant bridged LAN are separated into different VLANs that have different designated forwarders to achieve better load balancing.

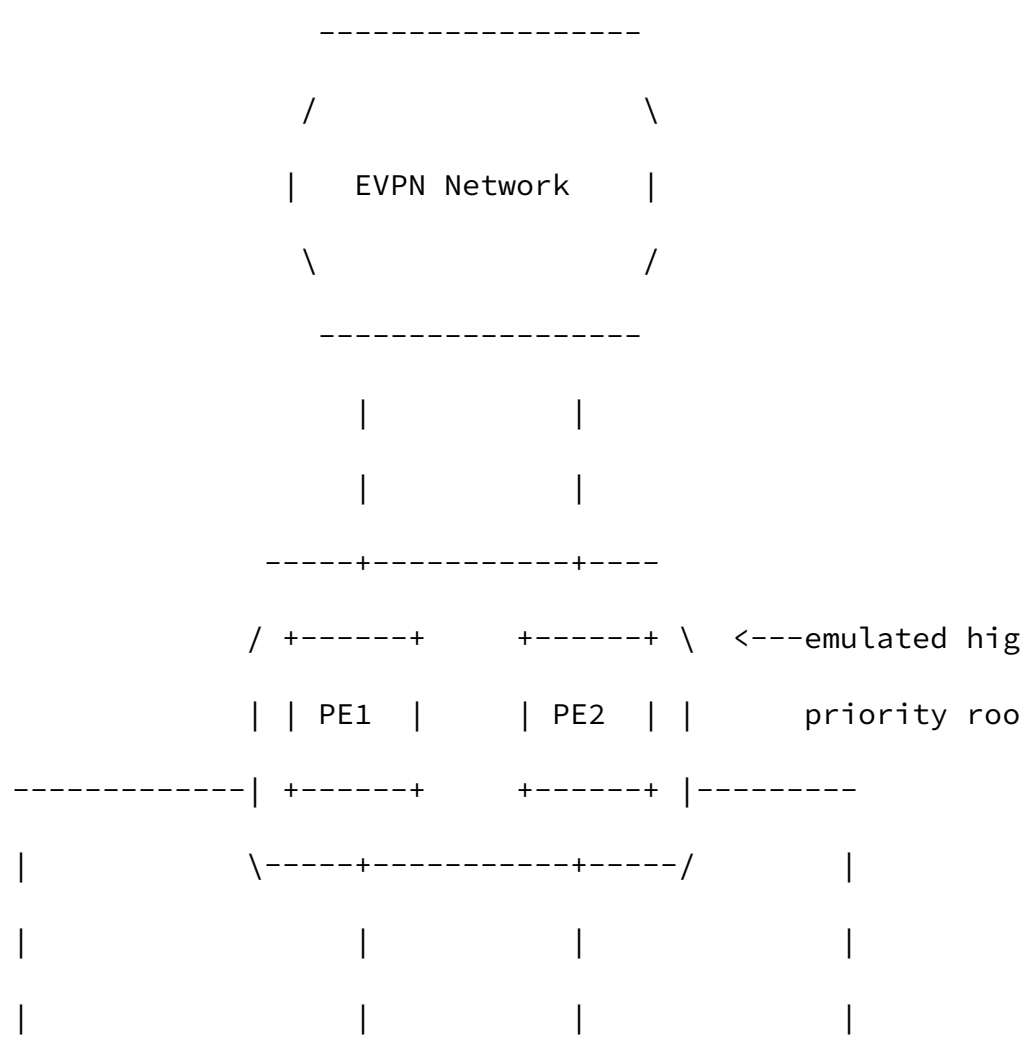
4. Active/Active MAC-based load balancing mechanism

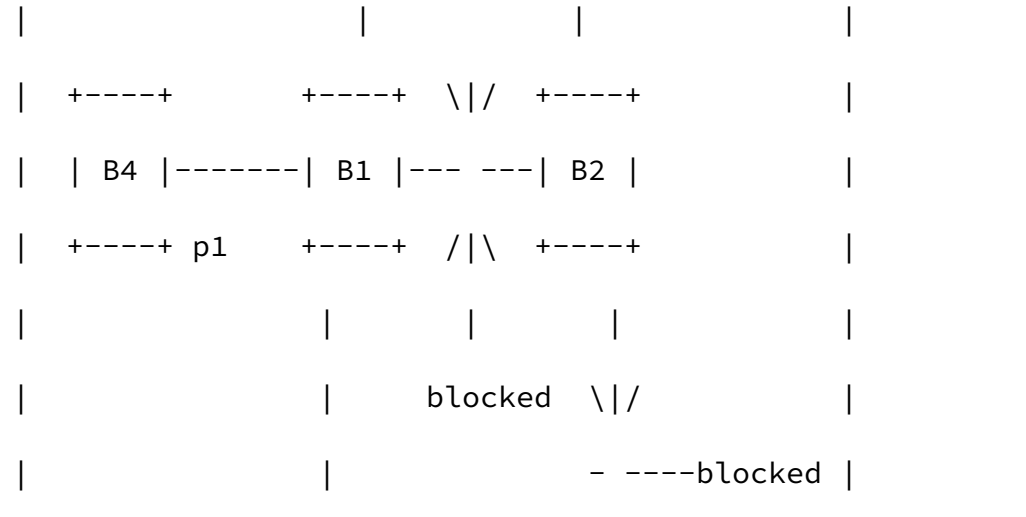
Active/Active MAC-based load balancing mechanism requires the PEs to participate in the control plane protocol of the bridged network. With this mechanism, loop avoidance and per-vlan MAC-based load balancing can be achieved. So it can achieve better load balancing than DF election, and is more applicable if all end stations in bridged network on the same VLAN may cause traffic congestion over the link to DF.

The following two solutions can be used to achieve active/active MAC-based load balancing. One is emulated MSTP root bridge solution; another is bridge control plane protocol tunneling solution. We will described them in the following subsections respectively.

4.1. Emulated MSTP root bridge solution




EVPN Network
PE1
PE2
<---emulated hig
priority roo


B4
B1
B2
p1
blocked
blocked

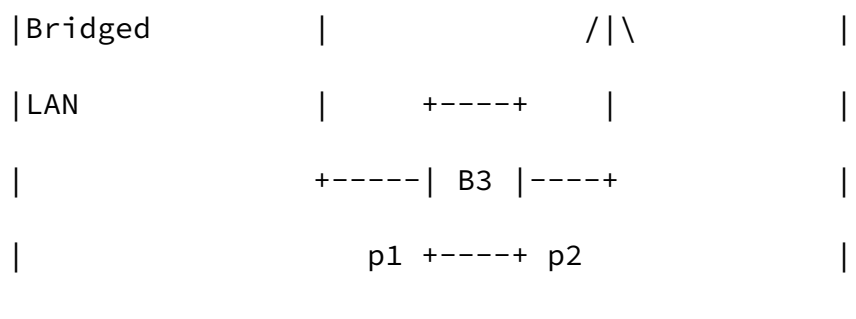


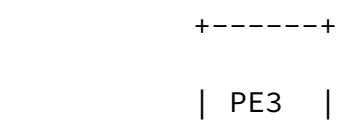
Figure 2 emulated MSTP root bridge solution

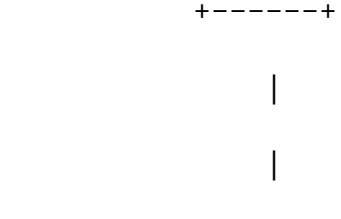
PE1 and PE2 act as an emulated MSTP root bridge. PE1 & PE2 use the same bridge ID to emit spanning tree BPDUs as the highest priority root Bx. All bridges in bridged network see PE1 and PE2 as single tree root. Therefore B1-B2 and B2-B3 links are blocked for loop avoidance by the spanning tree protocol.

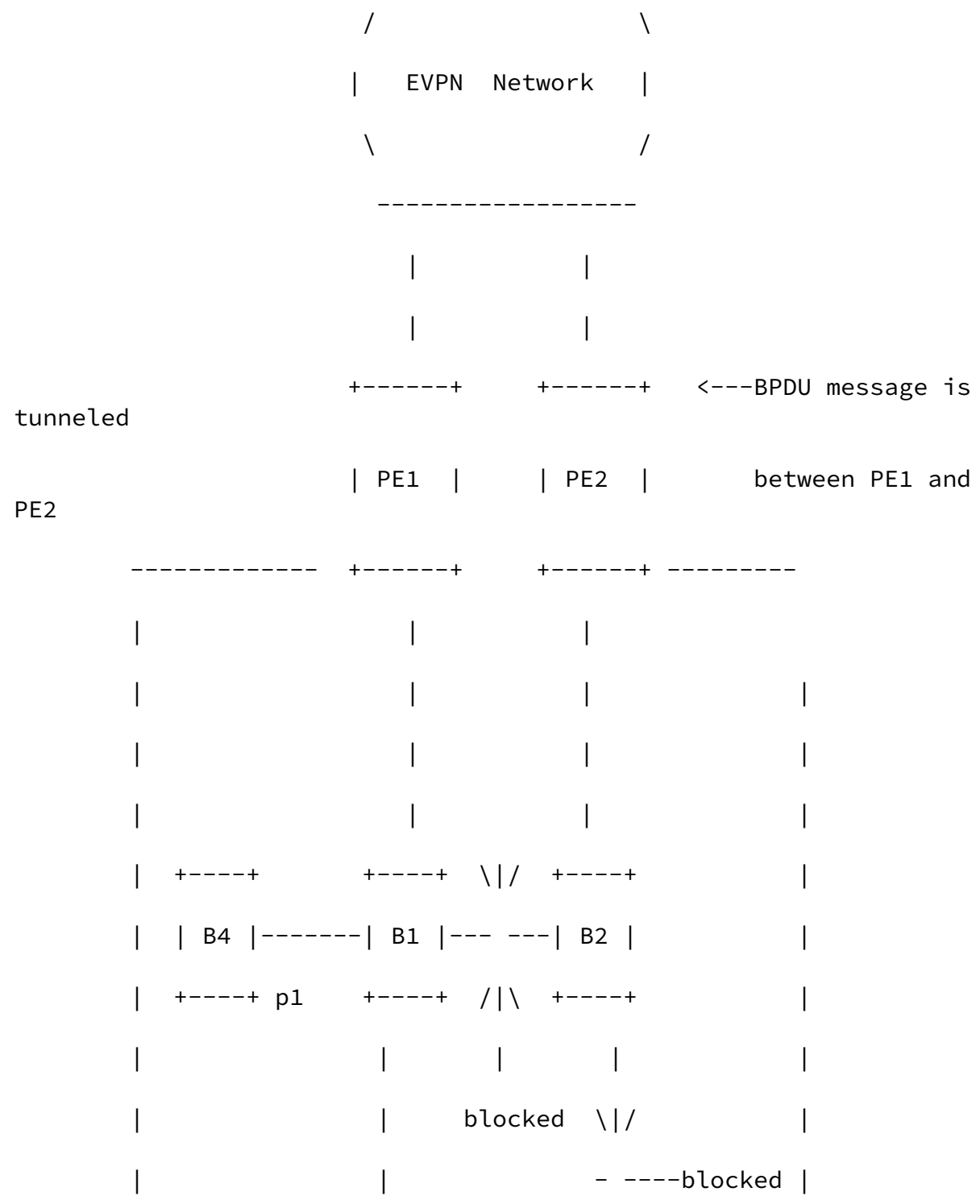
When B1-B3 link fails, alternate port p2 on B3 will start to send TC BPDU and go to forwarding state. PE2 receives TC BPDU from B2 sequentially. PE2 tunnel the TC BPDU to PE1. At the same time, PE2 notifies remote PE3 to flush the MAC table through corresponding Ethernet A-D route.

With this solution, PE1 and PE2 needs to tunnel TC BPDU to each other when topology change occurs in the local bridged network.

4.2. Bridge control plane protocol tunneling solution






EVPN Network
<---BPDU message is tunneled between PE1 and PE2
PE1
PE2
B4
B1
B2
p1
blocked
blocked

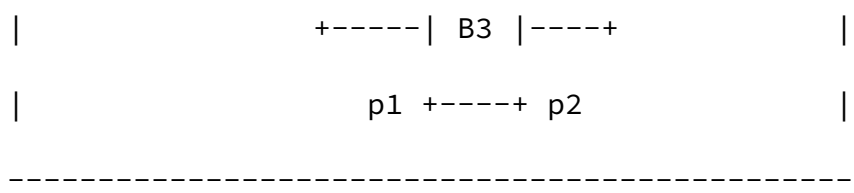


Figure 3 PE1 and PE2 act as normal MSTP bridge nodes

The solution described in the previous section is applicable for STP/MSTP domain. Now we are going to present another solution which can be used for both MSTP and G.8032 domain. The basic idea is to tunnel the control plane messages of local domain among the multi-homed PEs over EVPN network.

4.2.1. Scenario 1: Local bridged network is MSTP

PE1 and PE2 act as normal MSTP bridge nodes. MSTP root bridge can be PE or any switch in the bridged network. BPDU message can be sent through tunnel over EVPN network between PE1 and PE2. The tunnel can be MPLS P2P LSP, MPLS P2MP LSP, or NVO3 tunnel, etc. PE1 and PE2 regard the BPDU tunnel as normal physical link. To avoid BPDU tunnel blocked by MSTP, link cost of the tunnel should be set to 0 or minimum value in MSTP network. With such configuration, it is expected that the blocked port by MSTP protocol can never be the EVPN network facing port on PEs.

4.2.2. Scenario 2: Local bridged network is G.8032

Similarly, PE1 and PE2 act as normal G.8032 ring nodes. They support standard FDB MAC learning, forwarding, flush behavior and port blocking/unblocking mechanisms. G.8032 message can be sent through tunnel over EVPN network between PE1 and PE2. ring protection link(RPL) owner node can be PE or any switch in bridged network. If PE is RPL owner node, RPL can only be configured on access link and can never be configured on the EVPN network facing port on PEs.

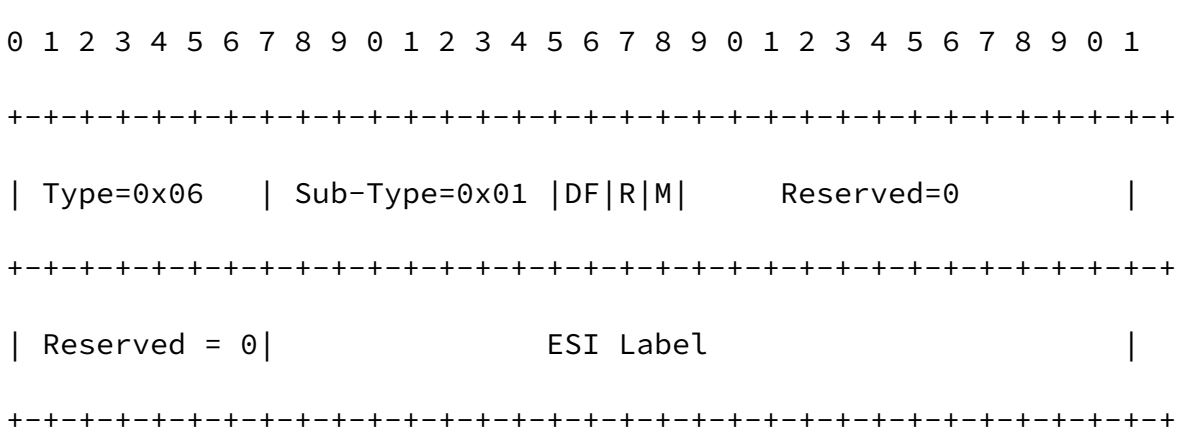
4.2.3. Fast convergence

For fast convergence, when a PE notice a topology change event, it should flush local MAC entries and notify the remote PE of the same EVPN instance to withdraw the corresponding Ethernet A-D route. The remote PE that received the withdrawal simply invalidates the MAC entries for that segment.

5. EVPN protocol extension

ESI Label Extended Community MUST be included in EVPN Ethernet A-D route. All-Active multi-homing or active-standby multi-homing mode is decided by the "Active-Standby" bit in the flags of the ESI Label Extended Community through DF mechanism.

ESI Label Extended Community should be extended to support the mechanisms illustrated in this document. "M" bit is introduced to indicate multi-homing mode of MAC-based all active without DF Election. DF selection procedures should be skipped if "M" bit is set to be 1. When remote PE receives Ethernet A-D route withdraw message, it simply invalidates the MAC entries for the segment that corresponding to the Ethernet A-D route.



DF: As defined in [EVPN]. It should be ignored if M bit is 1.

R: The bit is already defined as the "Root-Leaf" in [EVPN].

M: The bit is defined as "MAC-based all active without DF Election" and may be set to 1. The above "DF" bit is significant only when "M" bit is set to 0. A value of 1 for M bit means that multi-homed site uses MAC-based active-active access.

6. Security Considerations

TBD

7. IANA Considerations

TBD

7.1. Normative References

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

7.2. Informative References

[1] [EVPN-REQ] A. Sajassi, R. Aggarwal et. al., "Requirements for Ethernet VPN", draft-ietf-l2vpn-evpn-req-01.txt.

[2] [EVPN] Sajassi et al., "BGP MPLS Based Ethernet VPN", draft-ietf-l2vpn-evpn-00.txt, work in progress, February, 2012.

8. Acknowledgments

The authors wish to acknowledge the important contributions of Shunwan Zhuang.

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

Pei Xu
Huawei Technologies
101 Software Avenue,
Nanjing 210012
China
Phone: +86-25-56623590
Email: xupei@huawei.com