Network Working Group
Internet-Draft
Updates: 2544 (if approved)
Intended status: Informational
Expires: January 5, 2020

A. Morton
AT&T Labs
July 4, 2019

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.

Comments from Maciek and Vratko!

- Maciek's comments addressed in WG-00 version
- Some of Vratko's may benefit from discussion (today)

- Maciek asked to move "buffer size matters" to intro:
  - This is important, it's how the DUT stores packets during disruptions (like interrupts). DONE
- The goal is measuring (ingress) buffer size in front of HeaderProc, per DUT definition in section 3.
  - This works if there is no other buffering in the system under test.
  - Suggest to add a paragraph dictating the setup where no egress queue build up is possible.

2. Scope and Goals

skipping to change at *page 3, line 49*

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.

Somehow, this text (added in 05) was deleted in WG-00 ! Will be restored in WG-01 version!

[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). We present some of these details here.

The simplified model used in these calculations for the DUT includes a packet header processing function with limited rate of operation, as shown below:

```
                        |------------ DUT --------|
Generator -> Ingress -> Buffer -> HeaderProc -> Egress -> Receiver
```

So, in the back2back frame testing:

1. The Ingress burst arrives at Max Theoretical Frame Rate, and initially the frames are buffered

2. The packet header processing function (HeaderProc) operates at approximately the "Measured Throughput", removing frames from the buffer

3. Frames that have been processed are clearly not in the buffer, so the Corrected DUT buffer time equation (Section 5.4) estimates and removes the frames that the DUT forwarded on Egress during the burst.

Maciek's long, final comments listed areas of "noise" to the measurement:

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.

Maciek's long, final comments …., also mention the importance of explicitly using TST009 Binary Search w/Loss Verification algorithm, while recognizing other promising work-in-progress:

Corrected DUT Buffer Time =

$$= \text{Implied DUT Buffer Time} * \frac{\text{Measured Throughput}}{\text{Max Theoretical Frame Rate}}$$

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 aglorithm in [TST009] where applicable, and MUST be expressed in Frames per second in this equation.

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.

Broad Comments from Vratko:

- Plan to be addressed in WG-01 version
- A scenario occurring in practice has suspended processor, but the buffer is being filled more slowly <than b2b>, say at throughput rate.
- Deployers wishing to predict the time for the buffer to fill up can use this formula:

$$\text{Real Buffer Time} = \frac{\text{Corrected Buffer Time} * \text{B2B Frame Rate}}{\text{Real Frame Rate}}$$

- That is why reporting Corrected (instead of implied) buffer time is useful.
- !!! Add this calculation ? !!!
- Also, it would be nice to name the scenarios and rename the buffer times.
  - For example, "running buffer time" and "suspended buffer time"
  - (instead of "implied" and "corrected" respectively).
- Does this name change work for others?

Broad Comments from Vratko (2):

- B2B processor rate:
  - For the computation of the corrected buffer time to be correct, real processor frame processing rate (average during B2B test) should be used instead Measured Throughput in the 5.4 formula.
  - The process rate is not easy to measure directly, especially if the immediate rate varies over the duration of B2B traffic.
  - I agree that Throughput is a reasonable approximation, but there may be other quantities (e.g. FRMOL [1]), that are either a better approximation, or at least easier to measure. <Frame Rate at Maximum Offered Load, in RFC 2285>
  - Not sure how much attention other such quantities should get in the draft, as Throughput has the advantage of avoiding some frame sizes.
- ACM: I think we need some contributions (experiments) to show the value of this change

Broad Comments from Vratko (3):

- DUT vs SUT:
  - This is related to final items of 4. Prerequisities.
  - “Therefore, sources of packet loss that are un-related to consistent evaluation of buffer size SHOULD be identified and removed or mitigated.”
  - ACM: this is material just added in WG-00…
  - Do we have a separate document discussing differences between testing DUT and SUT?
  - Vratko: We should have.
  - Usually I prefer testing SUT (meaning no extra mitigations), but in this case, for the analysis of the three aforementioned scenarios to work correctly, we need to make reasonably sure the processor is not going to get suspended during B2B test.
  - Also, I agree that an average result of Binary Search with Loss Verification gives a more realistic process rate estimate than an average result without loss verification.
  - ACM: B2B Benchmarking for a single DUT retains needed simplicity!
  - WG ??

Next Steps

- Continue WG Discussion
- Address remaining comments
- Target WGLC at IETF-106…