

Network Working Group  
Internet-Draft  
Updates: ???? (if approved)  
Intended status: Informational  
Expires: May 6, 2021

A. Morton  
J. Uttaro  
AT&T Labs  
November 2, 2020

Benchmarks and Methods for Multihomed EVPN  
draft-morton-bmwg-multihome-evpn-04

## Abstract

Fundamental Benchmarking Methodologies for Network Interconnect Devices of interest to the IETF are defined in [RFC 2544](#). Key benchmarks applicable to restoration and multi-homed sites are in [RFC 6894](#). This memo applies these methods to Multihomed nodes implemented on Ethernet Virtual Private Networks (EVPN).

## Requirements Language

The key words "MUST", "MUST NOT", "REQUIRED", "SHALL", "SHALL NOT", "SHOULD", "SHOULD NOT", "RECOMMENDED", "NOT RECOMMENDED", "MAY", and "OPTIONAL" in this document are to be interpreted as described in [BCP 14](#) [[RFC2119](#)] [[RFC8174](#)] when, and only when, they appear in all capitals, as shown here.

## 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 May 6, 2021.

## Copyright Notice

Copyright (c) 2020 IETF Trust and the persons identified as the document authors. All rights reserved.

Internet-Draft

Multihomed EVPN

November 2020

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

<a href="#">1.</a>	Introduction . . . . .	<a href="#">2</a>
<a href="#">2.</a>	Scope and Goals . . . . .	<a href="#">3</a>
<a href="#">3.</a>	Motivation . . . . .	<a href="#">3</a>
<a href="#">4.</a>	Test Setups . . . . .	<a href="#">3</a>
<a href="#">4.1.</a>	Basic Configuration . . . . .	<a href="#">5</a>
<a href="#">5.</a>	Procedure for Full Mesh Throughput Characterization . . . . .	<a href="#">6</a>
<a href="#">5.1.</a>	Address Learning Phase . . . . .	<a href="#">6</a>
5.2.	Test for a Single Frame Size and Number of Unicast Flows	6
<a href="#">5.3.</a>	Detailed Procedure . . . . .	<a href="#">6</a>
<a href="#">5.4.</a>	Test Repetition . . . . .	<a href="#">7</a>
<a href="#">5.5.</a>	Benchmark Calculations . . . . .	<a href="#">7</a>
<a href="#">5.6.</a>	Reporting . . . . .	<a href="#">7</a>
<a href="#">6.</a>	Procedure for Mass Withdrawal Characterization . . . . .	<a href="#">7</a>
<a href="#">6.1.</a>	Address Learning Phase . . . . .	<a href="#">8</a>
<a href="#">6.2.</a>	Test for a Single Frame Size and Number of Flows . . . . .	<a href="#">8</a>
<a href="#">6.3.</a>	Test Repetition . . . . .	<a href="#">8</a>
<a href="#">6.4.</a>	Benchmark Calculations . . . . .	<a href="#">8</a>
<a href="#">7.</a>	Reporting . . . . .	<a href="#">9</a>
<a href="#">8.</a>	Security Considerations . . . . .	<a href="#">9</a>
<a href="#">9.</a>	IANA Considerations . . . . .	<a href="#">10</a>
<a href="#">10.</a>	Acknowledgements . . . . .	<a href="#">10</a>
<a href="#">11.</a>	References . . . . .	<a href="#">10</a>
<a href="#">11.1.</a>	Normative References . . . . .	<a href="#">10</a>
<a href="#">11.2.</a>	Informative References . . . . .	<a href="#">11</a>
	Authors' Addresses . . . . .	<a href="#">12</a>

## [1.](#) Introduction

The IETF's fundamental Benchmarking Methodologies are defined in[RFC2544], supported by the terms and definitions in [[RFC1242](#)], and [[RFC2544](#)] actually obsoletes an earlier specification, [[RFC1944](#)].

This memo recognizes the importance of Ethernet Virtual Private Network (EVPN) Multihoming connectivity scenarios, where a CE device is connected to 2 or more PEs using an instance of an Ethernet Segment.

In an all-active or Active-Active scenario, CE-PE traffic is load-balanced across two or more PEs.

Mass-withdrawal of routes may take place when an autodiscovery route is used on a per Ethernet Segment basis, and there is a link failure on one of the Ethernet Segment links (or when configuration changes take place).

Although EVPN depends on address-learning in the control-plane, the Ethernet Segment Instance is permitted to use "the method best suited to the CE: data-plane learning, IEEE 802.1x, the Link Layer Discovery Protocol (LLDP), IEEE 802.1aq, Address Resolution Protocol (ARP), management plane, or other protocols" [[RFC7432](#)].

This memo seeks to benchmark these important cases (and others).

## [2.](#) Scope and Goals

The scope of this memo is to define a method to unambiguously perform tests, measure the benchmark(s), and report the results for Capacity of EVPN Multihoming connectivity scenarios, and other key restoration activities (such as address withdrawal) covering link failure in the Active-Active scenario.

The goal is to provide more efficient test procedures where possible, and to expand reporting with additional interpretation of the results. The tests described in this memo address some key multihoming scenarios implemented on a Device Under Test (DUT) or System Under Test (SUT).

## [3.](#) Motivation

The Multihoming scenarios described in this memo emphasize features with practical value to the industry that have seen deployment. Therefore, these scenarios deserve further attention that follows from benchmarking activities and further study.

#### 4. Test Setups

For simple Capacity/Throughput Benchmarks, the Test Setup MUST be consistent with Figure 1 of [\[RFC2544\]](#), or Figure 2 when the tester's sender and receiver are different devices.

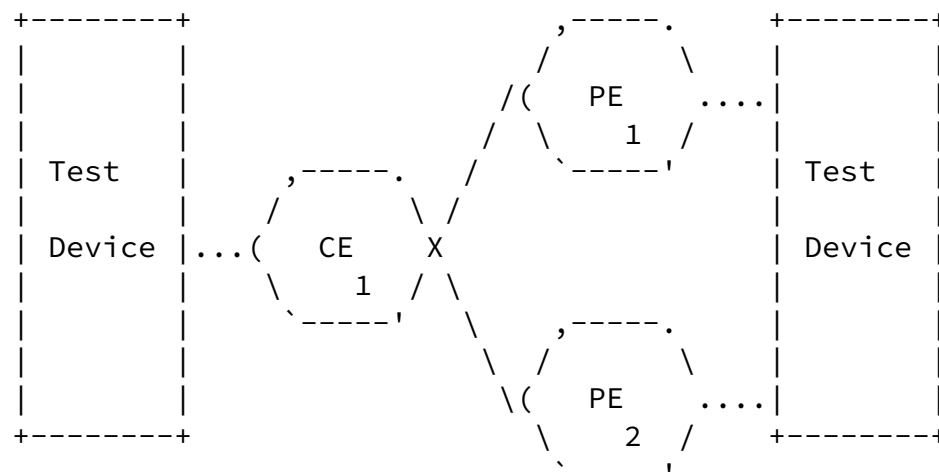


Figure 1 SUT for Throughput and other Ethernet Segment Tests

In Figure 1, the System Under Test (SUT) is comprised of a single CE device and two or more PE devices.

The tester SHALL be connected to all CE and every PE, and be capable of simultaneously sending and receiving frames on all ports with connectivity. The tester SHALL be capable of generating multiple flows (according to a 5-tuple definition, or any sub-set of the 5-tuple). The tester SHALL be able to control the IP capacity of sets of individual flows, and the presence of sets of flows on specific interface ports.

The tester SHALL be capable of generating and receiving a full mesh of Unicast flows, as described in [section 3.0 of \[RFC2889\]](#):

"In fully meshed traffic, each interface of a DUT/SUT is set up to both receive and transmit frames to all the other interfaces under test."

Other mandatory testing aspects described in [[RFC2544](#)] and [[RFC2889](#)] MUST be included, unless explicitly modified in the next section.

The ingress and egress link speeds and link layer protocols MUST be specified and used to compute the maximum theoretical frame rate when respecting the minimum inter-frame gap.

A second test case is where a BGP backbone implements MPLS-LDP to provide connectivity between multiple PE - ESI - CE locations.

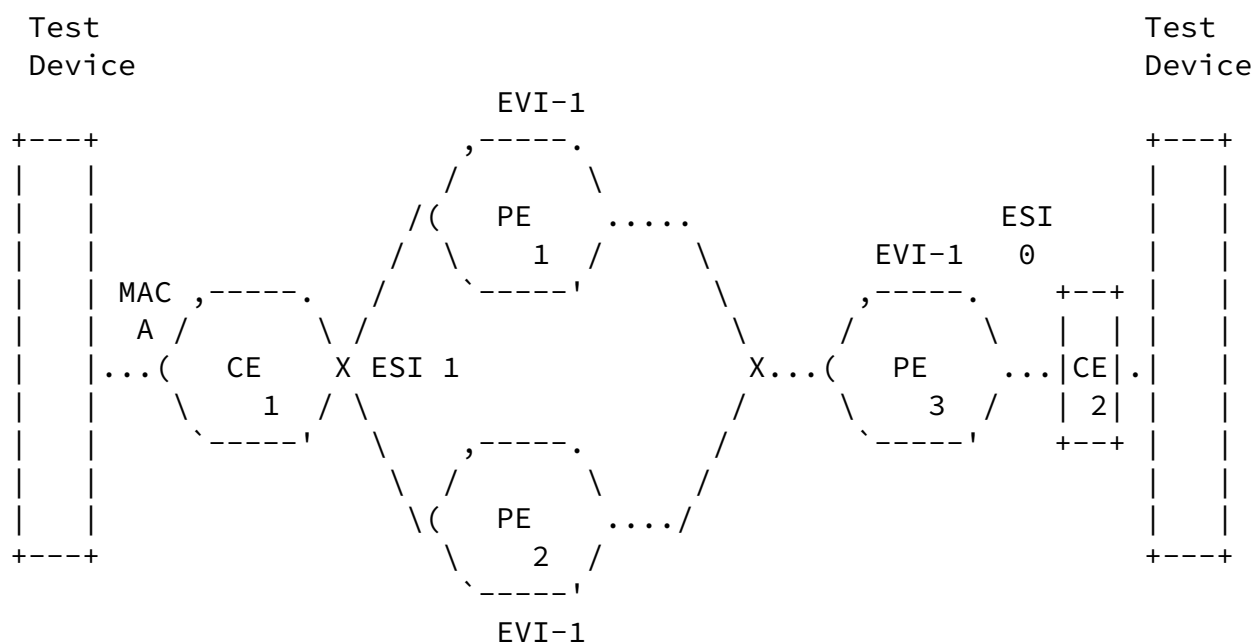


Figure 2 SUT with BGP & MPLS interconnecting multiple PE-ESI-CE locations

PE1 learns MAC A via data plane learning, PE1 and PE2 share ESI 1 (

Ethernet Segment Identifier ) and advertise an Ether A-D route with ESI 1 to PE3, PE1 also advertises MAC A to PE3. PE3 instantiates either Active/Backup or Active/Active towards PE1 and PE2 ( Assume PE1 is Active in Active/Backup scenario ) for MAC A.

All Link speeds MUST be reported, along with complete device configurations in the SUT and Test Device(s).

Additional Test Setups and configurations will be provided in this section, after review.

One capacity benchmark pertains to the number of ESIs that a network with multiple PE - ESI - CE locations can support.

#### [4.1.](#) Basic Configuration

This configuration serves as the base configuration for all test cases.

All routers except CE are configured with OSPF/IS-IS,LDP,MPLS,BGP with EVPN address family.

All routers except CE must have IBGP configured.

PE1,PE2,PE3 must be configured with an EVI context ( EVI 1 ).

PE1 and PE2 must be configured with a non-zero ESI indicating that the two VLANS coming from CE1 belong to the same ethernet segment ( ESI 1 ).

PE1 and PE2 are running Single Active mode of EVPN.

CE1 and CE2 are acting as bridges configured with VLANs that are configured on PE1, PE2, PE3.

In [[RFC2889](#)] procedures that follow, the test traffic will be bidirectional.

#### [5.](#) Procedure for Full Mesh Throughput Characterization

Objective: To characterize the ability of a DUT/SUT to process frames

between CE and one or more PEs in a multihomed connectivity scenario. Figure 1 gives the least-complex test setup. Figure 2 gives a possible alternative with full BGP and MPLS interconnection.

The Procedure follows.

#### [5.1.](#) Address Learning Phase

"For every address, learning frames MUST be sent to the DUT/SUT to allow the DUT/SUT to update its address tables properly." [[RFC2889](#)]

#### [5.2.](#) Test for a Single Frame Size and Number of Unicast Flows

Each trial in the test requires configuring a number of flows (from 100 to 100k) and a fixed frame size (64 octets to 128, 256, 512, 1024, 1280 and 1518 bytes, as per [[RFC2544](#)]). Frame formats MUST be specified, they are as described in [section 4 of \[RFC2889\]](#).

Only one of frame size and number of flows SHALL change for each test.

#### [5.3.](#) Detailed Procedure

The Procedure SHALL follow [section 5.1 of \[RFC2889\]](#).

Specifically, the Throughput measurement parameters found in [section 5.1.2 of \[RFC2889\]](#) SHALL be configured and reported with the results.

The procedure for transmitting Frames on each port is described in [section 5.1.3 of \[RFC2889\]](#) and SHALL be followed (adapting to the number of ports in the test setup).

Once the traffic is started, the procedure for Measurements described in [section 5.1.4 of \[RFC2889\]](#) SHALL be followed (adapting to the number of ports in the test setup). The section on Throughput measurement (5.1.4 of [[RFC2889](#)]) SHALL be followed.

In the case that one or more of the CE and PE are virtual implementations, then the search algorithm of [[TST009](#)] that provides consistent results when faced with host transient activity SHOULD be

used (Binary Search with Loss Verification).

#### [5.4.](#) Test Repetition

The test MUST be repeated N times for each frame size in the subset list, and each Throughput value made available for further processing (below).

#### [5.5.](#) Benchmark Calculations

For each Frame size and number of flows, calculate the following summary statistics for Throughput values over the N tests:

- o Average (Benchmark)
- o Minimum
- o Maximum
- o Standard Deviation

Comparison will determine how the load was balanced among PEs.

#### [5.6.](#) Reporting

The recommendation for graphical reporting provided in [Section 5.1.4 of \[RFC2889\]](#) SHOULD be followed, along with the specifications in [Section 7](#) below.

### [6.](#) Procedure for Mass Withdrawal Characterization

Objective: To characterize the ability of a DUT/SUT to process frames between CE and one or more PE in a multihomed connectivity scenario when a mass withdrawal takes place. Figure 2 gives the test setup.

The Procedure follows.

#### [6.1.](#) Address Learning Phase



"For every address, learning frames MUST be sent to the DUT/SUT to allow the DUT/SUT update its address tables properly." [[RFC2889](#)]

## [6.2.](#) Test for a Single Frame Size and Number of Flows

Each trial in the test requires configuring a number of flows (from 100 to 100k) and a fixed frame size (64 octets to 128, 256, 512, 1024, 1280 and 1518 bytes, as per [[RFC2544](#)]).

Only one of frame size and number of flows SHALL change for each test.

The Offered Load SHALL be transmitted at the Throughput level corresponding to the level previously determined for the selected Frame size and number of Flows in use (see [section 5](#)).

The Procedure SHALL follow [section 5.1 of \[RFC2889\]](#) (except there is no need to search for the Throughput level). See [section 5](#) above for additional requirements, especially [section 5.3](#).

When traffic has been sent for 5 seconds one of the CE-PE links on the ESI SHALL be disabled, and the time of this action SHALL be recorded for further calculations. For example, if the CE1 link to PE1 is disabled, this should trigger a Mass withdrawal of EVI-1 addresses, and the subsequent re-routing of traffic to PE2.

Frame losses are expected to be recorded during the restoration time. Time for restoration may be estimated as described in [section 3.5](#) of [[RFC6412](#)].

## [6.3.](#) Test Repetition

The test MUST be repeated N times for each frame size in the subset list, and each restoration time value made available for further processing (below).

## [6.4.](#) Benchmark Calculations

For each Frame size and number of flows, calculate the following summary statistics for Loss (or Time to return to Throughput level after restoration) values over the N tests:

- o Average (Benchmark)
- o Minimum

- o Maximum
- o Standard Deviation

## 7. Reporting

The results SHOULD be reported in the format of a table with a row for each of the tested frame sizes and Number of Flows. There SHOULD be columns for the frame size with number of flows, and for the resultant average frame count (or time) for each type of data stream tested.

The number of tests Averaged for the Benchmark, N, MUST be reported.

The Minimum, Maximum, and Standard Deviation across all complete tests SHOULD also be reported.

The Corrected DUT Restoration Time SHOULD also be reported, as applicable.

Frame Size, octets + # Flows	Ave Benchmark, fps, frames or time	Min,Max,StdDev	Calculated Time, Sec
64,100	26000	25500,27000,20	0.00004

### Throughput or Loss/Restoration Time Results

Static and configuration parameters:

Number of test repetitions, N

Minimum Step Size (during searches), in frames.

## 8. Security Considerations

Benchmarking activities as described in this memo are limited to technology characterization using controlled stimuli in a laboratory environment, with dedicated address space and the other constraints [RFC2544].

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. See [[RFC6815](#)].

Further, benchmarking is performed on a "black-box" basis, relying solely on measurements observable external to the DUT/SUT.

Special capabilities SHOULD NOT exist in the DUT/SUT specifically for benchmarking purposes. Any implications for network security arising from the DUT/SUT SHOULD be identical in the lab and in production networks.

## [9.](#) IANA Considerations

This memo makes no requests of IANA.

## [10.](#) Acknowledgements

Thanks to Sudhin Jacob for his review and comments on the bmwg-list.

Thanks to Aman Shaikh for sharing his comments on the draft directly with the authors.

## [11.](#) References

### [11.1.](#) Normative References

- [RFC1242] Bradner, S., "Benchmarking Terminology for Network Interconnection Devices", [RFC 1242](#), DOI 10.17487/RFC1242, July 1991, <<https://www.rfc-editor.org/info/rfc1242>>.
- [RFC1944] Bradner, S. and J. McQuaid, "Benchmarking Methodology for Network Interconnect Devices", [RFC 1944](#), DOI 10.17487/RFC1944, May 1996, <<https://www.rfc-editor.org/info/rfc1944>>.
- [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>>.

- [RFC2889] Mandeville, R. and J. Perser, "Benchmarking Methodology for LAN Switching Devices", [RFC 2889](#), DOI 10.17487/RFC2889, August 2000, <<https://www.rfc-editor.org/info/rfc2889>>.

- [RFC5180] Popoviciu, C., Hamza, A., Van de Velde, G., and D. Dugatkin, "IPv6 Benchmarking Methodology for Network Interconnect Devices", [RFC 5180](#), DOI 10.17487/RFC5180, May 2008, <<https://www.rfc-editor.org/info/rfc5180>>.
- [RFC6201] Asati, R., Pignataro, C., Calabria, F., and C. Olvera, "Device Reset Characterization", [RFC 6201](#), DOI 10.17487/RFC6201, March 2011, <<https://www.rfc-editor.org/info/rfc6201>>.
- [RFC6412] Poretsky, S., Imhoff, B., and K. Michielsen, "Terminology for Benchmarking Link-State IGP Data-Plane Route Convergence", [RFC 6412](#), DOI 10.17487/RFC6412, November 2011, <<https://www.rfc-editor.org/info/rfc6412>>.
- [RFC6815] Bradner, S., Dubray, K., McQuaid, J., and A. Morton, "Applicability Statement for [RFC 2544](#): Use on Production Networks Considered Harmful", [RFC 6815](#), DOI 10.17487/RFC6815, November 2012, <<https://www.rfc-editor.org/info/rfc6815>>.
- [RFC6985] Morton, A., "IMIX Genome: Specification of Variable Packet Sizes for Additional Testing", [RFC 6985](#), DOI 10.17487/RFC6985, July 2013, <<https://www.rfc-editor.org/info/rfc6985>>.
- [RFC8174] Leiba, B., "Ambiguity of Uppercase vs Lowercase in [RFC 2119](#) Key Words", [BCP 14](#), [RFC 8174](#), DOI 10.17487/RFC8174, May 2017, <<https://www.rfc-editor.org/info/rfc8174>>.

## [11.2](#). Informative References

[OPNFV-2017]

Cooper, T., Morton, A., and S. Rao, "Dataplane Performance, Capacity, and Benchmarking in OPNFV", June 2017,  
<<https://wiki.opnfv.org/download/attachments/10293193/VSPERF-Dataplane-Perf-Cap-Bench.pptx?api=v2>>.

[RFC8239] Avramov, L. and J. Rapp, "Data Center Benchmarking Methodology", [RFC 8239](#), DOI 10.17487/RFC8239, August 2017,  
<<https://www.rfc-editor.org/info/rfc8239>>.

Morton & Uttaro

Expires May 6, 2021

[Page 11]

---

Internet-Draft

Multihomed EVPN

November 2020

[TST009] Morton, R. A., "ETSI GS NFV-TST 009 V3.2.1 (2019-06), "Network Functions Virtualisation (NFV) Release 3; Testing; Specification of Networking Benchmarks and Measurement Methods for NFVI"", June 2019,  
<[https://www.etsi.org/deliver/etsi\\_gs/NFV-TST/001\\_099/009/03.01.01\\_60/gs\\_NFV-TST009v030101p.pdf](https://www.etsi.org/deliver/etsi_gs/NFV-TST/001_099/009/03.01.01_60/gs_NFV-TST009v030101p.pdf)>.

[VSPERF-b2b]

Morton, A., "Back2Back Testing Time Series (from CI)", June 2017, <[https://wiki.opnfv.org/display/vsperf/Traffic+Generator+Testing#TrafficGeneratorTesting-AppendixB:Back2BackTestingTimeSeries\(fromCI\)](https://wiki.opnfv.org/display/vsperf/Traffic+Generator+Testing#TrafficGeneratorTesting-AppendixB:Back2BackTestingTimeSeries(fromCI))>.

[VSPERF-BSLV]

Morton, A. and S. Rao, "Evolution of Repeatability in Benchmarking: Fraser Plugfest (Summary for IETF BMWG)", July 2018,  
<<https://datatracker.ietf.org/meeting/102/materials/slides-102-bmwg-evolution-of-repeatability-in-benchmarking-fraser-plugfest-summary-for-ietf-bmwg-00>>.

Authors' Addresses

Al Morton  
AT&T Labs

200 Laurel Avenue South  
Middletown,, NJ 07748  
USA

Phone: +1 732 420 1571  
Fax: +1 732 368 1192  
Email: [acm@research.att.com](mailto:acm@research.att.com)

Jim Uttaro  
AT&T Labs  
200 Laurel Avenue South  
Middletown,, NJ 07748  
USA

Email: [uttaro@att.com](mailto:uttaro@att.com)