Workgroup: BESS WG
Published: 22 August 2020
Intended Status: Standards Track
Expires: 23 February 2021
Authors: Y. Wang R. Chen
ZTE Corporation ZTE Corporation

# EVPN Egress Protection

## Abstract

A fast reroute framework for egress node protection is specified by [RFC8679] . But it cannot be applied to EVPN directly. This document specifies a mechanism to apply Egress Node Protection to EVPN nodes and apply Egress Link Protection to EVPN EAD/EVI routes.

In Section 6.2, this draft is compared with three other drafts. These drafts are:

*VTEP Group - [I-D.eastlake-bess-evpn-vxlan-bypass-vtep].
*DT2UL DX2L - [I-D.hu-bess-srv6-vpn-bypass-sid].
*Mirror SID - [I-D.ietf-rtgwg-srv6-egress-protection].

## 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 23 February 2021.

**Table of Contents**

## 1. Introduction

A principal feature of EVPN is the ability to support multi-homing from a customer equipment (CE) to multiple PE with Ethernet segment (ES) links. This draft specifies a VXLAN/SRv6 gateway mechanism to simplify PE processing in the multi-homed case and enhance EVPN convergency on egress failures.

### 1.1. Terminology and Acronyms

This document uses the following acronyms and terms:

*All-Active Redundancy Mode - When a device is multihomed to a group of two or more PEs and when all PEs in such redundancy group can forward traffic to/from the multihomed device or network for a given VLAN.

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

*BUM - Broadcast, Unknown unicast, and Multicast.

*CE - Customer Edge equipment.

*DCI - Data Center Interconnect.

*EELP bypass tunnel - Egress ESI Link Protection bypass tunnel - A tunnel used to reroute service packets upon an egress ESI link failure.

*Egress failure - An egress node failure or an egress link failure.

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

*Egress loopback - the loopback interface on the Egress router, whose IP address is the destination of the Egress-protected tunnel.

*Egress node failure - A failure of an egress router.

*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-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 EVI - An EVPN MAC-VRF or IP-VRF that is carried by an egress-protected tunnel and hence protected by a mechanism according to this framework.

*Egress-protecting tunnel - A VXLAN tunnel whose destination IP address is the same value as the Egress-protected tunnel. The Egress-protecting tunnel is constructed on the Protector not on the Egress router. The egress router of the egress-protecting tunnel is the protector. Note that from the view of the ingress router the egress-protecting tunnel and the egress-protected tunnel is the same tunnel.

*ESI - Ethernet Segment Identifier - A unique non-reserved identifier that identifies an Ethernet segment.

*NVE - Network Virtualization Edge.

*OPE - Originating PE - the original Router of an EVPN route.

*PE - Provider Edge equipment. Note that VTEP/NVE are also called as PE in this draft.

*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.

*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 is physically independent of the egress router.

*Protector loopback - the loopback interface on the Protector, whose IP address is the destination of the Egress-protected tunnel.

*Single-Active Redundancy Mode - When a device or a network is multihomed to a group of two or more PEs and when only a single PE in such a redundancy group can forward traffic to/from the multihomed device or network for a given VLAN.

*VTEP - VXLAN Tunnel End Point.

*VXLAN - Virtual eXtensible Local Area Network [RFC7348].

*GRT - Global Routing Table.

*EVPN SID - SRv6 SID for EVPN Instances, e.g. End.DT2M SID, End.DT2U SID, End.DX2 SID, End.DX2V SID.

*DF - Designated Forwarder.

*NDF - non-DF, non Designated-Forwarder.

*NDF-Bias - An exception for filtering bypassed BUM packets. It says that when an outgoing AC is a NDF on its ES, the bypass-BUM filter rules will not be applied for that AC.

## 2. Detailed Problem and Solution Requirement

### 2.1. Scenarios and Basic Settings

In the scenario illustrated in Figure 1, where an CE1 is dual-homed to PE1 and PE2 to access the VXLAN/SRv6 network, which enhances network access reliability. When one PE fails, services can be rapidly switched to the other PE, minimizing the impact on services.

As shown in Figure 1, the EVPN instance EVI1 has three PEs, PE1, PE2 and PE3. The PE address of PE1 is IP1 and the PE address of PE2 is IP2, the PE address of PE3 is IP3, they are three different IP addresses. The BGP update-source of PE1 is IP_N1, of PE2 is IP_N2, and of PE3 is IP_N3.

LOC1 is the prefix from which IP1 is allocated, LOC2 is the prefix from which IP2 is allocated, LOC3 is the prefix from which IP3 is allocated.

Note that IP_N1 must not be allocated in prefix LOC1, IP_N2 must not be allocated in prefix LOC2. But IP_N3 may be the same as IP3.

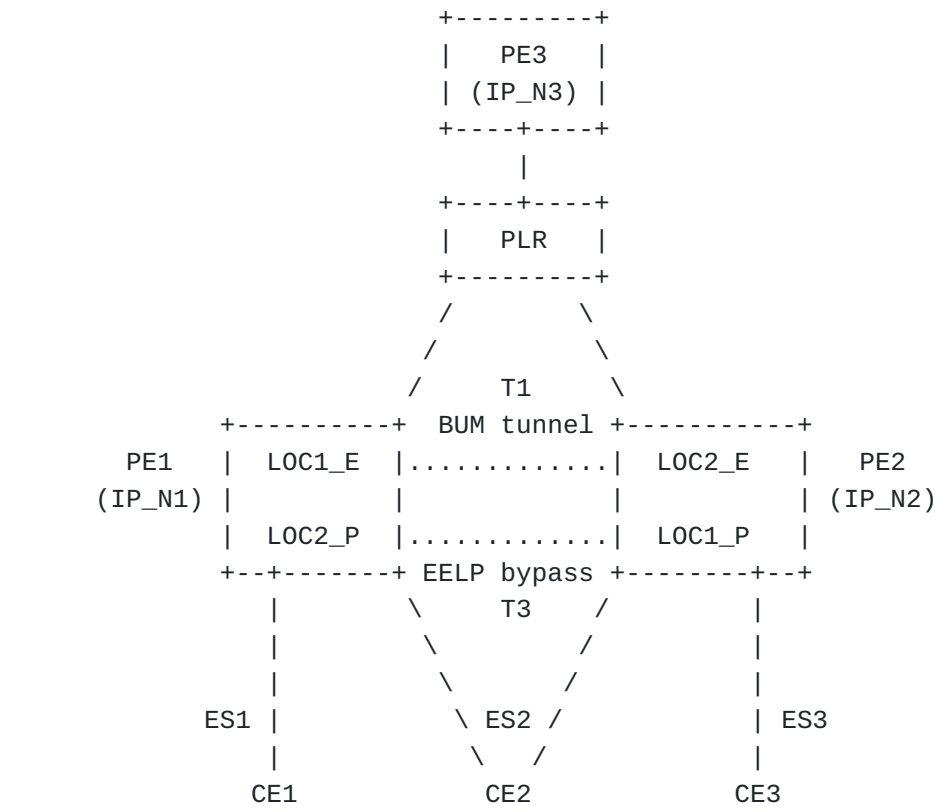


Figure 1: Egress Protection Scenario

Both PE1 and PE2 will advertise an IGP route for the prefix LOC1. The prefix LOC1 on PE1 is called LOC1_E, The prefix LOC1 on PE2 is called LOC1_P. Then we make the metric of LOC1_P lower than LOC1_E. Then we do the same for LOC2.

As a result, the PLR node will install an egress-FRR entry for LOC1 and LOC2. The primary egress for LOC1 is PE1, the backup egress for LOC1 is PE2. So when PE3 send a EVPN data packet to an IP address of LOC1, the PLR node will use PE1 as the primary path, and use PE2 as the backup path.

The different settings between VXLAN EVPN scenario and SRv6 EVPN scenario are described in the following list:

*In VXLAN EVPN scenario: LOC1_E is a loopback interface on PE1, LOC1_P is a loopback interface on PE2. Both of the two loopback interfaces are configured with the sam IP address IP1. LOC1 is the /32 or /128 prefix for IP1. So we also use LOC1 to stand for IP1 in VXLAN context in the remainder of this draft. The loopback interface for LOC1_E is IP1's egress loopback interface. The loopback interface for LOC1_P is IP1's protector loopback interface. Then we do the same for IP2.

*In SRv6 EVPN scenario: IP1 is the node SID of PE1, LOC1_E is the SRv6 locator for IP1. IP2 is the node SID of PE2, LOC2_E is the SRv6 locator for IP2. LOC1_P is the mirror of LOC1_E in PE2's GRT. LOC2_P is the mirror of LOC2_E in PE1's GRT.

#### 2.1.1. Common Problems

When PE2 receives an EVPN route R0 whose nexthop matches the prefix LOC1, PE2 may discard the route R0 because its nexthop is considered to be PE2's own address. Even though PE2 don't disccard R0, PE2 cannot use its nexthop to send an EVPN data packet to PE1.

Because that a destination IP within prefix LOC1 (in forms of LOC1_P) will be considered to be sent to PE2 itself. So we should use IP_N1 and IP_N2 to establish the bypass path between PE1 and PE2 instead of LOC1 and LOC2.

### 2.2. VXLAN-Specific Requirements

When PE2 receives the EVPN routes from PE3, only the VXLAN tunnel <LOC2_E, IP_N3> will be installed according to [RFC8365]. The VXLAN tunnel <LOC1_P, IP_N3> will not be installed. So when PE1 fails, although the packets to PE1 are fast-rerouted to PE2 by PLR, PE2 may discard these packets because of the absent of the corresponding VXLAN tunnel entity for their SIP and DIP.

### 2.3. SRv6-Specific Requirements

The PLR will not be expected to support any Segment Routing extensions at all, it is just asummed to be an ordinary IPv6 router.

When PE2 receives an EVPN data packet whose bottommost SID is an EVPN SID from LOC1. Although the EVPN SID can match the prefix LOC1_P on PE2, the EVPN data packet will be dropped because of the absence of SRv6 function indications.

### 2.4. ESI-Label-Specific Requirements

ESI-label is used for MPLS EVPN, SRv6 EVPN and Geneve EVPN, it is typically downstream-assigned in ingress-replication scenarios.

When a packet is received from a mirrored End.DT2M SID, the ESI-label in the SID's ARG.FE2 part have to be lookup in a label space that is different from the native ESI-label's label space. Otherwise the packet should be dropped. In multi-homing scenarios, the mirrored End.DT2M SIDs for differnt OPE must do ESI-label lookup in defferent label space, at the same time, the mirrored End.DT2M SIDs for the same OPE should do ESI-label lookup in the same label space.

But ESI labels are used only when two PEs from the same egress protection group send BUM packets to each other. In such case, they can use the second approach of [RFC8679] section 6 to transport these BUM packets. When that approach is used, the context-specific ESI-label lookup is typically not necessary, even if it is in the MPLS EVPN scenarios.

### 2.5. MPLS-Specific Requirements

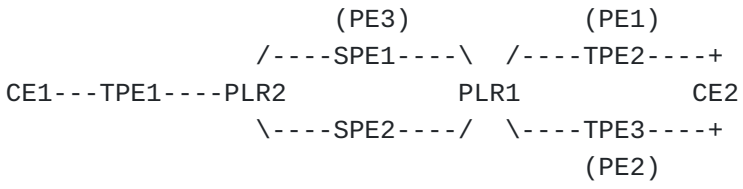


Figure 2: Anycast SPEs and Egress Protecting TPEs

The above figure is a combination of Figure 1 and [I-D.wang-bess-evpn-context-label]'s Figure 6. The TPE1/SPE1/SPE2/TPE2 above is the TPE1/SPE1/SPE2/TPE2 of [I-D.wang-bess-evpn-context-label]'s Figure 6, But TPE2 is also the PE1 of Figure 1, and TPE3 is the PE2, SPE1 is the PE3.

When TPE2 advertises an EVPN route (say R9), the same R9 will be advertised to both the two SPEs and TPE3. When TPE3 receives R9, they will do EVPN egress protection. When SPE1 or SPE2 receives the same R9, SPE1/SPE2 will advertise R9 to TPE1 with the same nexthop (the anycast tunnel address of SPE1 and SPE2) following [I-D.wang-bess-evpn-context-label]'s section 5.4.

Then the requirement here is clear that we want TPE2 use the same route attributes to advertise R9 to both the SPEs and the TPEs.

In addition, Note that when the BUM tunnel (T1) from PE1 (TPE2) to PE2 (TPE3) travels through the PLR1, and the second approach of [RFC8679] section 6 is used, and the PLR1 reroutes these packets (destined to PE2) to PE1 when PE2 fails, at that moment, PE1 should drop these packets because their EVI label are mirrored EVI labels (in context-specific label space) but their ESI labels are not absent.

## 3. Encoding the Originating Router's IP Address

This sections describe the extensions specified to meeting the requirements given in Section 2 and enhance VXLAN EVPN convergency.

This document reuse the OPE TLV defined in [I-D.heitz-bess-evpn-option-b] section 3. The OPE TLV carries the BGP update-source on corresponding PE. The PEs with egress protection procedures

described in this document will add the OPE TLV in the EVPN routes that they are about to advertise.

Note that the ESI label or leaf Label is not used in VXLAN packet, so the usage for OPE TLV here won't conflict with the usage in [I-D.heitz-bess-evpn-option-b].

# 4. Control Plane Processing

## 4.1. Common Procedures

We will discuss the common procedures for VXLAN EVPN and SRv6 EVPN first. Then we will discuss the procedures specific to VXLAN EVPN or SRv6 EVPN in the following sections.

Using the topology in Figure 1, the common procedures are described in the following list:

**[C1]** PE3 sends a MAC/IP route R1 and an IMET route R2 to PE1 and PE2. The nexthop of these routes is IP_N3 (we assume that IP_N3=IP3). PE3 won't add the OPE TLV to these routes because it works as a normal EVPN PE.

**[C2]** PE1 and PE2 receive R1 and R2 from PE3.

**[C3]** PE1 sends a MAC/IP route R4, an EAD/EVI route R5 and an IMET route R6 to PE2 and PE3. The nexthop of these routes is IP1. PE2 sends a MAC/IP route R7, an EAD/EVI route R8 and an IMET route R9 to PE1 and PE3. The nexthop of these route is IP2. PE1 and PE2 will both add the OPE TLV to these routes because they are configured with protector LOCs. The OPE TLV carries their BGP update-source IP address (IP_N1 or IP_N2).

**[C4]** When PE2 receives R4, R5 and R6 from PE1, it installs the bypass tunnel <IP_N2, IP_N1>. Because that their nexthops is an IP address within the prefix LOC1_P on PE2. The bypass tunnel <IP_N1, IP_N2> is called Egress ESI Link Protection (EELP) bypass tunnel. and PE2 will apply the egress link protection procedures to the received EAD/EVI route R5 following the second approach of [RFC8679] section 6. Please see [Section 5.2] for details.

Note that the MAC/IP entries from PE1 is installed in GRT by PE2 as mirrored entries.

The procedures when PE1 receives R7, R8 and R9 are simlar to the above.

**[C5]** When PE3 receives the EVPN routes from PE1/PE2, it will ignore the OPE TLV because the route's tunnel encapsulation is VXLAN or SRv6 and the nexthop is not a local address on PE3.

### 4.2. VXLAN-Specific Procedures

First of all, we assume that the VNI for the same EVI on PE1 and PE2 must be the same.

The VXLAN-specific procedures are defined in the following list:

**[V1]** In [C2], when PE1 receives the MAC/IP route from PE3, it constructs two VXLAN tunnels: <LOC1_E, IP_N3> and <LOC2_P, IP_N3>. Because it is configured with egress loopback and protector loopback.

**[V2]** In [C4], the bypass tunnel is a VXLAN tunnel.

### 4.3. SRv6-Specific Procedures

In VXLAN EVPN, the VXLAN tunnels are constructed for packet-validating purpose. In SRv6 EVPN, there aren't such packet-validating tunnels. So when a SRv6 PE receives EVPN routes from other PEs, no packet-validating tunnels will be installed.

But the bypass tunnels aren't constructed for packet-validating purpose, they are used to transport flows among the PEs of the same egress protection group. So the bypass tunnel between PE1 and PE2 must also be installed in SRv6 EVPN scenarios.

The SRv6-specific procedures are defined in the following list:

**[6A]** In [C4], The bypass tunnel is a SRv6 tunnel.

**[6B]** In [C4], When PE2 receives R4, R5 from PE1, it mirrors the remote EVPN-SID of these routes in the GRT. This is for the requirements from [Section 2.3].
Note that the End.DT2M SID of IMET route R5 SHOULD not be mirrored by default on PE2. The reason will be described in Section 5.1.1.

**[6C]** In [C4], PE2 may apply the egress link protection control-plane procedures to the received EAD/EVI route R5 following the first approach of [RFC8679] section 6. The corresponding data-plane details will be described in [UA-SID1].

### 4.4. MPLS-Specific Procedures

First of all, We reserve a portion of the label space for assignment by a central authority. We refer to this reserved portion as the "Domain-wide Common Block" (DCB) of labels. This is analogous to the DCB that is described in [I-D.wang-bess-evpn-context-label]'s section 5.4. The DCB is taken from the same label space that is used for downstream-assigned labels, but each PE would know not to allocate local labels from that space. A PE would know by provisioning which label from the DCB corresponds to itself, and each of other labels from the DCB correspond to each PE of the domain.

Note that the PEs don't have to know exactly which label correspond to a specified PE, They just need know which label is for itself, and other labels is not for itself.

The MPLS-specific procedures are defined in the following list:

**[M1]** In [C3], when TPE2(PE1) advertise R4/R5/R6, It following [I-D.wang-bess-evpn-context-label]'s section 5.4. This means that a Downstream-CLS ID EC will be advertised along with R4/R5/R6. And this EC carries the label (in DCB) that identifying TPE2(PE1) itself.

**[M2]** In [C4], when TPE3(PE2) receives R4/R5/R6, It install the mirrored ILM entry in a context-specific labels space (say CLS23). The CLS23 is identified by the Downstream-CLS ID EC (say CIL2) of R4/R5/R6. The mirrored ILM entry is called as a CLS-specific ILM entry (CLS-ILM).

**[M3]** In [C5], when SPE1(PE3) receives R4/R5/R6, It should impose the context-identifying label (CIL) carried in R4/R5/R6's Downstream-CLS ID EC onto the label stack following [I-D.wang-bess-evpn-context-label]'s section 5.4. That CIL is the outer label of the EVPN label of R4/R5/R6. In addition, SPE1(PE3) will aplly the proedures of [I-D.wang-bess-evpn-context-label]'s section 5.4 too. Although these procedures is not of EVPN egress protection schema, they share the same signalling with EVPN protection. This simplifies the signalling procedures, because there no longer will be a requirement to advertise different route attributes to different PEs.

## 5. Protection Procedures

This section describes how Layer 2 unicast and BUM (Broadcast, Unknown unicast, and Multicast) packet forwarding are protected. A description of how Layer 3 packet forwarding are protected will be provided in a furture version of this document.

### 5.1. EVPN Egress Node Protection (EENP)

The following two subsections discuss EENP procedures for BUM forwarding and Unicast Forwarding.

#### 5.1.1. BUM Forwarding Protection

##### 5.1.1.1. Bypass-BUM Filter

PE3 will do ingress replication to PE1 and PE2 for BUM packets, one copy for each PE. So BUM packets need not to be protected by the egress node protection mechanism. But there will be another issue along with the BUM packets. It is:

> PE1 and PE2 will receive a copy of BUM packet from PE3 separately, and the DF node for the <ESI2, EVI> will forward it to the CE node. When the non-DF node of them fails, the BUM packets destined to it will be re-routed to the other one, which will be the DF-node for that <ESI2, EVI>, so these BUM packets will be forwarded to CE2. But PE3 has sent another copy of BUM packets directly to the DF-node, and this copy will be forwarded to CE2 either. So CE2 will receive duplicated BUM packets from the DF-node after the FRR switch of PLR untill the global repair finishes.

In order to avoid the excessive BUM packets, some specific rules are defined in the following list:

* For VXLAN EVPN: The BUM packet received via LOC1_P or LOC2_P will be dropped.

* For SRv6 EVPN: Because the End.DT2M SIDs are not mirrored according to [Section 4.3], those duplicating BUM packets will be dropped due to the absence of SRv6 function indication.

The bypass-BUM-filter also insures that when PE1 send BUM packet to PE2 but PE2 fails, The FRR-switch on PLR node will not bring out duplicated BUM packets to PE1's local CEs. The tunnel address from PE1 to PE2 for BUM packets is called T1, The tunnel address from PE3 to PE2 for BUM packets is called T2, The bypass-BUM-filter also insures that T1 can be the same as T2.

Note that the BUM tunnel T1 may travel through PLR too.

##### 5.1.1.2. NDF-Bias Rules for Bypass-BUM filter

When PE1 node fails, the DF-election between PE1 and PE2 will be restarted, and the PLR will do FRR-switch. We assume that the PLR FRR-switch may be faster than the DF re-election. So if PE1 is the DF node for <ESI2, EVI1> before its failure, the rules that

described in [Section 5.1.1.1] will cause packet drop before the DF re-election finishes. Because that PE2 will be the non-DF (NDF) node for <ESI2, EVI1> at that time.

In order to accelerate the convergence of bypass-BUM packets, some specific rules are defined in the following list:

* For VXLAN EVPN: The BUM packet received via LOC1_P or LOC2_P will not be dropped when it is about to forwarded to an AC whose DF-role is NDF.

* For SRv6 EVPN: The End.DT2M SIDs need to be mirrored, those BUM packets received via a mirrored End.DT2M SID will not be dropped only when it is about to forwarded to an AC whose DF-role is NDF. Note that a single-homing AC is always considered as DF-role, so it will filter all bypass BUM packets.

  When BUM packets are received via a mirrored End.DT2M SID, and their Arg.FE2 parts are not empty, such BUM packets will be dropped. Because that such BUM packets just originated inside the same protection group of current PE node.

These rules are called as "NDF-Bias" rules in this draft.

### 5.1.2. Unicast Forwarding Protection

When PE1 fails, the data packets (destined to CE2) from PE3 to PE1 are fast- rerouted to PE2 by the PLR node in the underlay network, the PE2 won't discard these packets because of the existence of VXLAN tunnel<IP3, IP1_P> or mirrored EVPN SID or mirrored CLS-ILM on PE2 itself. The PE2 will forward them to CE.

Note that in MPLS scenario (Figure 2), SPE1(PE3) will impose CIL2 onto the label stack, so the PLR1 wouldn't impose CIL2 (in fact that CIL2 need not be advertised in the underlay network) again. PLR1 just do ordinary anycast FRR or TI-LFA.

## 5.2. Egress ESI Link Protection (EELP)

The EELP <ESI, EVI> forwarding entry on PE1 will take the ESI link as primary forwarding path, and take the EAD/EVI route from PE2 as backup forwarding path. This procedure follows the second approach of [RFC8679] section 6.

When the ESI2 link fails, the backup path will be activated on the result of a FRR switch by the overlay network.

Note that even when the ESI is All-Active redundancy mode the EELP will follow the FRR behavior. The EELP behavior is the same for All-Active redundancy mode and Single-Active redundancy mode.

When ESI is All-Active redundancy mode PE3 will performing overlay ECMP via EAD/EVI routes to PE1/PE2, When the ESI link on PE1 fails, PE1 will forwarding the packets via EELP bypass tunnel before PE3 delete the EAD/EVI routes. But the bypass forwarding is temporary, after PE3 delete the EAD/EVI routes upon the withdraw of the EAD/EVI route from PE1, there won't be any bypass forwarding again.

Given that the destination IP address of the EELP bypass tunnel from PE1 to PE2 is called T3, note that T3 may be not the same as T1 or T2 of section 5.1.1.1. When T3 is a different IP address, different forwarding behaviors can be applied. For example, T3 should not be protected by PLR's egress node protection procedures.

When T3 is different from T1/T2, [6C] can be used. Otherwise, only the second approach of [RFC8679] section 6 can be used if the travel from PE1 to PE2 passes the PLR node.

Note that when ESI is Single-Active redundancy mode, there is no importance for PE3 to use the EAD/EVI routes from PE1/PE2. But the EAD/EVI route is still useful between PE1 and PE2 for EELP procedures in Single- Active redundancy mode.

### 5.2.1. Source Squelching Rules

When a PE of an egress protection group receives packets from an EELP bypass tunnel, that PE MUST not send it to another PE in the same egress protection group over any of the bypass tunnels. But when a PE receives a VXLAN/SRv6 encapsulated data packet from an ordinary underlay destination IP address, that PE can bypass that packet following a bypass tunnel.

### 5.2.2. SRv6-specific EELP Rules

**[SID-Hiden1]** When the ESI2 link on PE1 fails, the bypassed EVPN packet's underlay destination IP adrress can't be the EVPN SID directly. The EVPN SID have to be hiden in the SRH header in order to avoid the problems described in [Section 2.1.1], even if the bypass tunnel is a SR-BE tunnel.

**[SID-Hiden2]** Note that there are no egress link protection for BUM packets, all the bypass-BUM packets are the result of egress node protection on the PLR.

But when CE1 requests CE3's MAC address, PE1 can't forward the ARP request to PE2 using SRv6 BE encapsualtion. Although these ARP packets are not bypassed packets, the EVPN SID have to be hiden in the SRH header too, in order to avoid the same problems as above.

**[UA-SID1]**

When [6C] is applied, PE1 will use the EVPN SID of itself to encapsulate the bypassed EVPN packet, not use the EVPN SID from the received EAD/EVI route.

When that bypassed EVPN packet is received by PE2, the packet will match a mirrored EVPN SID on PE2. So PE2 will know that the packet is a bypassed data packet. The bypassed data packets will be forwarded to local CEs only.

**[UA-SID2]** When [6C] is applied, the bypass tunnel's destination must be an IP-address for whom the PLR will not do egress node protection. Otherwise micro-loops will arise when the FRR-switch of that egress node protection is triggered on the PLR node.

#### 5.2.3. MPLS-specific EELP Rules

The first approach of [RFC8679] section 6 should be applied, and the bypass tunnel's destination must be an IP-address for whom the PLR1 will not do egress node protection. Otherwise micro-loops will arise when the FRR-switch for that egress node protection is triggered on the PLR node.

It means that the send TPE will impose the EVPN label and CIL of itself onto the label stack, on the failure of local ACs.

## 6. Comparison with Other Drafts

### 6.1. Questions

We compare this draft with three other drafts in this section. These drafts are:

*VTEP Group - [I-D.eastlake-bess-evpn-vxlan-bypass-vtep].
*DT2UL DX2L - [I-D.hu-bess-srv6-vpn-bypass-sid].
*Mirror SID - [I-D.ietf-rtgwg-srv6-egress-protection].

We use the following questions for these solutions to do the comparison:

**[Steady bypassing]**
*When AC fails but PE node still works well, will there be steady bypassing traffic?

**[Bypass-BUM filter]**
*Is Bypass-BUM filter rules supported by that solution? The "Bypass-BUM filter" rules are defined in [Section 5.1.1.1].

**[Node Protect]**
*Will Egress Node Protection be supported?

**[Link Protect]**

*Will Egress Link Protection be supported?

**[PLR SRv6-aware]**

*Must the PLR node be SRv6-aware? Note that the PLRs here meant only the PLRs for egress node failure.

**[Extra VPN SID]**

*Should Extra EVPN SID be configured (Like what have been done by "DT2UL DX2L" Solution) for that solution?

**[Special BGP]**

*Should special BGP extensions dedicated to SRv6 scenarios be implemented for that solution?

**[Special IGP]**

*Should special IGP extensions dedicated to SRv6 scenarios be implemented for that solution?

**[Implicit IP-VRF]**

*Are implicit IP-VRF instances (identified by "mirror SID") needed by that solution?

**[RT-1 Given up]**

*Is Ethernet A-D route per EVI have to be given up in that solution?

**[NDF-Bias Rules]**

*Is NDF-Bias rules supported by that solution? The "NDF-Bias" rules are defined in [Section 5.1.1.2]

**[VXLAN-Suitable]**

*Is that solution suitable for NVO3 EVPN?

**[ESI-L Mirroring]**

*Is that solution discussed the mirroring ESI-label?

## 6.2. Summary Comparisons

We place the detailed comparisons about the answers of these questions for each solution in separated sections, but we place the brief comparisons in the following table:

| Questions | This Draft | DT2UL | Mirror-SID | VTEP-Group |
|---|---|---|---|---|
| Steady-bypassing | No | No | No | Yes |
| Bypass-BUM-filter | Yes | No | No | No |
| Node-Protect | Yes | No | Yes | Yes |
| Link-Protect | Yes | Yes | No | Yes |
| PLR-SRv6-aware | No | No | Yes | N/A |

| Questions | This Draft | DT2UL | Mirror-SID | VTEP-Group |
|---|---|---|---|---|
| Extra-VPN-SID | No | Yes | No | N/A |
| Special-BGP | No | Yes | No | N/A |
| Special-IGP | No | No | Yes | N/A |
| Implicit-IP-VRF | No | No | Yes | N/A |
| RT-1 Given up | No | No | No | Yes |
| NDF-Bias-Rules | Yes | No | No | No |
| VXLAN-Suitable | Yes | No | No | Yes |
| ESIL-Mirroring | Yes | No | No | No |

Table 1: Solution Comparisons

### 6.3. Detailed Comparisons with VTEP Group

According to [I-D.eastlake-bess-evpn-vxlan-bypass-vtep], the following issues will arise:

*The VTEP group address is per PE-basis, not AC-basis. So when an AC of ESI2 (All-Active mode) fails on PE1 but PE1 itself still works well, PE3 will continue to load-balance 50% flows(to ES2) to PE1. These flows have to be bypassed on PE1 before that AC comes up.

*The steady bypassing can't be solved even if the EAD/EVI route is used. Because an EAD/EVI route whose nexthop is the VTEP group address will be load-balanced by the underlay network too. Such EAD/EVI route's original fundamental mission is destroyed by the VTEP group address.

But according to this draft, when ACs of ESI2 (All-Active mode) fails on PE1 but PE1 itself still works well, no steady bypassing traffic (to ES2) will arise.

### 6.4. Detailed Comparisons with DT2UL and DX2L

According to [I-D.hu-bess-srv6-vpn-bypass-sid], the following issues will arise:

*Each EVPN Instance will be configured with two VPN SID for unicast, one End.DT2U SID and one End.DT2UL SID. The End.DT2UL SID is an extra VPN SID.

*The End.DT2UL SID need specail BGP extensions to make it to be advertised.

*End.DT2UL/DX2L is only suitable for EELP, it can't be used in EENP scenarios.

According to this draft, Only one End.DT2U/DX2 SID need to be configured per EVI. No SRv6-dedicated BGP extension needed according to current draft.

### 6.5. Detailed Comparisons with Mirror SID

According to this draft, we don't expect the PLR to support SRv6 extensions, it can be just a simple IPv6 router. The mirrored locators and mirrored EVPN SIDs will be installed in GRT, not in a "context-specific routing space" which is identified by a "mirror-SID". The dataplane will be a little more simpler according to current draft. And the solution of current draft will be a common solution for SRv6 EVPNs and VXLAN EVPNs.

According to [I-D.ietf-rtgwg-srv6-egress-protection], the following issues will arise:

* The PLR node may support LFA and Remote-LFR, but it may don't support any SR-extensions. Even if the PLR node is SRv6-aware, it may support TI-LFA only. In such use case, mirror SID will not work.

* The locators LOC1 and LOC2 won't overlap with each other (If so, the problems that [I-D.eastlake-bess-evpn-vxlan-bypass-vtep] encountered will come up, the details see Section 6.3). So the mirrored locator LOC1_P don't have to be installed into different routing spaces. It means that the mirror-SID need not to identify an implicit IP-VRF instance which is called "context-specific routing space" in that draft.

  In fact, the PLRs will obtain no awareness of that whether the mirror-SID actually identifies the GRT or not. Any node SID that won't be mirrored by other PEs can be used as a mirror-SID.

* The mirror-SID is advertised in the underlay network, but the egress link protection (EELP in this draft) is processed in the overlay network. So [I-D.ietf-rtgwg-srv6-egress-protection] can't support egress link protection.

Two approaches for egress link protection is defined in [RFC8679] section 6. The second approach don't use the context-specific forwarding method even in MPLS dataplane. We prefer the second approach, because that it is more simpler. When using the second approach, the destination IP of the bypass-tunnel takes the mirror-SID's place in egress link protection procedures. But in egress node protection, the mirror-SID typically brings out no significant results.

It is very important to be noticed that [RFC8679] is written totally in MPLS environments. So it is necessary for [RFC8679] to assume

that the VPN labels are downstream-assigned dynamically. But the SRv6 locator is typically assigned by a centerlized authority. So it is not necessary for us to use a "context-specific forwarding table" again in SRv6 scenarios.

## 7. IANA Considerations

IANA Considerations for OPE TLV following [I-D.heitz-bess-evpn-option-b].

## 8. Security Considerations

This section will be added in future versions.

## 9. Acknowledgements

The authors would like to thank the following for their comments and review of this document:

Chunning Dai, Bing Song, Zheng Zhou.

## 10. References

### 10.1. Normative References

**[I-D.heitz-bess-evpn-option-b]**
Heitz, J., Sajassi, A., Drake, J., and J. Rabadan, "Multi-homing and E-Tree in EVPN with Inter-AS Option B", Work in Progress, Internet-Draft, draft-heitz-bess-evpn-option-b-01, 13 November 2017, <https://tools.ietf.org/html/draft-heitz-bess-evpn-option-b-01>.

**[I-D.ietf-bess-evpn-prefix-advertisement]**
Rabadan, J., Henderickx, W., Drake, J., Lin, W., and A. Sajassi, "IP Prefix Advertisement in EVPN", Work in Progress, Internet-Draft, draft-ietf-bess-evpn-prefix-advertisement-11, 18 May 2018, <https://tools.ietf.org/html/draft-ietf-bess-evpn-prefix-advertisement-11>.

**[I-D.ietf-bess-evpn-inter-subnet-forwarding]**
Sajassi, A., Salam, S., Thoria, S., Drake, J., and J. Rabadan, "Integrated Routing and Bridging in EVPN", Work in Progress, Internet-Draft, draft-ietf-bess-evpn-inter-subnet-forwarding-08, 5 March 2019, <https://tools.ietf.org/html/draft-ietf-bess-evpn-inter-subnet-forwarding-08>.

**[I-D.wang-bess-evpn-context-label]**
Wang, Y. and B. Song, "Context Label for MPLS EVPN", Work in Progress, Internet-Draft, draft-wang-bess-evpn-

context-label-03, 15 August 2020, <https://tools.ietf.org/html/draft-wang-bess-evpn-context-label-03>.

**[RFC8679]** Shen, Y., Jeganathan, M., Decraene, B., Gredler, H., Michel, C., and H. Chen, "MPLS Egress Protection Framework", RFC 8679, DOI 10.17487/RFC8679, December 2019, <https://www.rfc-editor.org/info/rfc8679>.

**[RFC7432]** Sajassi, A., Ed., Aggarwal, R., Bitar, N., Isaac, A., Uttaro, J., Drake, J., and W. Henderickx, "BGP MPLS-Based Ethernet VPN", RFC 7432, DOI 10.17487/RFC7432, February 2015, <https://www.rfc-editor.org/info/rfc7432>.

**[RFC8365]** Sajassi, A., Ed., Drake, J., Ed., Bitar, N., Shekhar, R., Uttaro, J., and W. Henderickx, "A Network Virtualization Overlay Solution Using Ethernet VPN (EVPN)", RFC 8365, DOI 10.17487/RFC8365, March 2018, <https://www.rfc-editor.org/info/rfc8365>.

### 10.2. Informative References

**[I-D.eastlake-bess-evpn-vxlan-bypass-vtep]**
Eastlake, D., Li, Z., and S. Zhuang, "EVPN VXLAN Bypass VTEP", Work in Progress, Internet-Draft, draft-eastlake-bess-evpn-vxlan-bypass-vtep-05, 27 April 2020, <https://tools.ietf.org/html/draft-eastlake-bess-evpn-vxlan-bypass-vtep-05>.

**[I-D.hu-bess-srv6-vpn-bypass-sid]**
Hu, C., "Enhance IPv6-Segment-Routing-based EVPN VPWS All Active Usage", Work in Progress, Internet-Draft, draft-hu-bess-srv6-vpn-bypass-sid-00, 2 July 2018, <https://tools.ietf.org/html/draft-hu-bess-srv6-vpn-bypass-sid-00>.

**[I-D.ietf-rtgwg-srv6-egress-protection]**
Hu, Z., Chen, H., Chen, H., Wu, P., Toy, M., Cao, C., Liu, L., and X. Liu, "SRv6 Path Egress Protection", Work in Progress, Internet-Draft, draft-ietf-rtgwg-srv6-egress-protection-00, 18 March 2020, <https://tools.ietf.org/html/draft-ietf-rtgwg-srv6-egress-protection-00>.

## Authors' Addresses

Yubao Wang
ZTE Corporation
No.68 of Zijinghua Road, Yuhuatai Distinct
Nanjing

China

Email: wang.yubao2@zte.com.cn

Ran Chen
ZTE Corporation
No. 50 Software Ave, Yuhuatai Distinct
Nanjing
China

Email: chen.ran@zte.com.cn