Benchmarking Methodology Working Group   G. Lencse
Internet-Draft   BUTE
Intended status: Informational   K. Shima
Expires: November 21, 2020   IIJ-II
                                                            May 20, 2020

An Upgrade to Benchmarking Methodology for Network Interconnect Devices

draft-lencse-bmwg-rfc2544-bis-00

Abstract

RFC 2544 has defined a benchmarking methodology for network interconnect devices. We recommend a few upgrades to it for producing more reasonable results. The recommended upgrades can be classified into two categories: the application of the novelties of RFC 8219 for the legacy RFC 2544 use cases and the following new ones. Checking a reasonably small timeout individually for every single frame in the throughput and frame loss rate benchmarking procedures. Performing a statistically relevant number of tests for all benchmarking procedures. Addition of an optional non-zero frame loss acceptance criterion for the throughput measurement procedure and defining its reporting format.

Status of This Memo

Copyright Notice

Table of Contents

1. Introduction

[RFC2544] has defined a benchmarking methodology for network interconnect devices. [RFC5180] addressed IPv6 specificities and also added technology updates, but declared IPv6 transition technologies out of its scope. [RFC8219] addressed the IPv6 transition technologies, and it added further measurement procedures (e.g. for packet delay variation (PDV) and inter packet delay variation (IPDV)). It has also recommended to perform multiple tests (at least 20), and it proposed median as summarizing function and 1st and 99th percentiles as the measure of variation of the results of the multiple tests. This is a significant change compared to [RFC2544], which always used only average as summarizing function. [RFC8219] also redefined the latency measurement procedure with the requirement of marking at least 500 frames with identifying tags for latency measurements, instead of using only a single one. However, all these improvements apply only for the IPv6 transition technologies, and no update was made to [RFC2544] / [RFC5180], which we believe to be desirable.

Moreover, [RFC8219] has reused the throughput and frame loss rate benchmarking procedures from [RFC2544] with no changes. When we tested their feasibility with a few SIIT [RFC7915] implementations, we have pointed out three possible improvements in [LEN2020A]:

- Checking a reasonably small timeout individually for every single frame with the throughput and frame loss rate benchmarking procedures.

- Performing a statistically relevant number of tests for these two benchmarking procedures.

- Addition of an optional non-zero frame loss acceptance criterion for the throughput benchmarking procedure and defining its reporting format.

1.1. 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 BCP14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

2. Recommendation to Backport the Novelties of RFC8219

Besides addressing IPv6 transition technologies, [RFC8219] has also made several technological upgrades reflecting the current state of the art of networking technologies and benchmarking. But all the novelties mentioned in Section 1 of this document currently apply only for the benchmarking of IPv6 transition technologies. We contend that they could be simply backported to the benchmarking of network interconnect devices. For example, siitperf [SIITPERF], our [RFC8219] compliant DPDK-based software Tester was designed for benchmarking different SIIT [RFC7915] (also called stateless NAT64) implementations, but if it is configured to have the same IP version on both sides, it can be used to test IPv4 or IPv6 (or dual stack) routers [LEN2020B]. We highly recommend the backporting of the latency, PDV and IPDV benchmarking measurement procedures of [RFC8219].

3. Improved Throughput and Frame Loss Rate Measurement Procedures using Individual Frame Timeout

The throughput measurement procedure defined in [RFC2544] only counts the number of the sent and received test frames, but it does not identify the test frames individually. On the one hand, this approach allows the Tester to send always the very same test frame to

the DUT, which was very likely an important advantage in 1999. However, on the other hand, thus the Tester cannot check if the order of the frames is kept, or if the frames arrive back to the Tester within a given timeout time. (Perhaps none of them was an issue of hardware based network interconnect devices in 1999. But today network packet forwarding and manipulation is often implemented in software having larger buffers and producing potentially higher latencies.)

Whereas real-time applications are obviously time sensitive, other applications like HTTP or FTP are often considered throughput hungry and time insensitive. However, we have demonstrated that when we applied 100ms delay to 1% of the test frames, the throughput of HTTP download dropped by more that 50% [LEN2020C]. Therefore, an advanced throughput measurement procedure that checks the timeout time for every single test frame may produce more reasonable results. We have shown that this measurement is now feasible [LEN2020B]. In this case, we used 64-bit integers to identify the test frames and measured the latency of the frames as required by the PDV measurement procedure in Section 7.3.1. of [RFC8219]. In our particular test, we used 10ms as frame timeout, which could be a suitable value, but we recommend further studies do determine the recommended timeout value.

We recommend that the reported results of the improved throughput and frame loss rate measurements SHOULD include the applied timeout value.

4. Requirement of Statistically Relevant Number of Tests

Section 4 of [RFC2544] says that: "Furthermore, selection of the tests to be run and evaluation of the test data must be done with an understanding of generally accepted testing practices regarding repeatability, variance and statistical significance of small numbers of trials." It is made a stronger requirement (by using a "MUST") in Section 3 of [RFC5180] stating that: "Test execution and results analysis MUST be performed while observing generally accepted testing practices regarding repeatability, variance, and statistical significance of small numbers of trials." But no practical guidelines are provided concerning the minimally necessary number of tests.

[RFC8219] mentions at four different places that the tests must be repeated at least 20 times. These places are the benchmarking procedures for:

- latency (Section 7.2)

- packet delay variation (Section 7.3.1)

- inter packet delay variation (Section 7.3.2)

- DNS64 performance (Section 9.2).

We believe that a similar guideline for the minimal number of tests would be helpful for the throughput and frame loss rate benchmarking procedures. We consider 20 as an affordable number of minimum repetitions of the frame loss rate measurements. However, as for throughput measurements, we contend that the binary search may require rather high number of steps in certain situations (e.g. tens of millions of frames per second rate and high resolution) that the requirement of at least 20 repetitions of the binary search would result in unreasonably high measurement execution times. Therefore, we recommend to use an algorithm that checks the statistical properties of the results of the tests and it may stop before 20 repetitions, if the results are consistent, but it may require more than 20 repetitions, if the results are scattered. (The algorithm is yet to be developed.)

5. An Optional Non-zero Frame Loss Acceptance Criterion for the Throughput Measurement Procedure

When we defined the measurement procedure for DNS64 performance in Section 9.2 of [RFC8219], we followed both spirit and wording of the [RFC2544] throughput measurement procedure including the requirement for absolutely zero packet loss. We have elaborated our underlying considerations in our research paper [LEN2017] as follows:

1. Our goal is a well-defined performance metric, which can be measured simply and efficiently. Allowing any packet loss would result in a need for scanning/trying a large range of rates to discover the highest rate of successfully processed DNS queries.

2. Even if users may tolerate a low loss rate (please note the DNS uses UDP with no guarantee for delivery), it cannot be arbitrarily high, thus, we could not avoid defining a limit. However, any other limits than zero percent would be hardly defensible.

3. Other benchmarking procedures use the same criteria of zero packet loss and this is the standard in IETF Benchmarking Methodology Working Group.

On the one hand, we still consider our arguments valid, however, on the other hand, we are aware of different arguments for the justification of an optional non-zero frame loss acceptance criterion, too:

- Frame loss is present in our networks from the very beginning and our applications are well prepared to handle frame loss. They can definitely tolerate some low frame loss rates like 0.01% (1 frame from 10,000 frames).

- It is a wide-spread practice among benchmarking professionals to allow a certain low rate of frame loss for a long time [TOL2001] and commercially available network performance testers allow to specify a parameter usually called as "Loss Tolerance" to express a zero or non-zero acceptance criterion for throughput measurements.

- Today network packet forwarding and manipulation is often implemented in software. They do not work the same as the hardware-based forwarding devices, and may be affected by other processes running in the same host hardware. So it is not feasible to require 0% of frame loss in such forwarding devices.

- Forwarding devices (especially but not necessarily only the software-based ones) may today also have larger buffers and thus they may produce potentially higher latencies. As we have shown in Section 3, late packets are not really useful for the applications, and thus they are to be considered as lost ones. For being strict with the latency during throughput measurements (e.g. 10ms timeout), we should make up with the loss tolerance to provide meaningful benchmarking results.

- Likely due to the high frame loss rates can be experienced in WiFi networks, the latest development direction of TCP congestion control algorithms considers loss no more a sign of congestion (e.g. TCP BBR).

So we felt the necessity of having options to allow frame loss. Therefore, we recommend that throughput measurement with some low tolerated frame loss rates like 0.001% or 0.01% be a recognized optional test for network interconnect devices. To avoid the possibility of gaming, our recommendation is that the results of such tests MUST clearly state the applied loss tolerance rate.

6. Acknowledgements

The authors would like to thank ... (TBD)

7. IANA Considerations

This document does not make any request to IANA.

8. Security Considerations

We have no further security considerations beyond that of [RFC8219]. Perhaps they should be cited here so that they be applied not only for the benchmarking of IPv6 transition technologies, but also for the benchmarking of all network interconnect devices.

9. References

9.1. 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>.

[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>.

[RFC7915] Bao, C., Li, X., Baker, F., Anderson, T., and F. Gont, "IP/ICMP Translation Algorithm", RFC 7915, DOI 10.17487/RFC7915, June 2016, <https://www.rfc-editor.org/info/rfc7915>.

[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>.

[RFC8219] Georgescu, M., Pislaru, L., and G. Lencse, "Benchmarking Methodology for IPv6 Transition Technologies", RFC 8219, DOI 10.17487/RFC8219, August 2017, <https://www.rfc-editor.org/info/rfc8219>.

9.2. Informative References

[LEN2017] Lencse, G., Georgescu, M., and Y. Kadobayashi, "Benchmarking Methodology for DNS64 Servers", Computer Communications, vol. 109, no. 1, pp. 162-175, DOI: 10.1016/j.comcom.2017.06.004, Sep 2017, <http://www.hit.bme.hu/~lencse/publications/ECC-2017-B-M-DNS64-revised.pdf>.

[LEN2020A]
Lencse, G. and K. Shima, "Performance analysis of SIIT implementations: Testing and improving the methodology", Computer Communications, vol. 156, no. 1, pp. 54-67, DOI: 10.1016/j.comcom.2020.03.034, Apr 2020, <http://www.hit.bme.hu/~lencse/publications/ECC-2020-SIIT-Performance-published.pdf>.

[LEN2020B]
Lencse, G., "Design and Implementation of a Software Tester for Benchmarking Stateless NAT64 Gateways", under second review in IEICE Transactions on Communications, <http://www.hit.bme.hu/~lencse/publications/IEICE-2020-siitperf-revised.pdf>.

[LEN2020C]
Lencse, G., Shima, K., and A. Kovacs, "Gaming with the Throughput and the Latency Benchmarking Measurement Procedures of RFC 2544", under review in International Journal of Advances in Telecommunications, Electrotechnics, Signals and Systems, <http://www.hit.bme.hu/~lencse/publications/IJATES2-2020-Gaming-RFC2544-for-review.pdf>.

[SIITPERF]
Lencse, G. and Y. Kadobayashi, "Siitperf: An RFC 8219 compliant SIIT (stateless NAT64) tester written in C++ using DPDK", source code, available from GitHub, 2019, <https://github.com/lencsegabor/siitperf>.

[TOL2001] Tolly, K., "The real meaning of zero-loss testing", IT World Canada, 2001, <https://www.itworldcanada.com/article/kevin-tolly-the-real-meaning-of-zero-loss-testing/33066>.

Appendix A. Change Log

A.1. 00

Initial version.

Authors' Addresses

Gabor Lencse
Budapest University of Technology and Economics
Magyar Tudosok korutja 2.
Budapest H-1117
Hungary

Email: lencse@hit.bme.hu

Keiichi Shima
IIJ Innovation Institute
Iidabashi Grand Bloom, 2-10-2 Fujimi
Chiyoda-ku, Tokyo 102-0071
Japan

Email: keiichi@iijlab.net