

Internet Engineering Task Force
Internet-Draft
Intended status: Informational
Expires: November 5, 2020

S. Jacob, Ed.
V. Nagarajan
Juniper Networks
May 4, 2020

Benchmarking Methodology for EVPN Multicasting
draft-vikjac-bmwg-evpnmultest-04

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 November 5, 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. Introduction	2
1.1. Requirements Language	2
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. Leave Latency of N Vlan in DUT	9
3.6. Join Latency of N vlans in DUT working EVPN AA mode	9
3.7. Leave Latency of DUT operating in EVPN AA	10
3.8. Join Latency with reception of Type 6 route	11
4. Link Flap	11
4.1. Packet Loss measurement in DUT due to CE link Failure	12
4.2. Core Link Failure in EVPN AA	12
4.3. Routing Failure in DUT operating in EVPN-VXLAN AA	13
5. High Availability	14
5.1. Routing Engine Fail over.	14
6. SOAK Test	14
6.1. Stability of the DUT with traffic.	15
7. Acknowledgments	15
8. IANA Considerations	15
9. Security Considerations	15
10. References	16
10.1. Normative References	16
10.2. Informative References	16
Appendix A. Appendix	16
Authors' Addresses	16

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.

EBGP: Exterior Border Gateway Protocol.

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 Visualization Overlay

RT Traffic Generator.

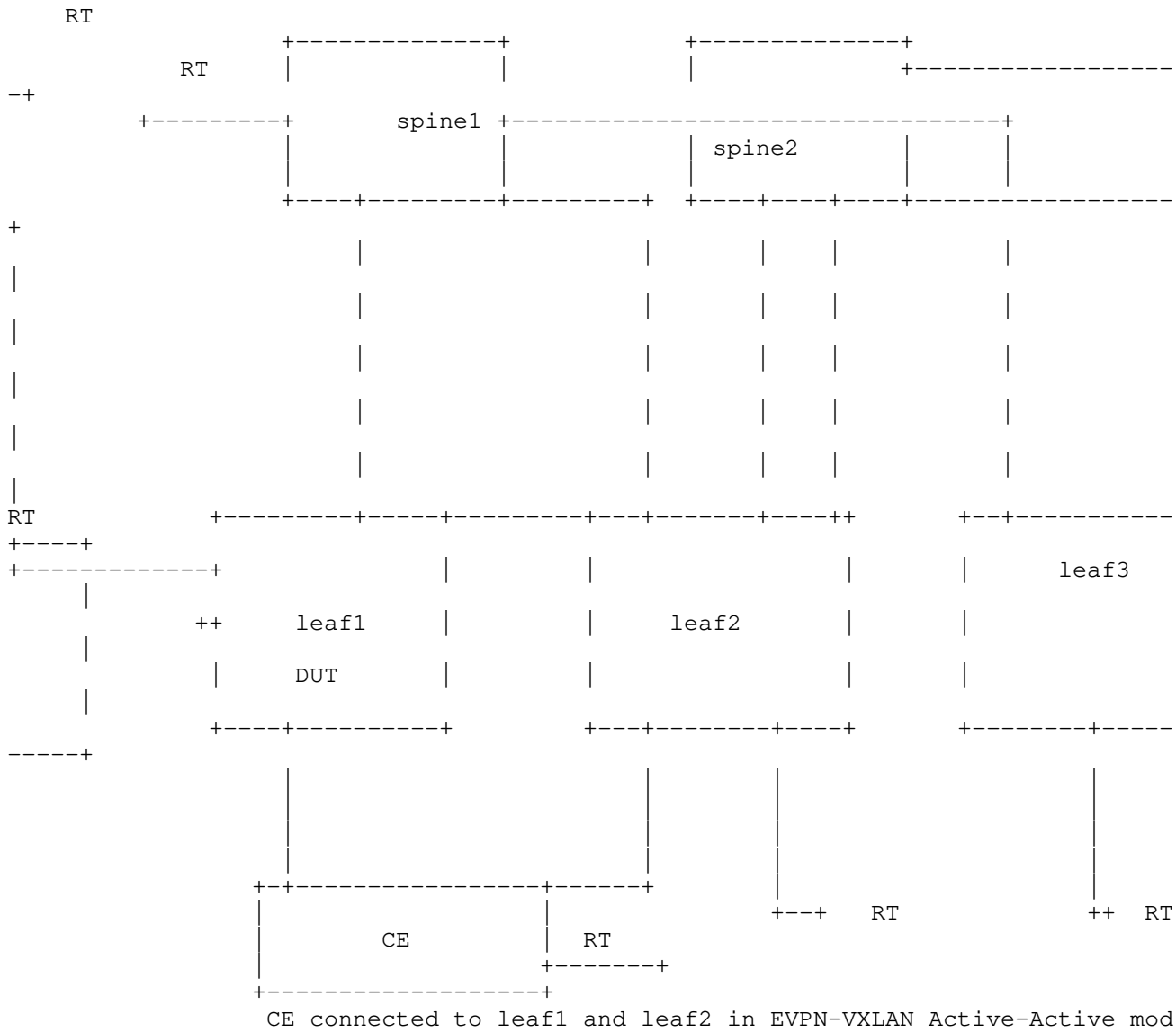
Sub Interface Each physical Interfaces is subdivided into Logical units.

VXLAN: Virtual Extensible LAN

2. Test Topology

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.

Topology Diagram



e.

Topology 1

Topology Diagram

Figure 1

Test Setup Configurations:

Leaf1, Leaf2,Leaf3 are configured with Exterior Border Gateway protocol as the underlay protocol. The routes are advertised over it. The EVPN signaling is enabled on it in order to have the overlay reachability. Leaves are configured with "N" EVPN-VXLAN EVI's. CE

is multi homed to leaf1 and leaf2. The Interface connecting to the CE is configured with ESI per interface or ESI per vlan. Leaf1 and leaf2 are running EVPN-VXLAN AA mode to CE.

Spine1,spine2 are configured with Exterior Border Gateway protocol as the underlay protocol. The routes are advertised over it. The EVPN signaling is enabled over it to have the overlay reachability. Spines are configured with "N" EVPN-VXLAN EVI's. Traffic generators are connected spine1,spine2. Spine1 and Spine2 work as single home EVPN-VXLAN EVI's.

CE is acting as bridge configured with multiple vlans,the same vlans are configured on leaf1 and leaf2. traffic generator is connected to CE. The traffic generator acts as sender or receiver of traffic.

Depending up on the test scenarios the traffic generators will be used to generate igmp membership report or multicast traffic.

The above configuration will be serving as the base configuration for all 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:

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

Topology : Topology 1

Procedure:

Confirm the DUT is up and running with EVPN-VXLAN.Traffic generator connected to leaf1 must send IGMP membership report for groups X1... Xn to a vlan present in leaf1,leaf2 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 by the DUT to learn the "X" IGMP membership report. The test is repeated for "N" times and the values are collected. The IGMP membership report learning rate is calculated by

averaging the values obtained from "N" samples. "N" is an arbitrary number to get a sufficient sample. The time measured for each sample is denoted by T1,T2...Tn.The measurement is carried out using external server which polls the DUT using automated scripts.

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

3.2. Flush Rate

Objective:

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

Topology : Topology 1

Procedure:

Confirm the DUT is up and running with EVPN-VXLAN.Traffic generator connected to the leaf1 must send IGMP membership report for groups X1... Xn to a vlan present in leaf1 which is a part of EVPN-VLXAN EVI. Stop the membership report from traffic generator. Measure the time taken to Flush X1..Xn (*,G) entries in the DUT.

Measurement :

Measure the time taken by the DUT to flush the "X" (*,G) entries The test is repeated for "N" times and the values are collected. The flush rate is calculated by averaging the values obtained from "N" samples. "N" is an arbitrary number to get a sufficient sample. The time measured for each sample is denoted by T1,T2...Tn.The measurement is carried out using external server which polls the DUT using automated scripts.

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

3.3. Leave Latency

Objective:

Measure the time taken by the DUT to stop forwarding the multicast traffic during the receipt of IGMP leave from RT.

Topology : Topology 1

Procedure:

Confirm the DUT is up and running with EVPN-VXLAN.Traffic generator connected to the leaf1 must send IGMP membership report for groups

X1... Xn to a vlan present in leaf1, leaf2 which is a part of EVPN-VLXAN EVI. Send multicast traffic from the RT port connected to spine1 to these groups requested by the leaf1. The leaf1 must receives multicast traffic. Send the IGMP leave message from the traffic generator to the leaf1. Measure the time taken by leaf1 to Flush X1..Xn (*,G) entries and stop forwarding the multicast traffic to RT.

Measurement :

Measure the time taken by the DUT to stop forwarding the multicast traffic. The test is repeated for "N" times and the values are collected. The leave latency is calculated by averaging the values obtained from "N" samples. "N" is an arbitrary number to get a sufficient sample. The time measured for each sample is denoted by T1, T2...Tn. The measurement is carried out using external server which polls the DUT using automated scripts.

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

3.4. Join Latency

Objective:

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

Topology : Topology 1

Procedure:

Confirm the DUT is up and running with EVPN-VXLAN. Ensure the route reachability. The RT port connected to spine1 acts the source of the multicast traffic. The RT port connected to leaf1 acts as receiver of multicast traffic. Send IGMP membership report for groups X1...Xn from RT port connected to leaf1. The leaf1 has N vlans subscribed to these groups. Send multicast traffic from source. Measure the time taken to forward the multicast traffic to the receiver.

Measurement :

Measure the time taken by the DUT to forward the multicast traffic to these "N" vlans. The test is repeated for "N" times and the values are collected. The join latency is calculated by averaging the values obtained from "N" samples. "N" is an arbitrary number to get a sufficient sample. The time measured for each sample is denoted by T1, T2...Tn. The measurement is carried out using external server which polls the DUT using automated scripts.

$$\text{Join Latency} = (T1+T2+..Tn)/N$$

3.5. Leave Latency of N Vlans in DUT

Objective:

Measure 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:

Confirm the DUT is up and running with EVPN-VXLAN. Ensure the route reachability. The RT port connected to spine1 acts the source of the multicast traffic. The RT port connected to leaf1 acts as receiver of multicast traffic. Send IGMP membership report for groups X1...Xn from RT port connected to leaf1. The leaf1 has N vlans subscribed to these groups. Send multicast traffic from source. Once the traffic is in steady state, send IGMP leave message to these groups. Once the leaf1 receiver the leave messages. it will flush the entries and stop forwarding the traffic to the receiver.

Measurement :

Measure the time taken by the DUT to stop forwarding the multicast traffic to these "N" vlans. The test is repeated for "N" times and the values are collected. The join latency is calculated by averaging the values obtained from "N" samples. "N" is an arbitrary number to get a sufficient sample. The time measured for each sample is denoted by T1,T2...Tn. The measurement is carried out using external server which polls the DUT using automated scripts.

$$\text{Leave Latency} = (T1+T2+..Tn)/N$$

3.6. Join Latency of N vlans in DUT working EVPN AA mode

Objective:

Measure 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:

Confirm the DUT is up and running with EVPN-VXLAN. Ensure the route reachability. The RT port connected to spine1 acts the source of the

multicast traffic. The RT port connected to CE acts as receiver of multicast traffic. leaf1 and leaf2 are multi homed EVPN-VXLAN EVI's running AA mode. The leaf1 and leaf2 have "N" vlans configured in EVPN-VXLAN EVI's, these vlans subscribe to multicast groups ranging from X1...Xn. Send IGMP membership report to these groups from RT connected to CE for these "N" vlans. Send multicast traffic from source to these groups. Measure time taken by the EVPN DF to forward the multicast traffic to the CE.

Measurement :

Measure the time taken by the EVPN DF to forward the multicast traffic for "N" vlans. The test is repeated for "N" times and the values are collected. The join latency is calculated by averaging the values obtained from "N" samples. "N" is an arbitrary number to get a sufficient sample. The time measured for each sample is denoted by T1,T2...Tn. The measurement is carried out using external server which polls the DUT using automated scripts.

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

3.7. Leave Latency of DUT operating in EVPN AA

Objective:

Measure 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:

Confirm the DUT is up and running with EVPN-VXLAN. Ensure the route reachability. The RT port connected to spine1 acts the source of the multicast traffic. The RT port connected to CE acts as receiver of multicast traffic. leaf1 and leaf2 are multi homed EVPN-VXLAN EVI's running AA mode. The leaf1 and leaf2 have "N" vlans configured in EVPN-VXLAN EVI's, these vlans subscribe to multicast groups ranging from X1...Xn. Send IGMP membership report to these groups from RT connected to CE for these "N" vlans. Send multicast traffic from source to these groups. Once traffic reaches steady state, send IGMP leave from RT connected to CE. Measure the time taken by the EVPN DF to stop forward the multicast traffic to the CE.

Measurement :

Measure the time taken by the EVPN DF to stop forward the multicast traffic for "N" vlans. The test is repeated for "N" times and the

values are collected. The leave latency is calculated by averaging the values obtained from "N" samples. "N" is an arbitrary number to get a sufficient sample. The time measured for each sample is denoted by T1, T2...Tn. The measurement is carried out using external server which polls the DUT using automated scripts.

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

3.8. Join Latency with reception of Type 6 route

Objective:

Measure the time takes to forward 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 = $(T1+T2+...Tn/N)$

4. Link Flap

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

Objective:

Measure the packet loss during the CE to DF(DUT) link failure.

Topology : Topology 1

Procedure:

Confirm the DUT is up and running with EVPN-VXLAN. Ensure the route reachability. The RT port connected to spine1 acts the source of the multicast traffic. The RT port connected to CE acts as receiver of multicast traffic. leaf1 and leaf2 are multi homed EVPN-VXLAN EVI's running AA mode. The leaf1 and leaf2 have "N" vlans configured in EVPN-VXLAN EVI's, these vlans subscribe to multicast groups ranging from X1...Xn. Send IGMP membership report to these groups from RT connected to CE for these "N" vlans. Send multicast traffic from source to these groups. The DF is the leaf1(DUT). Disable the link between DF and CE. Traffic switch to the new DF. Measure the loss of the traffic.

Measurement :

Measure the packet loss duration during the link disable. The test is repeated for "N" times and the values are collected. The packet loss duration is calculated by averaging the values obtained from "N" samples. "N" is an arbitrary number to get a sufficient sample. The time measured for each sample is denoted by T1,T2...Tn.

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

4.2. Core Link Failure in EVPN AA

Objective:

Measure the packet loss during the DF core failure

Topology : Topology 1

Procedure:

Confirm the DUT is up and running with EVPN-VXLAN. Ensure the route reachability. The RT port connected to spine1 acts the source of the multicast traffic. The RT port connected to CE acts as receiver of multicast traffic. leaf1 and leaf2 are multi homed EVPN-VXLAN EVI's

running AA mode. The leaf1 and leaf2 have "N" vlans configured in EVPN-VXLAN EVI's, these vlans subscribe to multicast groups ranging from X1...Xn. Send IGMP membership report to these groups from RT connected to CE for these "N" vlans. Send multicast traffic from source to these groups. The DF is the leaf1(DUT). Disable all the core links of DUT. Traffic switch to the new DF. Measure the loss of the traffic.

Measurement :

Measure the packet loss duration during the core link disable. The test is repeated for "N" times and the values are collected. The packet loss duration is calculated by averaging the values obtained from "N" samples. "N" is an arbitrary number to get a sufficient sample. The time measured for each sample is denoted by T1, T2...Tn.

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

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

Objective:

Measure the packet loss during the DF routing failure

Topology : Topology 1

Procedure:

Confirm the DUT is up and running with EVPN-VXLAN. Ensure the route reachability. The RT port connected to spine1 acts the source of the multicast traffic. The RT port connected to CE acts as receiver of multicast traffic. Leaf1 and leaf2 are multi homed EVPN-VXLAN EVI's running AA mode. The leaf1 and leaf2 have "N" vlans configured in EVPN-VXLAN EVI's, these vlans subscribe to multicast groups ranging from X1...Xn. Send IGMP membership report to these groups from RT connected to CE for these "N" vlans. Send multicast traffic from source to these groups. The DF is the leaf1(DUT). Perform restart routing DUT. Traffic switch to the new DF. Measure the loss of the traffic.

Measurement :

Measure the packet loss duration during the routing failure in DUT. The test is repeated for "N" times and the values are collected. The packet loss duration is calculated by averaging the values obtained from "N" samples. "N" is an arbitrary number to get a sufficient sample. The time measured for each sample is denoted by T1, T2...Tn.

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

5. High Availability

5.1. Routing Engine Fail over.

Objective:

Measure traffic loss during routing engine failover.

Topology : Topology 3

Procedure:

Confirm the DUT is up and running with EVPN-VXLAN. Ensure the route reachability. The RT port connected to spine1 acts the source of the multicast traffic. The RT port connected to CE acts as receiver of multicast traffic. leaf1 and leaf2 are multi homed EVPN-VXLAN EVI's running AA mode. The leaf1 and leaf2 have "N" vlans configured in EVPN-VXLAN EVI's, these vlans subscribe to multicast groups ranging from X1...Xn. Send IGMP membership report to these groups from RT connected to CE for these "N" vlans. Send multicast traffic from source to these groups. The DF is the leaf1(DUT). Perform routing engine failover in DUT. Traffic switch to the new DF. Measure the loss of the traffic.

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 properly 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 = $(T1+T2+...Tn)/N$

6. 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.

6.1. Stability of the DUT with traffic.

Objective:

Measure the stability of the DUT in a scaled environment with traffic.

Topology : Topology 3

Procedure:

Confirm the DUT is up and running with EVPN-VXLAN. Ensure the route reachability. The RT port connected to spine1 acts the source of the multicast traffic. The RT port connected to CE acts as receiver of multicast traffic. leaf1 and leaf2 are multi homed EVPN-VXLAN EVI's running AA mode. The leaf1 and leaf2 have "N" vlans configured in EVPN-VXLAN EVI's, these vlans subscribe to multicast groups ranging from X1...Xn. Send IGMP membership report to these groups from RT connected to CE for these "N" vlans. Send multicast traffic from source to these groups. The DF is the leaf1(DUT). Traffic will be forwarded to the CE by the DF. Run the traffic for "T" time interval.

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.

7. Acknowledgments

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

8. IANA Considerations

This memo includes no request to IANA.

9. 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.

10. References

10.1. Normative References

- [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>>.
- [RFC2544] Bradner, S. and J. McQuaid, "Benchmarking Methodology for Network Interconnect Devices", RFC 2544, DOI 10.17487/RFC2544, March 1999, <<https://www.rfc-editor.org/info/rfc2544>>.
- [RFC2899] Ginoza, S., "Request for Comments Summary RFC Numbers 2800-2899", RFC 2899, DOI 10.17487/RFC2899, May 2001, <<https://www.rfc-editor.org/info/rfc2899>>.

10.2. Informative References

- [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>>.

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