Benchmarking Methodology for EVPN Multicasting

draft-vikjac-bmwg-evpnmultest-03

Abstract

This document defines methodologies for benchmarking IGMP proxy performance over EVPN-VXLAN. IGMP proxy over EVPN is defined in draft-ietf-bess-evpn-IGMP-mld-proxy-02, and is being deployed in data center networks. Specifically this document defines the methodologies for benchmarking IGMP proxy convergence, leave latency, scale, core isolation, high availability and longevity.

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 June 15, 2020.

Copyright Notice

Copyright (c) 2019 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. Introduction ............................................. 2
   1.1. Requirements Language ............................ 3
   1.2. Terminologies ..................................... 3
2. Test Topology ........................................... 4
3. Test Cases ............................................... 6
   3.1. Learning Rate ..................................... 6
   3.2. Flush Rate ......................................... 7
   3.3. Leave Latency ...................................... 7
   3.4. Join Latency ....................................... 8
   3.5. Flush Rate of N vlans in DUT ....................... 8
   3.6. Leave Latency of N Vlans in DUT .................... 9
   3.7. Join Latency of N vlans in DUT working EVPN AA mode ... 10
   3.8. Flush Rate of DUT working EVPN AA ................. 10
   3.9. Leave Latency of DUT operating in EVPN AA ........... 11
   3.10. Join Latency with reception of Type 6 route .......... 11
4. Link Flap .................................................. 12
   4.1. Packet Loss measurement in DUT due to CE link Failure .. 12
   4.2. Core Link Failure in EVPN AA ........................ 13
   4.3. Routing Failure in DUT operating in EVPN-VXLAN AA .... 13
5. Scale Convergence ....................................... 14
   5.1. Core Link Failure. .................................. 14
6. High Availability ......................................... 15
   6.1. Routing Engine Fail over. ............................ 15
7. SOAK Test ................................................. 15
   7.1. Stability of the DUT with traffic. ................... 15
8. Acknowledgments .......................................... 16
9. IANA Considerations ...................................... 16
10. Security Considerations ................................. 16
11. References .............................................. 17
   11.1. Normative References ............................... 17
   11.2. Informative References ............................. 17
Appendix A. Appendix ........................................ 17
Authors’ Addresses ......................................... 17

1. Introduction

IGMP proxy over EVPN-VXLAN is defined in draft-ietf-bess-evpn-IGMP-mld-proxy-02, and is being deployed in data center networks. Specifically this document defines the methodologies for benchmarking IGMP proxy convergence, leave latency Scale, Core isolation, high availability and longevity.
1.1. 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 RFC 2119 [RFC2119].

1.2. Terminologies

All-Active Redundancy Mode: When all PEs attached to an Ethernet segment are allowed to forward known unicast traffic to/from that Ethernet segment for a given VLAN, then the Ethernet segment is defined to be operating in All-Active redundancy mode.

AA: All Active mode

CE: Customer Router/Devices/Switch.

DF: Designated Forwarder

DUT: Device under test.


Ethernet Segment (ES): When a customer site (device or network) is connected to one or more PEs via a set of Ethernet links, then that set of links is referred to as an ‘Ethernet segment’.

EVI: An EVPN instance spanning the leaf, spine devices participating in that EVPN.

EVPN: Ethernet Virtual Private Network

Ethernet Segment Identifier (ESI): A unique non-zero identifier that identifies an Ethernet segment is called an ‘Ethernet Segment Identifier’.

Ethernet Tag: An Ethernet tag identifies a particular broadcast domain, e.g., a VLAN. An EVPN instance consists of one or more broadcast domains.

Interface: Physical interface of a router/switch.

IGMP: Internet Group Management Protocol

IBGP: Interior Border Gateway Protocol

IRB: Integrated routing and bridging interface
MAC: Media Access Control addresses on a PE.

MLD: Multicast Listener Discovery

NVO: Network Virtualization Overlay

RT Traffic Generator.

Sub Interface Each physical Interfaces is subdivided into Logical units.

SA Single Active

Single-Active Redundancy Mode: When only a single PE, among all the PEs attached to an Ethernet segment, is allowed to forward traffic to/from that Ethernet segment for a given VLAN, then the Ethernet segment is defined to be operating in Single-Active redundancy mode.

VXLAN: Virtual Extensible LAN

2. Test Topology

EVPN Overlay Network running on leaf1, leaf2 leaf3, spine1 and spine 2 :

Topology Diagram
CE connected to leaf1 and leaf2 in EVPN-VXLAN Active-Active mode.

Topology 1

Topology Diagram

Figure 1

There are six routers in the topology. Leaf1, leaf2, leaf3, spine1, spine2 emulating a data center network. CE is a customer device connected to leaf1 and leaf2, it is configured with bridge domains in different vlans. The traffic generator is connected to CE, leaf1, leaf2, leaf3, spine1 and spine 2 to emulate multicast source and host generating IGMP join/leave.
All routers except CE are configured with EBGP for the underlay

All routers are configured with EVPN-VXLAN overlay

All leaves and spine must be configured "N" EVPN-VXLAN EVI’s

Leaf1 and Leaf2 must be configured with ESI per VLAN or ESI on Interface.

Leaf1 and leaf2 are running Active Active mode of EVPN-VXLAN.

CE is acting as bridge configured with vlans

Depends up on the test multicast traffic/host will be emulated by RT

The above configuration will serve as base configuration for all the test cases.

3. Test Cases

The following tests are conducted to measure the learning rate, leave rate, leave latency of IGMP messages which propagates in leaf and spine.

3.1. Learning Rate

Objective:

To Record the time taken to learn X1...Xn IGMP join generated by host/hosts.

Topology : Topology 1

Procedure:

Configure "N" EVPN-VXLAN EVI in leaf1, leaf2, leaf3, spine1 and spine2. Leaf1 and leaf2 are connected to CE which are working in EVPN-VXLAN AA mode. Send IGMP membership report for groups X1... Xn from RT to one vlan present in leaf1 which is a part of EVPN-VLXAN EVI.

Measure the time taken to learn X1..Xn (*,G) entries in the DUT.

Measurement:

Measure the time taken to learn the X1....Xn groups creating (*,G) entries in the DUT.

The test is repeated for "N" times and the values are collected. The time is calculated by averaging the values obtained from "N" samples.
Learning Rate = (T1+T2+..Tn)/N

3.2. Flush Rate

Objective:
To Record the time taken to clear the X1... Xn (*,G) entries in DUT.

Topology : Topology 1

Procedure:

Configure "N" EVPN-VXLAN EVI in leaf1, leaf2, leaf3, spine1 and spine2. Leaf1 and leaf2 are connected to CE which are working in EVPN AA mode. Send IGMP membership report for groups ranging from X1...Xn from RT to one vlan present in leaf1 which is a part of EVPN-VXLAN EVI. Then stop these IGMP membership report from RT.

Measurement :
Measure the time taken to flush these X1...Xn (*,G) entries in DUT.

The test is repeated for "N" times and the values are collected. The time is calculated by averaging the values obtained from "N" samples.

Flush Rate = (T1+T2+..Tn)/N

3.3. Leave Latency

Objective:
To Record the time taken by the DUT to stop forwarding the multicast traffic during the receipt of IGMP leave from RT.

Topology : Topology 1

Procedure:

Configure "N" EVPN-VXLAN EVI in leaf1, leaf2, leaf3, spine1 and spine2. Leaf1 and leaf2 are connected to CE which are working in EVPN AA mode. Send IGMP membership report for groups ranging from "X1....Xn" from RT to a vlan present in leaf1 which is a part of EVPN-VXLAN EVI. Then send traffic to these groups from spine1. Traffic flows from spine1 to leaf1. Send IGMP leave messages for these groups from RT to leaf1. Measure the time taken by the DUT to stop these multicast traffic to RT. This can be measure by the time taken to clear the (*,G) entries from the DUT.
Measurement:

Measure the time taken by DUT to clear the (*)G entries and stop forwarding the traffic.

The test is repeated for N times and the values are collected. The time is calculated by averaging the values obtained from N samples.

Leave Latency = (T1+T2+..Tn)/N

3.4. Join Latency

Objective:

To Record the time taken by the DUT to create IGMP entries for N vlans.

Topology: Topology 1

Procedure:

Configure "N" EVPN-VXLAN EVI’s in leaf1, leaf2, leaf3, spine1 and spine2. Leaf1 and leaf2 are connected to CE which are working in EVPN AA mode. Send IGMP membership report for groups ranging from X1...Xn for each vlan configured in leaf1 EVPN-VXLAN EVI’s from RT. Measure the time taken to learn these X1..Xn (*)G entries in the DUT for N vlans.

Measurement:

Measure the time taken to learn the X1....Xn groups creating (*)G entries in the DUT for N vlans.

The test is repeated for N times and the values are collected. The time is calculated by averaging the values obtained from N samples.

Join Latency = (T1+T2+..Tn)/N

3.5. Flush Rate of N vlans in DUT

Objective:

To Record the time taken to clear the X1... Xn (*)G entries in DUT for N vlans.

Topology: Topology 1

Procedure:
Configure "N" EVPN-VXLAN EVI in leaf1, leaf2, leaf3, spine1 and spine2. Leaf1 and leaf2 are connected to CE which are working in EVPN-VXLAN AA mode. Send IGMP membership report for groups ranging from X1...Xn for each vlan configured in leaf1 EVPN-VXLAN EVI’s from RT. Stop the IGMP membership report. Measure the time taken to flush these X1..Xn (*,G) entries in the DUT for N vlans.

Measurement:

Measure the time taken to flush these X1...Xn (*,G) entries in DUT.

The test is repeated for "N" times and the values are collected. The time is calculated by averaging the values obtained from "N" samples.

Flush rate for N vlans = (T1+T2+..Tn)/N

3.6. Leave Latency of N Vlans in DUT

Objective:

To record the time taken by the DUT to stop forwarding the multicast traffic to N vlans during the receipt of IGMP leave messages from RT.

Topology: Topology 1

Procedure:

Configure "N" EVPN-VXLAN in leaf1, leaf2, leaf3, spine1 and spine2. Leaf1 and leaf2 are connected to CE which are working in EVPN AA mode. Send IGMP membership report for groups ranging from X1...Xn for each vlan configured in leaf1 EVPN-VXLAN EVI’s from RT. Then send traffic to these groups from spine1. Traffic flows from spine1 to leaf1. Send the IGMP leave messages for these groups in all vlans. Measure the time taken by the DUT to stop the multicast traffic.

Measurement:

Measure the time taken by DUT to stop the multicast traffic flowing towards RT.

The test is repeated for "N" times and the values are collected. The time is calculated by averaging the values obtained from "N" samples.

Leave Latency = (T1+T2+..Tn)/N
3.7. Join Latency of N vlans in DUT working EVPN AA mode

Objective:

To Record the time taken to learn X1...Xn IGMP join generated by host/hosts located in N vlans in DUT operating in EVPN AA mode.

Topology : Topology 1

Procedure:

Configure "N" EVPN-VXLAN in leaf1, leaf2, leaf3, spine1 and spine2. Leaf1 and leaf2 are connected to CE which are working in EVPN AA mode. Configure N vlans in RT, these vlans must be present in leaf1, leaf2, then send IGMP membership report for the groups ranging from X1...Xn for these N vlans from RT to CE connected to leaf1 and leaf2 working EVPN AA mode. Measure the time taken to learn these X1..Xn (*,G) entries in the DUT for N vlans.

Measurement :

Measure the time taken to learn the X1....Xn groups by creating (*,G) entries in the DUT for N vlans.

The test is repeated for "N" times and the values are collected. The time is calculated by averaging the values obtained from "N" samples.

Join Latency = (T1+T2+..Tn)/N

3.8. Flush Rate of DUT working EVPN AA

Objective:

To Record the time taken to clear the X1... Xn (*,G) entries in DUT for N vlans.

Topology : Topology 1

Procedure:

Configure "N" EVPN-VXLAN in leaf1, leaf2, leaf3, spine1 and spine2. Leaf1 and leaf2 are connected to CE which are working in AA mode. Configure N vlans in RT, these vlans must be present in leaf1, then send IGMP join messages for groups ranging from X1...Xn for these N vlans from RT to CE which is connected to leaf1 and leaf2 working in EVPN AA mode. Then stop these IGMP messages.

Measurement :
Measure the time taken to flush these X1...Xn (*,G) entries in DUT.

The test is repeated for "N" times and the values are collected. The time is calculated by averaging the values obtained from "N" samples.

Flush Rate= (T1+T2+..Tn)/N

3.9. Leave Latency of DUT operating in EVPN AA

Objective:
To Record the time taken by the DUT to stop forwarding the multicast traffic to N vlans during the receipt of IGMP leave messages from RT.

Topology : Topology 1

Procedure:
Configure "N" EVPN-VXLAN in leaf1,leaf2,leaf3,spine1 and spine2. Leaf1 and leaf2 are connected to CE which are working in EVPN AA mode. Configure N vlans in RT which are present in leaf1, then send IGMP join messages from RT connected to CE for groups ranging from X1...Xn to these vlans. The CE in turn forwards the IGMP messages to leaf1 and leaf2 operating in EVPN AA mode. Then send traffic to these groups from spine1. Traffic flows from spine1 to CE. Send the IGMP leave messages for these groups in all vlans from RT connected to CE. Measure the time taken by the DUT to stop the traffic for these group flowing towards RT.

Measurement :
Measure the time taken by DUT to stop the multicast traffic flowing towards RT.

Repeat these test and plot the data. The test is repeated for "N" times and the values are collected. The time is calculated by averaging the values obtained from "N" samples.

Leave Latency = (T1+T2+..Tn/N)

3.10. Join Latency with reception of Type 6 route

Objective:
To record the time takes for forwarding the traffic by DUT after the receipt of type 6 join from peer MHPE in same ESI.

Topology : Topology 1
Procedure:

Configure "N" EVPN-VXLAN in leaf1, leaf2, leaf3, spine1 and spine2. Leaf1 and leaf2 are connected to CE which are working in EVPN AA mode. Configure N vlans in RT which are present in leaf1, then send IGMP join messages from RT connected to CE for groups ranging from X1...Xn to these vlans. The CE in turn forwards the IGMP messages to leaf2 operating in EVPN AA mode. leaf2 and leaf1 are working EVPN AA mode. Leaf 2 will send the type 6 join to the DUT (leaf 1). Then send traffic to these groups from spine1. Traffic flows from spine1 to CE. Measure the time taken by DUT to forward the traffic after the receipt of type 6 join from leaf1.

Measurement:

Measure the time taken by DUT to forward the multicast traffic flowing towards RT.

Repeat these tests and plot the data. The test is repeated for "N" times and the values are collected. The time is calculated by averaging the values obtained from "N" samples.

Time taken by DUT to forward the traffic towards RT in sec = (T1+T2+...Tn/N)

4. Link Flap

4.1. Packet Loss measurement in DUT due to CE link Failure

Objective:

To measure the packet loss during the CE to DF link failure.

Topology: Topology 1

Procedure:

Configure "N" EVPN-VXLAN in leaf1, leaf2, leaf3, spine1 and spine2. Leaf1 and leaf2 are connected to CE which are working in EVPN AA mode. Configure N vlans in RT which are present in leaf1, then send IGMP join messages from RT connected to CE for groups ranging from X1...Xn to these vlans. The CE in turn forwards the IGMP messages to leaf1 and leaf2 operating in EVPN AA mode. Then send traffic to these groups from spine1. Traffic flows from spine1 to CE. Fail the DF-CE link. The NON DF now will act as DF and start forwarding the multicast traffic.
Measurement:

Measure the multicast packet loss during the link failure. Repeat the test "N" times and plot the data. The packet loss is calculated by averaging the values obtained from "N" samples.

Packet loss in sec = (T1+T2+..Tn)/N

4.2. Core Link Failure in EVPN AA

Objective:

To measure the packet loss during the DF core failure

Topology: Topology 1

Procedure:

Configure "N" EVPN-VXLAN in leaf1, leaf2, leaf3, spine1 and spine2. Leaf1 and leaf2 are connected to CE which are working in EVPN AA mode. Configure N vlans in RT which are present in leaf1, then send IGMP join messages from RT connected to CE for groups ranging from X1...Xn to these vlans. The CE in turn forwards the IGMP messages to leaf1 and leaf2 operating in EVPN AA mode. Then send traffic to these groups from spine1. Traffic flows from spine1 to CE. Fail the DF core link. The NON DF now will act as the DF and starts forwarding the multicast traffic.

Measurement:

Measure the multicast packet loss during the link failure. Repeat the test "N" times and plot the data. The packet loss is calculated by averaging the values obtained from "N" samples.

Packet loss in sec = (T1+T2+..Tn)/N

4.3. Routing Failure in DUT operating in EVPN-VXLAN AA

Objective:

To measure the packet loss during the DF routing failure

Topology: Topology 1

Procedure:
Configure "N" EVPN-VXLAN in leaf1, leaf2, leaf3, spine1, and spine2. Leaf1 and leaf2 are connected to CE which are working in EVPN AA mode. Configure N vlans in RT which are present in leaf1, then send IGMP join messages from RT connected to CE for groups ranging from \( X_1 \ldots X_n \) to these vlans. The CE in turn forwards the IGMP messages to leaf1 and leaf2 operating in EVPN AA mode. Then send traffic to these groups from spine1. Traffic flows from spine1 to CE. Fail the DF by restart routing. The NON DF now will act as the DF and starts forwarding the multicast traffic.

Measurement:

Measure the multicast packet loss during the link failure. Repeat the test "N" times and plot the data. The packet loss is calculated by averaging the values obtained from "N" samples.

Packet loss in sec = \( \frac{T_1+T_2+\ldots+T_n}{N} \)

5. Scale Convergence

5.1. Core Link Failure.

Objective:

To Measure the convergence at a higher number of vlans and IGMP membership reports.

Topology: Topology 1

Procedure:

Configure "N" EVPN-VXLAN in leaf1, leaf2, leaf3, spine1, and spine2. Leaf1 and leaf2 are connected to CE which are working in EVPN AA mode. Configure N vlans in RT which are present in leaf1, then send IGMP membership report from RT connected to CE for groups ranging from \( X_1 \ldots X_n \) to these vlans. The CE in turn forwards the IGMP messages to leaf1 and leaf2 operating in EVPN AA mode. Then send traffic to these groups from spine1. Traffic flows from spine1 to CE. Fail the core link of DF. The NON DF now will act as DF and start forwarding the multicast traffic. The vlans and the multicast groups must be a higher value of \( N \) taken at random.

Measurement:

Measure the packet loss in seconds once the core link is restored. Repeat the test "N" times and plot the data. The packet loss is calculated by averaging the values obtained from "N" samples.
6. High Availability

6.1. Routing Engine Fail over.

Objective:
To record traffic loss during routing engine failover.

Topology : Topology 3

Procedure:

Configure "N" EVPN-VXLAN in leaf1,leaf2,leaf3,spine1 and spine2. Leaf1 and leaf2 are connected to CE which are working in EVPN AA mode. Configure N vlans in RT which are present in leaf1, then send IGMP membership report from RT connected to CE for groups ranging from X1...Xn to these vlans. The CE in turn forwards the IGMP messages to leaf1 and leaf2 operating in EVPN AA mode. Then send traffic to these groups from spine1. Traffic flows from spine1 to CE. Then perform a routing engine failure.

Measurement :

The expectation of the test is 0 traffic loss with no change in the DF role. DUT should not withdraw any routes. But in cases where the DUT is not property synchronized between master and standby, due to that packet loss are observed. In that scenario the packet loss is measured. The test is repeated for "N" times and the values are collected. The packet loss is calculated by averaging the values obtained by "N" samples.

Packet loss in sec = (T1+T2+..Tn)/N

7. SOAK Test

This is measuring the performance of DUT running with scaled configuration with traffic over a period of time "T". In each interval "t1" the parameters measured are CPU usage, memory usage, crashes.

7.1. Stability of the DUT with traffic.

Objective:
To measure the stability of the DUT in a scaled environment with traffic.

Topology: Topology 3

Procedure:

Configure "N" EVPN-VXLAN in leaf1, leaf2, leaf3, spine1 and spine2. Leaf1 and leaf2 are connected to CE which are working in EVPN AA mode.
Configure N vlans in RT which are present in leaf1, then send IGMP membership report from RT connected to CE for groups ranging from X1...Xn to these vlans. The CE in turn forwards the IGMP messages to leaf1 and leaf2 operating in EVPN AA mode. Then send traffic to these groups from spine1. Traffic flows from spine1 to CE.

Measurement:

Take the hourly reading of CPU, process memory. There should not be any leak, crashes, CPU spikes. The CPU spike is determined as the CPU usage which shoots at 40 to 50 percent of the average usage. The average value vary from device to device. Memory leak is determined by increase usage of the memory for EVPN-VPWS process. The expectation is under steady state the memory usage for EVPN-VXLAN, IGMP processes should not increase.

8. Acknowledgments

We would like to thank Al and Sarah for the support.

9. IANA Considerations

This memo includes no request to IANA.

10. Security Considerations

The benchmarking tests described in this document are limited to the performance characterization of controllers in a lab environment with isolated networks. The benchmarking network topology will be an independent test setup and MUST NOT be connected to devices that may forward the test traffic into a production network or misroute traffic to the test management network. Further, benchmarking is performed on a "black-box" basis, relying solely on measurements observable external to the controller. Special capabilities SHOULD NOT exist in the controller specifically for benchmarking purposes. Any implications for network security arising from the controller SHOULD be identical in the lab and in production networks.
11. References

11.1. Normative References


11.2. Informative References


Appendix A. Appendix

Authors’ Addresses

Sudhin Jacob (editor)
Juniper Networks
Bangalore, Karnataka  560103
India

Phone: +91 8061212543
Email: sjacob@juniper.net

Vikram Nagarajan
Juniper Networks
Bangalore, Karnataka  560103
India

Phone: +91 8061212543
Email: vikramna@juniper.net