BESS WG
Internet-Draft
Intended status: Standards Track
Expires: 22 May 2024

S. Dikshit
Aruba, HPE
G. Mishra
Verizon Inc.
S. Rao
S. Easale
A. Dahiya
Aruba, HPE

19 November 2023

EVPN Mpls Ping Extension
draft-saum-evpn-lsp-ping-extension-04

Abstract

In an EVPN or any other VPN deployment, there is an urgent need to tailor the reachability checks of the client nodes via off-box tools which can be triggered from a remote Overlay end-point or a centralized controller. There is also a ease of operability needed when the knowledge known is partial or incomplete. This document aims to address the limitation in current standards for doing so and provides solution which can be made standards in future. As an additional requirement, in network border routers, there are liaison/ dummy VRFs created to leak routes from one network/fabric to another. There are scenarios wherein an explicit reachability check for these type of VRFs is not possible with existing mpls-ping mechanisms. This draft intends to address this as well. Few of missing pieces are equally applicable to the native lsp ping as well.

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 22 May 2024.

Table of Contents

1. Important Terms

VTEP: Virtual Tunnel End Point or Vxlan Tunnel End Point

RD: Route Distinguisher

RT: Route Target

LSP: Label Switched Path

LER: Label Edge Router

LSR: Label Switch Router

NLRI: Network Layer Reachability Information

EVPN: Etherenet Virtual Private Network

2. Introduction

In an EVPN or any other VPN deployment, there is an urgent need to tailor the reachability checks of the client nodes via off-box tools which can be triggered from a remote Overlay end-point or a centralized controller and also customize check if the knowledge known is partial or incomplete. This document aims to address the limitation in current standards for doing so and provides solution which can be made standards in future. As an additional requirement, in network border routers, there are liaison/dummy VRFs created to leak routes from one network/fabric to another. There are scenarios wherein an explicit reachability check for these type of VRFs is not possible with existing mpls-ping mechanisms. This draft intends to address this as well.

3. Requirements Language

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].

When used in lowercase, these words convey their typical use in common language, and they are not to be interpreted as described in [RFC2119].

4. Problem Description

This document intends to solve multiple problems, all related to ease of serviceability, troubleshooting and provisioning. In a nut shell, the solution eases out the network management of overlay network with MPLS fabric for network operators and end-users. the following subsections detail out the problems at hand.

4.1. EVPN NLRI is a Complex String

For overlays like EVPN, where the NLRI key is complex to remember; the OAM ping (access) to a NLRI, may be difficult to achieve by providing the exact prefix key. For example, an EVPN NLRI index consists of list of following parameters, which typically are combined to be treated as long string index, comprising of, Route Type, RD, Ethernet Segment Index (ESI), Ethernet-Tag, IP-prefix, MAC-IP. Instead it will be easier, if the administrator remembers few or significant of the above information and remaining can be sent as wild-card or dont care values. For example, the OAM trigger for LSP-ping for a host 10.10.10.1 to a remote tunnel endpoint referred by IP address 1.1.1.1, can be initiated a combination of Route-Distinguisher, Ethernet Segment Index and Ethernet tag as wild card values, thus simplifying the OAM procedures.

The complex string problem is generic in nature and is applicable to other attributes like FEC, thus making this enhancement useful for underlay mpls ping as well.

4.1.1. FEC is a Complex String too

The complex string is similar to FEC carried in the ping packet. For example, RSVP IPv4 FEC carries attributes to complete the traffic engineering tuple index. While remembering the complete information may not be trivial for the operator. Hence partial information like Tunnel-ID and Destination IP address may be significant ones which can achieve the same check.

4.2. Partial Validation Support

The current set of OAM standards are built around validating the co-relation of control plane and dataplane information. For example, set of same-prefixes which are published by more than two external border routers, only one of them may make it to the Routing table of other routers (receiving these routes).

* The remote OAM check may want to check all the routes published into the routing table or may want to check all the routes in the protocol fib.

* This selective mechanism to fetch information is not supported for Overlays via standard OAM methods.

As mentioned above, the choice of validating control plane and dataplane for an NLRI ping is not in place in the EVPN(or any Overlay) OAM specifications [I-D.draft-ietf-bess-evpn-lsp-ping]. When the routing data is huge, and the control plane protocol are in

the middle of churn, it is difficult to ascertain if the remote network in remote site is in steady state or not. An overlay ping is should help validate only the data plane and forgo any control plane validation, so that the control plane churn is not adding to the CPU cycles for the routing or OAM entities like processes and daemons running on the remote vteps.

To extend this problem state further, when admin access to vtep (in a non-local operator domain) is not possible, control plane information can be obtained by leveraging the control plane options only. Thus providing a side-view of the protocol rib on the remote device.

This problem is also generic in nature and not restricted to EVPN or any other VPN NLRI per-se. Hence equally applicable to underlay or transport LSPs.

4.3. Reachability to Liaison VRFs

In a typical VPN deployments between branch offices, or a Datacenter deployment in an enterprise, be it MPLS or Vxlan fabric, the border routers of the fabric cater to terminating or relaying of multi-tenancy across fabric. That is, border routers are provisioned with routing and/or bridging-domains for clients while also extending it beyond the geography or site. The border routers are provisioned with stitching of inter-site tunnels/Overlays.

To simplify configuration and provisioning of overlays, a dedicated VRF is created to ensure all routes learnt from external network (from various client VRFs) over, lets say, BGP-MPLS L3VPN peering, can be de-multiplexed or leaked into a single VRF which is leveraged as a dedicated VRF for learnings from external network. This VRF is used by the intra fabric constructs as a client VRF. For example, in a Vxlan fabric, this is vrf is one of the tenant VRFs which a rightful mapping to EVPN constructs like EVI(for example VNI). This client VRF does not require any interface configuration, as the purpose of this VRF is to act as a liaison for the external routes.

Since there is no ip address(layer 3 interface) configured on this VRF, its not possible to check the state of the VRF on the border router via OAM methods. The state of VRF can be defined as following

* Working Configuration that is, VRF is operationally and administratively UP and WORKING

* Network Reachability, that is, VRF is reachable via remote fabric routers like Vteps or LSR or LER routers

* Existing OAM tools DO NOT provide enough ammunition to address this use case.

If there is no route leaked into the VRF, the BR will not form a tunnel with any other Vtep in the site. Hence an OAM check to reach out to the VRF will not work even though the VRF is up and working.

5. Solution(s)

The EVPN extension for MPLS OAM is being driven by [I-D.draft-ietf-bess-evpn-lsp-ping], and does not resolves the problem mentioned above.

This document proposes a three new TLVs which an Overlay OAM PDU like mpls ping, that can carry to fill up the gap with the rightful or optimal information to the remote tunnel end points

* dont care option

* mode of validation

* liaison vrf information.

These PDUs are described for an MPLS EVPN fabric, but can be generalized for any EVPN fabric per se

* Wild Card List TLV

* Validation TLV

* EVI Sub Tlv

5.1. Wild Card Tlv

The Wild Card Tlv addresses the problem described in section Section 4.1.

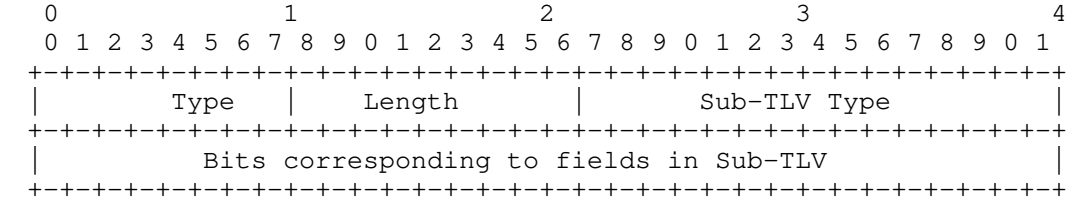
(1) It Carries the information regarding the fields (TLVs or sub TLVs), which need to be ignored on processing in mpls lsp ping PDU.

(2) For example, if an OAM ping to a prefix does not requires any RD (Route-Distinguisher) validation, then RD value, to be carried in IP prefix TLV; can be indicated as wild-card (dont care).

* The control-plane validation of the lsp-ping then should ignore the RD value in the TLV, and respond back as success even if there is atleast one NLRI which complies with other attributes (not set as wild card).

5.1.1. Description

The following diagram shows the wild-card list TLV and the following table, describe the fields, followed by the receive side processing



Field	Description
Type	Type field can be newly defined as a proprietary one.
Length	length of the TLV
Sub-TLV Type	Sub-TLV type value as defined in draft-ietf-bess-evpn-lsp-ping
Bitmap	Maps to field(s) inside Sub-TLV. The bit-map indicates which field(s) in the Sub-TLV type is carried as wild card.

Figure 1: Figure 1: WILD CARD TLV

NOTE: The bitmap for fields is very specific to the sub-tlv. The assumption is that there are no more than 32 unique fields carried in mpls ping packet across all sub tlvs. For example, in [I-D.draft-ietf-bess-evpn-lsp-ping], if for a EVPN MAC Sub-TLV, the RD is to set as wild card, then the Sub-TLV-Type carries a value 2 as defined in [RFC7432] and bitmap has 1st bit set indicating the 1st field of the TLV is RD.

5.1.2. Processing

(1) If the receiving BGP peer does not supports the wild-card list TLV,

* it ignores the TLV while processing other information carried in sub-TLVs

(2) If the receiving BGP peer support wild-card-list TLV but does not supports the wild-card ignorance of the field for validating the OAM request

(a) It responds back the error defined in [RFC4379]

(b) The error code which is to be leveraged is '2' which represent the error: 'One or more of the TLVs was not understood'.

(3) if the receiving BPG peer supports wild-card list TLV, then,

(a) it extracts the information and maps it to the corresponding fields in other sub-TLVs as carried in the OAM message (MPLS LSP ping or any other fabric OAM).

(b) It Ignores the value carried in those fields for performing Control-plane or Dataplane Validation.

(c) Then, responds back with appropriate messages with errors or otherwise as described in [I-D.draft-ietf-bess-evpn-lsp-ping].

5.2. Validation Scope Tlv

The validation Scope TLV addresses the problem mentioned in section Section 4.2.

(1) It defines the type validation to be done for the OAM mpls ping PDU at the receiving end before a response can be corroborated and sent back to the sender

(2) The validation types are defined as follows

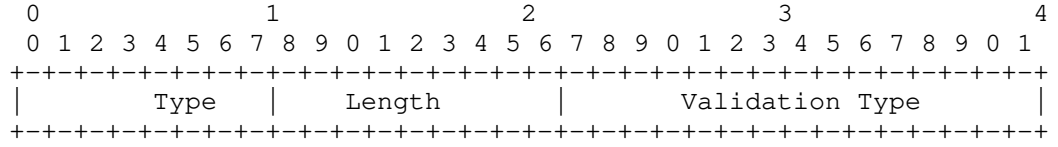
(a) Dataplane Validation: Validating the parameters which matter to the FIB (forwarding information base) or routing/switching/bridging table

(b) Control Plane Validation: Validating parameters which are matter to the protocol(s) producing those routes. For example, validating the carried parameters against the protocol(s) RIB (routing information base). This operation can be CPU intensive and can impact the control plane processing

(c) Both Control plane and Dataplane Validation: Typically performed to sanitize the network in a new-installation or post/pre upgrades when the network is in steady state and routers/switches in contention are not experiencing protocol churns.

5.2.1. Description

The following diagram shows the wild-card list TLV and the following table, describe the fields



Field	Description
Type	Type field can be newly defined as a proprietary one.
Length	Length of the TLV
Validation type	Three values for the validation as of now: 0 Both Control plane and Dataplane Validation (DEFAULT) 1 Only Control plane Validation 2 Only Data plane Validation

Figure 2: Figure 2: Validation Scope TLV

5.2.2. Processing

(1) If the receiving BGP peer does not supports the Validation TLV,

* it ignores the TLV while processing other information carried in sub-TLVs

* Alternatively, It responds back with the error defined in [RFC4379],

(2) If the receiving BGP peer does supports the Validation TLV but does not supports the non-default mode (1 and 2), it does the validation as described in the standard document, that is the default mode (both control plane and dataplane validation) in [I-D.draft-ietf-bess-evpn-lsp-ping].

(3) If receiving side supports Validation TLV and all its modes, it performs the validation only in the requested mode:

* Both Control plane and dataplane

* Only Control Plane

* Only Dataplane

5.3. EVI Sub Tlv

The EVI Sub Tlv addresses the issues mentioned in the section Section 4.3.

This solution proposes a new Object/TLV which carries the EVI (Virtual Network Identifier) information, thus ensuring that following tools and/or action-sets can be supported:

(1) Ping or path tracing to check the configuration of an EVI on a remote Vtep

(2) Ping to check VRF configuration (mapped to an EVI) on remote Vtep,

* even though no layer-3 configuration is enable against that VRF

(3) Ping to check VRF configuration (mapped to an EVI) on remote Vtep,

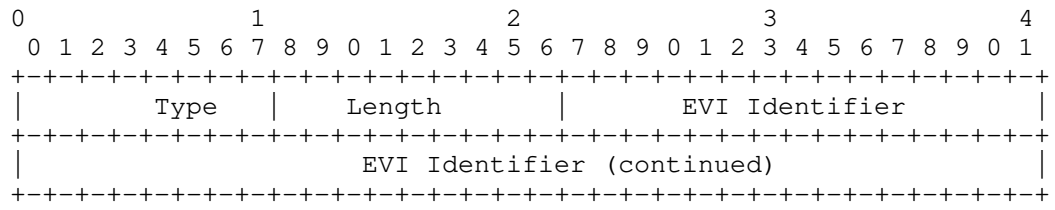
* For which EVPN tunnel not been provisioned yet

The EVI values carried in the EVI Sub TLV can be user-defined or derived from underlaying fabric idenfier for the EVI.

* For mpls fabric the EVI values can be MPLS labels (mapped to the VRFs), whereas,

* For other encapsulations like Vxlan (GUE, Geneve, GPE), the EVI value should be the VNI (mapped to the VRFs).

5.3.1. Description



Type: 1 Octet: Type field can be newly defined as a proprietary one.
Length: 1 Octet: Defines the length of the Value field.
Value: 6 Octets: EVI identifier.

Figure 3: Figure 2: Validation Scope TLV

This TLV aligns generically with any Overlay OAM-ping, agnostic to a fabric used in the deployment (Vxlan, MPLS, GUE, Geneve, GPE). This TLV can be integrated into OAM tools of any underlying fabric. For example, the EVI identifier for MPLS will be 4-octets. Hence length field will carry '4' as the length.

NOTE: Nil FEC described in [RFC8029], can also be leveraged for the ping when the underneath fabric is MPLS.

6. Backward Compatibility

Backward Compatibility for non-support nodes is as per the following standards already defined in [RFC7606], that, BGP speaker should discard the unsupported TLV types

7. Security Considerations

This document inherits all the security considerations discussed in [I-D.draft-ietf-bess-evpn-lsp-ping].

8. IANA Considerations

This document inherits all the IANA considerations discussed in [I-D.draft-ietf-bess-evpn-lsp-ping].

9. Acknowledgements

10. References

10.1. Normative References

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

10.2. Informative References

[I-D.draft-ietf-bess-evpn-lsp-ping]
Jain, P., "LSP-Ping Mechanisms for EVPN and PBB-EVPN", Work in Progress, Internet-Draft, 8029, February 2022, <https://www.ietf.org/archive/id/draft-ietf-bess-evpn-lsp-ping-07.txt>.

[I-D.draft-tissa-nvo3-oam-fm]
Senevirathne, T., "NVO3 Fault Managemen", Work in Progress, Internet-Draft, draft-tissa-nvo3-oam-fm-04, May 2017, <https://datatracker.ietf.org/doc/html/draft-tissa-nvo3-oam-fm-04>.

[RFC4379] Kompella, K., "Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures", RFC 4379, February 2006, <https://www.rfc-editor.org/rfc/rfc4379.html>.

[RFC7348] Mahalingam, M., "Virtual eXtensible Local Area Network (VXLAN): A Framework for Overlaying Virtualized Layer 2 Networks over Layer 3 Networks", RFC 7348, August 2014, <http://www.rfc-editor.org/rfc/rfc7348.txt>.

[RFC7432] Sajassi, A., "BGP MPLS-Based Ethernet VPN", RFC 7432, February 2015, <http://www.rfc-editor.org/rfc/rfc7432.txt>.

[RFC7606] Chen, E., "Revised Error Handling for BGP UPDATE Messages", RFC 7606, August 2015, <https://www.rfc-editor.org/rfc/rfc7606.html>.

[RFC8029] Kompella, K., "Detecting Multi-Protocol Label Switched (MPLS) Data Plane Failures", RFC 8029, February 2006, <https://www.rfc-editor.org/rfc/rfc8029.html>.

[RFC9014] Rabadan, J., "Interconnect Solution for Ethernet VPN (EVPN) Overlay Networks", RFC 9014, May 2021, <http://www.rfc-editor.org/rfc/rfc9014.txt>.

Authors' Addresses

Saumya Dikshit
Aruba Networks, HPE
Mahadevpura
Bangalore 560 048
Karnataka
India
Email: saumya.dikshit@hpe.com

Gyan Mishra
Verizon Inc.
Email: gyan.s.mishra@verizon.com

Srinath Rao
Aruba Networks, HPE
Email: srinath.krishnarao@hpe.com

Santosh Easale
Aruba Networks, HPE
Email: santosh.easale@hpe.com

Ashwini Dahiya
Aruba Networks, HPE
Email: ashwini.dahiya@hpe.com