BESS Working Group
Internet-Draft
Intended status: Standards Track
Expires: 21 September 2024

W. Wang
A. Wang
China Telecom
H. Wang
Huawei Technologies

20 March 2024

Layer-3 Accessible EVPN Services
draft-wang-bess-l3-accessible-evpn-06

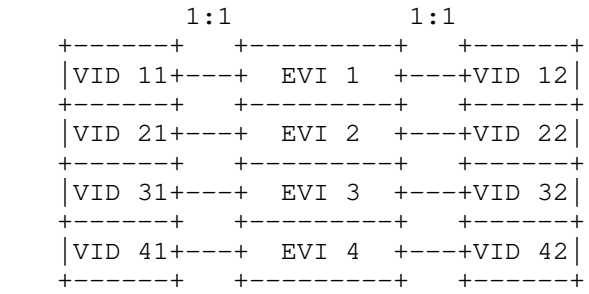
Abstract

This draft describes layer-3 accessible EVPN service interfaces, and proposes a new solution which can span layer-3 network. This solution allows each PE in EVPN network to maintain only one MAC-VRF.

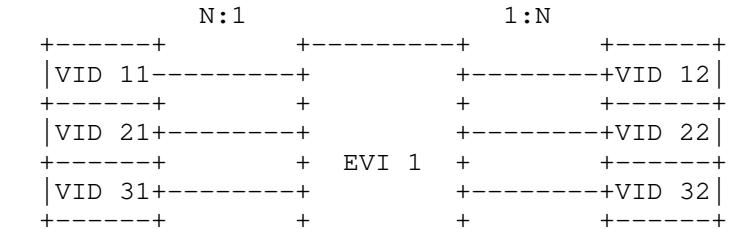
Table of Contents

1. Introduction

[RFC7432]defines three service interfaces for layer-2 accessible EVPN: VLAN-Based Service Interface, VLAN-Bundle Service Interface and VLAN-Aware Bundle Service Interface. These three types of service interfaces can realize the isolation of layer-2 traffic of customers in different ways, as shown in Figure 1.



VLAN-based Service Interface



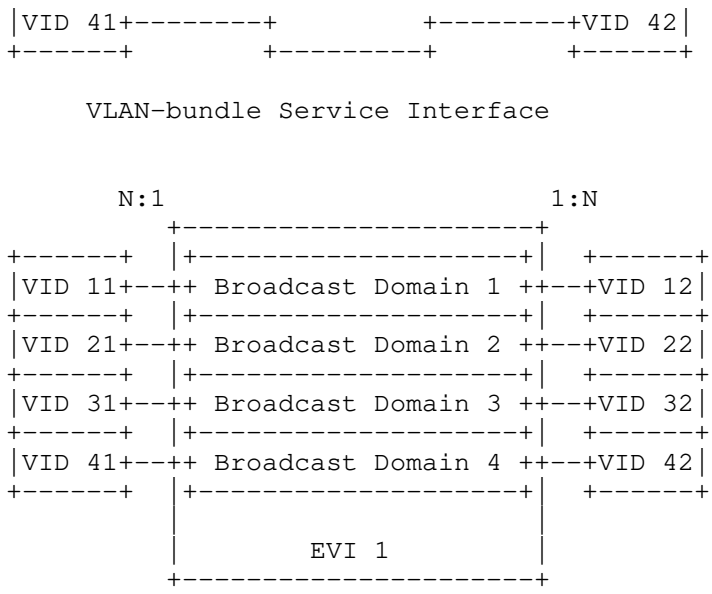


Figure 1: EVPN Service Interfaces Overview

For VLAN-based service interface, there is a one to one mapping between VID and EVI. Each EVI has a single broadcast domain so that traffic from different customers can be isolated.

For VLAN-bundle service interface, there is a N to one mapping between VID and EVI. Each EVI has a single broadcast domain, but the MAC address MUST be unique that can be used for customer traffic isolation.

For VLAN-aware bundle service interface, there is a N to one mapping between VID and EVI. Each EVI has multiple broadcast domains while the MAC address can overlap. One broadcast domain corresponds to one VID, which can be used to customer traffic isolation.

In the scenarios corresponding to these service interfaces, CE-PE should be placed in the same Layer-2 network. In most of provider network, CE-PE need to cross a Layer-3 network, then the above service interfaces should be extended to adapt to the layer-3 network.

In practical network operations, some users may require the L2--L3--L2 network connection method, as shown in Figure 2.

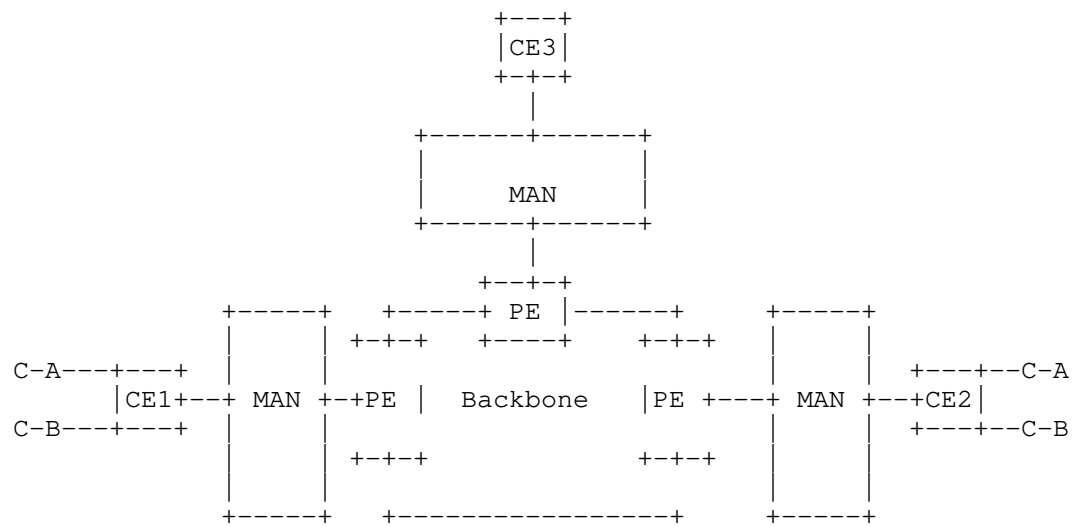


Figure 2: LSI-aware bundle service interface scenario

Assuming that the customer is a cross-regional bank, CE3 represents its headquarter site, while CE1 and CE2 are branch sites of the bank located in different cities. In this structure, the headquarters serves as the core of the network, responsible for the management and control of the overall network; while the branches serve as edge nodes of the network, responsible for collecting local business data and uploading it to the headquarters. Each site needs to connect to the backbone network through the metropolitan area network of its city for communication. To ensure data security, communication between branch sites needs to be routed through the headquarters site first, and end-to-end L2 communication is adopted between the branch sites and the headquarters site. The backbone is an EVPN domain, and MANs are L3 domains. The packets should be transmitted from CE to PE through VxLAN/IPSec tunnel. Due to the EVI cannot be transmitted in this scenario, we need an EVPN solution that can span the L3 network.

In this draft, we describe three layer-3 accessible interfaces for EVPN, the above problem can be solved by using these L3 accessible interfaces.

2. Conventions used in this document

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

The following terms are defined in this draft:

* CE: Client Edge

* EVPN: BGP/MPLS Ethernet VPN, defined in [RFC7432]

* IPSec: Internet Protocol Security, defined in [RFC4301]

* PE: Provider Edge

* SPI: Security Parameters Index, defined in [RFC4301]

* VNI: VXLAN Network Identifier (or VXLAN Segment ID), defined in [RFC7348]

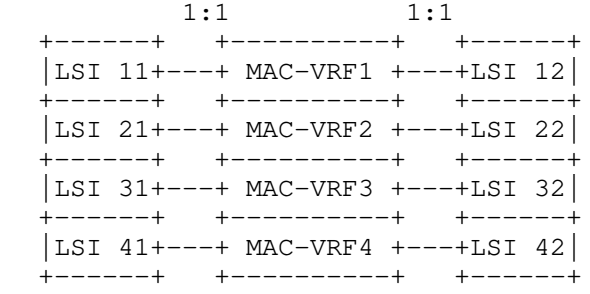
* VxLAN: Virtual eXtensible Local Area Network, defined in [RFC7348]

4. Service Interfaces in layer-3 accessible EVPN

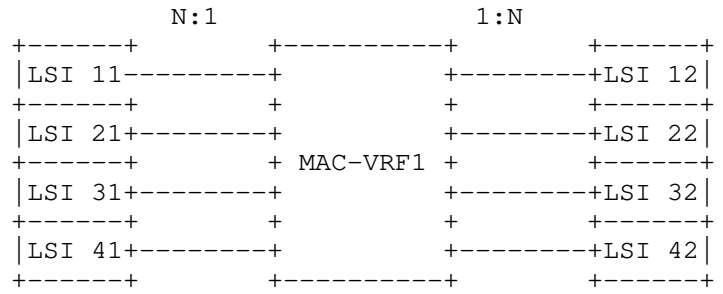
In most of provider network, CE-PE need to cross a Layer-3 network. With this scenario, service interfaces defined in [RFC7432] should be extended to adapt to the layer-3 network. To achieve the traffic isolation, tunnel encapsulation technologies can be used.

We define Logical Session Identifier(LSI) to distinguish the packets from different tunnels, which is related to VNI/SPI. The length of LSI is 16 bits.

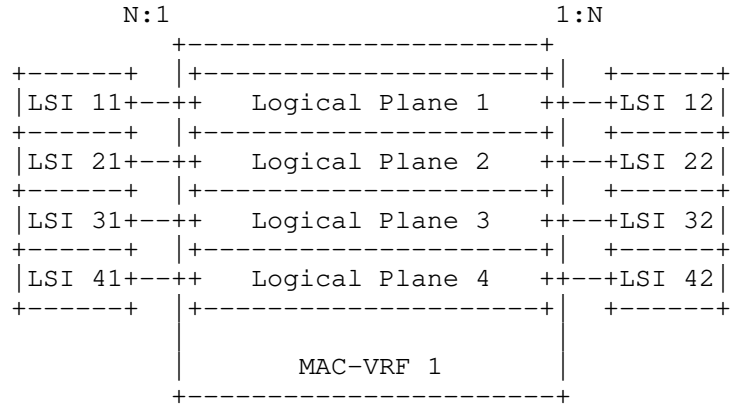
The layer-3 accessible interfaces for EVPN are shown in Figure 3, refer to [RFC7432]



LSI-based Service Interface



LSI-bundle Service Interface



LSI-Aware Bundle Service Interface

Figure 3: Layer-3 accessible EVPN Service Interfaces Overview

For LSI-based service interface, there is a one to one mapping between LSI and MAC-VRF. Each MAC-VRF has a single logical plane so that traffic from different customers can be isolated.

For LSI-bundle service interface, there is a N to one mapping between LSI and MAC-VRF. Each MAC-VRF has a single logical plane, but the MAC address MUST be unique that can be used for customer traffic isolation.

For LSI-aware bundle service interface, there is a N to one mapping between LSI and MAC-VRF. Each MAC-VRF has multiple logical planes while the MAC address can overlap. One logical plane corresponds to one LSI, which can be used to customer traffic isolation.

5. Solutions of LSI-aware bundle service interface

For the scenario shown in Figure 2, where Backbone is EVPN domain, and the MANs are Layer-3 network. There is a 1:1 mapping between LSI and VNI, LSIs are used for traffic isolation. If customers need end-to-end L2 data transmission, the use of LSI is similar to VLAN ID.

If each VNI has its own MAC-VRF, each PE and CE maintain an MAC-VRF for each deployment. The packets should be transmitted from CE to PE through VxLAN/IPSec tunnel. The EVI cannot be transmitted during this process, because it cannot be carried in VxLAN or IPSec header.

This problem can be solved by using LSI information to identify different customer routes / traffic. As described above, LSI can be generated by VNI/SPI, and there is a one to one mapping between LSI and VNI/SPI. PEs should maintain the mapping table of LSI and VNI/SPI, so that they can distinguish different customer routes / traffic. LSI information can be transmitted by using Ethernet Tag ID or a newly defined ESI type.

6. Protocol Extensions

6.1. Forwarding Plane

6.1.1. Extensions to VxLAN

When the forwarding plane uses VxLAN tunnel technologies, we should extend the VxLAN GPE header to carry the LSI information, the extentions to the VxLAN GPE header is shown in Figure 4:

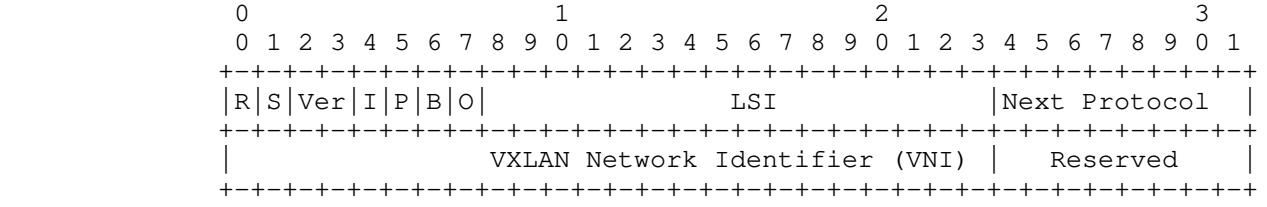


Figure 4: The extentions to VxLAN GPE header

We define a S bit. If S is set to 1, it means the field after O bit contains LSI information.

6.2. Control Plane

We proposed two methods to advertise LSI information in control plane:

* Reusing the Ethernet Tag ID field. This method requires the update of [I-D.ietf-bess-evpn-prefix-advertisement] (Etherenet Tag ID is set to 0 for route type 5), and may arise some confuse with the original definition of Ethernet Tag ID.

* Using the newly defined ESI type as shown in Figure 5. This method can preserve the original purpose of ESI definition (multi-homing).

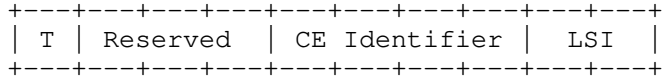


Figure 5: The format of new ESI type

Where:

* T (1 octet): specify the ESI Type. The recommended value is 0x06.

* CE Identifier (3 octets): the route ID/IPv4 address of CE.

* LSI (2 octets): the LSI information.

Since the length of LSI is 16 bits, while the length of Ethernet Tag ID and ESI are 80 bits and 32 bits, respectively. We can only use the lower 16 bits of Ethernet Tag ID / ESI field to carry LSI information, the other bits MUST set to 0.

7. Modification of MAC address storage mode on PE

LSI-aware bundle service interface also changes the storage mode of MAC address on PE, as shown in Figure 6.

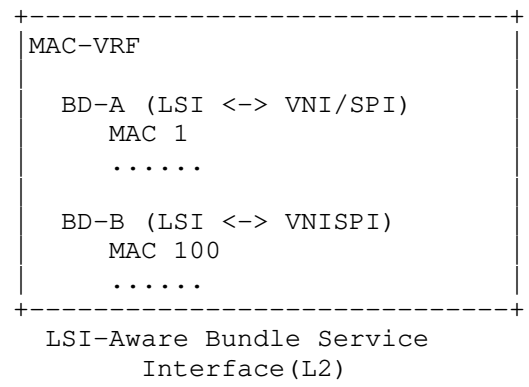


Figure 6: Modification of MAC/IP address storage mode on PE

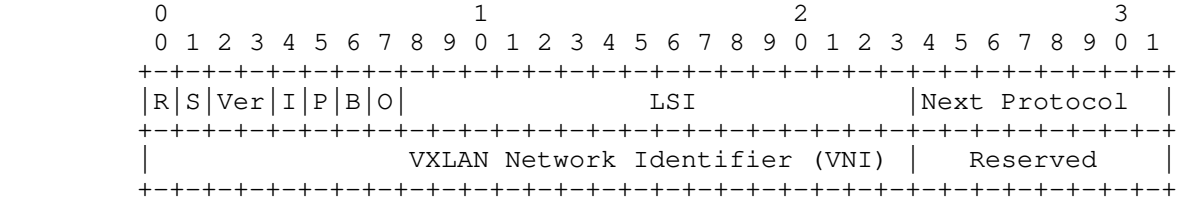
For end-to-end layer-2 data transmission, the storage mode of MAC address in MAC-VRF is similar to VLAN-aware bundle service, the only change is that different bridge domains are distinguished by LSI.

8. Security Considerations

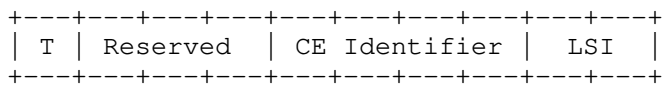
TBD

9. IANA Considerations

This draft extends the VxLAN GPE header, S bit of Flag and LSI field are added:



This draft also define a new ESI type:



10. Normative References

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

[I-D.ietf-bess-mvpn-evpn-aggregation-label]
Zhang, Z. J., Rosen, E. C., Lin, W., Li, Z., and I. Wijnands, "MVPN/EVPN Tunnel Aggregation with Common Labels", Work in Progress, Internet-Draft, draft-ietf-bess-mvpn-evpn-aggregation-label-14, 4 October 2023, <https://datatracker.ietf.org/doc/html/draft-ietf-bess-mvpn-evpn-aggregation-label-14>.

[RFC2119] Bradner, S., "Key words for use in RFCs to Indicate Requirement Levels", BCP 14, RFC 2119, DOI 10.17487/RFC2119, March 1997, <https://www.rfc-editor.org/info/rfc2119>.

[RFC2890] Dommety, G., "Key and Sequence Number Extensions to GRE", RFC 2890, DOI 10.17487/RFC2890, September 2000, <https://www.rfc-editor.org/info/rfc2890>.

[RFC4301] Kent, S. and K. Seo, "Security Architecture for the Internet Protocol", RFC 4301, DOI 10.17487/RFC4301, December 2005, <https://www.rfc-editor.org/info/rfc4301>.

[RFC7348] Mahalingam, M., Dutt, D., Duda, K., Agarwal, P., Kreeger, L., Sridhar, T., Bursell, M., and C. Wright, "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", RFC 7348, DOI 10.17487/RFC7348, August 2014, <https://www.rfc-editor.org/info/rfc7348>.

[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>.

Authors' Addresses

Wei Wang
China Telecom
Beiqijia Town, Changping District
Beijing
Beijing, 102209
China
Email: weiwang94@foxmail.com

Aijun Wang
China Telecom
Beiqijia Town, Changping District
Beijing
Beijing, 102209
China
Email: wangaj3@chinatelecom.cn

Haibo Wang
Huawei Technologies
Huawei Building, No.156 Beiqing Rd.
Beijing
Beijing, 100095
China
Email: rainsword.wang@huawei.com