                Redundancy Mechanism for Inter-domain VPLS Service

Abstract

   In many existing Virtual Private LAN Service (VPLS) inter-domain
   deployments (based on RFC 4762), pseudowire (PW) connectivity offers
   no Provider Edge (PE) node redundancy, or offers PE node redundancy
   with only a single domain. This deployment approach incurs a high
   risk of service interruption, since at least one domain will not
   offer PE node redundancy. This document describes an inter-domain
   VPLS solution that provides PE node redundancy across domains.

Status of This Memo

   This is an Internet Standards Track document.

   This document is a product of the Internet Engineering Task Force
   (IETF). It represents the consensus of the IETF community. It has
   received public review and has been approved for publication by the
   Internet Engineering Steering Group (IESG). Further information on
   Internet Standards is available in Section 2 of RFC 5741.

   Information about the current status of this document, any errata,
   and how to provide feedback on it may be obtained at
   http://www.rfc-editor.org/info/rfc7309.

Copyright Notice

Table of Contents

1. Introduction

   In many existing Virtual Private LAN Service (VPLS) deployments based
   on [RFC4762], pseudowire (PW) connectivity offers no Provider Edge
   (PE) node redundancy, or offers PE node redundancy with only a single
   domain. This deployment approach incurs a high risk of service
   interruption, since at least one domain will not offer PE node
   redundancy. This document describes an inter-domain VPLS solution
   that provides PE node redundancy across domains. The redundancy
   mechanism will provide PE node redundancy and link redundancy in both
   domains. The PE throughout the document refers to a routing and
   bridging capable PE defined in [RFC4762], Section 10. The domain in
   this document refers to an autonomous system (AS), or other
   administrative domains.

   The solution relies on the use of the Inter-Chassis Communication
   Protocol (ICCP) [RFC7275] to coordinate between the two redundant
   edge nodes, and use of 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 are introduced.
   This solution is a description of reusing existing protocol building
   blocks to achieve the desired function, but also defines
   implementation behavior 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 Autonomous System Border Router
   (ASBR) and associated physical links that connect the domains carry a
   multitude of services. As such, it is important to provide PE node
   and link redundancy, to ensure high service availability and meet the
   end customer service level agreements (SLAs).

   Several current deployments of inter-AS VPLS are implemented like
   inter-AS option A as described in [RFC4364], Section 10, where the
   Virtual Local Area Network (VLAN) is used to hand-off the services
   between two domains. In these deployments, PE node/link redundancy
   is achieved using Multi-Chassis Link Aggregation (MC-LAG) and ICCP
   [RFC7275]. 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 interconnection between the domains uses packet
over Synchronous Optical Networking (SONET) / Synchronous Digital
Hierarchy (SDH), 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.

Multiprotocol Border Gateway Protocol (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 [RFC7275] 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 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 ASBR. The ASBR then sends
traffic point to point 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
```
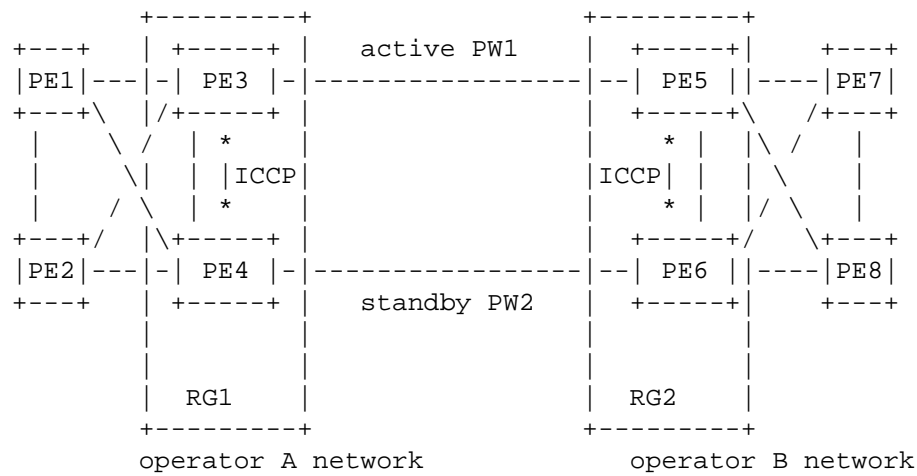
                             Figure 1

   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
```

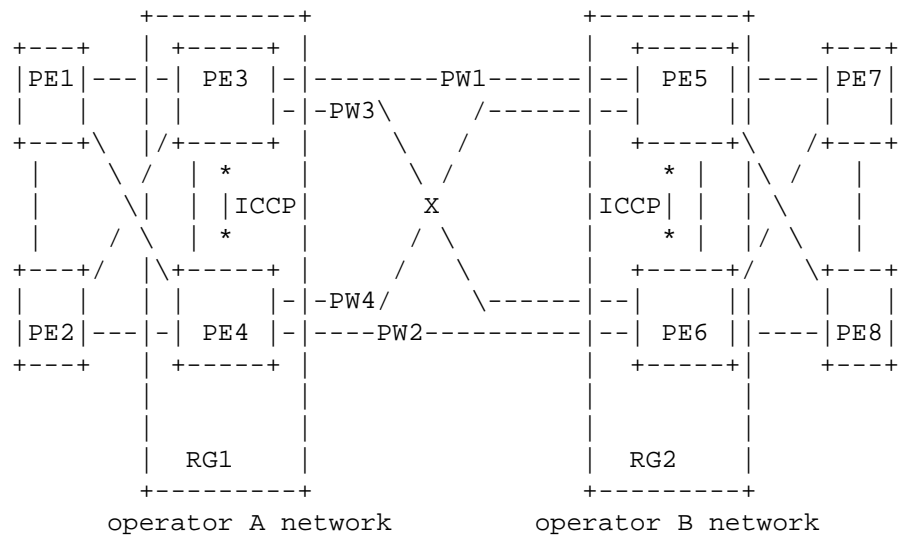   Figure 2

5. PW Redundancy Application Procedure for Inter-domain Redundancy

   PW redundancy application procedures are described in Section 9.1 of
   [RFC7275]. When a PE node encounters a failure, the other PE takes
   over. This document reuses the PW redundancy mechanism defined in
   [RFC7275], 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 that PEs ensure that the same mode be supported on the
   two ICCP peers in the same RG. This can be achieved using manual
   configuration at the ICCP peers. Other methods for ensuring
   consistency 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 a 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 SHOULD advertise 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 endpoints. The
   inter-domain VPLS relationship normally involves a contractual
   process between operators, and the configuration of PW roles forms
   part of this process. For example, 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 the same active/standby
   state. For example, in Figure 2, both PW1 and PW3 are connected to
   PE3 and should advertise active/standby state.

   For the 1:1 protection model, the operation would 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. 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 the 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 options A and B. The implementation MUST support option A
   and MAY support option B. Option B will be useful when the two
   legacy PEs in one domain do not support the function in this
   document. The two legacy PEs still need to support PW redundancy
   defined in [RFC6870] and be configured as slave node.

   For option A of the 3:1 protection model, the support of the Request
   Switchover status bit [RFC6870] is required. The operation is as
   follows:

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

    o PW redundancy mode: Independent mode with 'request switchover' bit
       support;

    o 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:

    o 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 PW status message with the
       'request switchover' bit set on that local standby inter-domain PW
       to the remote PE;

    o Local and remote PE SHOULD also update the new PW status to their
       ICCP peers, respectively, in Application Data Messages with the
       PW-RED Synchronization Request TLV for corresponding service, so
       as to synchronize the latest PW status on both PE sides.

    o While waiting for the acknowledgement, the PE that sends 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 SHOULD ignore the
       request and continue to wait for the acknowledgement from its peer
       in the remote domain. The PE with the lower IP address SHOULD
       clear the 'request switchover' bit and set the 'Preferential
       Forwarding' local status bit, and update the PW status to ICCP
       peer.

    o The remote PE receiving the 'request switchover' bit SHOULD
       acknowledge the request and activate the PW only when it is ready
       to take over as described in Section 5.1; otherwise, it SHOULD
       ignore the request.

    The PE node isolation failure and PE node failure is described in
    Section 5.1.

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

    o ICCP deployment option: ICCP is deployed on VPLS edge nodes in
       only one domain;

    o PW redundancy mode: master/slave only;

    o 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:

o The PE with higher PW priority should only enable one PW active,
   and the other PWs should be in the standby state.

o When the PE detects an active PW DOWN, it SHOULD enable the other
   local standby PW to be active with preference. Only when two
   inter-domain PWs connected to the PE are DOWN, the ICCP peer PE in
   the same RG SHOULD switchover as described in Section 5.1.

The PE node isolation failure and PE node failure are described in
Section 5.1.

6. Management Considerations

When deploying the inter-domain redundancy mechanism described in
this document, consistent provisioning is required for proper
operation. The two domains must both use the same use case
(Section 5.2 or Section 5.3). Within each section, all of the
described modes and options must be provisioned identically both
within each RG and between the RGs. Additionally, for the two-PWs
redundancy options defined in Section 5.2, the two operators must
also negotiate to configure same high/low PW priority at the two PW
endpoints. If the provisioning is inconsistent, then the inter-
domain redundancy mechanism may not work properly.

7. Security Considerations

Besides the security properties of [RFC7275] for the ICCP control
plane, and [RFC4762] and [RFC6870] for the PW control plane, this
document has additional security considerations for the ICCP control
plane.

In this document, ICCP is deployed between two PEs or ASBRs. The two
PEs or ASBRs should only be connected by a network that is well
managed and whose service levels and availability are highly
monitored. This should be ensured by the operator.

The state flapping on the inter-domain and intra-domain PW may cause
security threats or be exploited to create denial-of-service attacks.
For example, excessive PW 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. Acknowledgements

   The authors would like to thank Sami Boutros, Giles Heron, Adrian
   Farrel, Andrew G. Malis, and Stephen Kent for their valuable
   comments.

9. Contributors

   Daniel Cohn
   Email:daniel.cohn.ietf@gmail.com

   Yubao Wang
   ZTE Corporation

   Nanjing, China
   Email: wang.yubao@zte.com.cn

10. References

10.1. Normative references

   [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.

   [RFC7275]   Martini, L., Salam, S., Sajassi, A., Bocci, M.,
               Matsushima, S., and T. Nadeau, "Inter-Chassis
               Communication Protocol for Layer 2 Virtual Private Network
               (L2VPN) Provider Edge (PE) Redundancy", RFC 7275, June
               2014.

10.2. 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.

   [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.

   [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