

Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: May 18, 2018

Yimin Shen
Minto Jeyananth
Juniper Networks
Bruno Decraene
Orange
Hannes Gredler
RtBrick Inc
Carsten Michel
Deutsche Telekom
Huaimo Chen
Yuanlong Jiang
Huawei Technologies Co., Ltd.
November 14, 2017

MPLS Egress Protection Framework
draft-shen-mpls-egress-protection-framework-07

Abstract

This document specifies a fast reroute framework for protecting IP/MPLS services and MPLS transport tunnels against egress node and egress link failures. In this framework, the penultimate-hop router of an MPLS tunnel acts as the point of local repair (PLR) for egress node failure, and the egress router of the MPLS tunnel acts as the PLR for egress link failure. Each of them pre-establishes a bypass tunnel to a protector. Upon an egress node or link failure, the corresponding PLR performs local failure detection and local repair, by rerouting packets over the corresponding bypass tunnel. The protector in turn performs context label switching or context IP forwarding to send the packets to the ultimate service destination(s). This mechanism can be used to reduce traffic loss before global repair reacts to the failure and control plane protocols converge on the topology changes due to the failure. The framework is applicable to all types of IP/MPLS services and MPLS tunnels. Under the framework, service protocol extensions may be further specified to support service label distribution to the protector.

Status of This Memo

This Internet-Draft is submitted in full conformance with the provisions of BCP 78 and BCP 79.

Internet-Drafts are working documents of the Internet Engineering Task Force (IETF). Note that other groups may also distribute working documents as Internet-Drafts. The list of current Internet-Drafts is at <https://datatracker.ietf.org/drafts/current/>.

Internet-Drafts are draft documents valid for a maximum of six months and may be updated, replaced, or obsoleted by other documents at any time. It is inappropriate to use Internet-Drafts as reference material or to cite them other than as "work in progress."

This Internet-Draft will expire on May 18, 2018.

Copyright Notice

Copyright (c) 2017 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to BCP 78 and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in Section 4.e of the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1. Introduction	3
2. Specification of Requirements	5
3. Terminology	5
4. Requirements	7
5. Egress node protection	8
5.1. Reference topology	8
5.2. Egress node failure and detection	8
5.3. Protector and PLR	9
5.4. Protected egress	10
5.5. Egress-protected tunnel and service	11
5.6. Egress-protection bypass tunnel	11
5.7. Context ID, context label, and context based forwarding	12
5.8. Advertisement and path resolution for context ID	14
5.9. Egress-protection bypass tunnel establishment	15
5.10. Local repair on PLR	15
5.11. Service label distribution from egress router to protector	16
5.12. Centralized protector mode	16
6. Egress link protection	18
7. Global repair	21
8. Example: Layer-3 VPN egress protection	21
8.1. Egress node protection	23
8.2. Egress link protection	24
8.3. Global repair	24

9. IANA Considerations	24
10. Security Considerations	24
11. Acknowledgements	25
12. References	25
12.1. Normative References	25
12.2. Informative References	25
Authors' Addresses	26

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 can satisfy both individual and aggregate requirements (e.g. CoS, QoS) of these services. The egress router of the tunnel should host the corresponding service instances of the services. An MPLS service instance is responsible for forwarding service packets via an egress link to the service destination, based on a service label. An IP service instance is responsible for doing the same based on a service IP address. The egress link is often called a PE-CE (provider edge - customer edge) link or attachment circuit (AC).

Today, local repair based fast reroute mechanisms [RFC4090], [RFC5286], [RFC7490], [RFC7812] have been widely deployed to protect MPLS tunnels against transit link/node failures. They can achieve fast restoration of traffic in the order of tens of milliseconds. Local repair refers to the scenario where the router upstream to an anticipated failure (aka. PLR, i.e. point of local repair) pre-establishes a bypass tunnel to the router downstream of the failure (aka. MP, i.e. merge point), and pre-installs the forwarding state of the bypass tunnel in the data plane. The PLR also uses a rapid mechanism (e.g. link layer OAM, 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 describes a fast reroute framework for egress node and egress link protection. Similar to transit link/node protection, this framework 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 relies on a so-called "protector" to serve as the tailend 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 is doing local repair, the protector is responsible for

performing "context label switching" for rerouted MPLS service packets and "context IP forwarding" for rerouted IP service packets. Thus, the service packets can continue to reach service destinations with minimum disruption.

This framework considers an egress node failure as a failure of a tunnel, as well as a failure of all the services carried by the tunnel, because service packets can no longer reach the service instances on the egress router. Therefore, the framework addresses egress node protection at both 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, normally via two MPLS edge routers. One of them is the egress router of the service's transport tunnel, and the other is a backup egress router which hosts "backup service instances". In the "co-located" protector mode in this document, the backup egress router serves as a protector, and hence each backup service instance acts as a protection instance. In the "centralized" protector mode (Section 5.12), a protector and a backup egress router are decoupled, and each protection service instance and its corresponding backup service instance are hosted on separate routers.

The framework is described by mainly referring to P2P (point-to-point) tunnels. However, it is equally applicable to P2MP (point-to-multipoint), MP2P (multipoint-to-point) and MP2MP (multipoint-to-multipoint) tunnels, when a sub-LSP can be viewed as a P2P tunnel.

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, and a service label distribution protocol, and the service is transported over an MPLS tunnel of any type. The framework also assumes service labels to be downstream assigned, i.e. assigned by egress routers. Therefore, the framework is generally applicable to most existing and future services. Services which use upstream-assigned service labels are out of scope of this document and left for further 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 expects transport tunnels and bypass tunnels to be established by using the generic mechanisms provided by the protocols. On the other hand, it does not preclude future extensions to the protocols which may facilitate the procedures. One example of such extension is [RSVP-EP]. The framework may need extensions for

IGPs and service label distribution protocols, to support protection establishment and context label switching. This document provides guidelines for these extensions, but the specific details SHOULD be addressed in separate documents.

The framework is intended to complement control-plane convergence and global repair, which are traditionally used to recover networks from egress node and egress link failures. 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 as faster in traffic restoration, due to the nature of local failure detection and local repair. However, it is RECOMMENDED that the framework SHOULD be used in conjunction with control-plane convergence or global repair, in order to take the advantages of both approaches to achieve more effective protection. That is, the framework provides fast and temporary repair, and 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", "MAY", and "OPTIONAL" in this document are to be interpreted as described in RFC2119.

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 which 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 which 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 which is hosted by a backup egress router, and corresponding 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, corresponding 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, and 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 followings as the design requirements of this egress protection framework.

- o The framework must support P2P tunnels. It should equally support P2MP, MP2P and MP2MP tunnels, by treating each sub-LSP as a P2P tunnel.
- o 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 generic solution for environments where different types of services and tunnels may co-exist.
- o The framework must consider minimizing disruption during deployment. It should only involve routers close to egress, and be transparent to ingress routers and other transit routers.
- o In egress node protection, for scalability and performance reasons, a PLR must be agnostic to services and service labels, like PLRs in transit link/node protection. It must maintain bypass tunnels and bypass forwarding state on a per-transport-tunnel basis, rather than per-service-destination or per-service-label basis. It should also support bypass tunnel sharing between transport tunnels.
- o A PLR must be able to use its local visibility or information of routing and/or TE topology to compute or resolve a path for a bypass tunnel to a protector.
- o A protector must be able to perform context label switching for rerouted MPLS service packets, based on 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.

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

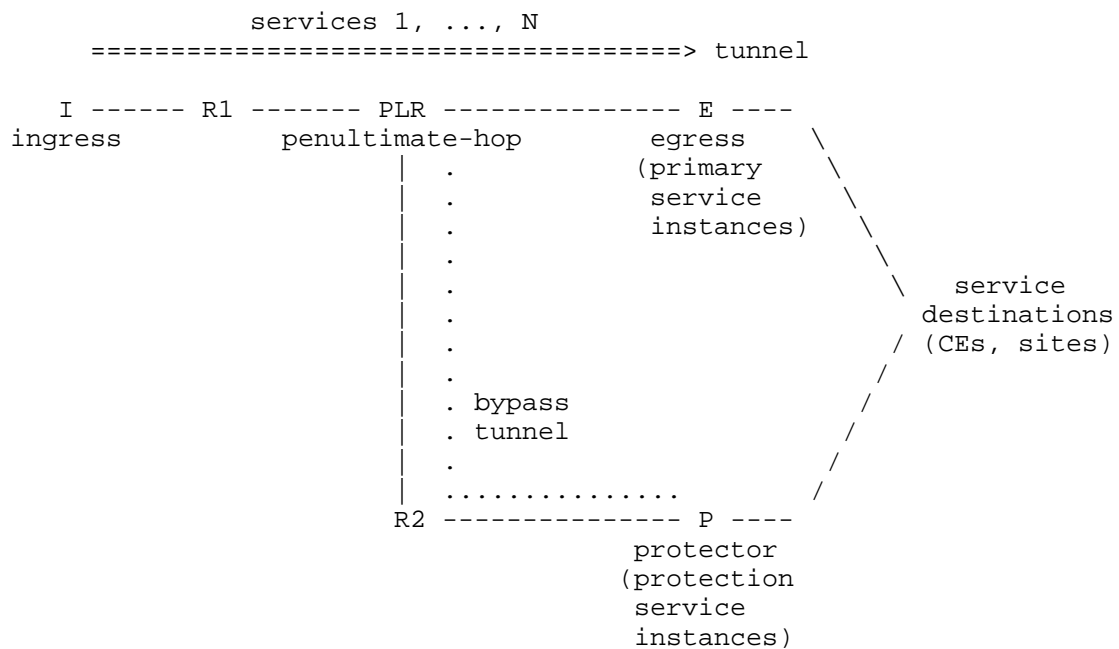


Figure 1

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 also means a service instance failure for each IP/MPLS service carried by the tunnel.

Ideally, an egress node failure can be detected by an adjacent router (i.e. PLR in this framework) using a node liveness detection

mechanism, or based on a collective failure of all the links to that node. However, the assumption is that the mechanisms SHOULD 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 mechanism are the key factors to decide its applicability in egress node protection. This document provides the following guidelines in this regard.

- o If the PLR has a reasonably fast 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.
- o If the PLR has a fast mechanism to detect a link failure and an egress node failure, but 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 global repair and control plane convergence to handle.
 2. It MAY set up egress node protection only, and treat a link failure as a trigger for the egress node protection. However, 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 tailend of a bypass tunnel, and performing context label switching and/or context IP forwarding for rerouted service packets.

A tunnel can be protected by only one protector at a given time. 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. The protector in turn forwards the rerouted service packets towards the ultimate service destinations. Specifically, it performs context label switching for MPLS service packets, based on 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 requires that each service destination MUST be dual-homed or have dual paths to the egress router and a backup egress router which serves as the protector. Each protection service instance on the protector relies on such connectivity to set up forwarding state for context label switching and/or 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 tailend 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 same 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).

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 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}. However, its path MUST be resolved and established with E as the physical tailend.

A service, which is associated with a protected egress {E, P}, is called an egress-protected service. The egress router E hosts the primary instance of the service, and the 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 may or may not share 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 "context ID" for 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 the 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}. However, due to its bypass nature, it MUST be resolved and established with P as the physical tailend and E as the node to avoid. The bypass tunnel MUST have the property that it MUST NOT be affected by any topology change caused by an egress node failure.

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}. For multiple egress-protected tunnels associated with a common protected egress {E, P}, there may be one or

multiple egress-protection bypass tunnels from one or multiple PLRs to the protector P, depending on the paths of the egress-protected tunnels.

5.7. Context ID, context label, and context based forwarding

In this framework, a globally unique IPv4/v6 address is assigned to a protected egress {E, P} to serve 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 node, 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 in the routing domain and the TE domain (Section 5.8), similar to an anycast address.

The main purpose of a context ID is to coordinate ingress router, egress router, PLR and protector in setting up egress protection. Given an egress-protected service associated with a protected egress {E, P}, its context ID is used as below:

- o 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 also attaches the context ID to the advertisement message. How the context ID is encoded in the messages is a choice of the service protocol, and may need protocol extensions to define a "context ID" object.
- o The ingress router uses the context ID as destination to establish or resolve an egress-protected tunnel. The ingress router then maps the service to the tunnel for transportation. In this process, the special 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, and behaves in the same manner as in establishing or resolving a regular transport tunnel, although the end result is an egress-protected tunnel.
- o 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 ISIS) or a topology-driven label distribution protocol (e.g. LDP). The PLR uses the context ID as destination to establish or resolve an egress-protection bypass tunnel to P while avoiding E.

- o P maintains a dedicated label space or a dedicated IP address space for E, depending on whether the service is MPLS or IP. This is referred to as "E's label space" or "E's IP address space", respectively. P uses the context ID to identify the space.
- o 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, attached with 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 interprets the service and constructs forwarding state for the route based on P's own connectivity to the service's destination.
- o 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".
- o 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 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 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 SID labels with the context label as the inner-most SID label (Section 5.9). In any case, the bypass tunnel is a UHP tunnel whose incoming label at P is the context label.
- o During local repair, all the service packets received by P on the bypass tunnel have the context label as top label. P first pops the context label. For an MPLS service packet, P further looks up the service label in E's label space indicated by the context label, which 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, which 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. Therefore, given a protected egress {E, P} and its context ID, E and P MUST coordinate the context ID in the routing domain and the TE domain via IGP advertisement. The context ID MUST be advertised in such a manner that all egress-protected tunnels MUST have E as tailend, and all egress-protection bypass tunnels MUST have P as tailend while avoiding E.

This document suggests two 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 should automatically follow the shortest IGP or TE paths to E. Each PLR will no longer view itself as a penultimate-hop, but rather 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 should actually be 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 a regular IP address. P advertises the context ID and the context label by using a "context ID label binding" advertisement. The advertisement MUST be understood by the PLR. In both routing domain and TE domain, the context ID is only reachable via E. This ensures that all egress-protected tunnels destined for the context ID should have E as tailend. 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 IGP mirroring context segment in [SR-ARCH], [SR-OSPF] and [SR-ISIS]. These IGP extensions are generic in nature, and hence can be used for egress protection purposes.

In a scenario where an egress-protected tunnel is an inter-area or inter-AS tunnel, its associated context ID MUST be propagated from the residing area/AS to the other areas/AS' via IGP or BGP, so that the ingress router of the tunnel can have the reachability to the context ID. The propagation process of the context ID SHOULD be the same as that of a regular IP address in an inter-area/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 to 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 which are commonly used in the computation of traffic engineering paths and loop-free alternate (LFA) paths. Since the context ID is advertised in the routing domain and the TE domain by IGP according to Section 5.8, the PLR should be able to resolve or establish such a bypass path with the protector as tailend. In some cases like the 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 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, and others.

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 via the egress-protection bypass tunnel. Service labels remain intact in MPLS service packets.

Label operation during the rerouting 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 a stack of SID labels of the bypass tunnel.

5.11. Service label distribution from egress router to protector

As mentioned in previous sections, 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 such kind of service labels 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 corresponding ingress router, attached with 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 protocol extensions may be needed on a per-protocol basis or per-service-type basis. The details of the extensions SHOULD be specified in separate documents. As an example, RFC 8104 specifies the LDP extensions for pseudowire services.

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 discussion has been focusing 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 the scenario where a protector and a backup egress router are decoupled.

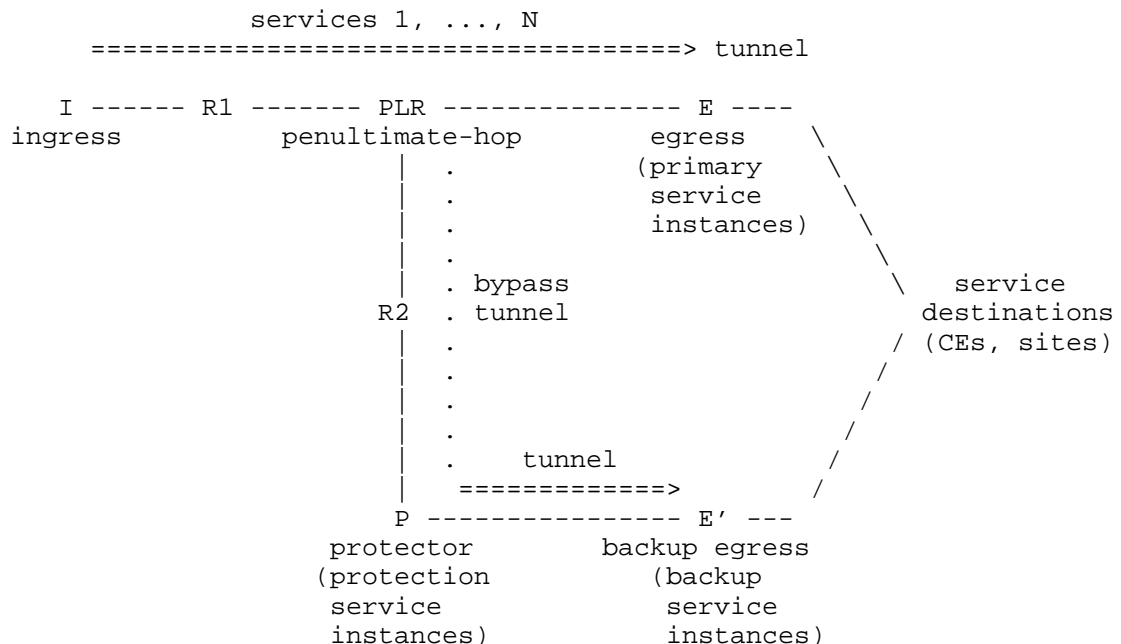


Figure 2

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

them to the service destination. Specifically, in the case of 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 accordingly sets up forwarding state for the label of the egress-protected service in the label space of the egress router.

The service label which the backup egress router advertises to the protector can be the same as the label which 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 is not usable for egress protection, because it will create a loop, which MUST be avoided. In this case, the backup egress router MUST advertise a unique service label for egress protection, and set its forwarding state to use the backup egress router's 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 PLR, by performing local failure detection and local repair. Specifically, the egress router pre-establishes an egress-protection bypass tunnel to a protector, and installs bypass forwarding state for the service label, pointing 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).

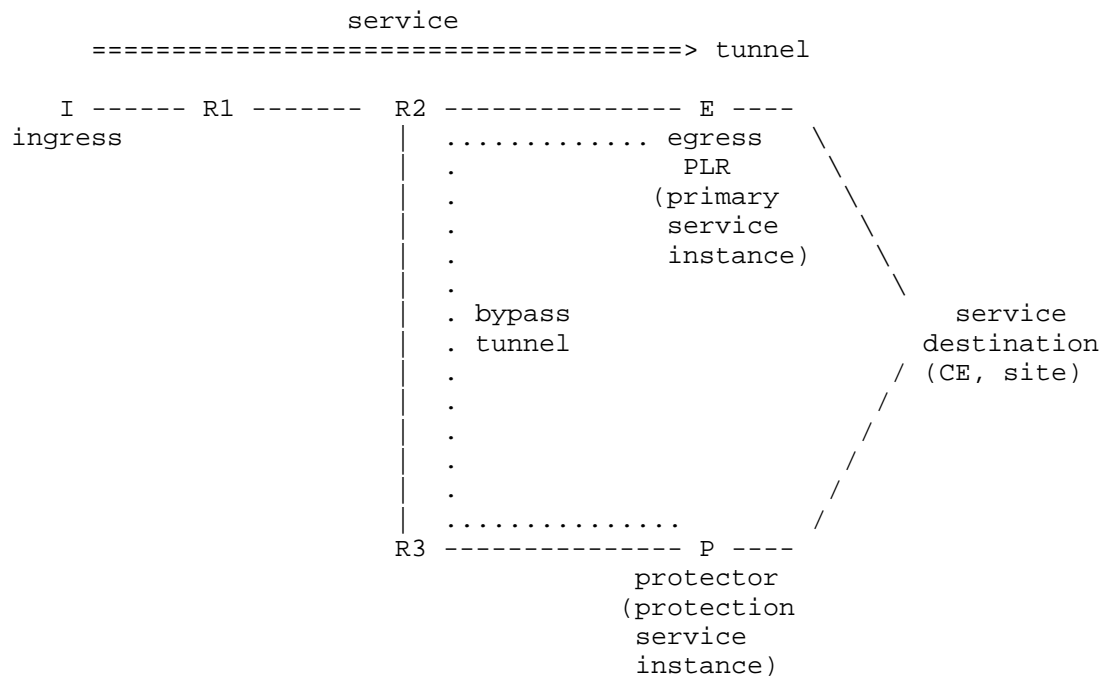


Figure 3

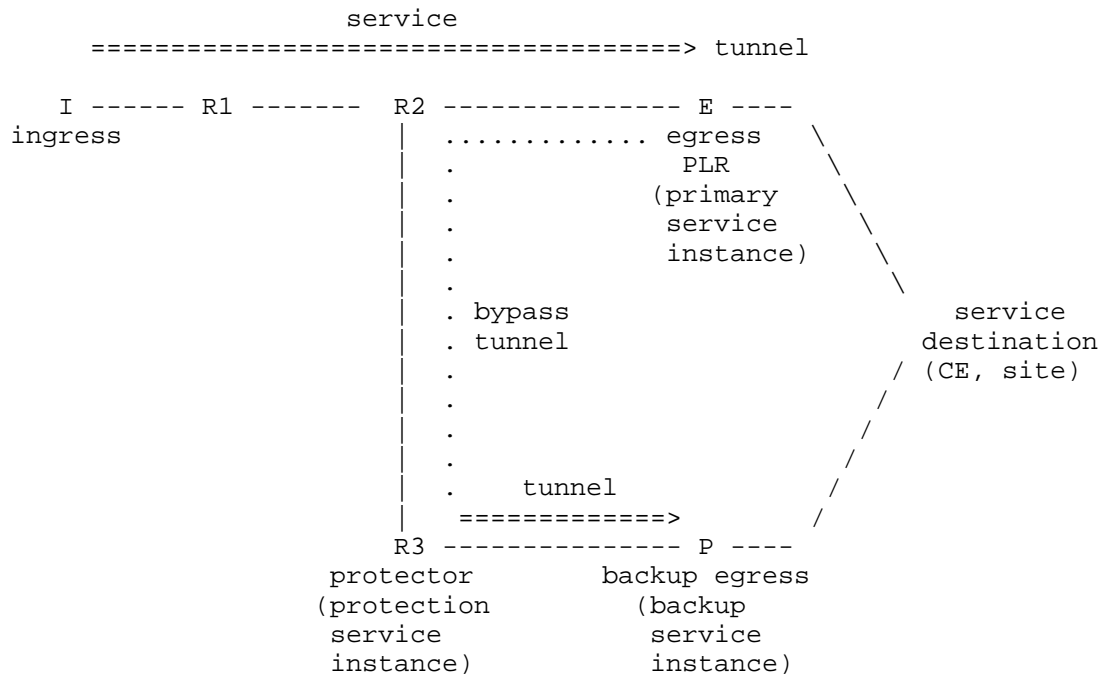


Figure 4

There are two approaches to set up the bypass forwarding state on the egress router, depending on whether the egress router knows the service label advertised 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, and may be used in parallel.

(1) The first approach applies when the egress router does not know the service label advertised 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 {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 and egress link protection in a consistent manner, and both the co-located protector mode and the centralized protector mode may be used (Figure-3 and Figure-4).

(2) The second approach applies when the egress router knows the service label advertised by the backup egress route, 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 should be the same router in the co-located protector mode but may be 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 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 {E, P}. The protector simply forwards rerouted service packets based on its own service label, rather than performing context label switching. With 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 bi-directionally. Therefore, an egress link failure may simultaneously be an ingress link failure for the traffic in the opposite direction. However, 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, it is RECOMMENDED that the traffic SHOULD be moved to an alternative tunnel or alternative services which are fully functional. This is referred to as global repair. Possible triggers of global repair include control plane notifications of tunnel and service status, end-to-end OAM and fault detection at tunnel or service levels, and others. The alternative tunnel and services may be pre-established as standby, or dynamically established as a result of the triggers or network protocol convergence.

8. Example: Layer-3 VPN egress protection

This section shows an example of egress protection for a layer-3 VPN.

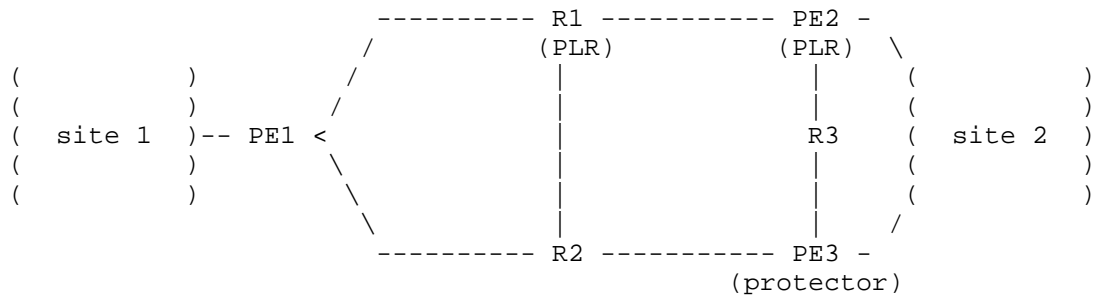


Figure 5

In this example, the site 1 (subnet 203.0.113.192/26) of a given VPN is attached to PE1, and site 2 (subnet 203.0.113.128/26) is dual-homed to PE2 and PE3. PE2 is the primary PE for site 2, and PE3 is the backup PE. Each PE hosts a VPN instance. R1 and R2 are transit routers in the MPLS network. The network uses OSPF as routing protocol, and RSVP-TE as tunnel signaling protocol. The PEs use BGP to exchange VPN prefixes and VPN labels between each other.

Using the framework in this document, the network assigns PE3 to be a protector for PE2 to protect the VPN traffic in the direction from site 1 to site 2. This is the co-located protector mode. Hence, PE2 and PE3 form a protected egress {PE2, PE3}. A context ID 198.51.100.1 is assigned to the protected egress {PE2, PE3}. The VPN instance on PE3 serves as a protection instance for the VPN instance on PE2. On PE3, a 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 the label 100 in its default MPLS forwarding table, with nexthop 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 per-VRF VPN label allocation mode. It assigns a single label 9000 to the VRF of the VPN. For a given VPN prefix 203.0.113.128/26 in site 2, PE2 advertises it along with the label 9000 and other attributes to PE1 and PE3 via BGP. In particular, the NEXT_HOP attribute is set to the context ID 198.51.100.1.

Similarly, PE3 also uses per-VRF VPN label allocation mode. It assigns a single label 10000 to the VRF of the VPN. For the VPN prefix 203.0.113.128/26 in site 2, PE3 advertises it along with the label 10000 and other attributes to PE1 and PE2 via BGP. In particular, the NEXT_HOP attribute is set to an IP address of PE3.

Upon receipt and acceptance of the BGP advertisement, PE1 uses the context ID 198.51.100.1 as destination to compute a TE path for an egress-protected tunnel. The resulted path is PE1->R1->PE2. PE1 then uses RSVP to signal the tunnel, with the context ID 198.51.100.1 as destination, and with the "node protection desired" flag set in the SESSION_ATTRIBUTE of RSVP Path message. Once the tunnel comes up, PE1 maps the VPN prefix 203.0.113.128/26 to the tunnel and installs a route for the prefix in the corresponding VRF. The route's nexthop is a push with the VPN label 9000, followed by a push with the outgoing label of the egress-protected tunnel.

Upon receipt of the above BGP advertisement from PE2, PE3 (i.e. the protector) recognizes the context ID 198.51.100.1 in the NEXT_HOP attribute, and installs a route for label 9000 in the label table pe2.mpls. PE3 sets the route's nexthop to a "protection VRF". This protection VRF contains IP routes corresponding to the IP prefixes in the dual-homed site 2, including 203.0.113.128/26. The nexthops of these routes MUST be based on PE3's connectivity with site 2, even if this connectivity is not the best path in PE3's VRF due to metrics (e.g. MED, local preference, etc.), and MUST NOT use any path traversing PE2. Note that the protection VRF is a logical concept, and it may simply be PE3's own VRF if the VRF satisfies the requirement.

8.1. 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 resulted bypass path is R1->R2->PE3. R1 then signals the path (i.e. egress-protection bypass tunnel), with 198.51.100.1 as 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 hence responds with the context label 100 in an RSVP Resv message.

After the egress-protection bypass tunnel comes up, R1 installs a bypass nexthop for the egress-protected tunnel. The bypass nexthop is a 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 nexthop to reroute VPN service packets. The packets will have the label of the bypass tunnel as outer label, and the VPN label 9000 as inner label. When the packets arrive at PE3, they will have the

context label 100 as outer label, and the VPN label 9000 as 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 IP packets will finally be forwarded to site 2 based on the protection VRF.

8.2. Egress link protection

PE2 serves as the PLR for egress link protection. It has already learned the VPN label 10000 from PE3, and hence it uses the approach (2) described in Section 6 to set up 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 destination. After the bypass tunnel comes up, PE2 installs a bypass nexthop for the VPN label 9000. The bypass nexthop is a label swap from the incoming label 9000 to the VPN label 10000 of PE3, followed by a label push with the outgoing label of the bypass tunnel.

When PE3 detects a failure of the egress link, it will invoke the above bypass nexthop to reroute VPN service packets. The packets will have the label of the bypass tunnel as outer label, and the VPN label 10000 as inner label. When the packets arrive at PE3, the VPN label 10000 will be popped, and the IP packets will be forwarded based on the VRF indicated by on the VPN label 10000.

8.3. Global repair

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

9. IANA Considerations

This document has no request for new IANA allocation.

10. Security Considerations

The framework in this document relies on fast reroute around a network failure. Specifically, service traffic is temporarily rerouted from a PLR to a protector. In the centralized protector mode, the traffic is further rerouted from the protector to a backup egress router. Such kind of fast reroute is planned and anticipated, and hence it should not be viewed as a new security threat.

The framework requires a service label distribution protocol to run between an egress router and a protector. The available security

measures of the protocol MAY be used to achieve a secured session between the two routers.

11. Acknowledgements

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

12. References

12.1. Normative References

- [SR-ARCH] Filsfils, C., Previdi, S., Decraene, B., Litkowski, S., and R. Shakir, "Segment Routing Architecture", draft-ietf-spring-segment-routing (work in progress), 2017.
- [SR-OSPF] Psenak, P., Previdi, S., Filsfils, C., Gredler, H., Shakir, R., Henderickx, W., and J. Tantsura, "OSPF Extensions for Segment Routing", draft-ietf-ospf-segment-routing-extensions (work in progress), 2017.
- [SR-ISIS] Previdi, S., Filsfils, C., Bashandy, A., Gredler, H., Litkowski, S., Decraene, B., and J. Tantsura, "IS-IS Extensions for Segment Routing", draft-ietf-isis-segment-routing-extensions (work in progress), 2017.

12.2. Informative References

- [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>>.
- [BGP-PIC] Bashandy, P., Filsfils, C., and P. Mohapatra, "BGP Prefix Independent Convergence", draft-ietf-rtgwg-bgp-pic-05.txt (work in progress), 2017.
- [RSVP-EP] Chen, H., Liu, A., Saad, T., Xu, F., Huang, L., and N. So, "Extensions to RSVP-TE for LSP Egress Local Protection", draft-ietf-teas-rsvp-egress-protection (work in progress), 2017.

Authors' Addresses

Yimin Shen
Juniper Networks
10 Technology Park Drive
Westford, MA 01886
USA

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

Minto Jeyananth
Juniper Networks
1133 Innovation Way
Sunnyvale, CA 94089
USA

Phone: +1 4089367563
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
Huawei Technologies Co., Ltd.

Email: huaimo.chen@huawei.com

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

Email: jiangyuanlong@huawei.com