Internet Engineering Task Force
Internet-Draft
Intended status: Standards Track
Expires: April 24, 2017

Yimin Shen
Minto Jeyananth
Juniper Networks
Bruno Decraene
Orange
Hannes Gredler
RtBrick Inc

October 21, 2016

MPLS Egress Protection Framework
draft-shen-mpls-egress-protection-framework-03

Abstract

This document specifies a fast reroute framework for protecting IP/MPLS services and MPLS transport tunnels against egress router failures. In this framework, the penultimate-hop router of an MPLS tunnel pre-establishes a bypass tunnel to a protector. Upon an egress router failure, the penultimate-hop router performs local failure detection and local repair, by rerouting traffic over the bypass tunnel. The protector in turn performs context label switching or context IP forwarding to send the traffic to 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 facilitate service label distribution to 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 http://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 April 24, 2017.

Table of Contents

1. Introduction

In MPLS networks, LSPs (label switched paths) 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, etc. In general, a tunnel may carry multiple services of one or multiple types, given that the tunnel can satisfy both individual and aggregate requirements (e.g. CoS, QoS) of these services. The egress router of the tunnel must host corresponding service instances for the services. An MPLS service instance is responsible for forwarding service packets to service destination based on a service label. An IP service instance is responsible for forwarding service packets to service destination based on IP header.

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 (aka. PLR, i.e. point of local repair) upstream adjacent to an anticipated failure pre-establishes a bypass tunnel to the router (aka. MP, i.e. merge point) downstream of the failure, 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, etc) 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 router protection. Similar to the transit link/node protection, this framework relies on local failure detection and local repair to be performed by a PLR, which is the penultimate-hop router of a tunnel. However, there is no MP in this case, because the tunnel does not have a router downstream of the egress router. Instead, this framework relies on a so-called "protector" to serve as the tailend of a bypass tunnel. The protector is a router that hosts some protection service instances and has its own connectivity to service destinations. When the PLR does local repair, the protector is responsible for performing context label switching for rerouted MPLS service packets based on service labels assigned by the egress router, and performing 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 router failure as a failure of a tunnel, as well as a failure of all the services carried by the tunnel, as service packets can no longer reach the service instances

on the egress router. Therefore, the framework addresses protection at both tunnel level and service level simultaneously.

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. In most discussions in this document, the backup egress router serves as a protector, and the service instance hosted on the router acts as a protection instance. In the centralized protector mode (Section 5.14), a protector and a backup egress router may be decoupled.

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 from traffic flow's perspective.

The framework is a multi-service and multi-transport framework. It is applicable to all existing and future types of MPLS tunnels and IP/MPLS services. It does not require extensions to signaling or label distribution protocols of MPLS tunnels, as tunnels and bypass tunnels are expected to be established by using generic mechanisms. It may need extensions for IGPs and service label distribution protocols, to facilitate protection establishment and context label switching. This document provides guidelines for these extensions, but the details should be addressed in separate documents.

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-protected tunnel - A tunnel whose egress router is protected by this framework.

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

Protector - A router acting as an alternate of an egress router and responsible for handling service traffic in the event of a failure of the egress router, as if it were the nominal egress router. It protects an egress-protected tunnel and hosts protection service instances for the egress-protected services carried by the tunnel.

PLR - A router at point of local repair, which is the penultimate-hop router on an 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 for an egress-protected tunnel. It also serves as the virtual location of service instances for 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 - An tunnel established from a PLR to a protector, bypassing the egress router of an egress-protected tunnel.

Protection service instance - A service instance hosted by a protector, protecting the corresponding service instance on an egress router.

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 requirements of the egress protection framework.

- The framework must be based on local failure detection and local repair, in a similar manner to transit link/node protection.

- 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 RSPV, LDP, BGP, IGP, segment routing, etc. It must also accommodate existing and future IP/MPLS services, including layer-2 VPNs, layer-3 VPNs, hierarchical LSP, etc. It must

provide a generic solution for environments where different types of services and transport tunnels co-exist.

- A PLR must be agnostic on services and service labels, like PLRs in the transit link/node protection. It must maintain bypass tunnels and bypass forwarding state on a per-transport-tunnel basis, rather than per-service 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 and/or TE domain to compute or resolve path for a bypass tunnel to a protector.

- 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.

- 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 (aka. end-to-end repair) and control plane convergence.

5. Theory of Operation

5.1. Reference model

This document refers to the following model when describing the framework.

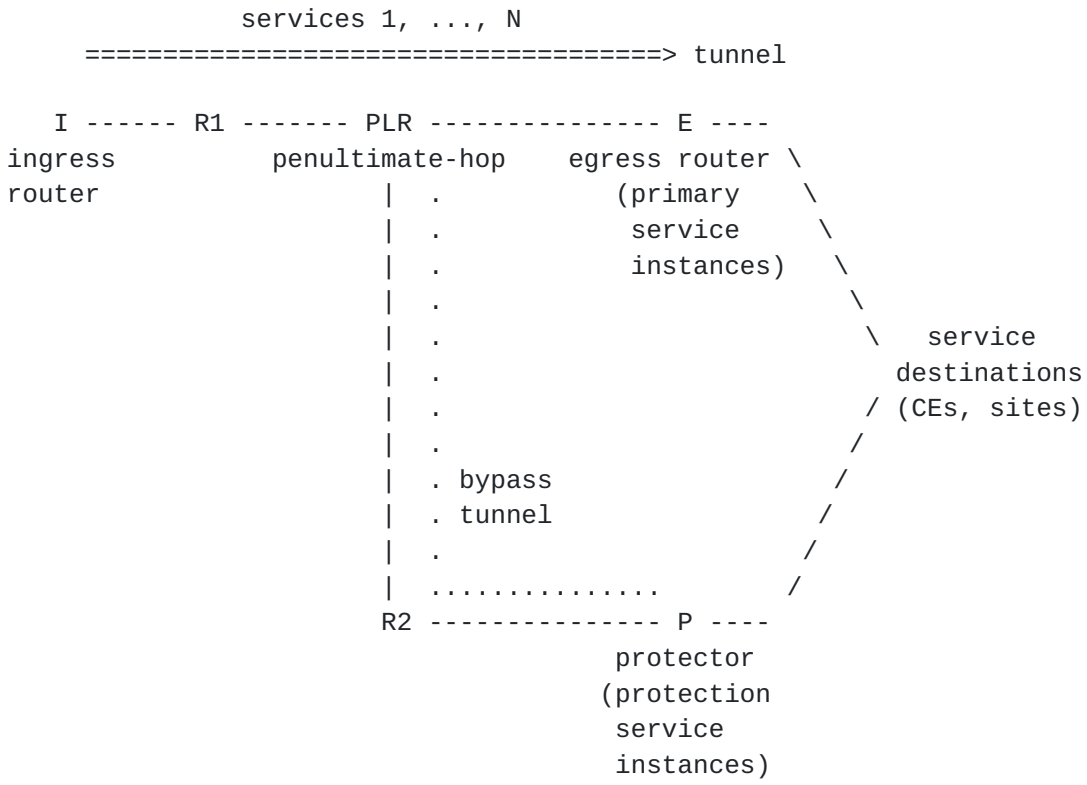

Figure 1

5.2. Egress failure

In this document, an egress failure refers to the node failure of an MPLS tunnel's egress router. At service level, it also means a service instance failure for each service carried by the tunnel.

All the failure detection mechanisms used by PLRs in transit link/node protection are applicable to egress failure detection. In a case where a PLR does not have a fast and reliable mechanism to detect a node failure or distinguish between a link failure and a node failure, it may conservatively treat a link failure as a node failure and trigger egress node protection.

5.3. Protector and PLR

In this framework, a router is assigned to the "protector" role to protect a tunnel and the services carried by the tunnel against an egress 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 which may have 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 failure, the PLR reroutes all the service packets received on the tunnel though the bypass tunnel to the protector. For MPLS service packets specifically, the PLR keeps service labels intact in the packets. The protector in turn forwards 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 failure. This also means that each service destination must be dual-homed or have dual paths to the egress router and 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 in the egress protection schema. 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 different protectors, e.g. P1, P2, etc, with each Pi protecting a subset of the tunnels. Thus, these routers form multiple protected egress', i.e. {E, P1} , {E, P2}, etc. 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 same protected egress {E, Pi}. This mapping must be ensured by the ingress router (Section 5.7).

Two routers X and Y may be protectors for each other's tunnels. In this case, they form two distinct protected egress {X, Y} and {Y, X}.

5.5. Egress-protected tunnel

A tunnel, which is associated with a protected egress {E, P}, is called an egress-protected tunnel. An egress-protected tunnel 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, these tunnels share the common egress router and protector, but may or may not share a common ingress router, a common path, or a common PLR.

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.

5.6. Egress-protected service

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.

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.

5.7. Egress-protected service to egress-protected tunnel mapping

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 5.9).

5.8. 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 tunnel 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 is not affected by any topology change caused by an egress 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 between multiple egress-protected tunnels, if they are 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.

5.9. 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" in this document, 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 routing domain and TE domain (Section 5.10).

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:

- If the service is an MPLS service, when E distributes the 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 dedicated "context ID" object.

- 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.

- 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 or topology-driven label distribution protocol. The PLR uses the context ID as destination to establish or resolve an egress-protection bypass tunnel to P while avoiding E.

- 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.

- 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 the MPLS forwarding table corresponding to E's label space. If the service is an IP service, P installs an IP route in the 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.

- P assigns a non-reserved label to the context ID. In the data plane, this label serves in the context ID's stead to indicate 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 signaled by RSVP, its destination must be the context ID, 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 labels with the context label as the inner-most label (Section 5.11). In any case, the bypass tunnel is a UHP tunnel whose incoming label at P is the context label.

- During local repair, all the service packets received by P on the bypass tunnel will have the context label as top label. P will first pop the context label. For MPLS service packets, P will further look up the service label in E's label space indicated by the context label. This is called context label switching. For IP service packets, P will look up the IP destination address in E's IP address space indicated by the context label. This is called context IP forwarding.

5.10. IGP advertisement and path resolution for context ID

Path resolution or computation for context ID is done on ingress router for egress-protected tunnel, and on PLR for egress-protection bypass tunnel. Therefore, given a protected egress {E, P} and its context ID, E and P must coordinate in IGP advertisement for the context ID in routing domain and TE domain. The context ID must be advertised in such a manner that any egress-protected tunnels MUST have E as tailend, and any 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 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 paths 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 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 may establish an egress-protection bypass tunnel in several manners (Section 5.11). The "context ID label binding" advertisement may use the IGP extensions for IGP mirroring context segment described in [SR-ARCH], [SR-OSPF] and [SR-ISIS].

5.11. 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 egress-protected tunnel. The PLR may or may not need to have the knowledge of 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). As the context ID is advertised in routing domain and TE domain by IGP according to Section 5.10, the PLR should be able to resolve or establish such a bypass tunnel 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 tunnel.

[2] It may be formed by a topology driven protocol (e.g. LDP). The protector binds the context label to the context ID as an IP prefix FEC.

[3] It may be constructed as a hierarchical tunnel. When the protector uses the alias mode (Section 5.10), 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, etc.

[4] It may be constructed by using segment routing. In this case, the protector uses the alias mode (Section 5.10), and advertises the context ID and context label binding as an IGP mirroring context segment. The PLR can then construct the bypass tunnel as a stack of labels, with the context label as the inner-most label.

5.12. Local Repair on PLR

A PLR is agnostic on services and services labels. This obviates the need to maintain bypass forwarding state on per-service basis, and allows bypass tunnel sharing between egress-protected tunnels. During local repair, the PLR simply reroutes all service packets received on a tunnel to the corresponding 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 in-label of the egress-protected tunnel to the out-label of the bypass tunnel. If the bypass tunnel is a hierarchical tunnel, the rerouting will involve swapping the in-label of the egress-protected tunnel to a context label, and pushing

the out-label of a regular bypass tunnel. If the bypass tunnel is constructed by segment routing, the rerouting will involve swapping the in-label of the egress-protected tunnel to a stack of labels, with a context label as the inner-most label.

5.13. 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 session of service label distribution protocol 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 installs the service label with the forwarding state in the label space of the egress router, as 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 are out of the scope of this framework, and SHOULD be specified in separate documents.

5.14. 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 the number of backup egress routers. As another scenario, a network may assign a single router to serve as a dedicated protector for all egress routers. This protector is topologically decoupled from backup egress routers, and is called a centralized protector.

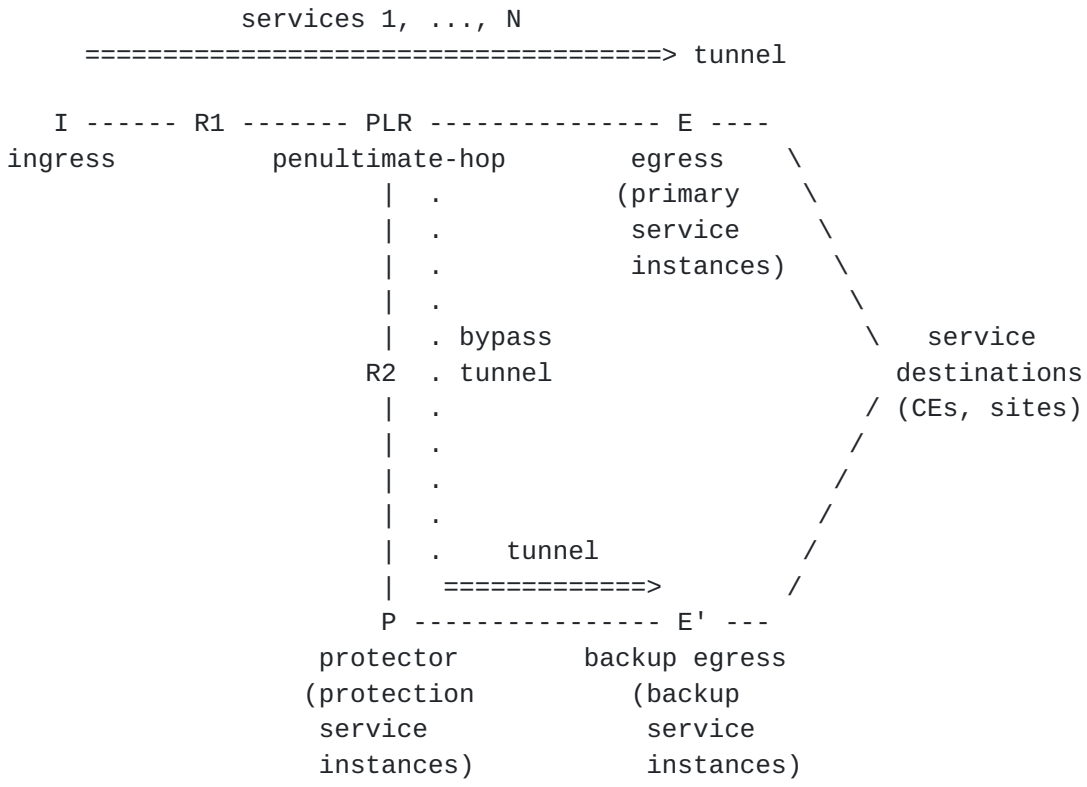

Figure 2

Like a co-located protector, a centralized protector hosts protection service instances, receives rerouted service traffic from PLR, and performs context label switching and/or context IP forwarding. For each service, instead of sending traffic directly to the service destination, the protector MUST send it over a transport tunnel to the corresponding backup egress router. The backup egress router in turn forwards the traffic to the service destination. Specifically, in the case of an MPLS service, the protector MUST swap the service label in each received packet to the service label of corresponding service advertised by the backup egress router, and then push a 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 session of service label distribution protocol between the backup egress router and the protector, in addition to the session between the egress router and the protector Section 5.13). Through this protocol, the backup egress router distributes its service label binding, the protected service FEC (which may be learned from configuration), and the context ID of the protected egress {E, P}. Based on this information, the protector associates the egress-protected service with the service on the backup egress router,

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 protected egress router E.

6. Global repair

The framework in this document provides a fast but temporary repair for traffic upon an egress failure. For permanent repair, it is RECOMMENDED that the traffic SHOULD be moved to an alternative tunnel or alternative services that are fully functional. This is referred to as global repair. Possible triggers of global repair include control plane notifications for tunnel and service status, end-to-end OAM and fault detection at tunnel and service levels, etc. These alternative tunnel and services may be pre-established backups, or newly established as a result of the triggers or network protocol convergence.

7. Example: Layer-3 VPN egress protection

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

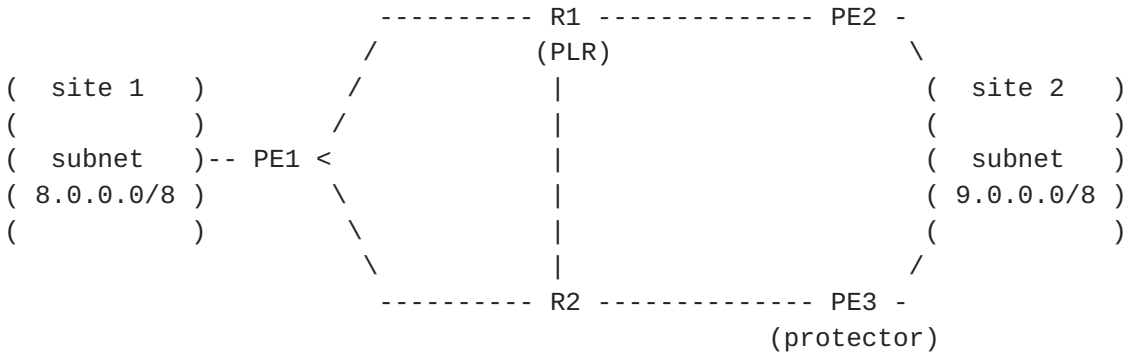

Figure 3

In this example, the site 1 of a given VPN is attached to PE1, and site 2 is dual-homed to PE2 and PE3. PE2 is the primary PE for site 2, and PE3 is the backup PE. Every 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. Hence, PE2 and PE3 form a protected egress {PE2,

PE3}. A context ID 1.1.1.1 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 routing domain and TE domain.

PE2 uses per-VRF VPN label allocation mode. It assigns a single label 9000 for the VRF of the VPN. For a given VPN prefix 9.0.0.0/8 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 1.1.1.1.

Upon receipt and acceptance of the BGP advertisement, PE1 uses the context ID 1.1.1.1 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 1.1.1.1 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 9.0.0.0/8 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 out-label of the egress-protected tunnel.

Upon receipt of the above BGP advertisement from PE2, PE3 (i.e. the protector) recognizes the context ID 1.1.1.1 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 9.0.0.0/8. The nexthops of these routes MUST be based on PE3's connectivity with site 2, 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.

R1, i.e. the penultimate-hop router of the egress-protected tunnel, serves as PLR. Based on the "node protection desired" flag and the destination address (i.e. context ID 1.1.1.1) of the tunnel, R1 computes a bypass path to 1.1.1.1 while avoiding PE2. The resulted bypass path is R1->R2->PE3. R1 then signals the path as an egress-protection bypass tunnel, with 1.1.1.1 as destination.

Upon receipt of RSVP Path message of the egress-protection bypass tunnel, PE3 recognizes the context ID 1.1.1.1 as the destination, and hence responds with the context label 100 in RSVP Resv message.

Once 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 in-label of the egress-protected tunnel to the out-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.

Eventually, global repair will take effect, as control plane protocols (BGP, OSPF, RSVP) 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.

8. IANA Considerations

This document has no request for new IANA allocation.

9. Security Considerations

This document does not introduce any security issues.

Note that the framework requires a label distribution protocol to run between an egress router and a protector, which is achievable in a secured fashion.

10. Acknowledgements

This document leverages work done by Yakov Rekhter, Kevin Wang, Zhaohui Zhang and several on MPLS egress protection.

11. References

11.1. 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), 2016.

[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), 2016.

[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), 2016.

11.2. 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, <http://www.rfc-editor.org/info/rfc4090>.

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

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

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