Networking Working Group
Internet-Draft
Intended status: Standards Track
Expires: October 12, 2014

Z. Liu
China Telecom
L. Jin

R. Chen
ZTE Corporation
D. Cai
S. Salam
Cisco

April 10, 2014

Redundancy provisioning for VPLS Inter-domain
draft-ietf-l2vpn-vpls-inter-domain-redundancy-05

Abstract

In many existing Virtual Private LAN Service (VPLS) deployments based on RFC 4762, inter-domain connectivity has been deployed without node redundancy, or with node redundancy in a single domain. This document describes a solution for inter-domain VPLS based on RFC 4762 with node and link redundancy in both domains.

Status of this Memo

Copyright Notice

Table of Contents

1. Introduction

In many existing Virtual Private LAN Service (VPLS) deployments based on [RFC4762], inter-domain connectivity has been deployed without node redundancy, or with node redundancy in a single domain. This document is to provide a service protection mechanism for inter-domain VPLS based on [RFC4762]. The protection mechanism will provide edge node redundancy and link redundancy in both domains. The domain in this document refers to autonomous system (AS), or other administrative domains.

The solution relies on the use of ICCP [I-D.ietf-pwe3-iccp] to coordinate between redundant edge nodes, and use of Pseudowire (PW) Preferential Forwarding Status Bit [RFC6870] to negotiate the PW status. There is no change to any protocol message formats and no new protocol options introduced. This solution is a description of reusing existing protocol building blocks to achieve the desired function, but also defines implementation behaviour necessary for the function to work.

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 offerings are widely deployed in service provider networks today. Typically, the ASBRs and associated physical links that connect the domains carry a multitude of services. As such, it is important to provide link and node redundancy, to ensure service high availability and meet end customer service level agreements(SLAs).

Several current deployments of inter-AS VPLS are implemented like Inter-AS option A in [RFC4364] section 10, where VLANs are used to hand-off the services between two domains. In these deployments, link/node redundancy is achieved using MC-LAG (Multi-Chassis Link Aggregation) and [I-D.ietf-pwe3-iccp]. This, however, places two restrictions on the interconnection: the two domains must be interconnected using Ethernet links, and the links must be homogeneous, i.e. of the same speed, in order to be aggregated. These two conditions cannot always be guaranteed in live deployments. For instance, there are many scenarios where the interconnect between the domains uses Packet over Sonet/SDH (POS), thereby ruling out the

applicability of MC-LAG as a redundancy mechanism. As such, from a technical point of view, it is desirable to use PWs to interconnect the VPLS domains, and to offer resiliency using PW redundancy mechanisms.

MP-BGP can be used for VPLS inter-domain protection, as described in [RFC6074], using either option B or option C inter-AS models. However, with this solution, the protection time relies on BGP control plane convergence. In certain deployments, with tight SLA requirements on availability, this mechanism may not provide the desired failover time characteristics. Furthermore, in certain situations MP-BGP is not deployed for VPLS. The redundancy solution described in this document reuses ICCP [I-D.ietf-pwe3-iccp] and PW redundancy [RFC6718] to provide fast convergence.

Furthermore, in the case where Label Switched Multicast is not used for VPLS multicast [RFC7117], the solution described here provides a better behavior compared to inter-AS option B: with option B, each Provider Edge (PE) must perform ingress replication to all other PEs in its local as well as the remote domain. Whereas, with the ICCP solution, the PE only replicates to local PEs and to the Autonomous System Border Router (ASBR). The ASBR then sends traffic P2P to the remote ASBR, and the remote ASBR replicates to its local PEs. As a result, the load of replication is distributed and is more efficient than option B.

Two PW redundancy modes defined in [RFC6718], namely independent mode and master/slave mode, are applicable in this solution. In order to maintain control plane separation between two domains, the independent mode is preferred by operators. The master/slave mode provides some enhanced capabilities and, hence, is included in this document.

4. Network Use Case

There are two network use cases for VPLS inter-domain redundancy: two-PWs redundancy case, and four-PWs redundancy case.

Figure 1 presents an example use case with two inter-domain PWs. PE3/PE4/PE5/PE6 may be ASBRs of their respective AS, or VPLS PEs within its own AS. PE3 and PE4 belong to one redundancy group (RG), and PE5 and PE6 belong to another RG. A deployment example of this use case is where there are only two physical links between two domains and PE3 is physically connected with PE5, and PE4 is physically connected with PE6.

```
                 +---------+                 +---------+
         +---+   | +-----+ |   active PW1    |  +-----+|    +---+
         |PE1|---|-| PE3 |-|-----------------|--| PE5 ||----|PE7|
         +---+\  |/+-----+ |                 |  +-----+\   /+---+
          |    \ /  | *    |                 |    * |  |\ /   |
          |     \|  | |ICCP|                 |ICCP| |  | \    |
          |    / \  | *    |                 |    * |  |/ \   |
         +---+/  |\+-----+ |                 |  +-----+/   \+---+
         |PE2|---|-| PE4 |-|-----------------|--| PE6 ||----|PE8|
         +---+   | +-----+ |   standby PW2   |  +-----+|    +---+
                 |         |                 |         |
                 |         |                 |         |
                 |  RG1    |                 |  RG2    |
                 +---------+                 +---------+
                 operator A network             operator B network
```

Figure 1

Figure 2 presents a four-PWs inter-domain VPLS redundancy use-case. PE3/PE4/PE5/PE6 may be ASBRs of their respective AS, or VPLS PEs within its own AS. A deployment example of this use case is where there are four physical links between two domains and four PEs are physically connected with each other with four links.

```
                  +---------+                 +---------+
          +---+   | +-----+ |                 |  +-----+|    +---+
          |PE1|---|-| PE3 |-|--------PW1------|--| PE5 ||----|PE7|
          |   |   | |     |-|-PW3\     /------|--|     ||    |   |
          +---+\  |/+-----+ |     \   /       |  +-----+\   /+---+
           |    \ /  | *    |      \ /        |    * |  |\ /   |
           |     \|  | |ICCP|       X         |ICCP| |  | \    |
           |    / \  | *    |      / \        |    * |  |/ \   |
          +---+/  |\+-----+ |     /   \       |  +-----+/   \+---+
          |   |   | |     |-|-PW4/     \------|--|     ||    |   |
          |PE2|---|-| PE4 |-|----PW2----------|--| PE6 ||----|PE8|
          +---+   | +-----+ |                 |  +-----+|    +---+
                  |         |                 |         |
                  |         |                 |         |
                  |  RG1    |                 |  RG2    |
                  +---------+                 +---------+
                operator A network         operator B network
```

Figure 2

5. PW redundancy application procedure for inter-domain redundancy

PW redundancy application procedures are described in section 9.1 of [I-D.ietf-pwe3-iccp]. When a PE node encounters a failure, the other PE takes over. This document reuses the PW redundancy mechanism defined in[I-D.ietf-pwe3-iccp], with new ICCP switchover conditions as specified in following section.

There are two PW redundancy modes defined in [RFC6870]: Independent mode and Master/Slave mode. For the inter-domain four-PW scenario, it is required for PEs to ensure that the same mode is supported on the two ICCP peers in the same RG. One method to ensure mode consistency is by manual operation. Other methods are also possible and are out of the scope of this document.

5.1. ICCP switchover condition

5.1.1. Inter-domain PW failure

When a PE receives advertisements from the active PE, in the same RG, indicating that all the inter-domain PW status has changed to DOWN/ STANDBY, then if it has the highest priority (after the advertising PE), it SHOULD advertise active state for all of its associated inter-domain PWs.

5.1.2. PE node isolation

When a PE detects failure of all PWs to the local domain, it SHOULD advertise standby state for all its inter-domain PWs to trigger remote PE to switchover.

5.1.3. PE node failure

When a PE node detects that the active PE, that is member of the same RG, has gone down, if the local PE has redundant PWs for the affected services and has the highest priority (after the failed PE), it advertises the active state for all associated inter-domain PWs.

5.2. Inter-domain redundancy with two-PWs

In this use case, it is recommended that the operation be as follows:

o ICCP deployment option: ICCP is deployed on VPLS edge nodes in both domains;

o PW redundancy mode: independent mode only;

o Protection architectures: 1:1(1 standby, 1 active).

The switchover rules described in section 5.1 apply. Before deploying this inter-domain VPLS, the operators should negotiate to

configure the same PW high/low priority at two PW end-points. The inter-domain VPLS relationship normally involves a contractual process between operators, and the configuration of PW roles forms part of this process. E.g, in figure 1, PE3 and PE5 MUST both have higher/lower priority than PE4 and PE6, otherwise both PW1 and PW2 will be in standby state.

5.3. Inter-domain redundancy with four-PWs

In this use case, there are two options to provide protection: 1:1 and 3:1 protection. The inter-domain PWs that connect to the same PE should have proper PW priority to advertise same active/standby state. E.g, in figure 2, both PW1 and PW3 connected to PE3 would advertise active/standby state.

For 1:1 protection model, the operation would be as follows:

- ICCP deployment option: ICCP is deployed on VPLS edge nodes in both domains;
- PW redundancy mode: independent mode only;
- Protection architectures: 1:1(1 standby, 1 active).

The switchover rules described in section 5.1 apply. In this case, the operators do not need to do any coordination of the inter-domain PW priority. The PE detecting one PW DOWN should set the other PW to STANDBY if available, and then synchronize the updated state to its ICCP peer. When a PE detects that the PWs from ICCP peer PE are DOWN or STANDBY, it should switchover as described in section 5.1.1.

There are two variants of the 3:1 protection model. We will refer to them as option A and B. For option A of the 3:1 protection model, the support of Request Switchover status bit [RFC6870] is required. The operation is as follows:

- ICCP deployment option: ICCP is deployed on VPLS edge nodes in both domains;
- PW redundancy mode: Independent mode with 'request switchover' bit support;
- Protection architectures: 3:1 (3 standby, 1 active).

In this case, the procedure on the PE for the PW failure is per section 6.3 of [RFC6870], and with the following additions:

- When the PE detects failure of the active inter-domain PW, it should switch to the other local standby inter-domain PW if available, and send an updated LDP pseudowire status message with the 'request switchover' bit set on that local standby inter-domain PW to the remote PE;
- Local and remote PE should also update the new PW status to their ICCP peers, respectively, in Application Data Messages with PW-RED Synchronization Request TLV for corresponding service, so as to

synchronize the latest PW status on both PE sides;

- While waiting for the acknowledgement, the PE that sent the 'request switchover' bit may receive a switchover request from its ICCP peer's PW remote endpoint by virtue of the ICCP synchronization. The PE MUST compare IP addresses with that PW remote peer. The PE with a higher IP address will ignore the request and continue to wait for the acknowledgement from its peer in the remote domain. The PE with the lower IP address MUST clear 'request switchover' bit and set 'Preferential Forwarding' local status bit, and update the PW status to ICCP peer.
- The remote PE receiving 'request switchover' bit will acknowledge the request and activate the PW only when it is ready to take over as described in section 5.1, otherwise, it MUST ignore the request.

The node isolation failure and node failure is described in section 5.1.

For option B of 3:1 protection model, master/slave mode support is required, and should be as follows:

- ICCP deployment option: ICCP is deployed on VPLS edge nodes in only one domain;
- PW redundancy mode: master/slave only;
- Protection architectures: 3:1 (3 standby, 1 active).

When master/slave PW redundancy mode is employed, the network operators of two domains must agree on which domain PEs will be master, and configure the devices accordingly. The inter-domain PWs that connect to one PE should have higher PW priority than the PWs on the other PE in the same RG. The procedure on the PE for PW failure is as follows:

- The PE with higher PW priority should only enable one PW active, and the other PWs standby.
- When the PE detects active PW DOWN, it should enable the other local standby PW to be active with preference. Only when two inter-domain PWs connect to the PE are DOWN, the ICCP peer PE in the same RG would switchover as described in section 5.1.

The node isolation failure and node failure is described in section 5.1.

6. Management Considerations

When deploying the inter-domain redundancy mechanism described in this document, some manual operation/negotiation is required to be done correctly and securely. E.g., each node within one RG should be configured with same redundancy mode; the two operators should

negotiate to configure same PW priority at two nodes. If the configuration consistency is broken, the inter-domain redundancy mechanism may not work properly.

7. Security Considerations

Besides the security properties of [I-D.ietf-pwe3-iccp], [RFC4762] and [RFC6870], this document will have additional security consideration.

ICCP is now deployed between two PEs or ASBRs, the two PEs or ASBRs should be connected by a well managed and highly monitored network. The LDP session could be secured with TCP Authentication Option [RFC5925]. This provides integrity and authentication for the ICCP messages. The LDP MD5 authentication key option, as described in section 2.9 of [RFC5036] MAY also be used.

The attention of implementers and deployers is drawn to [RFC6941] and [RFC6952] with special attention to the recommendation to use TCP-AO [RFC5925] for enhanced security of LDP sessions.

The activitiy on the inter-domain and intra-domain pseudowire may cause security threats or be exploited to create denial of service attackes. Excessive pseudowire state flapping (e.g., by malicious peer PE's implementation) may lead to excessive ICCP exchanges. Implementations SHOULD provide mechanisms to perform control-plane policing and mitigate such types of attacks.

8. IANA Consideration

No IANA allocation is required in this document.

9. Acknowledgements

The author would like to thank Sami Boutros, Giles Heron for their valuable comments.

10. Contributors

Daniel Cohn

Email:daniel.cohn.ietf@gmail.com

Yubao Wang
ZTE Corporation
Nanjing, China
Email: wang.yubao@zte.com.cn

11. References

11.1. Normative references

[I-D.ietf-pwe3-iccp]
Martini, L., Salam, S., Sajassi, A., and S. Matsushima, "Inter-Chassis Communication Protocol for L2VPN PE Redundancy", draft-ietf-pwe3-iccp-16 (work in progress), March 2014.

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

[RFC6870] Muley, P. and M. Aissaoui, "Pseudowire Preferential Forwarding Status Bit", RFC 6870, February 2013.

11.2. Informative references

[RFC4364] Rosen, E. and Y. Rekhter, "BGP/MPLS IP Virtual Private Networks (VPNs)", RFC 4364, February 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.

[RFC5925] Touch, J., Mankin, A., and R. Bonica, "The TCP Authentication Option", RFC 5925, June 2010.

[RFC6074] Rosen, E., Davie, B., Radoaca, V., and W. Luo, "Provisioning, Auto-Discovery, and Signaling in Layer 2 Virtual Private Networks (L2VPNs)", RFC 6074, January 2011.

[RFC6718] Muley, P., Aissaoui, M., and M. Bocci, "Pseudowire Redundancy", RFC 6718, August 2012.

[RFC6941] Fang, L., Niven-Jenkins, B., Mansfield, S., and R. Graveman, "MPLS Transport Profile (MPLS-TP) Security Framework", RFC 6941, April 2013.

[RFC6952] Jethanandani, M., Patel, K., and L. Zheng, "Analysis of BGP, LDP, PCEP, and MSDP Issues According to the Keying and Authentication for Routing Protocols (KARP) Design Guide", RFC 6952, May 2013.

[RFC7117] Aggarwal, R., Kamite, Y., Fang, L., Rekhter, Y., and C. Kodeboniya, "Multicast in Virtual Private LAN Service (VPLS)", RFC 7117, February 2014.

Authors' Addresses

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

Email: zhliu@gsta.com

Lizhong Jin
Shanghai
P.R.China

Email: lizho.jin@gmail.com

Ran Chen
ZTE Corporation
NO.19 East Huayuan Road
Haidian District Beijing 100191
P.R.China

Email: chen.ran@zte.com.cn

Dennis Cai
Cisco
3750 Cisco Way,
San Jose, California 95134
USA

Email: dcai@cisco.com

Samer Salam
Cisco
595 Burrard Street, Suite:2123
Vancouver, BC V7X 1J1
Canada

Email: ssalam@cisco.com