Benchmarking Methodology for EVPN Multicasting

draft-vikjac-bmgw-evpnmultest-02

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 December 28, 2019.

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 .................................................. 3
   1.1. Requirements Language ................................. 3
   1.2. Terminologies ....................................... 3
2. Test Topology .................................................. 4
3. Test Cases ..................................................... 6
   3.1. How long it takes to learn (X1...Xn) IGMP join messages in DUT ........................................... 6
   3.2. How long it takes to clear the (*,G) entries in the DUT. 7
   3.3. How long it takes the DUT to stop forwarding the traffic (Measuring the leave latency) ................... 7
   3.4. How long it takes to learn (X1...Xn) IGMP join messages for N vlans in DUT ....................................... 8
   3.5. How long it takes to clear the (*,G) entries in the DUT for N vlans .............................................. 9
   3.6. How long it takes the DUT to stop forwarding the traffic for N vlans (Measuring the leave latency) ........... 9
   3.7. How long it takes to learn (X1...Xn) IGMP join messages for N vlans in DUT working EVPN AA mode .............. 10
   3.8. How long it takes to clear the (*,G) entries for N vlans in DUT working EVPN AA ................................. 11
   3.9. How long it takes the DUT operating in EVPN AA to stop forwarding the traffic for N vlans (Measuring the leave latency) .................................................. 11
   3.10. How long does it take the DUT in EVPN AA to handle Join Timeout and stop forwarding ........................ 12
   3.11. How long does it take an Ingress to learn a remote Type-6 join, create state and forwarding .................. 13
4. Link Flap .......................................................... 13
   4.1. To Measure the multicast packet loss in EVPN AA scenario on a CE link failure ..................................... 14
   4.2. To Measure the multicast packet loss in EVPN AA scenario on a core failure .................................... 14
   4.3. To Measure the multicast packet loss in EVPN AA scenario on a routing failure ............................... 15
5. Scale Convergence ................................................. 15
   5.1. To measure the packet loss during the core link failure. .......................................................... 16
6. High Availability ................................................ 16
   6.1. To Record the whether there is traffic loss due to routing engine failover for redundancy test .................. 16
7. SOAK Test ....................................................... 17
   7.1. To Measure the stability of the DUT with scale and traffic ........................................................ 17
8. Acknowledgements ............................................... 18
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

Leaf A layer 2 or layer 3 capable device

Spine layer 3 capable device which is used to inter connect leaves

CE Customer Router/Devices/Switch.

AA EVPN Terminologies AA All-Active.

AC Attachment Circuit

RT Router Tester.

DUT Device under Test.

Sub Interface Each physical Interfaces is subdivided in to Logical units.

EVI EVPN Instances which will be running on sub interface or physical port of the provider Edge routers.

DF Designated Forwarder.

ESI Ethernet Segment Identifier.
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 AA 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 router tester 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 instances depending on the cases.

Leaf1 and Leaf2 must be configured with ESI per vlan or ESI on IFD.

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. How long it takes to learn (X1...Xn) IGMP join messages in DUT

Objective:
To Record the time taken to learn X1...Xn igmp join generated by host-hosts.

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 a vlan in RT which is present in leaf1 then send igmp join messages for groups X1...Xn from RT to this vlan present in leaf1.
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.
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.

Time taken by DUT to learn and create X1...Xn (*,G ) entries in DUT which is measured in sec = (T1+T2+..Tn/N)

3.2. How long it takes to clear the (*,G) entries in the DUT

Objective:

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

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 a vlan in RT which is present in leaf1, then send igmp join messages for groups ranging from X1...Xn from RT to this vlan present in leaf1 Then stop these igmp join messages from RT.

Measurement :

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

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.

Time taken by DUT to flush these X1...Xn (*,G) entries in sec = (T1+T2+..Tn/N)

3.3. How long it takes the DUT to stop forwarding the traffic(Measuring the 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 in leaf1, leaf2, leaf3, spine1 and spine2. Leaf1 and leaf2 are connected to CE which are working in EVPN AA mode.
mode. Configure a vlan in RT which is present in leaf1, then send igmp join from RT for this vlan to leaf1 for groups ranging from "X1...Xn". 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.

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.

Time taken by DUT to stop the traffic towards RT connected in leaf1 in sec = (T1+T2+..Tn/N)

3.4. How long it takes to learn (X1...Xn) IGMP join messages for N vlans in DUT

Objective:

To Record the time taken to learn X1...Xn IGMP join generated by host/hosts located in 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 EVPN AA mode. Configure N vlans in RT, these vlans must be present in leaf1, then send igmp join messages for the groups ranging from X1...Xn for these N vlans 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.

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.
3.5. How long it takes to clear the (*,G) entries in the DUT for N vlans

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. Then stop these IGMP messages.

Measurement :

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

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.

Time taken by DUT to flush these X1...Xn (*,G) entries in sec = (T1+T2+..Tn/N)

3.6. How long it takes the DUT to stop forwarding the traffic for N vlans (Measuring the leave latency)

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 for groups ranging from X1...Xn to these vlans
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.

Time taken by DUT to stop the traffic towards RT in sec = \( \frac{T_1+T_2+...T_n}{N} \)

**3.7. How long it takes to learn \((X_1...X_n)\) IGMP join messages for \(N\) vlans in DUT working EVPN AA mode**

**Objective:**

To record the time taken to learn \(X_1...X_n\) 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 join messages for the groups ranging from \(X_1...X_n\) for these \(N\) vlans from RT to CE connected to leaf1 and leaf2 working EVPN AA mode. Measure the time taken to learn these \(X_1...X_n\) \((*,G)\) entries in the DUT for \(N\) vlans.

**Measurement:**

Measure the time taken to learn the \(X_1...X_n\) groups by creating \((*,G)\) entries in the DUT for \(N\) vlans.

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.

Time taken by DUT to learn and create \(X_1...X_n\) \((*,G)\) entries for \(N\) vlans which is measured in sec = \(\frac{T_1+T_2+...T_n}{N}\)
3.8. How long it takes to clear the (*,G) entries for N vlans in 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.

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.

Time taken by DUT to flush these X1...Xn (*,G) entries in sec = \(\frac{T_1 + T_2 + ... + T_n}{N}\)

3.9. How long it takes the DUT operating in EVPN AA to stop forwarding the traffic for N vlans (Measuring the leave latency)

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.

Time taken by DUT to stop the traffic towards RT in sec = \( \frac{T1+T2+...Tn}{N} \)

3.10. How long does it take the DUT in EVPN AA to handle Join Timeout and stop forwarding

Objective:

To record the time takes for handling of Type-7 withdrawal and clearing the state and stop forwarding the traffic.

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. The igmp leave must reach the leaf1. It will send type 7 withdrawal to DUT working in EVPN AA. Measure the time taken by the DUT to stop the traffic flowing to CE. This time will give the leave latency due to type 7 withdrawal.

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.

Time taken by DUT to stop the traffic towards RT in sec = \( \frac{(T1+T2+..Tn)}{N} \)

3.11. How long does it take an Ingress to learn a remote Type-6 join, create state and forwarding

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 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.

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

4. Link Flap
4.1. To Measure the multicast packet loss in EVPN AA scenario on a 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. To Measure the multicast packet loss in EVPN AA scenario on a core failure

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 = \((T_1+T_2+\ldots+T_n)/N\)

4.3. To Measure the multicast packet loss in EVPN AA scenario on a routing failure

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 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 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 = \((T_1+T_2+\ldots+T_n)/N\)

5. Scale Convergence
5.1. To measure the packet loss during the core link failure.

Objective:

To Measure the convergence at a higher number of vlans and igmp joins.

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 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.

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

6. High Availability

6.1. To Record the whether there is traffic loss due to routing engine failover for redundancy test.

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
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. Then perform a routing engine failure.

Measurement:

There should be 0 traffic loss which is the ideal case, No change in the DF role. DUT should not withdraw any routes. 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)

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. To Measure the stability of the DUT with scale and 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 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.

Measurement:

Take the hourly reading of CPU, process memory. There should not be any leak, crashes, CPU spikes.
8. Acknowledgements

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

9. IANA Considerations

This memo includes no request to IANA.

10. Security Considerations

There is no additional consideration from RFC 6192.

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