Pseudowire (PW) Endpoint Fast Failure Protection

Abstract

   This document specifies a fast mechanism for protecting pseudowires
   (PWs) transported by IP/MPLS tunnels against egress endpoint
   failures, including egress attachment circuit (AC) failure, egress
   provider edge (PE) failure, multi-segment PW terminating PE failure,
   and multi-segment PW switching PE failure. Operating on the basis of
   multihomed customer edge (CE), redundant PWs, upstream label
   assignment, and context-specific label switching, the mechanism
   enables local repair to be performed by the router upstream adjacent
   to a failure. The router can restore a PW in the order of tens of
   milliseconds, by rerouting traffic around the failure to a protector
   through a pre-established bypass tunnel. Therefore, the mechanism
   can be used to reduce traffic loss before global repair reacts to the
   failure and the network converges on the topology changes due to the
   failure.

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 7841.

   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/rfc8104.

Table of Contents

1. Introduction

   Per [RFC3985], [RFC8077], and [RFC5659], a pseudowire (PW) or PW
   segment can be thought of as a connection between a pair of
   forwarders hosted by two PEs, carrying an emulated Layer 2 service
   over a packet switched network (PSN). In the single-segment PW
   (SS-PW) case, a forwarder binds a PW to an attachment circuit (AC).
   In the multi-segment PW (MS-PW) case, a forwarder on a terminating PE
   (T-PE) binds a PW segment to an AC, while a forwarder on a switching
   PE (S-PE) binds one PW segment to another PW segment. In each
   direction between the PEs, PW packets are transported by a PSN
   tunnel, which is also called a transport tunnel.

   In order to protect the PW service against network failures, it is
   necessary to protect every link and node along the entire data path.
   For the traffic in a given direction, this includes ingress AC,
   ingress (T-)PE, intermediate routers of the transport tunnel, S-PEs,
   egress (T-)PE, and egress AC. To minimize service disruption upon a
   failure, it is also desirable that each of these components is
   protected by a fast protection mechanism based on local repair. Such
   mechanisms generally involve a bypass path that is pre-computed and
   pre-installed in the data plane on the router upstream adjacent to an
   anticipated failure. This router is referred to as a "point of local
   repair" (PLR). The bypass path has the property that it can guide
   traffic around the failure, while remaining unaffected by the
   topology changes resulting from the failure. When the failure
   occurs, the PLR can invoke the bypass path to achieve fast
   restoration for the service.

   Today, fast protection against ingress AC failure and ingress (T-)PE
   failure can be achieved by using a multihomed CE and redundant ACs,
   such as a multi-chassis link aggregation group (MC-LAG). Fast
   protection against the failure of an intermediate router of the
   transport tunnel can be achieved through RSVP fast reroute [RFC4090]
   or IP/LDP fast reroute [RFC5286] [RFC5714]. However, there is no
   equivalent mechanism that can be used against an egress AC failure,
   an egress (T-)PE failure, or an S-PE failure. For these failures,
   service restoration has to rely on global repair or control-plane
   convergence. Global repair normally involves the ingress CE or the
   ingress (T-)PE switching traffic to an alternative path, based on
   remote failure detection via PW status notification, end-to-end
   Operations, Administration, and Maintenance (OAM), and others.
   Control-plane convergence relies on control protocols to react on the
   topology changes due to a failure. Compared to local repair, these
   mechanisms are relatively slow in reacting to a failure and restoring
   traffic.

This document addresses the above need. It specifies a fast
protection mechanism based on local repair to protect PWs against the
following endpoint failures:

a. Egress AC failure.

b. Egress PE failure: Link or node failure of an egress PE of an
    SS-PW or a T-PE of an MS-PW.

c. Switching PE failure: Link or node failure of an S-PE of an
    MS-PW.

The mechanism is applicable to LDP-signaled PWs. It is relevant to
networks with redundant PWs and multihomed CEs. It is designed on
the basis of MPLS upstream label assignment and context-specific
label switching [RFC5331]. Fast protection refers to its ability to
restore traffic in the order of tens of milliseconds. Compared with
global repair and control-plane convergence, this mechanism can
provide faster service restoration. However, it is intended to
complement these mechanisms, rather than replacing them, as networks
rely on them to eventually move traffic to fully functional
alternative paths.

2. Specification of Requirements

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. Reference Models for Egress Endpoint Failures

This document refers to the following topologies to describe egress
endpoint failures and protection procedures.

3.1. Single-Segment PW

```
                    |<------------- PW1 --------------->|

                - PE1 ------------- P1 --------------- PE2 -
               /                                             \
              /                                               \
          CE1                                                   CE2
              \                                               /
               \                                             /
                - PE3 ------------- P2 --------------- PE4 -

                    |<------------- PW2 --------------->|
```
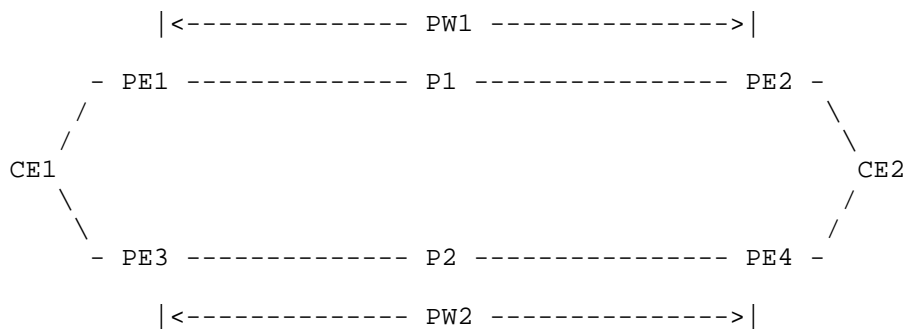
                               Figure 1

   In Figure 1, the IP/MPLS network consists of PE and P routers. It
   provides a PW service between CE1 and CE2. Each CE is multihomed via
   two ACs to two PEs. This forms two divergent paths between the CEs.
   The first path uses PW1 between PE1 and PE2, and the second path uses
   PW2 between PE3 and PE4. For clarity, the transport tunnels of the
   PWs and other links between the routers are not shown in this figure.

   In general, a CE may operate the ACs in two modes when sending
   traffic to the remote CE, i.e., active-standby mode and active-active
   mode.

   o In the active-standby mode, the CE chooses one AC as the active AC
      and the corresponding path as the active path, and it uses the
      other AC as the standby AC and the corresponding path as the
      standby path. The CE only sends traffic on the active AC as long
      as the active path is operational. The CE will only send traffic
      on the standby AC after it detects a failure of the active path.
      Note that the CE may receive traffic on the active or standby AC,
      depending on whether the remote CE chooses the same active path
      for the traffic of the reverse direction. In this document, even
      if both CEs choose the same active path, each CE should still
      anticipate receiving traffic on a standby AC, because the traffic
      may be redirected to the standby path by the fast protection
      mechanism.

   o In the active-active mode, the CE treats both ACs and their
      corresponding paths as active and sends traffic on both ACs in a
      load-balancing fashion. In the reverse direction, the CE may
      receive traffic on both ACs.

   The above modes assume the traffic to be data traffic, which is not
   bound to a specific AC. This does not include control-protocol

traffic between the CEs, when the CE-CE control-protocol sessions or
adjacencies established on the two ACs are considered as distinct
rather than having a primary and backup relationship. In general, a
dual-homed CE should not make any explicit or implicit assumptions
regarding the specific AC from which it receives packets from the
remote CE.

For either mode, when considering the traffic flowing in a given
direction over an active path, this document views the ACs, PEs, and
PWs as serving primary or backup roles. In particular, the ACs, PEs,
and PWs along this active path have primary roles, while those along
the other path have backup roles. Note that in the active-active
mode, each AC, PE, and PW on an active path has a primary role and
also a backup role protecting the other path, which is also active.

For Figure 1, the following roles are assumed for the traffic going
from CE1 to CE2 via PW1.

    Primary ingress AC: CE1-PE1

    Primary ingress PE: PE1

    Primary PW: PW1

    Primary egress PE: PE2

    Primary egress AC: PE2-CE2

    Backup ingress AC: CE1-PE3

    Backup ingress PE: PE3

    Backup PW: PW2

    Backup egress PE: PE4

    Backup egress AC: PE4-CE2

Based on this schema, this document describes egress endpoint
failures and the fast protection mechanism on the per-active-path and
per-direction basis. In this case, an egress AC failure refers to
the failure of the AC PE2-CE2, and an egress node failure refers to
the failure of PE2. The ultimate goal is that when a failure occurs,
the traffic should be locally repaired, so that it can eventually
reach CE2 via the backup egress PE (PE4) and the backup egress AC
(PE4-CE2).

Subsequent to the local repair, either the current active path should
heal after the control plane converges on the new topology or the
ingress CE should switch traffic from the primary path to the backup
path, depending on the failure scenario. In the latter case, the
ingress CE may perform the path switchover triggered by end-to-end
OAM (in-band or out-band), PW status notification, CE-PE control
protocols (e.g., the Link Aggregation Control Protocol (LACP)), and
others. In the active-standby mode, this will promote the standby
path to a new active path. In the active-active mode, it will make
the other active path carry all the traffic between the two CEs. In
any case, this phase of restoration falls into the control-plane
convergence and global repair category; hence, it is out of the scope
of this document. The purpose of the fast protection mechanism in
this document is to reduce traffic loss before this phase of
restoration takes place.

Note that in Figure 1, if the traffic in the reverse direction (i.e.,
from CE2 to CE1) traverses the AC CE2-PE2 and PE2 as an active path,
the failure of PE2 and the failure of the AC PE2-CE2 will be
considered as ingress failures of the traffic. If CE2 can detect the
failures, it may protect the traffic by switching it to the backup
path via the AC CE2-PE4 and PE4. However, this is categorized as
ingress endpoint failure protection; hence, it is not handled by the
mechanism described in this document.

Figure 2 shows another possible scenario, where CE1 is single-homed
to PE1, while CE2 remains multihomed to PE2 and PE4. From the
perspective of egress endpoint protection for the traffic going from
CE1 to CE2 over PW1, this scenario is the same as the scenario shown
in Figure 1.

```
               |<------------- PW1 --------------->|

                 ------------ P1 --------------- PE2 -
               /                                        \
              /                                          \
       CE1 -- PE1                                          CE2
              \                                          /
               \                                        /
                 ------------ P2 --------------- PE4 -

       |<------------- PW2 -------------->|
```
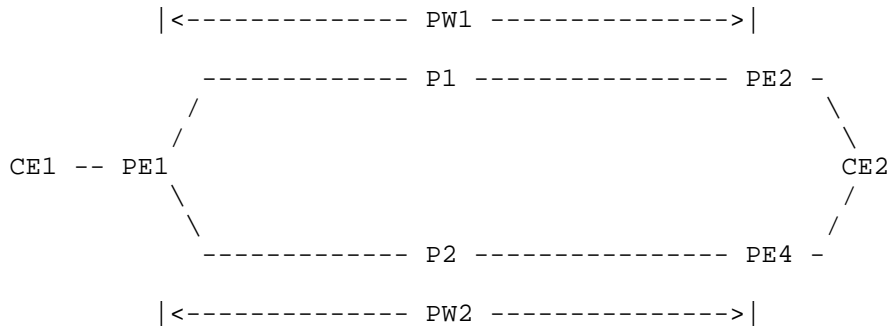
                            Figure 2

For clarity, primary egress AC, primary egress PE, backup egress AC,
and backup egress PE may simply be referred to as primary AC, primary
PE, backup AC, and backup PE, respectively, when the context of a
discussion is egress endpoint.

3.2. Multi-Segment PW

```
                |<-------------- PW1 --------------->|
                |<----- SEG1 ----->|<----- SEG2 ----->|

           - TPE1 ------------- SPE1 -------------- TPE2 -
          /                                              \
         /                                                \
       CE1                                                CE2
         \                                                /
          \                                              /
           - TPE3 ------------- SPE2 -------------- TPE4 -

                |<----- SEG3 ----->|<----- SEG4 ----->|
                |<-------------- PW2 --------------->|
```
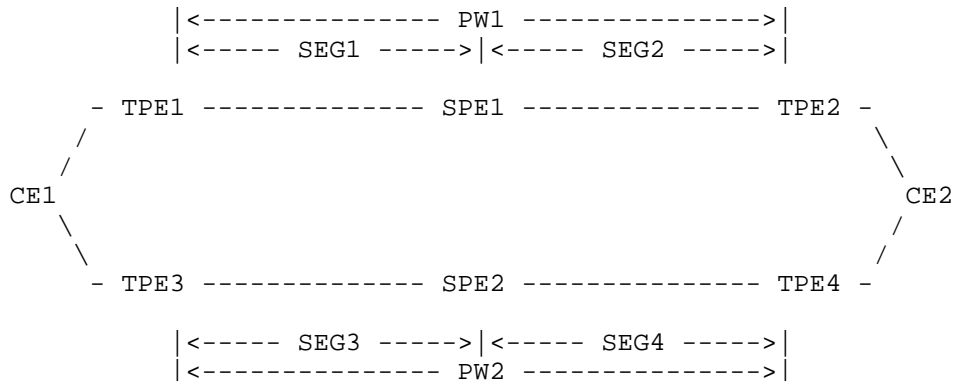
                            Figure 3

Figure 3 shows a topology that is similar to Figure 1 but in an MS-PW
environment. PW1 and PW2 are both MS-PWs. PW1 is established
between TPE1 and TPE2 and switched between segments SEG1 and SEG2 at
SPE1. PW2 is established between TPE3 and TPE4 and switched between
segments SEG3 and SEG4 at SPE2. CE1 is multihomed to TPE1 and TPE3.
CE2 is multihomed to TPE2 and TPE4. For clarity, the transport
tunnels of the PW segments are not shown in this figure.

In this document, the following primary and backup roles are assigned
for the traffic going from CE1 to CE2:

     Primary ingress AC: CE1-TPE1

     Primary ingress T-PE: TPE1

     Primary PW: PW1

     Primary S-PE: SPE1

     Primary egress T-PE: TPE2

     Primary egress AC: TPE2-CE2

     Backup ingress AC: CE1-TPE3

        Backup ingress T-PE: TPE3

        Backup PW: PW2

        Backup S-PE: SPE2

        Backup egress T-PE: TPE4

        Backup egress AC: TPE4-CE2

   In this case, an egress AC failure refers to the failure of the AC
   TPE2-CE2. An egress node failure refers to the failure of TPE2. An
   S-PE failure refers to the failure of SPE1.

   For consistency with the SS-PW scenario, primary T-PEs and primary
   S-PEs may simply be referred to as primary PEs in this document,
   where specifics are not required. Similarly, backup T-PEs and backup
   S-PEs may be referred to as backup PEs.

4. Theory of Operation

   The fast protection mechanism in this document provides three types
   of protection for PWs, corresponding to the three types of failures
   described in Section 1:

   a. Egress AC protection

   b. Egress (T-)PE node protection

   c. S-PE node protection

4.1. Applicability

   The mechanism is applicable to LDP-signaled PWs in an environment
   where an egress CE is multihomed to a primary PE and a backup PE and
   there exists a backup PW, as described in Section 3. The procedure
   for S-PE node protection is applicable when there exists a backup
   S-PE on the backup PW.

   The mechanism assumes IP/MPLS transport tunnels and is applicable to
   tunnels with single path and equal-cost multipaths (ECMPs). As an
   example of ECMPs, imagine a tunnel carrying one or multiple PWs and
   traversing a router with ECMPs to a primary PE. The ECMPs consist of
   at least one direct link to the PE and some multi-hop paths to the
   PE. Due to the direct link, the router is considered as a
   penultimate hop of the tunnel and can perform local detection and
   repair for an egress node failure. The router normally uses a
   hashing algorithm to distribute PW packets over the ECMPs, on a

per-PW or per-flow basis. Upon a failure of the direct link, i.e.,
transit link failure, the router removes the link from the hashing
algorithm, which automatically redistributes the traffic of the link
to the other paths of the ECMPs, achieving local repair. This
scenario is not the focus of this document. Upon a failure of the
PE, i.e., egress node failure, the router SHOULD perform local repair
by rerouting the entire traffic of the ECMPs, as the failure will
affect every path. If the router does not have a fast or reliable
mechanism to detect the egress node failure, it is RECOMMENDED that
the router SHOULD treat the failure of the direct link as an egress
node failure.

The mechanism is applicable to both best-effort and traffic
engineering (TE) transport tunnels. For TE transport tunnels that
require bandwidth protection, TE bypass tunnels with reserved
bandwidth MAY be used to avoid congestion for rerouted traffic.

It is also RECOMMENDED that the mechanism SHOULD be used in
conjunction with global repair and control-plane convergence, in such
a manner that the mechanism temporarily repairs a failed path by
using a bypass tunnel, and global repair and control-plane
convergence eventually move traffic to a fully functional alternative
path.

4.2. Local Repair

The fast protection ability of the mechanism comes from local repair
performed by routers upstream adjacent to failures. Each of these
routers is referred to as a PLR. A PLR MUST be able to detect a
failure by using a rapid mechanism, such as physical-layer failure
detection, Bidirectional Forwarding Detection (BFD) [RFC5880],
Seamless BFD (S-BFD) [RFC7880], and others. In anticipation of the
failure, the PLR MUST also pre-establish a bypass tunnel to a
"protector" and pre-install a bypass route for the bypass tunnel in
the data plane. The protector is either a backup PE or a router that
will forward traffic to a backup PE. The bypass tunnel MUST have the
property that it will not be affected by the topology changes due to
the failure. Specifically, it MUST NOT traverse the primary PE or
the penultimate link of the protected transport tunnel or share any
shared risk link groups (SRLGs) with the penultimate link. Upon
detecting the failure, the PLR invokes the bypass route in the data
plane and reroutes PW traffic to the protector through the bypass
tunnel. The protector in turn sends the traffic to the target CE.
This procedure is referred to as local repair.

Different routers may serve as PLR and protector in different
scenarios.

   o In egress AC protection, the PLR is the primary PE, and the
      protector is the backup PE (Figure 4).

```
                |<------------- PW1 --------------->|

              - PE1 ------------- P1 --------------- PE2 -
             /                                       PLR  \
            /                                         |     \
        CE1                                     bypass|      CE2
            \                                         |     /
             \                                        |    /
              - PE3 ------------- P2 --------------- PE4 -
                                                    protector

                |<------------- PW2 --------------->|
```
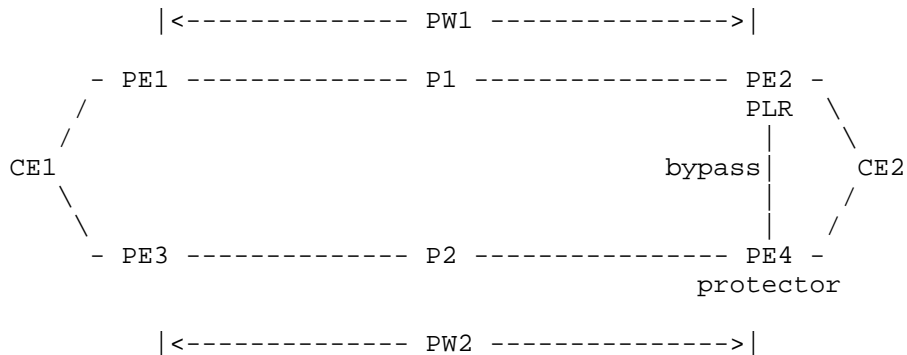
                            Figure 4

   o In egress PE node protection, the PLR is the penultimate hop
      router of the transport tunnel of the primary PW, and the
      protector is the backup PE (Figure 5).

```
                |<------------- PW1 --------------->|

              - PE1 ------------- P1 ------- P3 ----- PE2 -
             /                              PLR \       \
            /                                    \       \
        CE1                              bypass\           CE2
            \                                    \       /
             \                                    \     /
              - PE3 ------------- P2 --------------- PE4 -
                                                    protector

                |<------------- PW2 --------------->|
```
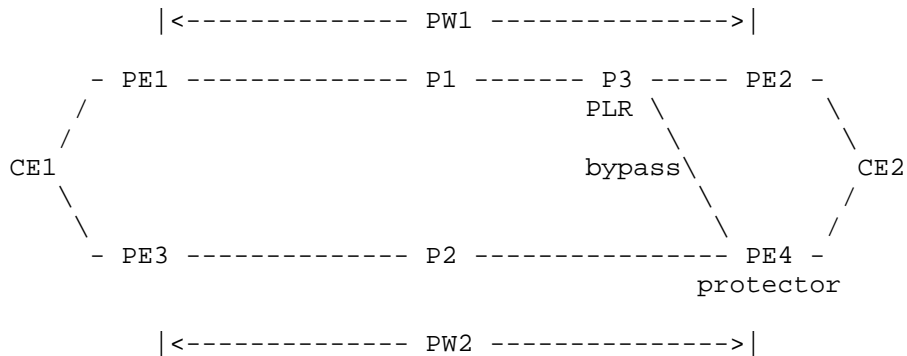
                            Figure 5

   o In S-PE node protection, the PLR is the penultimate hop router of
      the transport tunnel of the primary PW segment, and the protector
      is the backup S-PE (Figure 6).

```
              |<--------------- PW1 --------------->|
              |<----- SEG1 ----->|<----- SEG2 ----->|

          - TPE1 ----- P1  ----- SPE1 ------------- TPE2 -
         /             PLR \                              \
        /                   \                              \
     CE1               bypass\                              CE2
        \                     \                            /
         \                     \                          /
          - TPE3 -------------- SPE2 ------------- TPE4 -
                              protector

              |<----- SEG3 ----->|<----- SEG4 ----->|
              |<-------------- PW2 --------------->|
```
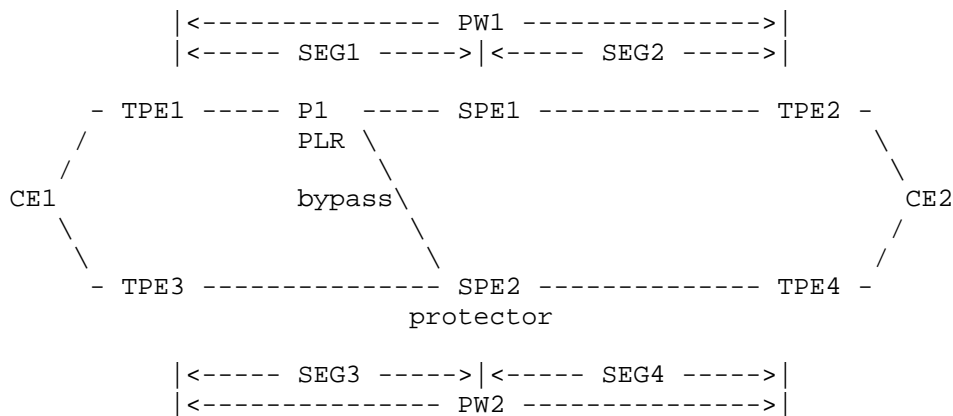
                              Figure 6

   In egress AC protection, a PLR realizes its role based on
   configuration of a "context identifier", which is introduced in this
   document (Section 4.3). The PLR establishes a bypass tunnel to the
   protector in the same fashion as a normal PSN tunnel.

   In egress PE and S-PE node protection, a PLR is a transit router on
   the transport tunnel, and it normally does not have knowledge of the
   PW(s) carried by the transport tunnel. In this document, the PLR
   simply computes and establishes a node-protection bypass tunnel in
   the same fashion as the normal IP/MPLS node protection, except that
   with the notion of the context identifier, the bypass tunnel will be
   established from the PLR to the protector (Section 4.6). Conversely,
   when the router is no longer a PLR for egress PE or S-PE node
   protection due to a change in network topology or the transport
   tunnel's path, the router should revert to the role of regular
   transit router, including PLR for transit link and node protection.

   In local repair, a PLR simply switches all the traffic received on
   the transport tunnel to the bypass tunnel. This requires that the
   protector given by the bypass tunnel MUST be intended for all the PWs
   carried by the transport tunnel. This is achieved by the ingress PE
   using a context identifier to associate a PW with the specific pair
   of {primary PE, protector} and map the PW to a transport tunnel
   destined for the same {primary PE, protector}. The ingress PE MAY
   map multiple PWs to the transport tunnel, if they share the {primary
   PE, protector} in common.

   In local repair, the PLR keeps the PW label intact in packets. This
   obviates the need for the PLR to maintain bypass routes on a per-PW
   basis and allows bypass tunnel sharing between PWs. On the other
   hand, this imposes a requirement on the protector that it MUST be
   able to forward the packets based on a PW label that is assigned by
   the primary PE and ensure that the traffic MUST reach the target CE
   via a backup path. From the protector's perspective, this PW label
   is an upstream assigned label [RFC5331]. To achieve this, the
   protector MUST learn the PW label from the primary PE prior to the
   failure and install a proper forwarding state for the PW label in a
   dedicated label space associated with the primary PE. During local
   repair, the protector MUST perform PW label lookup in this label
   space.

   The previous examples have shown the scenarios where the protectors
   are backup (T-/S-)PEs. It is also possible that a protector is a
   dedicated router to serve such a role, separate from the backup
   (T-/S-)PE. During local repair, the PLR still reroutes traffic to
   the protector through a bypass tunnel. The protector then forwards
   the traffic to the backup (T-/S-)PE, which further forwards the
   traffic to the target CE via a backup AC or a backup PW segment.
   More detail is included in Section 4.4.

4.3. Context Identifier

   A protector may protect multiple primary PEs. The protector MUST
   maintain a separate label space for each primary PE. Likewise, the
   PWs terminated on a primary PE may be protected by multiple
   protectors, each for a subset of the PWs. In any case, a given PW
   MUST be associated with one and only one pair of {primary PE,
   protector}.

   This document introduces the notion of a context identifier to
   facilitate protection establishment. A context identifier is an
   IPv4/v6 address assigned to each ordered pair of {primary PE,
   protector}. The address MUST be globally unique or unique in the
   address space of the network where the primary PE and the protector
   reside.

4.3.1. Semantics

   The semantics of a context identifier is twofold:

   o A context identifier identifies a primary PE and an associated
      protector. It represents the primary PE as the PW destination on
      a per-protector basis. A given primary PE may be protected by
      multiple protectors, each for a subset of the PWs terminated on
      the primary PE. A distinct context identifier MUST be assigned to
      each {primary PE, protector} pair.

      The ingress PE of a PW learns the context identifier of the PW's
      {primary PE, protector} from the primary PE via the Interface_ID
      TLV [RFC3471] [RFC3472] in the LDP Label Mapping message of the
      PW. The ingress PE then sets up or resolves a transport tunnel
      with the context identifier, rather than a private IP address of
      the primary PE, as the destination. This destination not only
      makes the transport tunnel reach the primary PE but also conveys
      the identity of the protector to the PLR, which MUST use the
      context identifier as the destination for the bypass tunnel to the
      protector. The ingress PE MUST map only the PWs terminated by the
      exact primary PE and protected by the exact protector to the
      transport tunnel.

   o A context identifier indicates the primary PE's label space on the
      protector. The protector may protect PWs for multiple primary
      PEs. For each primary PE, it MUST maintain a separate label space
      to store the PW labels assigned by that primary PE. It associates
      a PW label with a label space via the context identifier of the
      {primary PE, protector}, as below.

      In addition to the normal LDP PW signaling, the primary PE MUST
      have a targeted LDP session with the protector and advertise PW
      labels to the protector via LDP Label Mapping messages
      (Section 6). The primary PE MUST attach the context identifier to
      each message. Upon receiving the message, the protector MUST
      install the advertised PW label in the label space identified by
      the context identifier.

      When a PLR sets up or resolves a bypass tunnel to the protector,
      it MUST use the context identifier rather than a private IP
      address of the protector as the destination. The protector MUST
      use the bypass tunnel, either the MPLS tunnel label or the IP
      tunnel destination address, as the pointer to the corresponding
      label space. The protector MUST forward PW packets received on
      the bypass tunnel based on label lookup in that label space.

4.3.2. FEC

   In an MPLS network, a context identifier represents a Forwarding
   Equivalence Class (FEC) for transport tunnels and bypass tunnels
   destined for it. For example, it may be encoded in an LDP Prefix FEC
   Element or in the "tunnel endpoint address" of an RSVP Session
   object. The FEC is associated with a unique forwarding state on PLRs
   and the protector, which cannot be shared with other FECs. Some MPLS
   protocols (e.g., LDP) support FEC aggregation [RFC3031]. In this
   case, FEC aggregation MUST NOT be applied to a context identifier's
   FEC, and every router MUST assign a unique label to the FEC.

4.3.3. IGP Advertisement and Path Computation

   Using a context identifier as the destination for both the transport
   tunnel and bypass tunnel requires coordination between the primary PE
   and the protector in IGP advertisement of the context identifier in
   the routing domain and TE domain. The context identifier should be
   advertised in such a way that all the routers on the tunnels MUST be
   able to independently reach the following common view of paths:

   o The transport tunnel MUST have the primary PE as the path
      endpoint.

   o The bypass tunnel MUST have the protector as the path endpoint.
      In egress PE and S-PE node protection, the path MUST avoid the
      primary PE.

   There are generally two categories of approaches to achieve the
   above:

   o The first category does not require an ingress PE or a PLR to have
      knowledge of the PW egress endpoint protection schema. It does
      not require any IGP extension for context identifier
      advertisement. A context identifier is advertised by the primary
      PE and the protector as an address reachable via both routers.
      The ingress PE and the PLR can compute paths by using a normal
      method, such as Dijkstra, constrained shortest path first (CSPF),
      Loop-Free Alternate (LFA) [RFC5286], and Maximally Redundant Tree
      (MRT) [RFC7812]. One example is to advertise a context identifier
      as a virtual proxy node connected to the primary PE and the
      protector, with the link between the proxy node and the primary PE
      having a more preferable IGP and TE metric than the link between
      the proxy node and the protector. The transport tunnel will
      follow the shortest path or a TE path to the primary PE and be
      terminated by the primary PE. The PLR will no longer view itself
      as a penultimate hop of the transport tunnel, but rather two hops
      away from the proxy node, via the primary PE. Hence, a node-

          protection bypass tunnel will be available via the protector to
          the proxy node, but it will actually be terminated by the
          protector.

     o The second category requires a PLR to have knowledge of the PW
        egress endpoint protection schema. The primary PE advertises the
        context identifier as a regular IP address, while the protector
        advertises it by using an explicit "context identifier object",
        which MUST be understood by the PLR. The context identifier
        object requires an IGP extension. In both the routing domain and
        the TE domain, the context identifier is only reachable via the
        primary PE. This ensures that the transport tunnel is terminated
        by the primary PE. The PLR views itself as the penultimate hop of
        the transport tunnel, and based on the IGP context identifier
        object, it establishes or resolves a bypass tunnel to the
        advertiser (i.e., the protector), while avoiding the primary PE.

   The mechanism in this document intends to be flexible on the approach
   used by a network, as long as it satisfies the above requirements for
   the transport tunnel path and bypass tunnel path. In theory, the
   network can use one approach for context ID X and another approach
   for context ID Y. For a given context ID, all relevant routers,
   including primary PE, protector, and PLR, must support and agree on
   the chosen approach. The coordination between the routers can be
   achieved by configuration.

4.4. Protection Models

   There are two protection models based on the location of a protector:
   the co-located protector model and the centralized protector model.
   A network MAY use either model or both.

4.4.1. Co-located Protector

   In this model, the protector is a backup PE that is directly
   connected to the target CE via a backup AC, or it is a backup S-PE on
   a backup PW. That is, the protector is co-located with the backup
   (S-)PE. Examples of this model are shown in Figures 4, 5, and 6 in
   Section 4.2.

   In egress AC protection and egress PE node protection, when a
   protector receives traffic from the PLR, it forwards the traffic to
   the CE via the backup AC. This is shown in Figure 7, where PE2 is
   the PLR for egress AC failure, P3 is the PLR for PE2 failure, and PE4
   (backup PE) is the protector.

```
                |<-------------- PW1 --------------->|

           - PE1 ------------- P1 ------- P3 ----- PE2 ----
          /                            PLR \    PLR     \
         /                                  \    |       \
      CE1                              bypass\   |bypass  CE2
         \                                    \  |       /
          \                                    \ |      /
           - PE3 ------------- P2 --------------- PE4 ----
                                                 protector

                |<------------- PW2 --------------->|
```
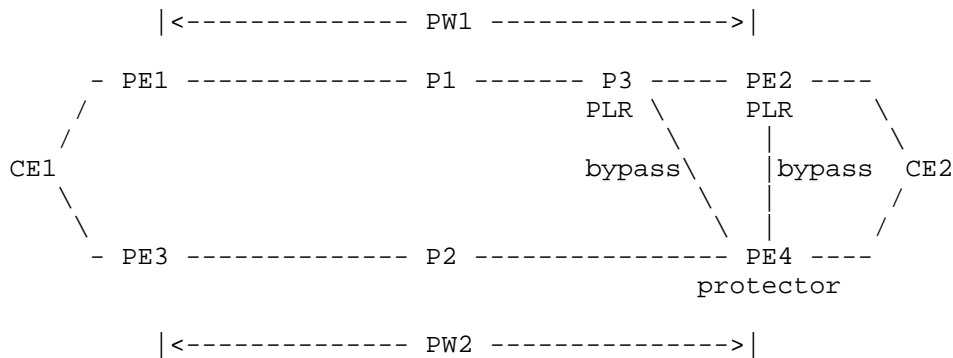
                            Figure 7

   In S-PE node protection, when a protector receives traffic from the
   PLR, it forwards the traffic over the next segment of the backup PW.
   The T-PE of the backup PW in turn forwards the traffic to the CE via
   a backup AC. This is shown in Figure 8, where P1 is the PLR for SPE1
   failure, and SPE2 (backup S-PE) is the protector for SPE1. SPE2
   receives traffic from P1, swaps SEG1's label to SEG4's label, and
   forwards the traffic over a transport tunnel to TPE4.

```
                |<--------------- PW1 --------------->|
                |<----- SEG1 ----->|<----- SEG2 ----->|

           - TPE1 ----- P1  ----- SPE1 ------------- TPE2 -
          /              PLR \                            \
         /                    \                            \
      CE1                bypass\                            CE2
         \                      \                           /
          \                      \                         /
           - TPE3 -------------- SPE2 ------------- TPE4 -
                            protector

                |<----- SEG3 ----->|<----- SEG4 ----->|
                |<-------------- PW2 --------------->|
```
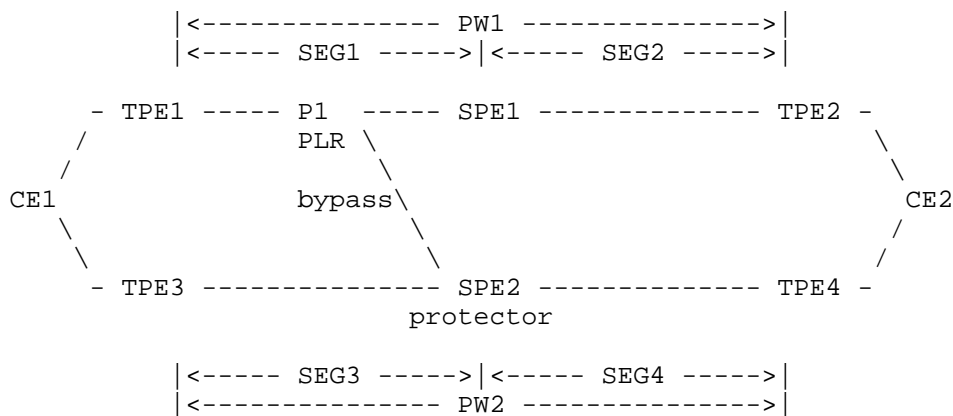
                            Figure 8

   In the co-located protector model, the number of context identifiers
   needed by a network is the number of distinct {primary PE, backup PE}
   pairs. From the perspective of scalability, the model is suitable
   for networks where the number of primary PEs and the average number
   of backup PEs per primary PE are both relatively low.

4.4.2. Centralized Protector

   In this model, the protector is a dedicated P router or PE router
   that serves the role. In egress AC protection and egress PE node
   protection, the protector may or may not be a backup PE directly
   connected to the target CE. In S-PE node protection, the protector
   may or may not be a backup S-PE on the backup PW.

   In egress AC protection and egress PE node protection, if the
   protector is not directly connected to the CE, it forwards the
   traffic to a backup PE, which in turn forwards the traffic to the CE
   via a backup AC. This is shown in Figure 9, where the protector
   receives traffic from P3 (PLR for egress PE failure) or PE2 (PLR for
   egress AC failure), swaps PW1's label to PW2's label, and forwards
   the traffic via a transport tunnel to PE4 (backup PE). The protector
   may be protecting other PWs and other primary PEs as well; for
   clarity, this is not shown in the figure.

```
            |<------------ PW1 --------------->|

        - PE1 ------------ P1 ------- P3 ----- PE2 --
       /                              PLR \    PLR    \
      /                                    \    /      \
     /                               bypass\  /bypass \
    /                                        \ /        \
   CE1                                     protector      CE2
    \                                          \          /
     \                              transport\         /
      \                                tunnel \      /
       \                                       \    /
        - PE3 ------------ P2 ----------------PE4 --

            |<------------ PW2 --------------->|
```
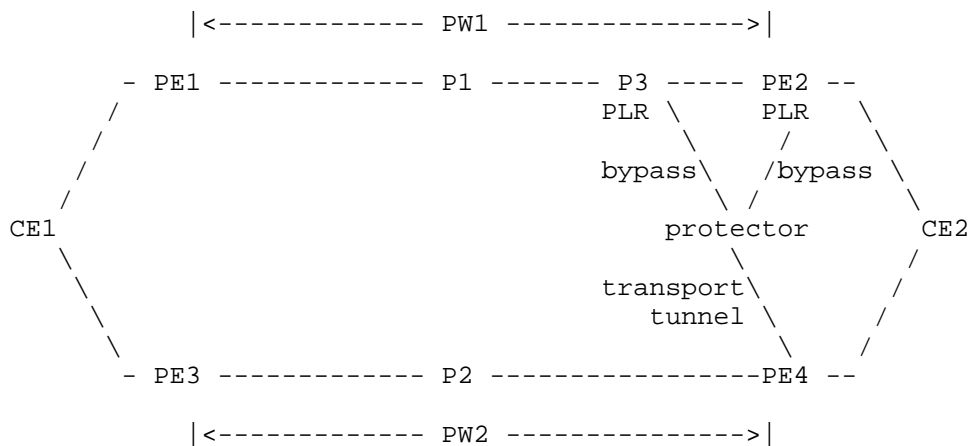
                              Figure 9

   In S-PE node protection, if the protector is not a backup S-PE, it
   forwards the traffic to the backup S-PE, which in turn forwards the
   traffic over the next segment of the backup PW. Finally, the T-PE of
   the backup PW forwards the traffic to the CE via the backup AC. This
   is shown in Figure 10, where the protector receives traffic from P1
   (PLR), swaps SEG1's label to SEG3's label, and forwards the traffic
   via a transport tunnel to SPE2 (backup S-PE). SPE2 in turn performs
   MS-PW switching from SEG3's label to SEG4's label and forwards the
   traffic over a transport tunnel to TPE4 (backup T-PE). The protector
   may be protecting other PW segments and other primary S-PEs as well;
   for clarity, this is not shown in the figure.

```
                   |<-------------- PW1 --------------->|
                   |<----- SEG1 ----->|<----- SEG2 ----->|

             - TPE1 ----- P1  ----- SPE1 ------------- TPE2 -
            /              PLR \                              \
           /                    \                              \
          /               bypass\                               \
         /                       \                               \
       CE1                      protector                        CE2
         \                         \                             /
          \               transport\                            /
           \                 tunnel \                          /
            \                        \                        /
             - TPE3 -------------- SPE2 ------------- TPE4 -

                   |<----- SEG3 ----->|<----- SEG4 ----->|
                   |<-------------- PW2 --------------->|
```
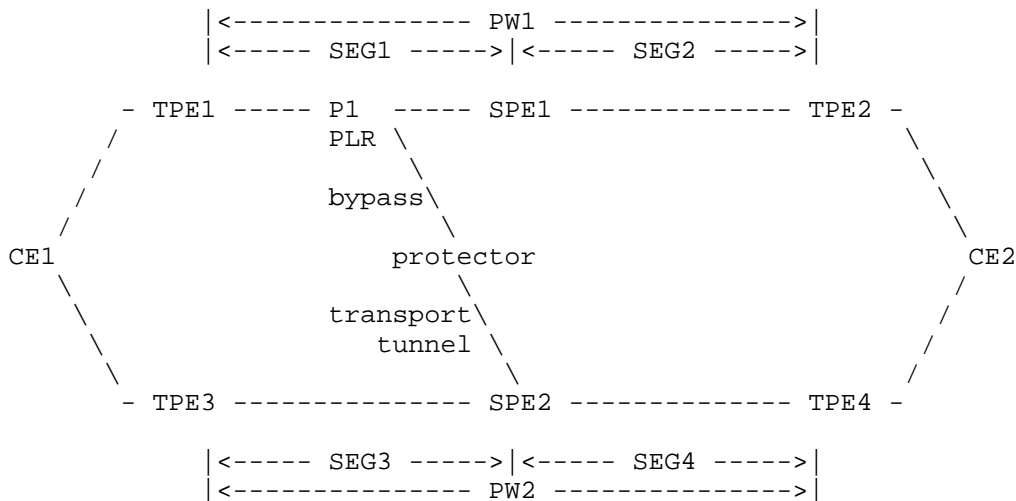
                              Figure 10

   The centralized protector model allows multiple primary PEs to share
   one protector. Each primary PE may need only one protector.
   Therefore, the number of context identifiers needed by a network may
   be bound to the number of primary PEs.

4.5. Transport Tunnel

   A PW is associated with a pair of {primary PE, protector}, which is
   represented by a unique context identifier. The ingress PE of the PW
   sets up or resolves a transport tunnel by using the context
   identifier rather than a private IP address of the primary PE as the
   destination. This not only ensures that the PW is transported to the
   primary PE but also facilitates bypass tunnel establishment at PLR,
   because the context identifier contains the identity of the protector
   as well. This is also the case for a multi-segment PW, where the
   ingress PE and egress PE are T-/S-PEs.

   An ingress PE learns the association between a PW and a context
   identifier from the primary PE, which MUST advertise the context
   identifier as a "third-party next hop" via the IPv4/v6 Interface_ID
   TLV [RFC3471] [RFC3472] in the LDP Label Mapping message of the PW.

   In an ECMP scenario, a transport tunnel may have multiple penultimate
   hop routers. Each of them SHOULD act as a PLR independently. Also
   in an ECMP scenario, a penultimate hop router may have ECMPs to the
   primary PE. At least one path of the ECMPs must be a direct link to
   the primary PE, qualifying the router as a penultimate hop. The
   other paths of the ECMPs may be direct links or indirect paths to the

primary PE. In egress PE node protection and S-PE node protection,
when a node failure is detected, or a link failure is detected on a
direct link and treated as a node failure, the penultimate hop router
SHOULD act as a PLR and reroute the entire traffic of the ECMPs to
the protector.

4.6. Bypass Tunnel

   A PLR may protect multiple PWs associated with one or multiple pairs
   of {primary PE, protector}. The PLR MUST establish a bypass tunnel
   to each protector for each context identifier associated with that
   protector. The destination of the bypass tunnel MUST be the context
   identifier (Section 4.3.1). Since the PLR is a transit router of the
   transport tunnel, it SHOULD derive the context identifier from the
   destination of the transport tunnel.

   For example, in Figures 7 and 9, a bypass tunnel is established from
   PE2 (PLR for egress AC failure) to the protector, and another bypass
   tunnel is established from P3 (PLR for egress node failure) to the
   protector. In Figures 8 and 10, a bypass tunnel is established from
   P1 (PLR for S-PE failure) to the protector.

   In local repair, a PLR reroutes traffic to the protector through a
   bypass tunnel, with the PW label intact in the packets. This
   normally involves pushing a label to the label stack, if the bypass
   tunnel is an MPLS tunnel, or pushing an IP header to the packets, if
   the bypass tunnel is an IP tunnel. Upon receipt of the packets, the
   protector forwards them based on the PW label. Specifically, the
   protector uses the bypass tunnel as a context to determine the
   primary PE's label space. If the bypass tunnel is an MPLS tunnel,
   the protector should have assigned a non-reserved label to the bypass
   tunnel; hence, this label can serve as the context. This label is
   also called a "context label", as it is actually bound to the context
   identifier. If the bypass tunnel is an IP tunnel, the context
   identifier should be the destination address of the IP header.

   To be useful for local repair, a bypass tunnel MUST have the property
   that it is not affected by any topology changes caused by the
   failure. It MUST NOT traverse the primary PE or the penultimate link
   of the transport tunnel, or share any SRLG with the penultimate link.
   A bypass tunnel may be a TE tunnel with reserved bandwidth to avoid
   traffic congestion for rerouted traffic. A bypass tunnel should
   remain effective during local repair, until the traffic is moved to
   an alternative path, i.e., either the same PW over a fully functional
   transport tunnel or another fully functional PW.

   There is little or no benefit to protect a bypass tunnel. Therefore,
   a bypass tunnel SHOULD NOT be protected against a transit link
   failure, transit node failure, or egress node failure.

4.7. Examples of Forwarding State

   This section provides some detailed examples of forwarding state on
   the PLR, protector, and other relevant routers.

   A protector learns PW labels from all the primary PEs that it
   protects (Section 6.2) and maintains the PW labels in separate label
   spaces on a per-primary-PE basis. In the control plane, each label
   space is identified by the context identifier of the corresponding
   {primary PE, protector}. In the forwarding plane, the label space is
   indicated by the bypass tunnel(s) destined for the context
   identifier.

4.7.1. Co-located Protector Model

   In Figure 11, PE4 is a co-located protector that protects PW1 against
   egress AC failure and egress node failure. It maintains a label
   space for PE2, which is identified by the context identifier of {PE2,
   PE4}. It learns PW1's label from PE2 and installs a forwarding entry
   for the label in that label space. The next hop of the forwarding
   entry indicates a label pop with an outgoing interface pointing to
   the backup AC PE4-CE2.
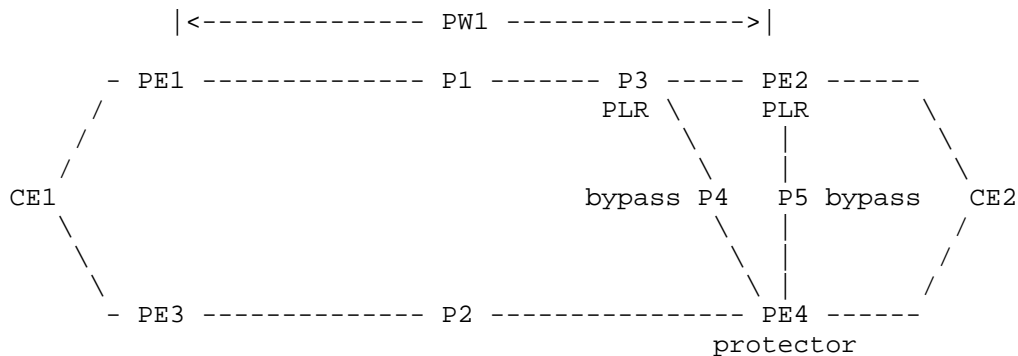
```
            |<------------- PW1 --------------->|

       - PE1 ------------- P1 ------- P3 ----- PE2 ------
      /                              PLR \     PLR       \
     /                                    \     |         \
    /                                      \    |          \
  CE1                                 bypass P4  P5 bypass   CE2
    \                                      \    |          /
     \                                      \   |         /
      \                                      \  |        /
       - PE3 ------------- P2 --------------- PE4 ------
                                             protector

            |<------------- PW2 --------------->|
```

```
         PW1's label assigned by PE2: 100
         PW2's label assigned by PE4: 200
         On P3: </t>
             Incoming label of transport tunnel to PE2: 1000
             Outgoing label of transport tunnel to PE2: implicit null
             Outgoing label of bypass tunnel to PE4: 2000
         On PE2:
             Outgoing label of bypass tunnel to PE4: 3000
         On PE4:
             Context label (incoming label of bypass tunnels): 999

         Forwarding state on P3:
         label 1000 -- primary next hop: pop, to PE2
                       backup next hop:  swap 2000, to P4

         Forwarding state on PE2:
         label 100 -- primary next hop: pop, to CE2
                      backup next hop:  push 3000, to P5

         Forwarding state on P4:
         label 2000 -- next hop: swap 999, to PE4

         Forwarding state on P5:
         label 3000 -- next hop: swap 999, to PE4

         Forwarding state on PE4:
         label 200 -- next hop: pop, to CE2
         label 999 -- next hop: label table of PE2's label space

         Label table of PE2's label space on PE4:
         label 100 -- next hop: pop, to CE2
```

                            Figure 11

In Figure 12, SPE2 is a co-located protector that protects PW1
against S-PE failure. It maintains a label space for SPE1, which is
identified by the context identifier of {SPE1, SPE2}. It learns
SEG1's label from SPE1 and installs a forwarding entry in the label
space. The next hop of the forwarding entry indicates a label swap
to SEG4's label and a label push with the label of a transport tunnel
to TPE4.

```
         |<--------------- PW1 --------------->|
         |<----- SEG1 ----->|<----- SEG2 ----->|

      - TPE1 ----- P1  ----- SPE1 --- P3 ------- TPE2 -
     /              PLR \                          \
    /                    \                          \
  CE1                bypass P2                        CE2
    \                    \                          /
     \                    \                        /
      - TPE3 -------------- SPE2 --- P4 ------- TPE4 -
                      protector

         |<----- SEG3 ----->|<----- SEG4 ----->|
         |<-------------- PW2 --------------->|
```

```
        SEG1's label assigned by SPE1: 100
        SEG2's label assigned by TPE2: 200
        SEG3's label assigned by SPE2: 300
        SEG4's label assigned by TPE4: 400
        On P1:
           Incoming label of transport tunnel to SPE1: 1000
           Outgoing label of transport tunnel to SPE1: implicit null
           Outgoing label of bypass tunnel to SPE2: 2000
        On SPE1:
           Outgoing label of transport tunnel to TPE2: 3000
        On SPE2:
           Outgoing label of transport tunnel to TPE4: 4000
           Context label (incoming label of bypass tunnel): 999


        Forwarding state on P1:
        label 1000 -- primary next hop: pop, to SPE1
                      backup next hop:  swap 2000, to P2


        Forwarding state on SPE1:
        label 100 -- next hop: swap 200, push 3000, to P3


        Forwarding state on P2:
        label 2000 -- next hop: swap 999, to SPE2


        Forwarding state on SPE2:
        label 300 -- next hop: swap 400, push 4000, to P4
        label 999 -- next hop: label table of SPE1's label space


        Label table of SPE1's label space on SPE2:
        label 100 -- next hop: swap 400, push 4000, to P4
```

                              Figure 12

4.7.2. Centralized Protector Model

   In the centralized protector model, for each primary PW of which the
   protector is not a backup (S-)PE, the protector MUST also learn the
   label of the backup PW from the backup (S-)PE (Section 6.3). This is
   the backup (S-)PE that the protector will forward traffic to. The
   protector MUST install a forwarding entry with a label swap from the
   primary PW's label to the backup PW's label and a label push with the
   label of a transport tunnel to the backup (S-)PE.

   In Figure 13, the protector is a centralized protector that protects
   PW1 against egress AC failure and egress node failure. It maintains
   a label space for PE2, which is identified by the context identifier
   of {PE2, protector}. It learns PW1's label from PE2 and PW2's label
   from PE4. It installs a forwarding entry for PW1's label in the
   label space. The next hop of the forwarding entry indicates a label
   swap to PW2's label and a label push with the label of a transport
   tunnel to PE4.

```
                  |<------------- PW1 --------------->|

          - PE1 ------------- P1 ------- P3 ------ PE2 ----
         /                             PLR \      PLR      \
        /                                   \    /          \
       /                              bypass P5  P6 bypass   \
      /                                        \  /           \
     /                                          \/             \
    CE1                                      protector          CE2
     \                                          \              /
      \                               transport  \            /
       \                                tunnel P7  \         /
        \                                           \       /
         \                                           \     /
          - PE3 ------------- P2 ---------------- PE4 ----

                  |<------------- PW2 --------------->|

        PW1's label assigned by PE2: 100
        PW2's label assigned by PE4: 200
        On P3:
            Incoming label of transport tunnel to PE2: 1000
            Outgoing label of transport tunnel to PE2: implicit null
            Outgoing label of bypass tunnel to protector: 2000
        On PE2:
            Outgoing label of bypass tunnel to protector: 3000
        On protector:
            Context label (incoming label of bypass tunnels): 999
            Outgoing label of transport tunnel to PE4: 4000
```

```
        Forwarding state on P3:
        label 1000 -- primary next hop: pop, to PE2
                      backup next hop:  swap 2000, to P5


        Forwarding state on PE2:
        label 100 -- primary next hop: pop, to CE2
                     backup next hop:  push 3000, to P6


        Forwarding state on P5:
        label 2000 -- next hop: swap 999, to protector


        Forwarding state on P6:
        label 3000 -- next hop: swap 999, to protector


        Forwarding state on P7:
        label 4000 -- next hop: pop, to PE4


        Forwarding state on PE4:
        label 200 -- next hop: pop, to CE2


        Forwarding state on protector:
        label 999 -- next hop: label table of PE2's label space


        Label table of PE2's label space on protector:
        label 100 -- next hop: swap 200, push 4000, to P7
```
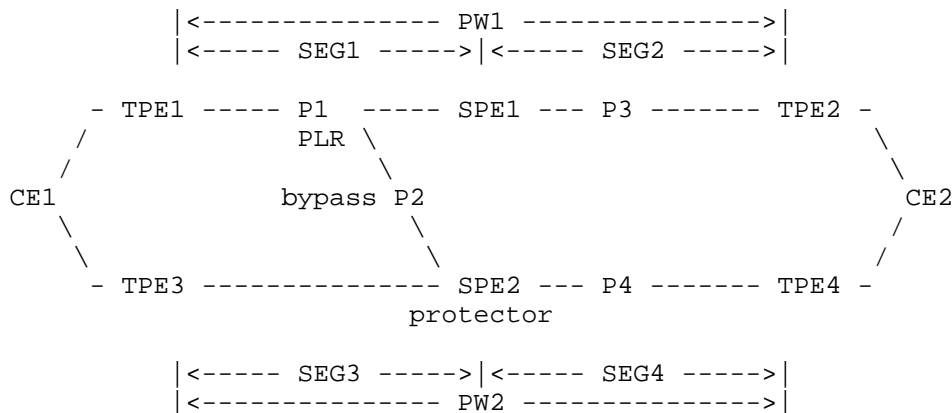
                             Figure 13

   In Figure 14, the protector is a centralized protector that protects
   the PW segment SEG1 of PW1 against the node failure of SPE1. It
   maintains a label space for SPE1, which is identified by the context
   identifier of {SPE1, protector}. It learns SEG1's label from SPE1
   and SEG3's label from SPE2. It installs a forwarding entry for
   SEG1's label in the label space. The next hop of the forwarding
   entry indicates a label swap to SEG3's label and a label push with
   the label of a transport tunnel to TPE4.

```
             |<-------------- PW1 --------------->|
             |<----- SEG1 ----->|<----- SEG2 ----->|


        - TPE1 ----- P1 ----- SPE1 --- P2 -------- TPE2 -
       /             PLR \                              \
      /                   \                              \
     /              bypass P4                             \
    /                       \                              \
   /                         \                              \
  CE1                     protector                         CE2
   \                         \                              /
    \                         \                            /
     \              transport P5                          /
      \                 tunnel \                         /
       \                        \                       /
        - TPE3 -------------- SPE2 --- P3 -------- TPE4 -


             |<----- SEG3 ----->|<----- SEG4 ----->|
             |<-------------- PW2 --------------->|
```

```
         SEG1's label assigned by SPE1: 100
         SEG2's label assigned by TPE2: 200
         SEG3's label assigned by SPE2: 300
         SEG4's label assigned by TPE4: 400
         On P1:
            Incoming label of transport tunnel to SPE1: 1000
            Outgoing label of transport tunnel to SPE1: implicit null
            Outgoing label of bypass tunnel to protector: 2000
         On SPE1:
            Outgoing label of transport tunnel to TPE2: 3000
         On SPE2:
            Outgoing label of transport tunnel to TPE4: 4000
         On protector:
            Context label (incoming label of bypass tunnel): 999
            Outgoing label of transport tunnel to SPE2: 5000

         Forwarding state on P1:
         label 1000 -- primary next hop: pop, to SPE1
                       backup next hop:  swap 2000, to P4

         Forwarding state on SPE1:
         label 100 -- next hop: swap 200, push 3000, to P2

         Forwarding state on P4:
         label 2000 -- next hop: swap 999, to protector

         Forwarding state on P5:
         label 5000 -- next hop: pop, to SPE2
```

        Forwarding state on SPE2:
        label 300 -- next hop: swap 400, push 4000, to P3

        Forwarding state on protector:
        label 999 -- next hop: label table of SPE1's label space

        Label table of SPE1's label space on protector:
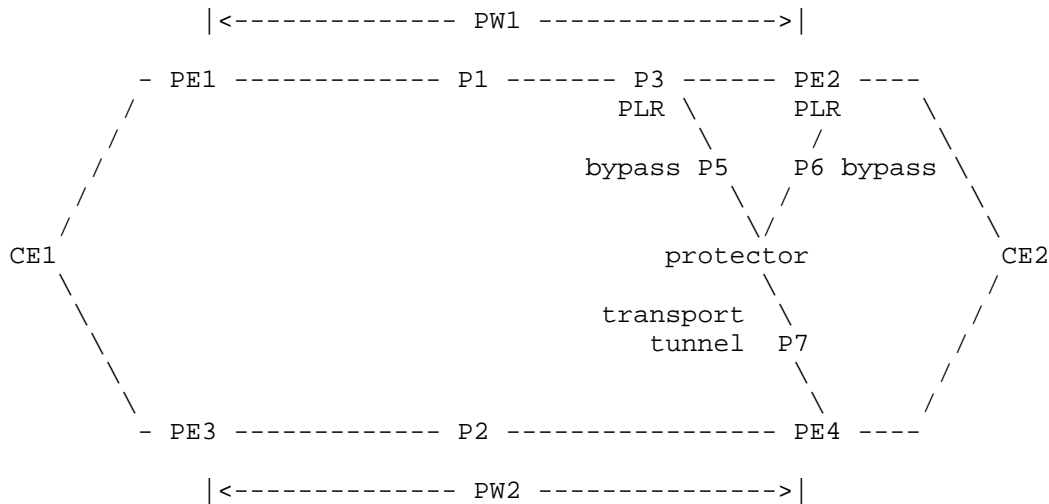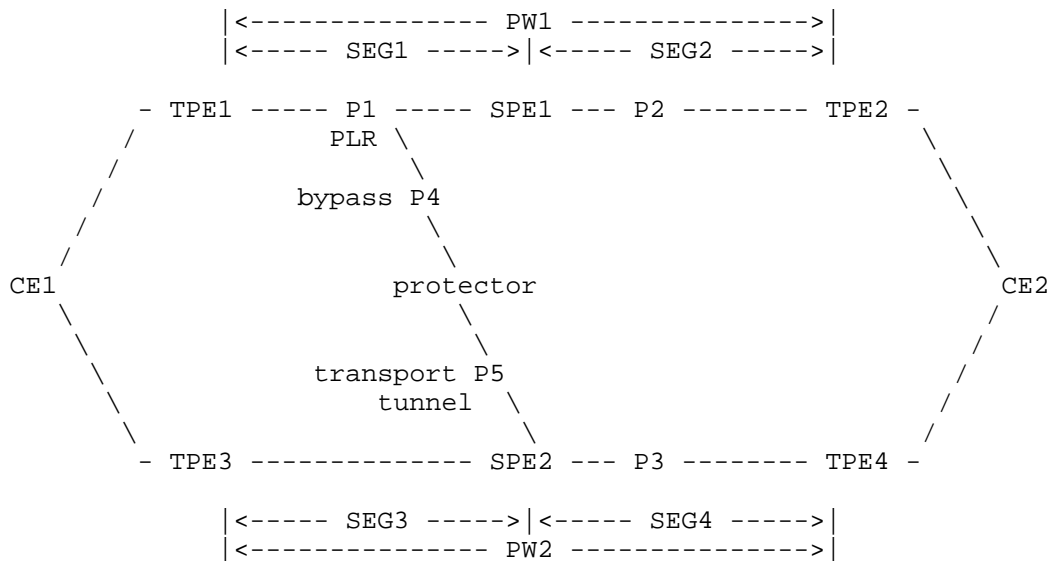   label 100 -- next hop: swap 300, push 5000, to P5

                            Figure 14

5. Restorative and Revertive Behaviors

   Subsequent to local repair, there are three strategies for a network
   to restore traffic to a fully functional alternative path:

   o Global repair

      If the ingress CE is multihomed (Figure 1), it MAY switch the
      traffic to the backup AC, which is bound to the backup PW.
      Alternatively, if the ingress PE hosts a backup PW (Figure 2), the
      ingress PE MAY switch the traffic to the backup PW. These
      procedures are referred to as global repair. Possible triggers of
      global repair include PW status notification, Virtual Circuit
      Connectivity Verification (VCCV) [RFC5085] [RFC5885], BFD,
      end-to-end OAM between CEs, and others.

   o Control-plane convergence

      In egress PE node protection and S-PE node protection, it is
      possible that the failure is limited to the link between the PLR
      and the primary PE, whereas the primary PE is still operational.
      In this case, the PLR or an upstream router on the transport
      tunnel MAY reroute the tunnel around the link via an alternative
      path to the primary PE. Thus, the transport tunnel can heal and
      continue to carry the PW to the primary PE. This procedure is
      driven by control-plane convergence on the new topology.

   o Local reversion

      The PLR MAY move traffic back to the primary PW, after the failure
      is resolved. In egress AC protection, upon detecting that the
      primary AC is restored, the PLR MAY start forwarding traffic over
      the AC again. Likewise, in egress PE node protection and S-PE
      node protection, upon detecting that the primary PE is restored,
      the PLR MAY re-establish the transport tunnel to the primary PE
      and move the traffic from the bypass tunnel back to the transport
      tunnel. These procedures are referred to as local reversion.

It is RECOMMENDED that the fast protection mechanism SHOULD be used
in conjunction with global repair. Particularly in the case of
egress PE and S-PE node failures, if the ingress PE or the protector
loses communication with the egress PE or S-PE for an extensive
period of time, the LDP session may go down. Consequently, the
ingress PE may bring down the primary PW completely, or the protector
may remove the forwarding entry of the primary PW label. In either
case, the service will be disrupted. In other words, although the
mechanism can temporarily repair traffic, control-plane state may
eventually expire if the failure persists. Therefore, global repair
SHOULD take place in a timely manner to move traffic to a fully
functional alternative path.

Control-plane convergence may automatically happen as control-plane
protocols react to the new topology. However, it is only applicable
to the specific link failure scenario described above.

Local reversion is optional. In the circumstances where the failure
is caused by resource flapping, local reversion MAY be dampened to
limit potential disruption. Local reversion MAY be disabled
completely by configuration.

6. LDP Extensions

As described in previous sections, a targeted LDP session MUST be
established between each pair of primary PE and protector. The
primary PE sends a Label Mapping message over this session to
advertise primary PW labels to the protector. In the centralized
protector model, a targeted LDP session MUST also be established
between a backup (S-)PE and the protector. The backup PE sends a
Label Mapping message over this session to advertise backup PW labels
to the protector.

To support the procedures, this document defines a new "Protection
FEC Element" TLV. The Label Mapping messages of both the LDP
sessions above MUST carry this TLV to identify a primary PW.
Specifically, in the centralized protector model, the Protection FEC
Element TLV advertised by a backup (S-)PE MUST match the one
advertised by the primary PE, so that the protector can associate the
primary PW's label with the backup PW's label and perform a label
swap. The backup (S-)PE builds such a Protection FEC Element TLV
based on local configuration.

This document also defines a new "Egress Protection Capability" TLV
as a new type of Capability Parameter TLV [RFC5561], to allow a
protector to announce its capability of processing the above
Protection FEC Element TLV and performing context-specific label
switching for PW labels.

The procedures in this section are only applicable if the protector
advertises the Egress Protection Capability TLV, the primary PE
supports the advertisement of the Protection FEC Element TLV, and in
the centralized protector model, the backup PE also supports the
advertisement of the Protection FEC Element TLV.

6.1. Egress Protection Capability TLV

A protector MUST advertise the Egress Protection Capability TLV in
its Initialization message and Capability message over the LDP
session with a primary PE. In the centralized protector model, the
protector MUST also advertise the TLV over the LDP session with a
backup PE. The TLV carries one or multiple context identifiers. To
the primary PE, the TLV MUST carry the context identifier of the
{primary PE, protector}. In the centralized protector model, the TLV
MUST carry multiple context identifiers to the backup PE, one for
each {primary PE, protector} where the backup PE serves as a backup
for the primary PE. This TLV MUST NOT be advertised by the primary
PE or the backup PE to the protector.

The processing of the Egress Protection Capability TLV by a receiving
router MUST follow the procedures defined in [RFC5561]. In
particular, the router MUST advertise PW information to the protector
by using the Protection FEC Element TLV, only after it has received
the Egress Protection Capability TLV from the protector. It MUST
validate each context identifier included in the TLV and advertise
the information of only the PWs that are associated with the context
identifier. It MUST withdraw previously advertised Protection FEC
TLVs, when the protector has withdrawn a previously advertised
context identifier or the entire Egress Protection Capability TLV via
a Capability message.

The encoding of the Egress Protection Capability TLV is defined
below. It conforms to the format of Capability Parameter TLV
specified in [RFC5561].

```
  0                   1                   2                   3
  0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |U|F| Egress Protection (0x0974)|               Length          |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
 |S| Reserved    |                                               |
 +-+-+-+-+-+-+-+-+                                               |
 |                                                               |
 |                                                               |
 ~                 Capability Data = context identifier(s)       ~
 |                                                               |
 |                                               +-+-+-+-+-+-+-+-+
 |                                               |
 +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

                              Figure 15

   The U-bit MUST be set to 1, so that a receiver MUST silently ignore
   this TLV if unknown to it and continue processing the rest of the
   message.

   The F-bit MUST be set to 0 since this TLV is sent only in
   Initialization and Capability messages, which are not forwarded.

   The TLV code point is 0x0974.

   The S-bit indicates whether the sender is advertising (S=1) or
   withdrawing (S=0) the capability.

   The Capability Data field is encoded with the context identifiers of
   the {primary PE, protector} pairs for which the advertiser is the
   protector.

6.2. PW Label Distribution from Primary PE to Protector

   A primary PE MUST advertise a primary PW's label to a protector by
   sending a Label Mapping message. The message includes a Protection
   FEC Element TLV (see Section 6.4 for encoding) and an Upstream-
   Assigned Label TLV [RFC6389] encoded with the PW's label. The
   combination of the Protection FEC Element TLV and the PW label
   represents the primary PE's forwarding state for the PW. The Label
   Mapping message MUST also carry an IPv4/v6 Interface_ID TLV [RFC3471]
   [RFC6389] encoded with the context identifier of the {primary PE,
   protector}.

   The protector that receives this Label Mapping message MUST install a
   forwarding entry for the PW label in the label space identified by
   the context identifier. The next hop of the forwarding entry MUST
   ensure that packets are sent towards the target CE via a backup AC or

a backup (S-)PE, depending on the protection scenario. The protector
MUST silently discard a Label Mapping message if the included context
identifier is unknown to it.

6.3. PW Label Distribution from Backup PE to Protector

In the centralized protector model, a backup PE MUST advertise a
backup PW's label to the protector by sending a Label Mapping
message. The message includes a Protection FEC Element TLV and a
Generic Label TLV encoded with the backup PW's label. This
Protection FEC Element MUST be identical to the Protection FEC
Element TLV that the primary PE advertises to the protector
(Section 6.2). This is achieved through configuration on the backup
PE. The context identifier MUST NOT be encoded in an Interface_ID
TLV in this message.

The protector that receives this Label Mapping message MUST associate
the backup PW with the primary PW, based on the common Protection FEC
Element TLV. It MUST distinguish between the Label Mapping message
from the primary PE and the Label Mapping message from the backup PE
based on the respective presence and absence of a context identifier
in the Interface_ID TLV. It MUST install a forwarding entry for the
primary PW's label in the label space identified by the context
identifier. The next hop of the forwarding entry MUST indicate a
label swap to the backup PW's label, followed by a label push or IP
header push for a transport tunnel to the backup PE.

6.4. Protection FEC Element TLV

   The Protection FEC Element TLV has type 0x83. Its format is defined
   below:

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Type(0x83)   |   Reserved    | Encoding Type |    Length     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                                                               |
|                                                               |
~                       PW Information                          ~
|                                                               |
|                       +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                       |
+-+-+-+-+-+-+-+-+-+-+-+-+
```
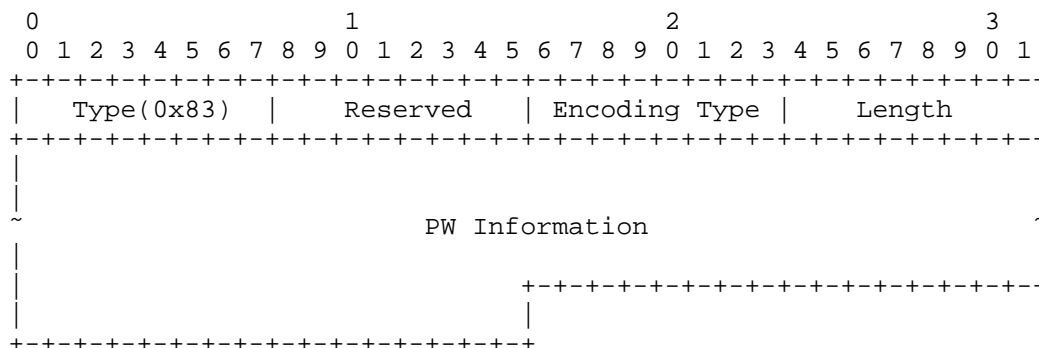
                             Figure 16

   - Encoding Type

      Type of encoding format of the PW Information field. The
      following types are defined, corresponding to the PWid FEC Element
      and Generalized PWid FEC Element defined in [RFC8077].

      1 - PWid FEC Element with IPv4 PE addresses (Section 6.4.1).

      2 - Generalized PWid FEC Element with IPv4 PE addresses
          (Section 6.4.2).

      3 - PWid FEC Element with IPv6 PE addresses (Section 6.4.3).

      4 - Generalized PWid FEC Element with IPv6 PE addresses
          (Section 6.4.4).

   - Length

      Length of the PW Information field in octets.

   - PW Information

      Field of variable length that specifies a PW.

6.4.1. Encoding Format for PWid with IPv4 PE Addresses

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |   Type(0x83)  |   Reserved    |  Enc Type(1)  |  Length(20)   |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                   Ingress PE IPv4 Address                     |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                    Egress PE IPv4 Address                     |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                          Group ID                            |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                           PW ID                              |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |C|          PW Type            |           Reserved           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

                             Figure 17

   - Ingress PE IPv4 Address

      IPv4 address of the ingress PE of PW.

   - Egress PE IPv4 Address

      IPv4 address of the egress PE of PW.

   - Group ID

      An arbitrary 32-bit value that represents a group of PWs and that
      is used to create groups in the PW space.

   - PW ID

      A non-zero 32-bit connection ID that, together with the PW Type
      field, identifies a particular PW.

   - Control word bit (C)

      A bit that flags the presence of a control word on this PW. If
      C = 1, control word is present; if C = 0, control word is not
      present.

   - PW Type

      A 15-bit quantity that represents the type of PW.

6.4.2. Encoding Format for Generalized PWid with IPv4 PE Addresses

```
     0                   1                   2                   3
     0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |  Type(0x83)   |    Reserved   | Enc Type(2)   |    Length     |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                     Ingress PE IPv4 Address                   |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |                     Egress PE IPv4 Address                    |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |C|          PW Type            |            Reserved           |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |   AGI Type    |     Length    |            AGI Value          |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    ~                       AGI Value (contd.)                      ~
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |   AII Type    |     Length    |            SAII Value         |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    ~                       SAII  Value (contd.)                    ~
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    |   AII Type    |     Length    |            TAII Value         |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
    ~                       TAII Value (contd.)                     ~
    |                                                               |
    +-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```
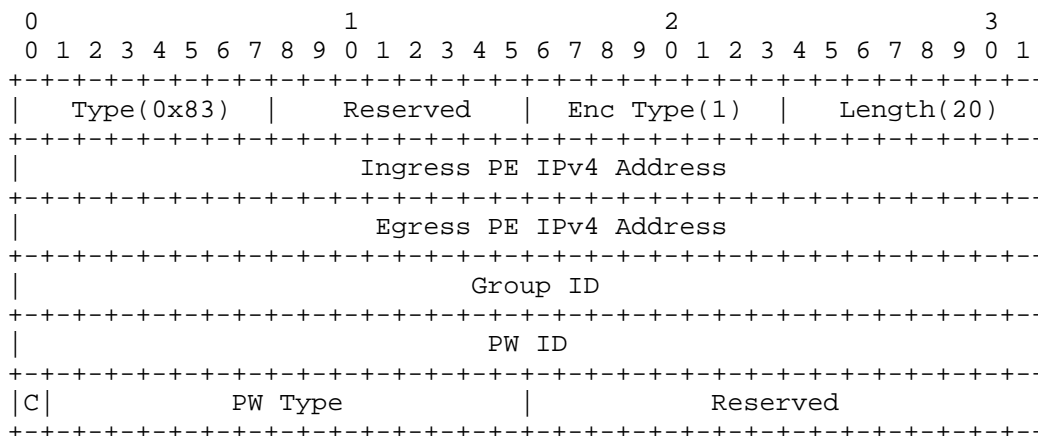
Figure 18

   - Ingress PE IPv4 Address

      IPv4 address of the ingress PE of PW.

   - Egress PE IPv4 Address

      IPv4 address of the egress PE of PW.

   - Control word bit (C)

      A bit that flags the presence of a control word on this PW. If
      C = 1, control word is present; if C = 0, control word is not
      present.

   - PW Type

      A 15-bit quantity that represents the type of PW.

   - AGI Type, Length, AGI Value

      Attachment Group Identifier of PW.

   - AII Type, Length, SAII Value

      Source Attachment Individual Identifier of PW.

   - AII Type, Length, TAII Value

      Target Attachment Individual Identifier of PW.

6.4.3. Encoding Format for PWid with IPv6 PE Addresses

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|  Type(0x83)   |    Reserved   |  Enc Type(3)  |  Length(44)   |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                     Ingress PE IPv6 Address                   ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                     Egress PE IPv6 Address                    ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                         Group ID                             |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|                          PW ID                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|C|           PW Type         |              Reserved          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

                             Figure 19

   - Ingress PE IPv6 Address

      IPv6 address of the ingress PE of PW; 16 octets.

   - Egress PE IPv6 Address

      IPv6 address of the egress PE of PW; 16 octets.

   - Group ID

      An arbitrary 32-bit value that represents a group of PWs and that
      is used to create groups in the PW space.

   - PW ID

      A non-zero 32-bit connection ID that, together with the PW Type
      field, identifies a particular PW.

   - Control word bit (C)

      A bit that flags the presence of a control word on this PW. If
      C = 1, control word is present; if C = 0, control word is not
      present.

   - PW Type

      A 15-bit quantity that represents the type of PW.

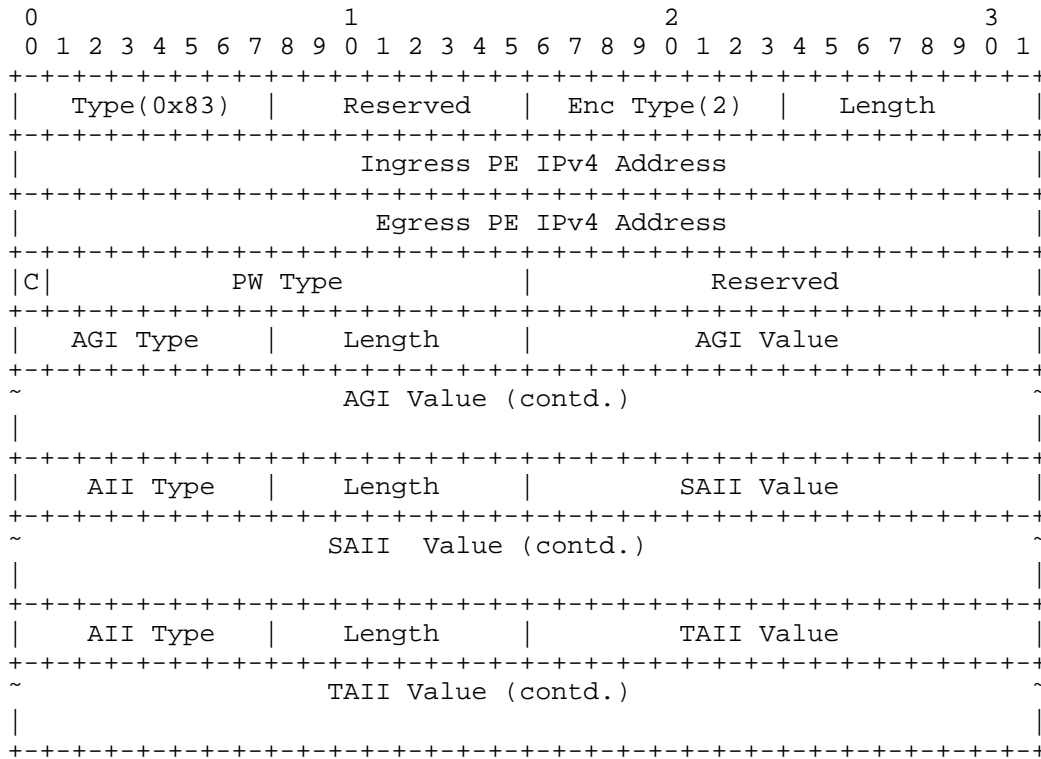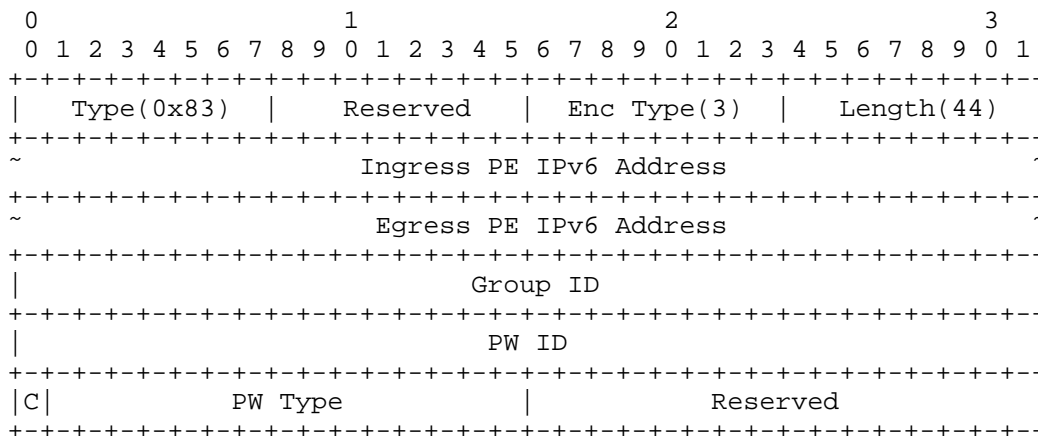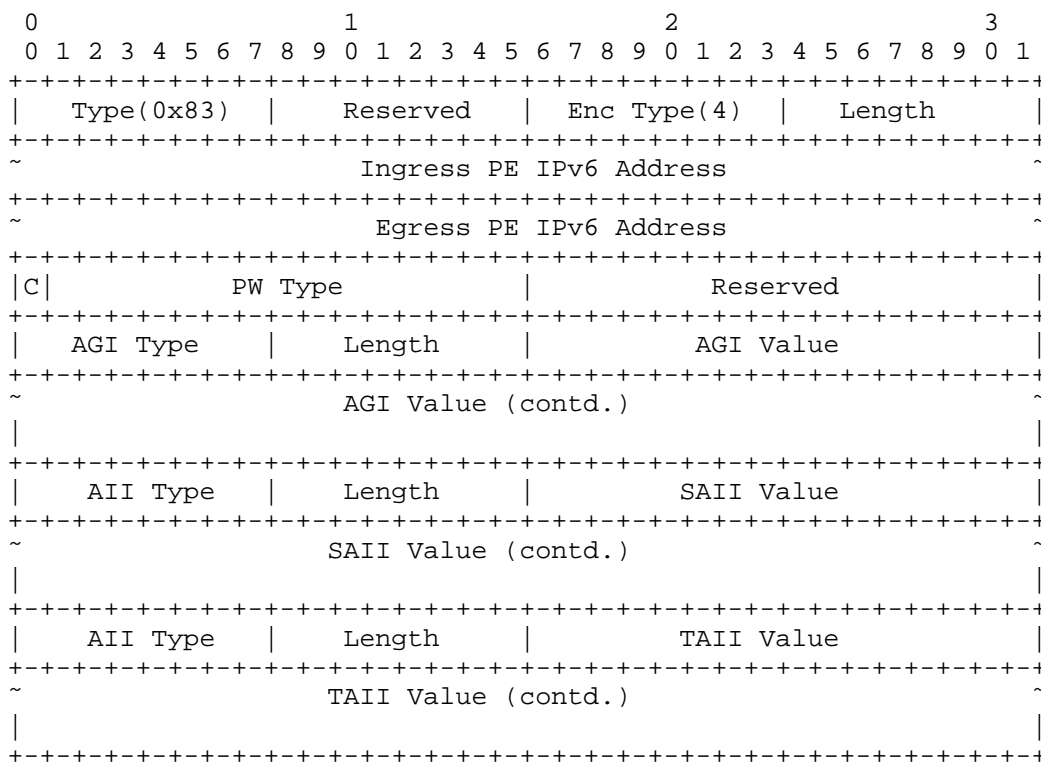6.4.4. Encoding Format for Generalized PWid with IPv6 PE Addresses

```
 0                   1                   2                   3
 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   Type(0x83)  |    Reserved   |  Enc Type(4)  |    Length     |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                     Ingress PE IPv6 Address                   ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                     Egress PE IPv6 Address                    ~
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|C|         PW Type             |            Reserved           |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   AGI Type    |    Length     |            AGI Value          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                       AGI Value (contd.)                      ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   AII Type    |    Length     |           SAII Value          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                      SAII Value (contd.)                      ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
|   AII Type    |    Length     |           TAII Value          |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
~                      TAII Value (contd.)                      ~
|                                                               |
+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+-+
```

                               Figure 20

- Ingress PE IPv6 Address

  IPv6 address of the ingress PE of PW; 16 octets.

- Egress PE IPv6 Address

  IPv6 address of the egress PE of PW; 16 octets.

- Control word bit (C)

  A bit that flags the presence of a control word on this PW. If
  C = 1, control word is present; if C = 0, control word is not
  present.

- PW Type

  A 15-bit quantity that represents the type of PW.

- AGI Type, Length, AGI Value

  Attachment Group Identifier of PW.

- AII Type, Length, SAII Value

  Source Attachment Individual Identifier of PW.

- AII Type, Length, TAII Value

  Target Attachment Individual Identifier of PW.

7. IANA Considerations

   This document defines a new Egress Protection Capability TLV in
   Section 6. IANA has assigned value 0x0974 for it in the "TLV Type
   Name Space" registry.

   This document defines a new Protection FEC Element TLV in Section 6.
   IANA assigned value 0x83 for it in the "Forwarding Equivalence Class
   (FEC) Type Name Space" registry per RFC 7358. IANA has updated the
   registry to also reference this document.

8. Security Considerations

   In this document, PW traffic can be temporarily rerouted by a PLR to
   a protector. In the centralized protector scenario, the traffic can
   be further rerouted to a backup PE. In the control plane, there is a
   targeted LDP session between a primary PE and a protector. In the
   centralized protector scenario, there is also a targeted LDP session
   between a backup PE and a protector. In all scenarios, traffic
   rerouting via PLRs, protectors, and backup PEs is planned and
   anticipated, and backup PEs can be used anyway to host PWs and LDP
   sessions. Hence, the rerouted traffic and the LDP sessions
   introduced in this document should not be viewed as a new security
   threat.

   In general, [RFC5920] describes the security framework for MPLS
   networks. [RFC3209] describes the security considerations for RSVP
   Label Switched Paths (LSPs). [RFC5036] describes the security
   considerations for the base LDP specification. [RFC5561] describes
   the security considerations that apply when using the LDP capability
   mechanism. All these security frameworks and considerations apply to
   this document as well.

9. References

9.1. Normative References

   [RFC2119] Bradner, S., "Key words for use in RFCs to Indicate
              Requirement Levels", BCP 14, RFC 2119,
              DOI 10.17487/RFC2119, March 1997,
              <http://www.rfc-editor.org/info/rfc2119>.

   [RFC8077] Martini, L., Ed. and G. Heron, Ed., "Pseudowire Setup and
              Maintenance Using the Label Distribution Protocol (LDP)",
              STD 84, RFC 8077, DOI 10.17487/RFC8077, February 2017,
              <http://www.rfc-editor.org/info/rfc8077>.

   [RFC5331] Aggarwal, R., Rekhter, Y., and E. Rosen, "MPLS Upstream
              Label Assignment and Context-Specific Label Space",
              RFC 5331, DOI 10.17487/RFC5331, August 2008,
              <http://www.rfc-editor.org/info/rfc5331>.

   [RFC5561] Thomas, B., Raza, K., Aggarwal, S., Aggarwal, R., and JL.
              Le Roux, "LDP Capabilities", RFC 5561,
              DOI 10.17487/RFC5561, July 2009,
              <http://www.rfc-editor.org/info/rfc5561>.

   [RFC3471] Berger, L., Ed., "Generalized Multi-Protocol Label
              Switching (GMPLS) Signaling Functional Description",
              RFC 3471, DOI 10.17487/RFC3471, January 2003,
              <http://www.rfc-editor.org/info/rfc3471>.

   [RFC3472] Ashwood-Smith, P., Ed. and L. Berger, Ed., "Generalized
              Multi-Protocol Label Switching (GMPLS) Signaling
              Constraint-based Routed Label Distribution Protocol (CR-
              LDP) Extensions", RFC 3472, DOI 10.17487/RFC3472, January
              2003, <http://www.rfc-editor.org/info/rfc3472>.

   [RFC6389] Aggarwal, R. and JL. Le Roux, "MPLS Upstream Label
              Assignment for LDP", RFC 6389, DOI 10.17487/RFC6389,
              November 2011, <http://www.rfc-editor.org/info/rfc6389>.

   [RFC4090] Pan, P., Ed., Swallow, G., Ed., and A. Atlas, Ed., "Fast
              Reroute Extensions to RSVP-TE for LSP Tunnels", RFC 4090,
              DOI 10.17487/RFC4090, May 2005,
              <http://www.rfc-editor.org/info/rfc4090>.

   [RFC5286] Atlas, A., Ed. and A. Zinin, Ed., "Basic Specification for
              IP Fast Reroute: Loop-Free Alternates", RFC 5286,
              DOI 10.17487/RFC5286, September 2008,
              <http://www.rfc-editor.org/info/rfc5286>.

   [RFC7812] Atlas, A., Bowers, C., and G. Enyedi, "An Architecture for
              IP/LDP Fast Reroute Using Maximally Redundant Trees (MRT-
              FRR)", RFC 7812, DOI 10.17487/RFC7812, June 2016,
              <http://www.rfc-editor.org/info/rfc7812>.

   [RFC3031] Rosen, E., Viswanathan, A., and R. Callon, "Multiprotocol
              Label Switching Architecture", RFC 3031,
              DOI 10.17487/RFC3031, January 2001,
              <http://www.rfc-editor.org/info/rfc3031>.

9.2. Informative References

   [RFC5920] Fang, L., Ed., "Security Framework for MPLS and GMPLS
              Networks", RFC 5920, DOI 10.17487/RFC5920, July 2010,
              <http://www.rfc-editor.org/info/rfc5920>.

   [RFC3985] Bryant, S., Ed. and P. Pate, Ed., "Pseudo Wire Emulation
              Edge-to-Edge (PWE3) Architecture", RFC 3985,
              DOI 10.17487/RFC3985, March 2005,
              <http://www.rfc-editor.org/info/rfc3985>.

   [RFC3209] Awduche, D., Berger, L., Gan, D., Li, T., Srinivasan, V.,
              and G. Swallow, "RSVP-TE: Extensions to RSVP for LSP
              Tunnels", RFC 3209, DOI 10.17487/RFC3209, December 2001,
              <http://www.rfc-editor.org/info/rfc3209>.

   [RFC5036] Andersson, L., Ed., Minei, I., Ed., and B. Thomas, Ed.,
              "LDP Specification", RFC 5036, DOI 10.17487/RFC5036,
              October 2007, <http://www.rfc-editor.org/info/rfc5036>.

   [RFC5659] Bocci, M. and S. Bryant, "An Architecture for Multi-
              Segment Pseudowire Emulation Edge-to-Edge", RFC 5659,
              DOI 10.17487/RFC5659, October 2009,
              <http://www.rfc-editor.org/info/rfc5659>.

   [RFC5714] Shand, M. and S. Bryant, "IP Fast Reroute Framework",
              RFC 5714, DOI 10.17487/RFC5714, January 2010,
              <http://www.rfc-editor.org/info/rfc5714>.

   [RFC5880] Katz, D. and D. Ward, "Bidirectional Forwarding Detection
              (BFD)", RFC 5880, DOI 10.17487/RFC5880, June 2010,
              <http://www.rfc-editor.org/info/rfc5880>.

   [RFC5085] Nadeau, T., Ed. and C. Pignataro, Ed., "Pseudowire Virtual
              Circuit Connectivity Verification (VCCV): A Control
              Channel for Pseudowires", RFC 5085, DOI 10.17487/RFC5085,
              December 2007, <http://www.rfc-editor.org/info/rfc5085>.

   [RFC5885] Nadeau, T., Ed. and C. Pignataro, Ed., "Bidirectional
              Forwarding Detection (BFD) for the Pseudowire Virtual
              Circuit Connectivity Verification (VCCV)", RFC 5885,
              DOI 10.17487/RFC5885, June 2010,
              <http://www.rfc-editor.org/info/rfc5885>.

   [RFC7880] Pignataro, C., Ward, D., Akiya, N., Bhatia, M., and S.
              Pallagatti, "Seamless Bidirectional Forwarding Detection
              (S-BFD)", RFC 7880, DOI 10.17487/RFC7880, July 2016,
              <http://www.rfc-editor.org/info/rfc7880>.

Authors' Addresses

   Yimin Shen
   Juniper Networks
   10 Technology Park Drive
   Westford, MA 01886
   United States of America

   Phone: +1 9785890722
   Email: yshen@juniper.net

   Rahul Aggarwal
   Arktan, Inc.

   Email: raggarwa_1@yahoo.com

   Wim Henderickx
   Nokia
   Copernicuslaan 50
   2018 Antwerp
   Belgium

   Email: wim.henderickx@nokia.com

   Yuanlong Jiang
   Huawei Technologies Co., Ltd.
   Bantian, Longgang district
   Shenzhen 518129
   China

   Email: jiangyuanlong@huawei.com