Abstract

EVPN is designed to provide a better VPLS service than [RFC4761] and [RFC4762], and EVPN indeed introduced many new features which couldn’t be achieved in those old VPLS implementions. But EVPN didn’t inherit all features of old VPLS, and a few issues arises for EVPN only.

Some of these issues can be imputed to the MP2P nature of EVPN labels. The PW label in old VPLS is a label for P2P VC, so it contains more context than a identifier in dataplane for it’s VSI instance. But the EVPN label just identifies it’s VSI instnace and it can’t stand for the ingress PE in dataplane. So the following issues arises with MPLS EVPN service:

MPLS EVPN statistics can’t be done per ingress PE.

MPLS EVPN can’t support hub/spoke use case which the spoke PE can only connect to each other by the hub PE.

MPLS EVPN can’t support AR REPLICATOR.

MPLS EVPN can’t support anycast SR-MPLS tunnel on the SPE nodes.

This document introduces a compound label stack to take advantage of both P2P VC and MP2P evpn labels.

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 July 20, 2020.

Copyright Notice

Copyright (c) 2020 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. Problem Statement ............................................. 2
2. Context VC Infrastructure ................................. 3
   2.1. Context VC Signalling .................................. 4
      2.1.1. Kompella Signalling for context VC ................. 4
      2.1.2. SR-MPLS signalling for context VC .................. 4
3. Solutions ...................................................... 5
   3.1. Solution for spoke PE isolating on hub PE .............. 5
   3.2. Solution for per ingress statistics .................... 6
   3.3. Solution for AR REPLICATOR in MPLS EVPN ............... 6
   3.4. Solution for anycast tunnel usage on SPE .............. 6
4. Security Considerations ....................................... 7
5. IANA Considerations ......................................... 7
6. Normative References ......................................... 7
Author’s Address ................................................ 8

1. Problem Statement

EVPN is designed to provide a better VPLS service than RFC4761/ RFC4762, and EVPN indeed introduced many new features which couldn’t be achieved in those old VPLS implemention.But EVPN didn’t inherit all features of old VPLS, and a few issues arises for EVPN only.

Some of these issues can be imputed to the MP2P nature of EVPN labels.The PW label in old VPLS is a label for P2P VC, so it contains more context than a identifier in dataplane for it’s VSI instance.But the EVPN label just identifies it’s VSI instnace and it can’t stand
for the ingress PE in dataplane. So the following issues arises with MPLS EVPN service:

MPLS EVPN statistics can’t be done per ingress PE. All flows from remote PEs share the same statistics on egress PE, because they share the same EVPN label and the egress PE can’t pick them out in the dataplane.

MPLS EVPN can’t support hub/spoke usecase, where the spoke PEs can only connect to each other through the hub PE. Especially when at least two of the spoke PEs are connected to a common route reflector.

MPLS EVPN can’t work as an AR-REPLICATOR. Because the AR-REPLICATOR will apply replication for the ingress AR-LEAF too, but a packet shoud not be sent back to the AR-LEAF where it is received from.

MPLS EVPN SPE cannot make use of SR-MPLS anycast tunnel because the two SPEs of the anycast tunnel will assign different EVPN labels for the same EVPN route.

So this document introduces an compound label stack to take advantage of both P2P VC and MP2P evpn labels.

2. Context VC Infrastructure

In order to add as much context as old VPLS to EVPN data packet, We can construct a infrastructure by a full-mesh of context VCs among the EVPN PEs.

Take the context VCs between PE-i and PE-j as an example, VC-ij is the context VC from PE-i to PE-j, and VC-ji is the context VC from PE-j to PE-i. The VC-ij identifies the PE-i node on PE-j. The VC-ji identifies PE-j node on PE-i. The VC-label for VC-ij is called as L-ij, and the VC-label for VC-ji is called as L-ji.

So the PE-i can push the L-ij in the EVPN data packet for PE-j to distinguish the packet of PE-i from other data packets. Because the L-ij identifies the ingress PE of the data packet. In other words, the context VC is dedicated to identifies the context for a data packet while the EVPN label still identifies the EVPN instance.
Figure 1: Encapsulation of Context VC Label for EVPN Payload

Note that typically a context VC can be shared by all the EVPN instances between it’s ingress PE and egress PE. In other words, we don’t have to construct a dedicated mesh of context VCs for each specified EVPN service. So we called the shared context VCs as a common infrastructure for those EVPN services.

2.1. Context VC Signalling

The VCs of a context VC infrastructure are set up by a context VC container, the container implements a VC signalling to set up the VCs. There are two existing signalling protocol can be reused to set up context VCs for a context VC container.

2.1.1. Kompella Signalling for context VC

The signalling used by a Kompella VPLS instance per [RFC4761] can also be used by a context VC container.

Different from the Kompella VPLS instance, a context VC container only use the signalling to set up the context VCs. They are the same in signalling but different in dataplane. Take the PW between PE-i and PE-j as an example, it is constructed by VC-ij and VC-ji, and none of the two context VCs will identify a MAC-VRF. In other words the PW is a context PW.

Note that the context VC containers don’t have a MAC-VRF or a MAC-table, they are just containers for context VC.

2.1.2. SR-MPLS signalling for context VC

SR-MPLS signalling is very similar to the singleton pattern of Kompella VPLS, in spite of their different data plane and service procedure. The SID is similar to the VE-ID, the SRGB is similar to the label block.
So the constructed LSPs of the SR-MPLS signalling can be reinterpreted as context VCs in another label space named S. These context VCs use the same label values as those SR-LSPs but they are constructed at the same time in different label spaces. Take the VC-ij as an example, its label value L-ij is the same as the SID label for PE-i in PE-j’s SRGB. But the VC-ij are constructed in the context label space S which is identified by a static label. It is not constructed in the same label space with that SID label.

The context VC signalling may be [RFC8665], [RFC8666], [RFC8667]. The context VC may be established along with SR-LSPs.

+---------------------------------+                     +---------------------------------+
|  underlay ethernet header       |                     |  overlay ethernet or IP header  |
|                                 |                     +---------------------------------+
|  PSN tunnel label               |                     |
|                                 +---------------------------------+
|  EVPN label                     |                     |
|                                 +---------------------------------+
|  Static Label for Label Space S |                     |
|                                 +---------------------------------+
|  Context VC Label               |                     |
|                                 +---------------------------------+

Figure 2: Encapsulation of Context VC Label in Context Label Space

Note that the static label S is the context label for L-ij, while the L-ij is the context label for the data packet.

3. Solutions

3.1. Solution for spoke PE isolating on hub PE

    PEs1--------RR1--------PEh--------RR2--------PEs3
      /                        /
    PEs2--------/

Figure 3: Hub PE and Spoke PEs

Now take above use case for example, there are three spoke PEs and one hub PE. The spoke PEs are PE1, PE2 and PE3. The hub PE is PEh. Two of the spoke PEs (PEs1 and PEs2) are connected to the same RR group and the third one connects to another RR group.

Although we can advertise different EVPN labels for different RR groups, we can’t advertise different EVPN labels for PE1 and PE2.
But PEh can request PEs1 or PEs2 to push the label of the context VC from it to PEh. Benefit from the context VC label, PEh can distinguish where the packet from, in other words, PEh can decide where the packet can’t be sent to.

The signaling for the hub PE to request the spoke PE to push the context VC label will be added in future versions.

Note that although PEs1 and PEs2 can receive EVPN routes from each other they won’t import these routes because of the hub/spoke behaviors.

3.2. Solution for per ingress statistics

This section will be added in future versions.

3.3. Solution for AR REPLICATOR in MPLS EVPN

This section will be added in future versions.

3.4. Solution for anycast tunnel usage on SPE

```
/--------SPE1-------\
 TPE1            TPE2
 \--------SPE2-------/
```

Figure 4: SPE with Anycast Tunnel

Now take above use case for example, the two SPEs are the egress nodes of an anycast SR-MPLS tunnel. The anycast SR-MPLS tunnel is used to transport flows from TPE1 to either SPE1 or SPE2 according to load balancing procedures. So SPE1 and SPE2 have to advertise the same EVPN label independently for a given EVPN route.

In fact, SPE1 and SPE2 can simply inherit the EVPN label from TPE2, and they advertise it to TPE1 along with a context VC label. The context VC label is for the context VC from TPE2 to SPE1 or SPE2. We can make the VC labels from TPE2 to SPE1 and SPE2 have the same value through configuring.

And the label stack on the anycast SR-MPLS tunnel is constructed as the following:
Figure 5: Encapsulation of Context VC Label for EVPN Label

Note that the context VC is also constructed in a context label space, the label space is identified by a static label. And the context label space is identified by the same label on all PEs of the service domain. So the label stacks on the anycast tunnel are the same for SPE1 and SPE2.

SPE1/SPE2 will perform ILM lookup for the EVPN label in the label space identified by the context VC label.

4. Security Considerations

This section will be added in future versions.

5. IANA Considerations

There is no IANA consideration.

6. Normative References


Author’s Address

Yubao Wang
ZTE Corporation
No. 50 Software Ave, Yuhuatai Distinct
Nanjing
China

Email: wang.yubao2@zte.com.cn