Networking Working Group
Internet-Draft
Intended status: Informational
Expires: September 10, 2012

Z. Liu
China Telecom
L. Jin
R. Chen
ZTE Corporation

March 9, 2012

Node redundancy provisioning for VPLS Inter-domain draft-liu-l2vpn-vpls-inter-domain-redundancy-02

Abstract

In many VPLS deployment based on [RFC4762], inter-domain has been deployed without node redundancy, or only with node redundancy in one domain. This document describes how to deploy inter-domain VPLS based on [RFC4762] with node redundancy in both domain. The draft reuses the existing protocols without introducing any new protocols.

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 September 10, 2012.

Table of Contents

1. Introduction

In many VPLS deployment based on [RFC 4762], inter-domain has been deployed without node redundancy, or only with node redundancy in one domain. This document describes how to deploy inter-domain VPLS based on [RFC 4762] with node redundancy in both domain. The draft reuses the existing protocols without introducing any new protocols. The domain in this document refers to AS, or other administrative domain.

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 RFC2119.

3. Motivation

Inter-AS VPLS has now been wildly deployed between two providers. Usually, the physical link and ASBR between the two providers would carry many kinds of service, then it is important to provider link and node redundancy for such kind of inter-AS service to ensure high availability.

Some current high availability deployments of inter-AS VPLS are provided by MC-LAG (Multi-Chassis Link Aggregation) and [I-D.ietf-pwe3-iccp], but there is a pre-condition that the interconnected link between the two providers are Ethernet link. There are also many interconnection cases between two providers to use POS (Packet over Sonet/SDH) link on which MC-LAG cannot be enabled. Moreover, it is also required for the VPLS between two providers to ensure bandwidth control, QoS, MAC address control and Broadcast/Multicast traffic control. Then from the technical point of view, it is necessary to use PW to interconnect the two VPLS in its corresponding providers, and also to provide link/node redundancy to ensure high availability.

4. Redundancy scenario with ICCP

The following figure presents a typical inter-AS VPLS deployment topology. PE3 and PE4 are the VPLS edge nodes in network of operator A, and PE5 and PE6 are the VPLS edge nodes in network of operator B. The PE3/PE4/PE5/PE6 may be ASBR of the AS, or VPLS PE within its own AS.

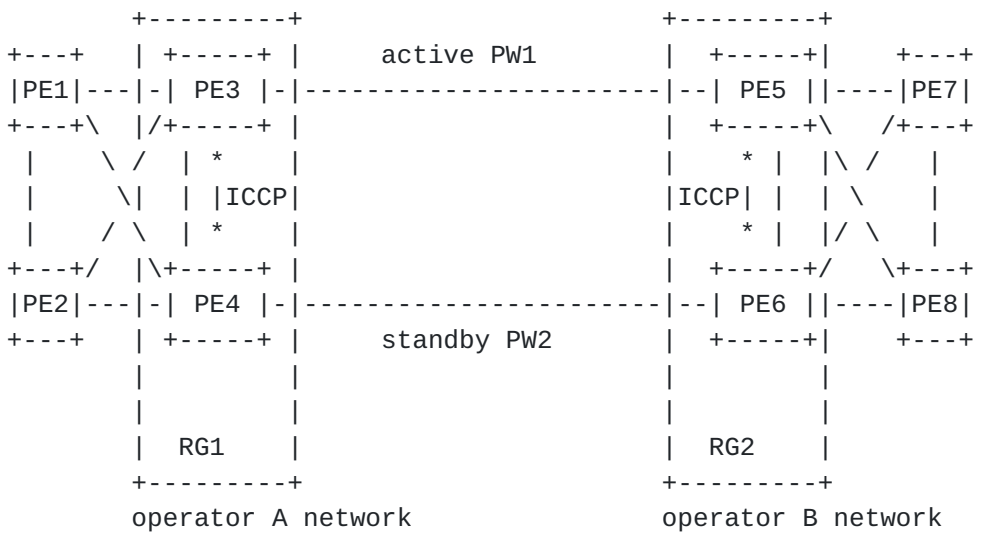

Figure 1

When inter-AS VPLS is deployed with node redundancy on both AS side, node redundancy protocol ICCP[I-D.ietf-pwe3-iccp]SHOULD be implemented on the VPLS edge nodes of the AS, e.g, ICCP should be running between PE3 and PE4, PE5 and PE6.

There are several deployment scenarios for inter-domain VPLS:

- ICCP deployment option: ICCP is deployed on VPLS edge nodes in one domain, or in both domain;
- PW redundancy mode: independent or master/slave;

From the operator's point of view, it is important to keep the technical balance and technical independence between the two operators. One operator will not highly rely on the other operator's technical choice for inter-domain VPLS node redundancy. Then it is highly recommended to be the deployment scenario as follows:

- ICCP deployment option: ICCP is deployed on VPLS edge nodes in both domain;
- PW redundancy mode: independent only;

And this draft will only focus on the above deployment option, other options are out of the scope.

5. Node redundancy for VPLS Inter-domain

The PEs in the RG are required to run an inter-chassis communication protocol ([I-D.ietf-pwe3-iccp]) in order to select which pseudowire(s) should be in active/standby state for a given VPLS service instance.

The procedures to select active/standby pseudowire(s):

- o The PEs in the RG enable ICCP[I-D.ietf-pwe3-iccp].
- o The PEs should establish a PW-RED application connection using the mechanism described in [I-D.ietf-pwe3-iccp], section 9.1.1.
- o When the PW-RED application connection first comes up, Each PE MUST advertise it local PW configuration to other PEs that are members of the same RG. As part of the configuration information, the PE should advertise a PW priority value that is used to determine the precedence of a given pseudowire.
- o Pseudowire Status Synchronization. In order to synchronize pseudowire state, "PW-RED State TLV" is send whenever the pseudowire state changes on a PE. The PE MAY re-advertise its PW-RED state in an unsolicited/solicited manner, the detailed mechanism is described in [I-D.ietf-pwe3-iccp], section 9.1.3.

The PEs SHOULD then use PW redundancy bit [I-D.ietf-pwe3-redundancy-bit] or basic PW status bit [RFC4447] to advertise the outcome of the arbitration to the peer PE(s).

Before deploying inter-domain VPLS, the operator MUST negotiate to configure same PW priority at two end-points. If different PW priority value is configured at the two PW end-points, e.g, PE3 and PE5 for PW1, and PE4 and PE6 for PW2 in figure 1, it is possible to select PE3 and PE6 as active for the two domain, then both PW1 and PW2 will be standby according to the independent mode in [I-D.ietf-pwe3-redundancy-bit].

6. MAC Withdraw procedure in VPLS Inter-domain

It MAY be desirable to remove or unlearn MAC addresses that have been dynamically learned for faster convergence. This is accomplished by sending an LDP Address Withdraw Message. PE SHOULD not advertise MAC Address Withdraw message from one domain to the other.

Correspondingly, VPLS PE that connects another domain SHOULD also reject any MAC Address Withdraw message received from that domain.

In figure 1, when the active PW failure is detected by the PE3, it will trigger MAC Address Withdraw message into the full mesh. By default, as per the processing rules defined in [RFC4762], upon PE4 activates the standby PW, it will also send a MAC Address Withdraw message. There would be two copies of MAC Address Withdraw message received by each PE, which would make the network convergence worse.

What's more, there are two MAC withdraw capabilities defined, positive MAC withdraw (Flush-all-but-mine, defined in [RFC4762] and [I-D.ietf-l2vpn-vpls-ldp-mac-opt]), and negative MAC withdraw (Flush-

all-from-me,defined in [I-D.ietf-l2vpn-vpls-ldp-mac-opt]). If the two PE support only positive MAC withdraw, then PE4 is required to send MAC withdraw message when PW switching. While if two PE support both positive & negative MAC withdraw or support only negative MAC withdraw capability, then PE3 is required to send MAC withdraw message when PE switching.

In order to determine which PE to send MAC withdraw message is most appropriate, we introduce an MAC withdraw notification TLV in ICCP PW-RED application to negotiate the MAC withdraw capability.

6.1. MAC withdraw notification TLV format

The MAC withdraw notification TLV is describe as below:

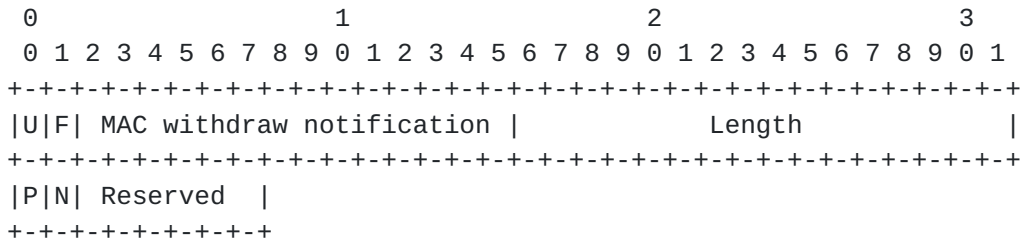

Figure 2

U-bit: Unknown bit. This bit SHOULD be set to 1 (ignore if not understood)

F-bit: Forward bit. This bit SHOULD be set to 0 (do not forward if not understood)

MAC withdraw notification TLV: It is requested in the IANA allocation.

P-bit: Used to indicate whether the node supports the positive (N=0, Flush-all-but-mine) MAC withdraw capability. P=1 indicates that the node has the capability to send the positive MAC withdraw message.

N-bit: Used to indicate whether the node supports the negative (N=1, Flush-all-from-me) MAC withdraw capability. N=1 indicates that the node has the capability to send the negative MAC withdraw message.

Reserved: Reserved for future use.

MAC withdraw notification TLV is advertised to a LDP ICCP peer if there is at least one RG enabled on the local PE, and this TLV should be carried in "RG Application Data" message.

6.2. Optimized MAC Withdraw processing

After receiving the MAC withdraw capability through MAC withdraw notification TLV, the PE should process as below:

- If the former active PE & standby PE support only the positive MAC withdraw capability, then former standby PE will trigger MAC withdraw with positive MAC withdraw message [I-D.ietf-l2vpn-vpls-ldp-mac-opt] to other PEs that in the same AS when active PW failures.
- If the former active PE & standby PE support only negative MAC withdraw capability or support both positive & negative MAC withdraw, then former active PE sends MAC Address withdraw message with negative MAC withdraw message [I-D.ietf-l2vpn-vpls-ldp-mac-opt] to other PEs that in the same AS when active PW failures.
- The former standby PE may send MAC Address withdraw message with positive MAC withdraw message [I-D.ietf-l2vpn-vpls-ldp-mac-opt] to other PEs that in the same AS when the active PE failures.

7. Load Balancing

It is recommended to configure different PW priority values for different VPLS instance, then the active PW of different VPLS will be running on different PEs, to provide load balancing between the two PE in one domain.

8. Security Considerations

This section will be added in a future version.

9. IANA Consideration

This document creates a new "ICC RG parameter type" (MAC withdraw notification TLV) that is allocated by IANA, and a value of 0x0020 is suggested for assignment with this TLV.

10. Normative references

[I-D.ietf-l2vpn-vpls-ldp-mac-opt]
Dutta, P., Balus, F., Stokes, O., and G. Calvignac, "LDP

Extensions for Optimized MAC Address Withdrawal in H-VPLS", draft-ietf-l2vpn-vpls-ldp-mac-opt-05 (work in progress), October 2011.

[I-D.ietf-pwe3-iccp]
Martini, L., Salam, S., Sajassi, A., Bocci, M., Matsushima, S., and T. Nadeau, "Inter-Chassis Communication Protocol for L2VPN PE Redundancy", draft-ietf-pwe3-iccp-05 (work in progress), April 2011.

[I-D.ietf-pwe3-redundancy]
Muley, P., Aissaoui, M., and M. Bocci, "Pseudowire Redundancy", draft-ietf-pwe3-redundancy-06 (work in progress), February 2012.

[I-D.ietf-pwe3-redundancy-bit]
Muley, P. and M. Aissaoui, "Pseudowire Preferential Forwarding Status Bit", draft-ietf-pwe3-redundancy-bit-06 (work in progress), February 2012.

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

[RFC4447] Martini, L., Rosen, E., El-Aawar, N., Smith, T., and G. Heron, "Pseudowire Setup and Maintenance Using the Label Distribution Protocol (LDP)", RFC 4447, April 2006.

[RFC4762] Lasserre, M. and V. Kompella, "Virtual Private LAN Service (VPLS) Using Label Distribution Protocol (LDP) Signaling", RFC 4762, January 2007.

[RFC5036] Andersson, L., Minei, I., and B. Thomas, "LDP Specification", RFC 5036, October 2007.

Authors' Addresses

Zhihua Liu
China Telecom
109 Zhongshan Ave.
Guangzhou 510630
P.R.China

Email: zhliu@gsta.com

Lizhong Jin
ZTE Corporation
889 Bibo Road
Shanghai 201203
P.R.China

Email: lizhong.jin@zte.com.cn

Ran Chen
ZTE Corporation
68 Zijinghua Road
Nanjing 210012
P.R.China

Email: chen.ran@zte.com.cn