Updates for the Back-to-back Frame Benchmark in RFC 2544

draft-ietf-bmwg-b2b-frame-00

Abstract

Fundamental Benchmarking Methodologies for Network Interconnect Devices of interest to the IETF are defined in RFC 2544. This memo updates the procedures of the test to measure the Back-to-back frames Benchmark of RFC 2544, based on further experience.

This memo updates Section 26.4 of RFC 2544.

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 January 5, 2020.
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]. Over time, the benchmarking community has updated [RFC2544] several times, including the Device Reset Benchmark [RFC6201], and the important Applicability Statement [RFC6815] concerning use outside the Isolated Test Environment (ITE) required for accurate benchmarking. Other specifications implicitly update [RFC2544], such as the IPv6 Benchmarking Methodologies in [RFC5180].
Recent testing experience with the Back-to-back Frame test and Benchmark in Section 26.4 of [RFC2544] indicates that an update is warranted [OPNFV-2017] [VSPERF-b2b]. In particular, analysis of the results indicates that buffers size matters when compensating for disruptions in the software packet processor, and this finding increases the importance of the Back-to-back frame characterization described here. This memo describes additional rationale and provides the updated method.

[RFC2544] provides its own Requirements Language consistent with [RFC2119], since [RFC1944] predates [RFC2119]. Thus, the requirements presented in this memo are expressed in [RFC2119] terms, and intended for those performing/reporting laboratory tests to improve clarity and repeatability, and for those designing devices that facilitate these tests.

2. Scope and Goals

The scope of this memo is to define an updated method to unambiguously perform tests, measure the benchmark(s), and report the results for Back-to-back Frames (presently described Section 26.4 of [RFC2544]).

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 the cases where the maximum frame rate of a single ingress port cannot be transferred to an egress port loss-free (for some frame sizes of interest).

[RFC2544] Benchmarks rely on test conditions with constant frame sizes, with the goal of understanding what network device capability has been tested. Tests with the smallest size stress the header processing capacity, and tests with the largest size stress the overall bit processing capacity. Tests with sizes in-between may determine the transition between these two capacities. However, conditions simultaneously sending multiple frame sizes, such as those described in [RFC6985], MUST NOT be used in Back-to-back Frame testing.

Section 3 of [RFC8239] describes buffer size testing for physical networking devices in a Data Center. The [RFC8239] methods measure buffer latency directly with traffic on multiple ingress ports that overload an egress port on the Device Under Test (DUT), and are not subject to the revised calculations presented in this memo. Likewise, the methods of [RFC8239] SHOULD be used for test cases where the egress port buffer is the known point of overload.
3. Motivation

Section 3.1 of [RFC1242] describes the rationale for the Back-to-back Frames Benchmark. To summarize, there are several reasons that devices on a network produce bursts of frames at the minimum allowed spacing, and it is therefore worthwhile to understand the Device Under Test (DUT) limit on the length of such bursts in practice. Also, [RFC1242] states:

"Tests of this parameter are intended to determine the extent of data buffering in the device."

After this test was defined, there have been occasional discussions of the stability and repeatability of the results, both over time and across labs. Fortunately, the Open Platform for Network Function Virtualization (OPNFV) VSPERF project’s Continuous Integration (CI) testing routinely repeats Back-to-back Frame tests to verify that test functionality has been maintained through development of the test control programs. These tests were used as a basis to evaluate stability and repeatability, even across lab set-ups when the test platform was migrated to new DUT hardware at the end of 2016.

When the VSPERF CI results were examined [VSPERF-b2b], several aspects of the results were considered notable:

1. Back-to-back Frame Benchmark was very consistent for some fixed frame sizes, and somewhat variable for others.

2. The Back-to-back Frame length reported for large frame sizes was unexpectedly long, and no explanation or measurement limit condition was indicated.

3. Calculation of the extent of buffer time in the DUT helped to explain the results observed with all frame sizes (for example, some frame sizes cannot exceed the frame header processing rate of the DUT and therefore no buffering occurs, therefore the results depended on the test equipment and not the DUT).

4. It was found that the actual buffer time in the DUT could be estimated using results from the Throughput tests conducted according to Section 26.1 of [RFC2544], because it appears that the DUT’s frame processing rate may tend to increase the estimate.

Further, if the Throughput tests of Section 26.1 of [RFC2544] are conducted as a prerequisite test, the number of frame sizes required for Back-to-back Frame Benchmarking can be reduced to one or more of the small frame sizes, or the results for large frame sizes can be
noted as invalid in the results if tested anyway (these are the frame
sizes for which the back-to-back frame rate cannot exceed the exceed
the frame header processing rate of the DUT and no buffering occurs).

[VSPERF-b2b] provides the details of the calculation to estimate the
actual buffer storage available in the DUT, using results from the
Throughput tests for each frame size, and the maximum theoretical
frame rate for the DUT links (which constrain the minimum frame
spacing). Knowledge of approximate buffer storage size (in time or
bytes) may be useful to estimate whether frame losses will occur if
DUT forwarding is temporarily suspended in a production deployment,
due to an unexpected interruption of frame processing (an
interruption of duration greater than the estimated buffer would
certainly cause lost frames).

The presentation of OPNFV VSPERF evaluation and development of
enhanced search alogorithms [VSPERF-BSLV] was discussed at IETF-102.
The enhancements are intended to compensate for transient inerrrupts
that may cause loss at near-Throughput levels of offered load.
Subsequent analysis of the results indicates that buffers within the
DUT can compensate for some interrupts, and this finding increases
the importance of the Back-to-back frame characterization described
here.

4. Prerequisites

The Test Setup MUST be consistent with Figure 1 of [RFC2544], or
Figure 2 when the tester’s sender and recover are different devices.
Other mandatory testing aspects described in [RFC2544] 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.

The test results for the Throughput Benchmark conducted according to
Section 26.1 of [RFC2544] for all [RFC2544]-RECOMMENDED frame sizes
MUST be available to reduce the tested frame size list, or to note
invalid results for individual frame sizes (because the burst length
may be essentially infinite for large frame sizes).

Note that:

- the Throughput and the Back-to-back Frame measurement
collection traffic characteristics (unidirectional or bi-
directional) MUST match.
the Throughput measurement MUST be under zero-loss conditions, according to Section 26.1 of [RFC2544].

The Back-to-back Benchmark described in Section 3.1 of [RFC1242] MUST be measured directly by the tester, where buffer size is inferred from packet loss measurements. Therefore, sources of packet loss that are un-related to consistent evaluation of buffer size SHOULD be identified and removed or mitigated. Example sources include:

- On-path active components that are external to the DUT
- Operating system environment interrupting DUT operation
- Shared resource contention between the DUT and other off-path component(s), impacting DUT’s behaviour, sometimes called the "noisy neighbour" problem.

Mitigations applicable to some of the sources above are discussed in Section 5.2, with the other measurement requirements described below in Section 5.

5. Back-to-back Frames

Objective: To characterize the ability of a DUT to process back-to-back frames as defined in [RFC1242].

The Procedure follows.

5.1. Preparing the list of Frame sizes

From the list of RECOMMENDED Frame sizes (Section 9 of [RFC2544]), select the subset of Frame sizes whose measured Throughput was less than the maximum theoretical Frame Rate. These are the only Frame sizes where it is possible to produce a burst of frames that cause the DUT buffers to fill and eventually overflow, producing one or more discarded frames.

5.2. Test for a Single Frame Size

Each trial in the test requires the tester to send a burst of frames (after idle time) with the minimum inter-frame gap, and to count the corresponding frames forwarded by the DUT.

The duration of the trial MUST be at least 2 seconds, to allow DUT buffers to deplete.
If all frames have been received, the tester increases the length of the burst according to the search algorithm and performs another trial.

If the received frame count is less than the number of frames in the burst, then the limit of DUT processing and buffering may have been exceeded, and the burst length is determined by the search algorithm for the next trial.

Classic search algorithms have been adapted for use in benchmarking, where the search requires discovery of a pair of outcomes, one with no loss and another with loss, at load conditions within the acceptable tolerance. Also for conditions encountered when benchmarking the Infrastructure for Network Function Virtualization require algorithm enhancement. Fortunately, the adaptation of Binary Search, and an enhanced Binary Search with Loss Verification have been specified in [TST009]. These algorithms (see clause 12.3) can easily be used for Back-to-back Frame benchmarking by replacing the Offered Load level with burst length in frames. [TST009] Annex B describes the theory behind the enhanced Binary Search algorithm.

There is also promising work-in-progress that may prove useful in for Back-to-back Frame benchmarking. [I-D.vpolak-mkonstan-bmwg-mlrsearch] and [I-D.vpolak-bmwg-plrsearch] are two such examples.

Either the [TST009] Binary Search or Binary Search with Loss Verification algorithms MUST be used, and input parameters to the algorithm(s) MUST be reported.

The Back-to-back Frame value is the longest burst of frames that the DUT can successfully process and buffer without frame loss, as determined from the series of trials. The tester may impose a (configurable) minimum step size for burst length, and the step size MUST be reported with the results (as this influences the accuracy and variation of test results).

5.3. Test Repetition

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

5.4. Benchmark Calculations

For each Frame size, calculate the following summary statistics for Back-to-back Frame values over the N tests:
Further, calculate the Implied DUT Buffer Time and the Corrected DUT Buffer Time in seconds, as follows:

**Implied DUT Buffer Time** =

\[
\frac{\text{Average num of Back-to-back Frames}}{\text{Max Theoretical Frame Rate}}
\]

The formula above is simply expressing the Burst of Frames in units of time.

The next step is to apply a correction factor that accounts for the DUT’s frame forwarding operation during the test (assuming a simple model of the DUT composed of a buffer and a forwarding function).

**Corrected DUT Buffer Time** =

\[
\frac{\text{Measured Throughput}}{\text{Max Theoretical Frame Rate}} = \text{Implied DUT Buffer Time} \times \frac{\text{Measured Throughput}}{\text{Max Theoretical Frame Rate}}
\]

where:

1. The "Measured Throughput" is the [RFC2544] Throughput Benchmark for the frame size tested, as augmented by methods including the Binary Search with Loss Verification algorithm in [TST009] where applicable, and MUST be expressed in Frames per second in this equation.

2. The "Max Theoretical Frame Rate" is a calculated value for the interface speed and link layer technology used, and MUST be expressed in Frames per second in this equation.

The term on the far right in the formula for Corrected DUT Buffer Time accounts for all the frames in the Burst that were transmitted by the DUT *while the Burst of frames were sent in*. So, these frames are not in the Buffer and the Buffer size is more accurately estimated by excluding them.
6. Reporting

The back-to-back results SHOULD be reported in the format of a table with a row for each of the tested frame sizes. There SHOULD be columns for the frame size and for the resultant average frame count 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 Buffer Time SHOULD also be reported.

If the tester operates using a maximum burst length in frames, then this maximum length SHOULD be reported.

+--------------+----------------+----------------+------------------+
| Frame Size,  | Ave B2B        | Min,Max,StdDev | Corrected Buff   |
| octets       | Length, frames |                | Time, Sec        |
+--------------+----------------+----------------+------------------+
| 64           | 26000          | 25500,27000,20 | 0.00004          |
+--------------+----------------+----------------+------------------+

Back-to-Back Frame Results

Static and configuration parameters:

Number of test repetitions, N

Minimum Step Size (during searches), in frames.

7. 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 of [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.

8. IANA Considerations

This memo makes no requests of IANA.

9. Acknowledgements

Thanks to Trevor Cooper, Sridhar Rao, and Martin Klozik of the VSPERF project for many contributions to the testing [VSPERF-b2b]. Yoshiaki Itou has also investigated the topic, and made useful suggestions. Maciek Konstantyowicz also provided many comments and suggestions based on his extensive integration testing and resulting search algorithm proposals - the most up-to-date feedback possible.

10. References

10.1. Normative References


10.2. Informative References


"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>.

[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)>.

[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>.

Author’s Address

Al Morton
AT&T Labs
200 Laurel Avenue South
Middletown, NJ 07748
USA
Phone: +1 732 420 1571
Fax: +1 732 368 1192
Email: acmorton@att.com