Internet Engineering Task Force (IETF)
Request for Comments: 8679
Category: Standards Track
ISSN: 2070-1721

Y. Shen
M. Jeyananth
Juniper Networks
B. Decraene
Orange
H. Gredler
RtBrick Inc.
C. Michel
Deutsche Telekom
H. Chen
Futurewei

December 2019

MPLS Egress Protection Framework

Abstract

This document specifies a fast reroute framework for protecting IP/MPLS services and MPLS transport tunnels against egress node and egress link failures. For each type of egress failure, it defines the roles of Point of Local Repair (PLR), protector, and backup egress router and the procedures of establishing a bypass tunnel from a PLR to a protector. It describes the behaviors of these routers in handling an egress failure, including local repair on the PLR and context-based forwarding on the protector. The framework can be used to develop egress protection mechanisms to reduce traffic loss before global repair reacts to an egress failure and control-plane protocols converge on the topology changes due to the egress failure.

Table of Contents

1. Introduction

2. Specification of Requirements
3. Terminology
4. Requirements
5. Egress Node Protection
  5.1. Reference Topology
  5.2. Egress Node Failure and Detection
  5.3. Protector and PLR
  5.4. Protected Egress
  5.5. Egress-Protected Tunnel and Service
  5.6. Egress-Protection Bypass Tunnel
  5.7. Context ID, Context Label, and Context-Based Forwarding
  5.8. Advertisement and Path Resolution for Context ID
  5.9. Egress-Protection Bypass Tunnel Establishment
  5.10. Local Repair on PLR
  5.11. Service Label Distribution from Egress Router to Protector
  5.12. Centralized Protector Mode
6. Egress Link Protection
7. Global Repair
8. Operational Considerations
9. General Context-Based Forwarding
10. Example: Layer 3 VPN Egress Protection
  10.1. Egress Node Protection
  10.2. Egress Link Protection
  10.3. Global Repair
  10.4. Other Modes of VPN Label Allocation
11. IANA Considerations
12. Security Considerations
13. References
  13.1. Normative References
  13.2. Informative References
Acknowledgements
Authors' Addresses

1. Introduction

In MPLS networks, Label Switched Paths (LSPs) are widely used as transport tunnels to carry IP and MPLS services across MPLS domains. Examples of MPLS services are Layer 2 VPNs, Layer 3 VPNs, hierarchical LSPs, and others. In general, a tunnel may carry multiple services of one or multiple types, if the tunnel satisfies both individual and aggregate requirements (e.g., Class of Service (CoS) and QoS) of these services. The egress router of the tunnel hosts the service instances of the services. An MPLS service instance forwards service packets via an egress link to the service destination, based on a service label. An IP service instance does the same, based on an IP service address. The egress link is often called a Provider Edge - Customer Edge (PE-CE) link or Attachment Circuit (AC).

Today, local-repair-based fast reroute mechanisms (see [RFC4090], [RFC5286], [RFC7490], and [RFC7812]) have been widely deployed to protect MPLS tunnels against transit link/node failures, with traffic restoration time in the order of tens of milliseconds. Local repair refers to the scenario where the router upstream to an anticipated failure, a.k.a., PLR, pre-establishes a bypass tunnel to the router downstream of the failure, a.k.a., Merge Point (MP), pre-installs the forwarding state of the bypass tunnel in the data plane, and uses a rapid mechanism (e.g., link-layer Operations, Administration, and Maintenance (OAM), Bidirectional Forwarding Detection (BFD), and others) to locally detect the failure in the data plane. When the failure occurs, the PLR reroutes traffic through the bypass tunnel to the MP, allowing the traffic to continue to flow to the tunnel's egress router.

This document specifies a fast reroute framework for egress node and egress link protection. Similar to transit link/node protection,

this framework also relies on a PLR to perform local failure detection and local repair. In egress node protection, the PLR is the penultimate hop router of a tunnel. In egress link protection, the PLR is the egress router of the tunnel. The framework further uses a so-called "protector" to serve as the tail end of a bypass tunnel. The protector is a router that hosts "protection service instances" and has its own connectivity or paths to service destinations. When a PLR does local repair, the protector performs "context label switching" for rerouted MPLS service packets and "context IP forwarding" for rerouted IP service packets, to allow the service packets to continue to reach the service destinations.

This framework considers an egress node failure as a failure of a tunnel and a failure of all the services carried by the tunnel as service packets that can no longer reach the service instances on the egress router. Therefore, the framework addresses egress node protection at both the tunnel level and service level, simultaneously. Likewise, the framework considers an egress link failure as a failure of all the services traversing the link and addresses egress link protection at the service level.

This framework requires that the destination (a CE or site) of a service MUST be dual-homed or have dual paths to an MPLS network, via two MPLS edge routers. One of the routers is the egress router of the service's transport tunnel, and the other is a backup egress router that hosts a "backup service instance". In the "co-located" protector mode in this document, the backup egress router serves as the protector; hence, the backup service instance acts as the protection service instance. In the "centralized" protector mode (Section 5.12), the protector and the backup egress router are decoupled, and the protection service instance and the backup service instance are hosted separately by the two routers.

The framework is described by mainly referring to point-to-point (P2P) tunnels. However, it is equally applicable to point-to-multipoint (P2MP), multipoint-to-point (MP2P), and multipoint-to-multipoint (MP2MP) tunnels, as the sub-LSPs of these tunnels can be viewed as P2P tunnels.

The framework is a multi-service and multi-transport framework. It assumes a generic model where each service is comprised of a common set of components, including a service instance, a service label, a service label distribution protocol, and an MPLS transport tunnel. The framework also assumes that the service label is downstream assigned, i.e., assigned by an egress router. Therefore, the framework is generally applicable to most existing and future services. However, there are services with certain modes, where a protector is unable to pre-establish the forwarding state for egress protection, or a PLR is not allowed to reroute traffic to other routers in order to avoid traffic duplication, e.g., the broadcast, multicast, and unknown unicast traffic in Virtual Private LAN Service (VPLS) and Ethernet VPN (EVPN). These cases are left for future study. Services that use upstream-assigned service labels are also out of scope of this document and left for future study.

The framework does not require extensions for the existing signaling and label distribution protocols (e.g., RSVP, LDP, BGP, etc.) of MPLS tunnels. It assumes that transport tunnels and bypass tunnels are to be established by using the generic procedures provided by the protocols. On the other hand, it does not preclude extensions to the protocols that may facilitate the procedures. One example of such extension is [RFC8400]. The framework does see the need for extensions of IGPs and service label distribution protocols in some procedures, particularly for supporting protection establishment and context label switching. This document provides guidelines for these extensions, but it leaves the specific details to separate documents.

The framework is intended to complement control-plane convergence and global repair. Control-plane convergence relies on control protocols to react on the topology changes due to a failure. Global repair relies on an ingress router to remotely detect a failure and switch traffic to an alternative path. An example of global repair is the BGP prefix independent convergence mechanism [BGP-PIC] for BGP-established services. Compared with these mechanisms, this framework is considered faster in traffic restoration, due to the nature of local failure detection and local repair. It is RECOMMENDED that the framework be used in conjunction with control-plane convergence or global repair, in order to take the advantages of both approaches. That is, the framework provides fast and temporary repair, while control-plane convergence or global repair provides ultimate and permanent repair.

2. Specification of Requirements

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

3. Terminology

Egress router:
: A router at the egress endpoint of a tunnel. It hosts service instances for all the services carried by the tunnel and has connectivity with the destinations of the services.

Egress node failure:
: A failure of an egress router.

Egress link failure:
: A failure of the egress link (e.g., PE-CE link, attachment circuit) of a service.

Egress failure:
: An egress node failure or an egress link failure.

Egress-protected tunnel:
: A tunnel whose egress router is protected by a mechanism according to this framework. The egress router is hence called a protected egress router.

Egress-protected service:
: An IP or MPLS service that is carried by an egress-protected tunnel and hence protected by a mechanism according to this framework.

Backup egress router:
: Given an egress-protected tunnel and its egress router, this is another router that has connectivity with all or a subset of the destinations of the egress-protected services carried by the egress-protected tunnel.

Backup service instance:
: A service instance that is hosted by a backup egress router and corresponds to an egress-protected service on a protected egress router.

Protector:
: A role acted by a router as an alternate of a protected egress router, to handle service packets in the event of an egress failure. A protector may be physically co-located with or

decoupled from a backup egress router, depending on the co-located or centralized protector mode.

Protection service instance:
A service instance hosted by a protector that corresponds to the service instance of an egress-protected service on a protected egress router. A protection service instance is a backup service instance, if the protector is co-located with a backup egress router.

PLR:
A router at the point of local repair. In egress node protection, it is the penultimate hop router on an egress-protected tunnel. In egress link protection, it is the egress router of the egress-protected tunnel.

Protected egress {E, P}:
A virtual node consisting of an ordered pair of egress router E and protector P. It serves as the virtual destination of an egress-protected tunnel and as the virtual location of the egress-protected services carried by the tunnel.

Context identifier (ID):
A globally unique IP address assigned to a protected egress {E, P}.

Context label:
A non-reserved label assigned to a context ID by a protector.

Egress-protection bypass tunnel:
A tunnel used to reroute service packets around an egress failure.

Co-located protector mode:
The scenario where a protector and a backup egress router are co-located as one router; hence, each backup service instance serves as a protection service instance.

Centralized protector mode:
The scenario where a protector is a dedicated router and is decoupled from backup egress routers.

Context label switching:
Label switching performed by a protector in the label space of an egress router indicated by a context label.

Context IP forwarding:
IP forwarding performed by a protector in the IP address space of an egress router indicated by a context label.

4. Requirements

This document considers the following as the design requirements of this egress protection framework.

* The framework must support P2P tunnels. It should equally support P2MP, MP2P, and MP2MP tunnels, by treating each sub-LSP as a P2P tunnel.

* The framework must support multi-service and multi-transport networks. It must accommodate existing and future signaling and label-distribution protocols of tunnels and bypass tunnels, including RSVP, LDP, BGP, IGP, Segment Routing, and others. It must also accommodate existing and future IP/MPLS services, including Layer 2 VPNs, Layer 3 VPNs, hierarchical LSP, and others. It MUST provide a general solution for networks where different types of services and tunnels co-exist.

* The framework must consider minimizing disruption during deployment. It should only involve routers close to the egress and be transparent to ingress routers and other transit routers.

* In egress node protection, for scalability and performance reasons, a PLR must be agnostic to services and service labels. It must maintain bypass tunnels and bypass forwarding state on a per-transport-tunnel basis rather than on a per-service-destination or per-service-label basis. It should also support bypass tunnel sharing between transport tunnels.

* A PLR must be able to use its local visibility or information of routing or TE topology to compute or resolve a path for a bypass tunnel.

* A protector must be able to perform context label switching for rerouted MPLS service packets, based on a service label(s) assigned by an egress router. It must be able to perform context IP forwarding for rerouted IP service packets, in the public or private IP address space used by an egress router.

* The framework must be able to work seamlessly with transit link/node protection mechanisms to achieve end-to-end coverage.

* The framework must be able to work in conjunction with global repair and control-plane convergence.

5. Egress Node Protection

5.1. Reference Topology

This document refers to the following topology when describing the procedures of egress node protection.

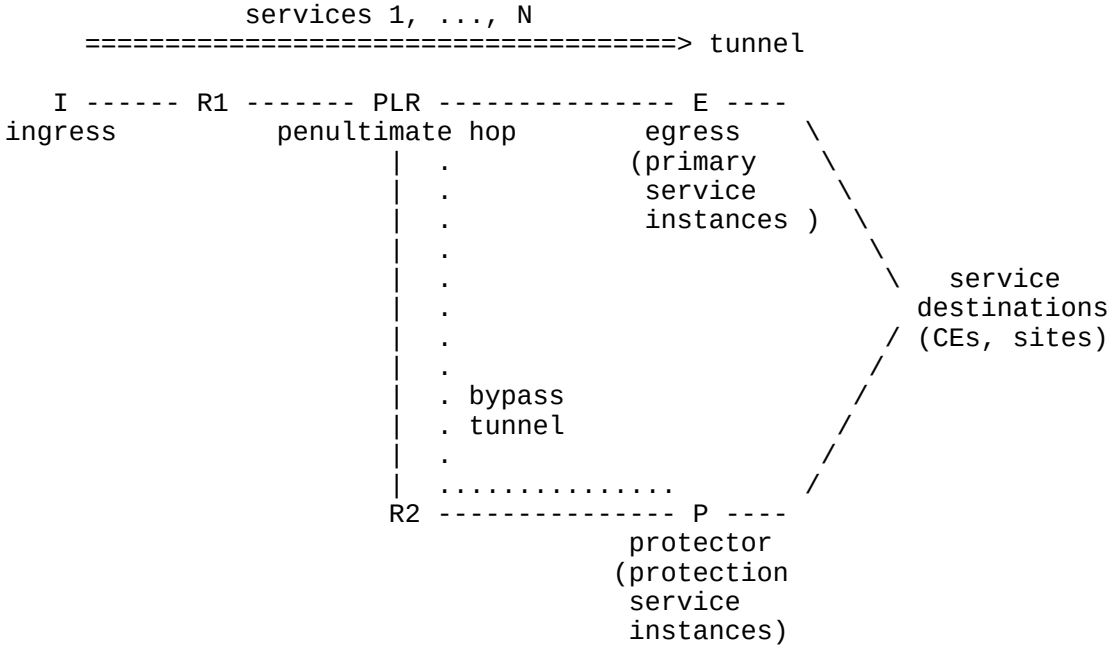


Figure 1

5.2. Egress Node Failure and Detection

An egress node failure refers to the failure of an MPLS tunnel's egress router. At the service level, it is also a service instance failure for each IP/MPLS service carried by the tunnel.

An egress node failure can be detected by an adjacent router (i.e., PLR in this framework) through a node liveness detection mechanism or a mechanism based on a collective failure of all the links to that node. The mechanisms MUST be reasonably fast, i.e., faster than control-plane failure detection and remote failure detection. Otherwise, local repair will not be able to provide much benefit compared to control-plane convergence or global repair. In general, the speed, accuracy, and reliability of a failure detection mechanism are the key factors to decide its applicability in egress node protection. This document provides the following guidelines for network operators to choose a proper type of protection on a PLR.

* If the PLR has a mechanism to detect and differentiate a link failure (of the link between the PLR and the egress node) and an egress node failure, it SHOULD set up both link protection and egress node protection and trigger one and only one protection upon a corresponding failure.

* If the PLR has a fast mechanism to detect a link failure and an egress node failure, but it cannot distinguish them, or if the PLR has a fast mechanism to detect a link failure only, but not an egress node failure, the PLR has two options:

  1. It MAY set up link protection only and leave the egress node failure to be handled by global repair and control-plane convergence.

  2. It MAY set up egress node protection only and treat a link failure as a trigger for the egress node protection. The assumption is that treating a link failure as an egress node failure MUST NOT have a negative impact on services. Otherwise, it SHOULD adopt the previous option.

5.3. Protector and PLR

A router is assigned to the "protector" role to protect a tunnel and the services carried by the tunnel against an egress node failure. The protector is responsible for hosting a protection service instance for each protected service, serving as the tail end of a bypass tunnel, and performing context label switching and/or context IP forwarding for rerouted service packets.

A tunnel is protected by only one protector. Multiple tunnels to a given egress router may be protected by a common protector or different protectors. A protector may protect multiple tunnels with a common egress router or different egress routers.

For each tunnel, its penultimate hop router acts as a PLR. The PLR pre-establishes a bypass tunnel to the protector and pre-installs bypass forwarding state in the data plane. Upon detection of an egress node failure, the PLR reroutes all the service packets received on the tunnel through the bypass tunnel to the protector. For MPLS service packets, the PLR keeps service labels intact in the packets. In turn, the protector forwards the service packets towards the ultimate service destinations. Specifically, it performs context label switching for MPLS service packets, based on the service labels assigned by the protected egress router; it performs context IP forwarding for IP service packets, based on their destination addresses.

The protector MUST have its own connectivity with each service destination, via a direct link or a multi-hop path, which MUST NOT traverse the protected egress router or be affected by the egress node failure. This also means that each service destination MUST be dual-homed or have dual paths to the egress router and a backup egress router that may serve as the protector. Each protection

service instance on the protector relies on such connectivity to set up forwarding state for context label switching and context IP forwarding.

5.4. Protected Egress

This document introduces the notion of "protected egress" as a virtual node consisting of the egress router E of a tunnel and a protector P. It is denoted by an ordered pair of {E, P}, indicating the primary-and-protector relationship between the two routers. It serves as the virtual destination of the tunnel and the virtual location of service instances for the services carried by the tunnel. The tunnel and services are considered as being "associated" with the protected egress {E, P}.

A given egress router E may be the tail end of multiple tunnels. In general, the tunnels may be protected by multiple protectors, e.g., P1, P2, and so on, with each Pi protecting a subset of the tunnels. Thus, these routers form multiple protected egresses, i.e., {E, P1}, {E, P2}, and so on. Each tunnel is associated with one and only one protected egress {E, Pi}. All the services carried by the tunnel are then automatically associated with the protected egress {E, Pi}. Conversely, a service associated with a protected egress {E, Pi} MUST be carried by a tunnel associated with the protected egress {E, Pi}. This mapping MUST be ensured by the ingress router of the tunnel and the service (Section 5.5).

The two routers X and Y may be protectors for each other. In this case, they form two distinct protected egresses: {X, Y} and {Y, X}.

5.5. Egress-Protected Tunnel and Service

A tunnel, which is associated with a protected egress {E, P}, is called an egress-protected tunnel. It is associated with one and only one protected egress {E, P}. Multiple egress-protected tunnels may be associated with a given protected egress {E, P}. In this case, they share the common egress router and protector, but they may or may not share a common ingress router or a common PLR (i.e., penultimate hop router).

An egress-protected tunnel is considered as logically "destined" for its protected egress {E, P}. Its path MUST be resolved and established with E as the physical tail end.

A service, which is associated with a protected egress {E, P}, is called an egress-protected service. Egress router E hosts the primary instance of the service, and protector P hosts the protection instance of the service.

An egress-protected service is associated with one and only one protected egress {E, P}. Multiple egress-protected services may be associated with a given protected egress {E, P}. In this case, these services share the common egress router and protector, but they may or may not be carried by a common egress-protected tunnel or a common ingress router.

An egress-protected service MUST be mapped to an egress-protected tunnel by its ingress router, based on the common protected egress {E, P} of the service and the tunnel. This is achieved by introducing the notion of a "context ID" for a protected egress {E, P}, as described in Section 5.7.

5.6. Egress-Protection Bypass Tunnel

An egress-protected tunnel destined for a protected egress {E, P} MUST have a bypass tunnel from its PLR to protector P. This bypass

tunnel is called an egress-protection bypass tunnel. The bypass tunnel is considered as logically "destined" for the protected egress {E, P}. Due to its bypass nature, it MUST be established with P as the physical tail end and E as the node to avoid. The bypass tunnel MUST NOT be affected by the topology change caused by an egress node failure; thus, it MUST contain a property that protects it from this scenario.

An egress-protection bypass tunnel is associated with one and only one protected egress {E, P}. A PLR may share an egress-protection bypass tunnel for multiple egress-protected tunnels associated with a common protected egress {E, P}.

5.7. Context ID, Context Label, and Context-Based Forwarding

In this framework, a globally unique IPv4 or IPv6 address is assigned as the identifier of the protected egress {E, P}. It is called a "context ID" due to its specific usage in context label switching and context IP forwarding on the protector. It is an IP address that is logically owned by both the egress router and the protector. For the egress router, it indicates the protector. For the protector, it indicates the egress router, particularly the egress router's forwarding context. For other routers in the network, it is an address reachable via both the egress router and the protector (Section 5.8), similar to an anycast address.

The main purpose of a context ID is to coordinate the ingress router, egress router, PLR, and protector to establish egress protection. The procedures are described below, given an egress-protected service associated with a protected egress {E, P} with a context ID.

* If the service is an MPLS service, when E distributes a service label binding message to the ingress router, E attaches the context ID to the message. If the service is an IP service, when E advertises the service destination address to the ingress router, E attaches the context ID to the advertisement message. The service protocol chooses how the context ID is encoded in the messages. A protocol extension of a "context ID" object may be needed, if there is no existing mechanism for this purpose.

* The ingress router uses the service's context ID as the destination to establish or resolve an egress-protected tunnel. The ingress router then maps the service to the tunnel for transportation. The semantics of the context ID is transparent to the ingress router. The ingress router only treats the context ID as an IP address of E, in the same manner as establishing or resolving a regular transport tunnel.

* The context ID is conveyed to the PLR by the signaling protocol of the egress-protected tunnel or learned by the PLR via an IGP (i.e., OSPF or IS-IS) or a topology-driven label distribution protocol (e.g., LDP). The PLR uses the context ID as the destination to establish or resolve an egress-protection bypass tunnel to P while avoiding E.

* P maintains a dedicated label space and a dedicated IP address space for E. They are referred to as "E's label space" and "E's IP address space", respectively. P uses the context ID to identify the label space and IP address space.

* If the service is an MPLS service, E also distributes the service label binding message to P. This is the same label binding message that E advertises to the ingress router, which includes the context ID. Based on the context ID, P installs the service label in an MPLS forwarding table corresponding to E's label space. If the service is an IP service, P installs an IP route in

an IP forwarding table corresponding to E's IP address space. In either case, the protection service instance on P constructs the forwarding state for the label route or IP route based on P's own connectivity with the service's destination.

* P assigns a non-reserved label to the context ID. In the data plane, this label represents the context ID and indicates E's label space and IP address space. Therefore, it is called a "context label".

* The PLR may establish the egress-protection bypass tunnel to P in several manners. If the bypass tunnel is established by RSVP, the PLR signals the bypass tunnel with the context ID as the destination, and P binds the context label to the bypass tunnel. If the bypass tunnel is established by LDP, P advertises the context label for the context ID as an IP prefix Forwarding Equivalence Class (FEC). If the bypass tunnel is established by the PLR in a hierarchical manner, the PLR treats the context label as a one-hop LSP over a regular bypass tunnel to P (e.g., a bypass tunnel to P's loopback IP address). If the bypass tunnel is constructed by using Segment Routing, the bypass tunnel is represented by a stack of Segment Identifier (SID) labels with the context label as the inner-most SID label (Section 5.9). In any case, the bypass tunnel is an ultimate hop-popping (UHP) tunnel whose incoming label on P is the context label.

* During local repair, all the service packets received by P on the bypass tunnel have the context label as the top label. P first pops the context label. For an MPLS service packet, P looks up the service label in E's label space indicated by the context label. Such kind of forwarding is called context label switching. For an IP service packet, P looks up the IP destination address in E's IP address space indicated by the context label. Such kind of forwarding is called context IP forwarding.

5.8. Advertisement and Path Resolution for Context ID

Path resolution and computation for a context ID are done on ingress routers for egress-protected tunnels and on PLRs for egress-protection bypass tunnels. Given a protected egress {E, P} and its context ID, E and P MUST coordinate on the reachability of the context ID in the routing domain and the TE domain. The context ID MUST be advertised in such a manner that all egress-protected tunnels MUST have E as the tail end, and all egress-protection bypass tunnels MUST have P as the tail end while avoiding E.

This document suggests three approaches:

1. The first approach is called "proxy mode". It requires E and P, but not the PLR, to have the knowledge of the egress protection schema. E and P advertise the context ID as a virtual proxy node (i.e., a logical node) connected to the two routers, with the link between the proxy node and E having more preferable IGP and TE metrics than the link between the proxy node and P. Therefore, all egress-protected tunnels destined for the context ID will automatically follow the IGP or TE paths to E. Each PLR will no longer view itself as a penultimate hop but rather as two hops away from the proxy node, via E. The PLR will be able to find a bypass path via P to the proxy node, while the bypass tunnel is actually terminated by P.

2. The second approach is called "alias mode". It requires P and the PLR, but not E, to have the knowledge of the egress protection schema. E simply advertises the context ID as an IP address. P advertises the context ID and the context label

by using a "context ID label binding" advertisement. In both the routing domain and TE domain, the context ID is only reachable via E. Therefore, all egress-protected tunnels destined for the context ID will have E as the tail end. Based on the "context ID label binding" advertisement, the PLR can establish an egress-protection bypass tunnel in several manners (Section 5.9). The "context ID label binding" advertisement is defined as the IGP Mirroring Context segment in [RFC8402] and [RFC8667]. These IGP extensions are generic in nature and hence can be used for egress protection purposes. It is RECOMMENDED that a similar advertisement be defined for OSPF as well.

3. The third approach is called "stub link mode". In this mode, both E and P advertise the context ID as a link to a stub network, essentially modeling the context ID as an anycast IP address owned by the two routers. E, P, and the PLR do not need to have the knowledge of the egress protection schema. The correctness of the egress-protected tunnels and the bypass tunnels relies on the path computations for the anycast IP address performed by the ingress routers and PLR. Therefore, care MUST be taken for the applicability of this approach to a network.

This framework considers the above approaches as technically equal and the feasibility of each approach in a given network as dependent on the topology, manageability, and available protocols of the network. For a given context ID, all relevant routers, including the primary PE, protector, and PLR, MUST support and agree on the chosen approach. The coordination between these routers can be achieved by configuration.

In a scenario where an egress-protected tunnel is an inter-area or inter-Autonomous-System (inter-AS) tunnel, its associated context ID MUST be propagated by IGP or BGP from the original area or AS to the area or AS of the ingress router. The propagation process of the context ID SHOULD be the same as that of an IP address in an inter-area or inter-AS environment.

5.9. Egress-Protection Bypass Tunnel Establishment

A PLR MUST know the context ID of a protected egress {E, P} in order to establish an egress-protection bypass tunnel. The information is obtained from the signaling or label distribution protocol of the egress-protected tunnel. The PLR may or may not need to have the knowledge of the egress-protection schema. All it does is set up a bypass tunnel to a context ID while avoiding the next-hop router (i.e., egress router). This is achievable by using a constraint-based computation algorithm similar to those commonly used for traffic engineering paths and Loop-Free Alternate (LFA) paths. Since the context ID is advertised in the routing domain and in the TE domain by IGP according to Section 5.8, the PLR is able to resolve or establish such a bypass path with the protector as the tail end. In the case of proxy mode, the PLR may do so in the same manner as transit node protection.

An egress-protection bypass tunnel may be established via several methods:

1. It may be established by a signaling protocol (e.g., RSVP), with the context ID as the destination. The protector binds the context label to the bypass tunnel.

2. It may be formed by a topology-driven protocol (e.g., LDP with various LFA mechanisms). The protector advertises the context ID as an IP prefix FEC, with the context label bound to it.

3. It may be constructed as a hierarchical tunnel. When the protector uses the alias mode (Section 5.8), the PLR will have the knowledge of the context ID, context label, and protector (i.e., the advertiser). The PLR can then establish the bypass tunnel in a hierarchical manner, with the context label as a one-hop LSP over a regular bypass tunnel to the protector's IP address (e.g., loopback address). This regular bypass tunnel may be established by RSVP, LDP, Segment Routing, or another protocol.

5.10. Local Repair on PLR

In this framework, a PLR is agnostic to services and service labels. This obviates the need to maintain bypass forwarding state on a per-service basis and allows bypass tunnel sharing between egress-protected tunnels. The PLR may share an egress-protection bypass tunnel for multiple egress-protected tunnels associated with a common protected egress {E, P}. During local repair, the PLR reroutes all service packets received on the egress-protected tunnels to the egress-protection bypass tunnel. Service labels remain intact in MPLS service packets.

Label operation performed by the PLR depends on the bypass tunnel's characteristics. If the bypass tunnel is a single level tunnel, the rerouting will involve swapping the incoming label of an egress-protected tunnel to the outgoing label of the bypass tunnel. If the bypass tunnel is a hierarchical tunnel, the rerouting will involve swapping the incoming label of an egress-protected tunnel to a context label and pushing the outgoing label of a regular bypass tunnel. If the bypass tunnel is constructed by Segment Routing, the rerouting will involve swapping the incoming label of an egress-protected tunnel to a context label and pushing the stack of SID labels of the bypass tunnel.

5.11. Service Label Distribution from Egress Router to Protector

When a protector receives a rerouted MPLS service packet, it performs context label switching based on the packet's service label, which is assigned by the corresponding egress router. In order to achieve this, the protector MUST maintain the labels of egress-protected services in dedicated label spaces on a per-protected-egress {E, P} basis, i.e., one label space for each egress router that it protects.

Also, there MUST be a service label distribution protocol session between each egress router and the protector. Through this protocol, the protector learns the label binding of each egress-protected service. This is the same label binding that the egress router advertises to the service's ingress router, which includes a context ID. The corresponding protection service instance on the protector recognizes the service and resolves forwarding state based on its own connectivity with the service's destination. It then installs the service label with the forwarding state in the label space of the egress router, which is indicated by the context ID (i.e., context label).

Different service protocols may use different mechanisms for such kind of label distribution. Specific extensions may be needed on a per-protocol or per-service-type basis. The details of the extensions should be specified in separate documents. As an example, the LDP extensions for pseudowire services are specified in [RFC8104].

5.12. Centralized Protector Mode

In this framework, it is assumed that the service destination of an

egress-protected service MUST be dual-homed to two edge routers of an MPLS network. One of them is the protected egress router, and the other is a backup egress router. So far in this document, the focus of discussion has been on the scenario where a protector and a backup egress router are co-located as one router. Therefore, the number of protectors in a network is equal to the number of backup egress routers. As another scenario, a network may assign a small number of routers to serve as dedicated protectors, each protecting a subset of egress routers. These protectors are called centralized protectors.

Topologically, a centralized protector may be decoupled from all backup egress routers, or it may be co-located with one backup egress router while decoupled from the other backup egress routers. The procedures in this section assume that a protector and a backup egress router are decoupled.

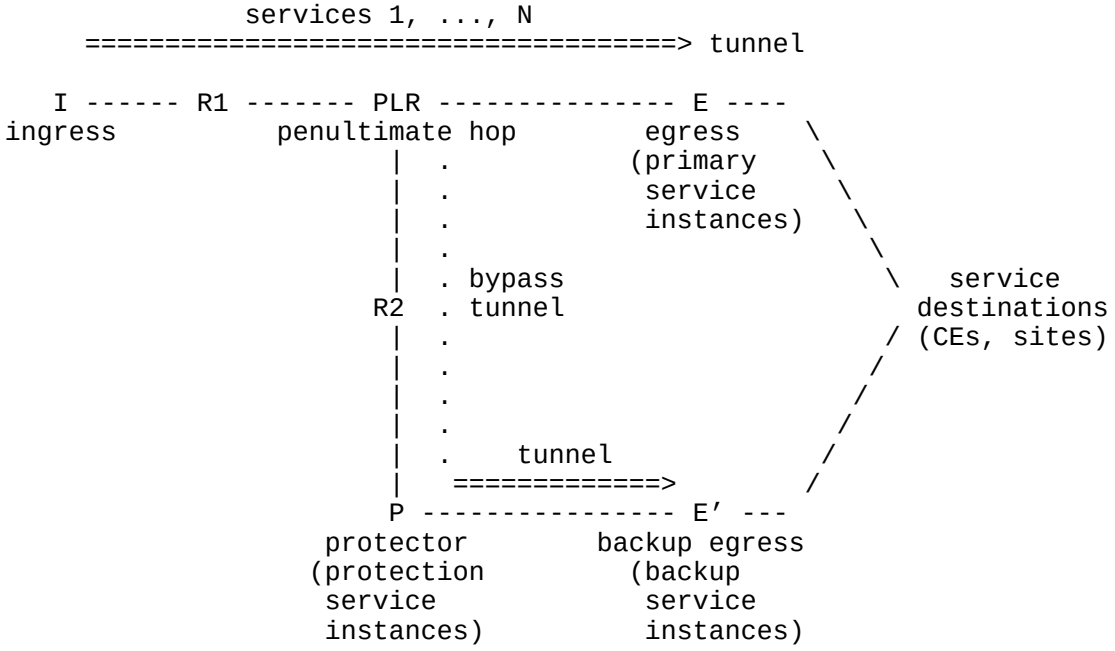


Figure 2

Like a co-located protector, a centralized protector hosts protection service instances, receives rerouted service packets from PLRs, and performs context label switching and/or context IP forwarding. For each service, instead of sending service packets directly to the service destination, the protector MUST send them via another transport tunnel to the corresponding backup service instance on a backup egress router. The backup service instance in turn forwards the service packets to the service destination. Specifically, if the service is an MPLS service, the protector MUST swap the service label in each received service packet to the label of the backup service advertised by the backup egress router, and then push the label (or label stack) of the transport tunnel.

In order for a centralized protector to map an egress-protected MPLS service to a service hosted on a backup egress router, there MUST be a service label distribution protocol session between the backup egress router and the protector. Through this session, the backup egress router advertises the service label of the backup service, attached with the FEC of the egress-protected service and the context ID of the protected egress {E, P}. Based on this information, the protector associates the egress-protected service with the backup service, resolves or establishes a transport tunnel to the backup egress router, and sets up forwarding state for the label of the egress-protected service in the label space of the egress router.

The service label that the backup egress router advertises to the protector can be the same as the label that the backup egress router advertises to the ingress router(s), if and only if the forwarding state of the label does not direct service packets towards the protected egress router. Otherwise, the label MUST NOT be used for egress protection, because it would create a loop for the service packets. In this case, the backup egress router MUST advertise a unique service label for egress protection and set up the forwarding state of the label to use the backup egress router's own connectivity with the service destination.

6. Egress Link Protection

Egress link protection is achievable through procedures similar to that of egress node protection. In normal situations, an egress router forwards service packets to a service destination based on a service label, whose forwarding state points to an egress link. In egress link protection, the egress router acts as the PLR and performs local failure detection and local repair. Specifically, the egress router pre-establishes an egress-protection bypass tunnel to a protector and sets up the bypass forwarding state for the service label to point to the bypass tunnel. During local repair, the egress router reroutes service packets via the bypass tunnel to the protector. The protector in turn forwards the packets to the service destination (in the co-located protector mode, as shown in Figure 3) or forwards the packets to a backup egress router (in the centralized protector mode, as shown in Figure 4).

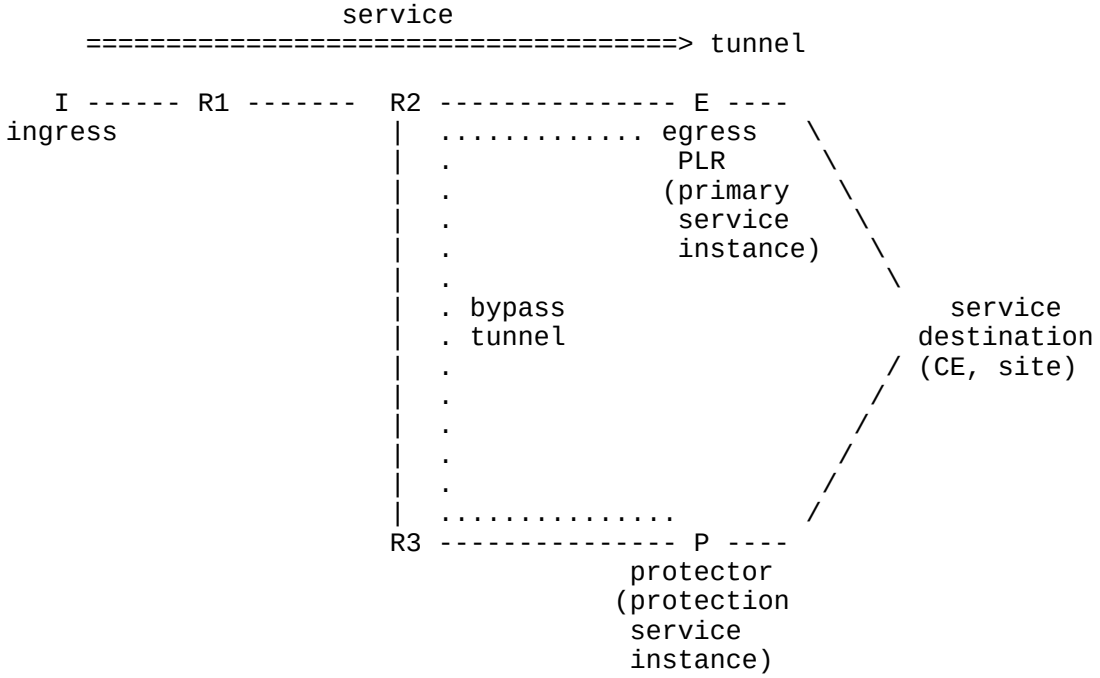
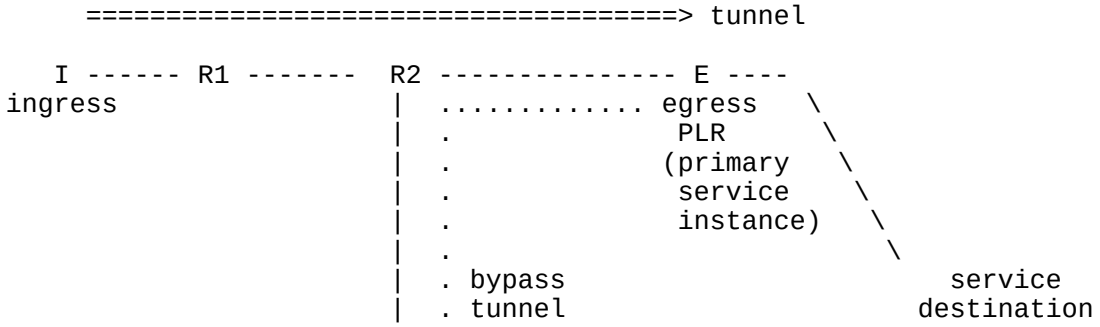


Figure 3



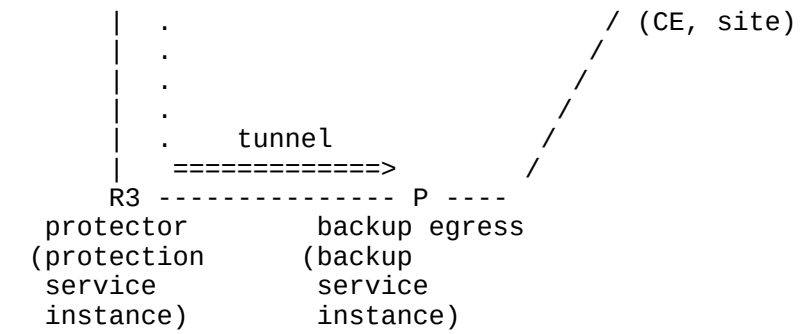


Figure 4

There are two approaches for setting up the bypass forwarding state on the egress router, depending on whether the egress router knows the service label allocated by the backup egress router. The difference is that one approach requires the protector to perform context label switching, and the other one does not. Both approaches are equally supported by this framework.

1. The first approach applies when the egress router does not know the service label allocated by the backup egress router. In this case, the egress router sets up the bypass forwarding state as a label push with the outgoing label of the egress-protection bypass tunnel. Rerouted packets will have the egress router's service label intact. Therefore, the protector MUST perform context label switching, and the bypass tunnel MUST be destined for the context ID of the protected egress {E, P} and established as described in Section 5.9. This approach is consistent with egress node protection. Hence, a protector can serve in egress node protection and egress link protection in a consistent manner, and both the co-located protector mode and the centralized protector mode are supported (see Figures 3 and 4).

2. The second approach applies when the egress router knows the service label allocated by the backup egress router, via a label distribution protocol session. In this case, the backup egress router serves as the protector for egress link protection, regardless of the protector of egress node protection, which will be the same router in the co-located protector mode but a different router in the centralized protector mode. The egress router sets up the bypass forwarding state as a label swap from the incoming service label to the service label of the backup egress router (i.e., protector), followed by a push with the outgoing label (or label stack) of the egress link protection bypass tunnel. The bypass tunnel is a regular tunnel destined for an IP address of the protector, instead of the context ID of the protected egress {E, P}. The protector simply forwards rerouted service packets based on its own service label rather than performing context label switching. In this approach, only the co-located protector mode is applicable.

Note that for a bidirectional service, the physical link of an egress link may carry service traffic bidirectionally. Therefore, an egress link failure may simultaneously be an ingress link failure for the traffic in the opposite direction. Protection for ingress link failure SHOULD be provided by a separate mechanism and hence is out of the scope of this document.

7. Global Repair

This framework provides a fast but temporary repair for egress node and egress link failures. For permanent repair, the services

affected by a failure SHOULD be moved to an alternative tunnel, or replaced by alternative services, which are fully functional. This is referred to as global repair. Possible triggers of global repair include control-plane notifications of tunnel status and service status, end-to-end OAM and fault detection at the tunnel and service level, and others. The alternative tunnel and services may be pre-established in standby state or dynamically established as a result of the triggers or network protocol convergence.

8. Operational Considerations

When a PLR performs local repair, the router SHOULD generate an alert for the event. The alert may be logged locally for tracking purposes, or it may be sent to the operator at a management station. The communication channel and protocol between the PLR and the management station may vary depending on networks and are out of the scope of this document.

9. General Context-Based Forwarding

So far, this document has been focusing on the cases where service packets are MPLS or IP packets, and protectors perform context label switching or context IP forwarding. Although this should cover most common services, it is worth mentioning that the framework is also applicable to services or sub-modes of services where service packets are Layer 2 packets or encapsulated in non-IP and non-MPLS formats. The only specific in these cases is that a protector MUST perform context-based forwarding based on the Layer 2 table or corresponding lookup table, which is indicated by a context ID (i.e., context label).

10. Example: Layer 3 VPN Egress Protection

This section shows an example of egress protection for Layer 3 IPv4 and IPv6 VPNs.

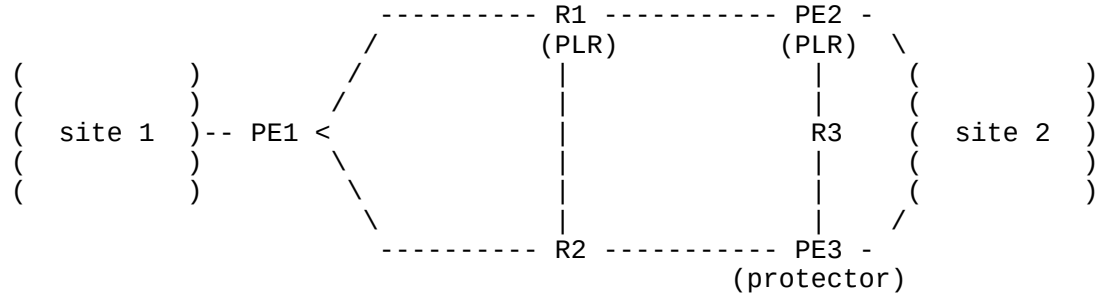


Figure 5

In this example, the core network is IPv4 and MPLS. Both of the IPv4 and IPv6 VPNs consist of sites 1 and 2. Site 1 is connected to PE1, and site 2 is dual-homed to PE2 and PE3. Site 1 includes an IPv4 subnet 203.0.113.64/26 and an IPv6 subnet 2001:db8:1:1::/64. Site 2 includes an IPv4 subnet 203.0.113.128/26 and an IPv6 subnet 2001:db8:1:2::/64. PE2 is the primary PE for site 2, and PE3 is the backup PE. Each of PE1, PE2, and PE3 hosts an IPv4 VPN instance and an IPv6 VPN instance. The PEs use BGP to exchange VPN prefixes and VPN labels between each other. In the core network, R1 and R2 are transit routers, OSPF is used as the routing protocol, and RSVP-TE is used as the tunnel signaling protocol.

Using the framework in this document, the network assigns PE3 to be the protector of PE2 to protect the VPN traffic in the direction from site 1 to site 2. This is the co-located protector mode. PE2 and PE3 form a protected egress {PE2, PE3}. Context ID 198.51.100.1 is assigned to the protected egress {PE2, PE3}. (If the core network is

IPv6, the context ID would be an IPv6 address.) The IPv4 and IPv6 VPN instances on PE3 serve as protection instances for the corresponding VPN instances on PE2. On PE3, context label 100 is assigned to the context ID, and a label table pe2.mpls is created to represent PE2's label space. PE3 installs label 100 in its MPLS forwarding table, with the next hop pointing to the label table pe2.mpls. PE2 and PE3 are coordinated to use the proxy mode to advertise the context ID in the routing domain and the TE domain.

PE2 uses the label allocation mode per Virtual Routing and Forwarding (VRF) for both of its IPv4 and IPv6 VPN instances. It assigns label 9000 to the IPv4 VRF, and label 9001 to the IPv6 VRF. For the IPv4 prefix 203.0.113.128/26 in site 2, PE2 advertises it with label 9000 and NEXT_HOP 198.51.100.1 to PE1 and PE3 via BGP. Likewise, for the IPv6 prefix 2001:db8:1:2::/64 in site 2, PE2 advertises it with label 9001 and NEXT_HOP 198.51.100.1 to PE1 and PE3 via BGP.

PE3 also uses per-VRF VPN label allocation mode for both of its IPv4 and IPv6 VPN instances. It assigns label 10000 to the IPv4 VRF and label 10001 to the IPv6 VRF. For the prefix 203.0.113.128/26 in site 2, PE3 advertises it with label 10000 and NEXT_HOP as itself to PE1 and PE2 via BGP. For the IPv6 prefix 2001:db8:1:2::/64 in site 2, PE3 advertises it with label 10001 and NEXT_HOP as itself to PE1 and PE2 via BGP.

Upon receipt of the above BGP advertisements from PE2, PE1 uses the context ID 198.51.100.1 as the destination to compute a path for an egress-protected tunnel. The resultant path is PE1->R1->PE2. PE1 then uses RSVP to signal the tunnel, with the context ID 198.51.100.1 as the destination, with the "node protection desired" flag set in the SESSION_ATTRIBUTE of the RSVP Path message. Once the tunnel comes up, PE1 maps the VPN prefixes 203.0.113.128/26 and 2001:db8:1:2::/64 to the tunnel and installs a route for each prefix in the corresponding IPv4 or IPv6 VRF. The next hop of route 203.0.113.128/26 is a push of the VPN label 9000, followed by a push of the outgoing label of the egress-protected tunnel. The next hop of route 2001:db8:1:2::/64 is a push of the VPN label 9001, followed by a push of the outgoing label of the egress-protected tunnel.

Upon receipt of the above BGP advertisements from PE2, PE3 recognizes the context ID 198.51.100.1 in the NEXT_HOP attribute and installs a route for label 9000 and a route for label 9001 in the label table pe2.mpls. PE3 sets the next hop of route 9000 to the IPv4 protection VRF and the next hop of route 9001 to the IPv6 protection VRF. The IPv4 protection VRF contains the routes to the IPv4 prefixes in site 2. The IPv6 protection VRF contains the routes to the IPv6 prefixes in site 2. The next hops of these routes must be based on PE3's connectivity with site 2, even if the connectivity may not have the best metrics (e.g., Multi-Exit Discriminator (MED), local preference, etc.) to be used in PE3's own VRF. The next hops must not use any path traversing PE2. Note that the protection VRFs are a logical concept, and they may simply be PE3's own VRFs if they satisfy the requirement.

10.1. Egress Node Protection

R1, i.e., the penultimate hop router of the egress-protected tunnel, serves as the PLR for egress node protection. Based on the "node protection desired" flag and the destination address (i.e., context ID 198.51.100.1) of the tunnel, R1 computes a bypass path to 198.51.100.1 while avoiding PE2. The resultant bypass path is R1->R2->PE3. R1 then signals the path (i.e., egress-protection bypass tunnel), with 198.51.100.1 as the destination.

Upon receipt of an RSVP Path message of the egress-protection bypass tunnel, PE3 recognizes the context ID 198.51.100.1 as the destination

and responds with context label 100 in an RSVP Resv message.

After the egress-protection bypass tunnel comes up, R1 installs a bypass next hop for the egress-protected tunnel. The bypass next hop is a label swap from the incoming label of the egress-protected tunnel to the outgoing label of the egress-protection bypass tunnel.

When R1 detects a failure of PE2, it will invoke the above bypass next hop to reroute VPN packets. Each IPv4 VPN packet will have the label of the bypass tunnel as outer label, and the IPv4 VPN label 9000 as inner label. Each IPv6 VPN packet will have the label of the bypass tunnel as the outer label and the IPv6 VPN label 9001 as the inner label. When the packets arrive at PE3, they will have the context label 100 as the outer label and the VPN label 9000 or 9001 as the inner label. The context label will first be popped, and then the VPN label will be looked up in the label table pe2.mpls. The lookup will cause the VPN label to be popped and the IPv4 and IPv6 packets to be forwarded to site 2 based on the IPv4 and IPv6 protection VRFs, respectively.

10.2. Egress Link Protection

PE2 serves as the PLR for egress link protection. It has already learned PE3's IPv4 VPN label 10000 and IPv6 VPN label 10001. Hence, it uses approach (2) as described in Section 6 to set up the bypass forwarding state. It signals an egress-protection bypass tunnel to PE3, by using the path PE2->R3->PE3, and PE3's IP address as the destination. After the bypass tunnel comes up, PE2 installs a bypass next hop for the IPv4 VPN label 9000 and a bypass next hop for the IPv6 VPN label 9001. For label 9000, the bypass next hop is a label swap to label 10000, followed by a label push with the outgoing label of the bypass tunnel. For label 9001, the bypass next hop is a label swap to label 10001, followed by a label push with the outgoing label of the bypass tunnel.

When PE2 detects a failure of the egress link, it will invoke the above bypass next hop to reroute VPN packets. Each IPv4 VPN packet will have the label of the bypass tunnel as the outer label and label 10000 as the inner label. Each IPv6 VPN packet will have the label of the bypass tunnel as the outer label and label 10001 as the inner label. When the packets arrive at PE3, the VPN label 10000 or 10001 will be popped, and the exposed IPv4 and IPv6 packets will be forwarded based on PE3's IPv4 and IPv6 VRFs, respectively.

10.3. Global Repair

Eventually, global repair will take effect, as control-plane protocols converge on the new topology. PE1 will choose PE3 as a new entrance to site 2. Before that happens, the VPN traffic has been protected by the above local repair.

10.4. Other Modes of VPN Label Allocation

It is also possible that PE2 may use per-route or per-interface VPN label allocation mode. In either case, PE3 will have multiple VPN label routes in the pe2.mpls table, corresponding to the VPN labels advertised by PE2. PE3 forwards rerouted packets by popping a VPN label and performing an IP lookup in the corresponding protection VRF. PE3's forwarding behavior is consistent with the above case where PE2 uses per-VRF VPN label allocation mode. PE3 does not need to know PE2's VPN label allocation mode or construct a specific next hop for each VPN label route in the pe2.mpls table.

11. IANA Considerations

This document has no IANA actions.

12. Security Considerations

The framework in this document involves rerouting traffic around an egress node or link failure, via a bypass path from a PLR to a protector, and ultimately to a backup egress router. The forwarding performed by the routers in the data plane is anticipated, as part of the planning of egress protection.

Control-plane protocols MAY be used to facilitate the provisioning of the egress protection on the routers. In particular, the framework requires a service label distribution protocol between an egress router and a protector over a secure session. The security properties of this provisioning and label distribution depend entirely on the underlying protocol chosen to implement these activities. Their associated security considerations apply. This framework introduces no new security requirements or guarantees relative to these activities.

Also, the PLR, protector, and backup egress router are located close to the protected egress router, which is normally in the same administrative domain. If they are not in the same administrative domain, a certain level of trust MUST be established between them in order for the protocols to run securely across the domain boundary. The basis of this trust is the security model of the protocols (as described above), and further security considerations for inter-domain scenarios should be addressed by the protocols as a common requirement.

Security attacks may sometimes come from a customer domain. Such attacks are not introduced by the framework in this document and may occur regardless of the existence of egress protection. In one possible case, the egress link between an egress router and a CE could become a point of attack. An attacker that gains control of the CE might use it to simulate link failures and trigger constant and cascading activities in the network. If egress link protection is in place, egress link protection activities may also be triggered. As a general solution to defeat the attack, a damping mechanism SHOULD be used by the egress router to promptly suppress the services associated with the link or CE. The egress router would stop advertising the services, essentially detaching them from the network and eliminating the effect of the simulated link failures.

From the above perspectives, this framework does not introduce any new security threat to networks.

13. References

13.1. 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, <https://www.rfc-editor.org/info/rfc2119>.

[RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in RFC 2119 Key Words", BCP 14, RFC 8174, DOI 10.17487/RFC8174, May 2017, <https://www.rfc-editor.org/info/rfc8174>.

[RFC8402] Filsfils, C., Ed., Previdi, S., Ed., Ginsberg, L., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", RFC 8402, DOI 10.17487/RFC8402, July 2018, <https://www.rfc-editor.org/info/rfc8402>.

[RFC8667] Previdi, S., Ginsberg, L., Filsfils, C., Bashandy, A., Gredler, H., and B. Decraene, "IS-IS Extensions for

Segment Routing", RFC 8667, DOI 10.17487/RFC8667, December 2019, <https://www.rfc-editor.org/info/rfc8667>.

13.2. Informative References

[BGP-PIC] Bashandy, A., Filsfils, C., and P. Mohapatra, "BGP Prefix Independent Convergence", Work in Progress, Internet-Draft, draft-ietf-rtgwg-bgp-pic-10, 2 October 2019, <https://tools.ietf.org/html/draft-ietf-rtgwg-bgp-pic-10>.

[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, <https://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, <https://www.rfc-editor.org/info/rfc5286>.

[RFC7490] Bryant, S., Filsfils, C., Previdi, S., Shand, M., and N. So, "Remote Loop-Free Alternate (LFA) Fast Reroute (FRR)", RFC 7490, DOI 10.17487/RFC7490, April 2015, <https://www.rfc-editor.org/info/rfc7490>.

[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, <https://www.rfc-editor.org/info/rfc7812>.

[RFC8104] Shen, Y., Aggarwal, R., Henderickx, W., and Y. Jiang, "Pseudowire (PW) Endpoint Fast Failure Protection", RFC 8104, DOI 10.17487/RFC8104, March 2017, <https://www.rfc-editor.org/info/rfc8104>.

[RFC8400] Chen, H., Liu, A., Saad, T., Xu, F., and L. Huang, "Extensions to RSVP-TE for Label Switched Path (LSP) Egress Protection", RFC 8400, DOI 10.17487/RFC8400, June 2018, <https://www.rfc-editor.org/info/rfc8400>.

Acknowledgements

This document leverages work done by Yakov Rekhter, Kevin Wang, and Zhaohui Zhang on MPLS egress protection. Thanks to Alexander Vainshtein, Rolf Winter, Lizhong Jin, Krzysztof Szarkowicz, Roman Danyliw, and Yuanlong Jiang for their valuable comments that helped to shape this document and improve its clarity.

Authors' Addresses

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

Phone: +1 978 589 0722
Email: yshen@juniper.net

Minto Jeyananth
Juniper Networks
1133 Innovation Way
Sunnyvale, CA 94089
United States of America

Phone: +1 408 936 7563
Email: minto@juniper.net

Bruno Decraene
Orange

Email: bruno.decraene@orange.com

Hannes Gredler
RtBrick Inc.

Email: hannes@rtbrick.com

Carsten Michel
Deutsche Telekom

Email: c.michel@telekom.de

Huaimo Chen
Futurewei
Boston, MA
United States of America

Email: Huaimo.chen@futurewei.com