Internet Engineering Task Force (IETF)
Request for Comments: 8239
Category: Informational
ISSN: 2070-1721

L. Avramov
Google
J. Rapp
VMware

August 2017

Data Center Benchmarking Methodology

Abstract

The purpose of this informational document is to establish test and evaluation methodology and measurement techniques for physical network equipment in the data center. RFC 8238 is a prerequisite for this document, as it contains terminology that is considered normative. Many of these terms and methods may be applicable beyond the scope of this document as the technologies originally applied in the data center are deployed elsewhere.

Status of This Memo

This document is not an Internet Standards Track specification; it is published for informational purposes.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Not all documents approved by the IESG are a candidate for any level of Internet Standard; see Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at http://www.rfc-editor.org/info/rfc8239.

Table of Contents

1. Introduction

Traffic patterns in the data center are not uniform and are constantly changing. They are dictated by the nature and variety of applications utilized in the data center. They can be largely east-west traffic flows (server to server inside the data center) in one data center and north-south (from the outside of the data center to the server) in another, while others may combine both. Traffic patterns can be bursty in nature and contain many-to-one, many-to-many, or one-to-many flows. Each flow may also be small and latency sensitive or large and throughput sensitive while containing a mix of UDP and TCP traffic. All of these can coexist in a single cluster and flow through a single network device simultaneously. Benchmarking tests for network devices have long used [RFC1242], [RFC2432], [RFC2544], [RFC2889], and [RFC3918], which have largely been focused around various latency attributes and throughput [RFC2889] of the Device Under Test (DUT) being benchmarked. These standards are good at measuring theoretical throughput, forwarding rates, and latency under testing conditions; however, they do not represent real traffic patterns that may affect these networking devices.

Currently, typical data center networking devices are characterized by:

- High port density (48 ports or more).
- High speed (currently, up to 100 GB/s per port).
- High throughput (line rate on all ports for Layer 2 and/or Layer 3).
- Low latency (in the microsecond or nanosecond range).
- Low amount of buffer (in the MB range per networking device).
- Layer 2 and Layer 3 forwarding capability (Layer 3 not mandatory).

This document provides a methodology for benchmarking data center physical network equipment DUTs, including congestion scenarios, switch buffer analysis, microburst, and head-of-line blocking, while also using a wide mix of traffic conditions. [RFC8238] is a prerequisite for this document, as it contains terminology that is considered normative.

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 BCP 14 [RFC2119] [RFC8174] when, and only when, they appear in all capitals, as shown here.

1.2. Methodology Format and Repeatability Recommendation

The following format is used in Sections 2 through 6 of this document:

- Objective
- Methodology
- Reporting Format

For each test methodology described in this document, it is critical that repeatability of the results be obtained. The recommendation is to perform enough iterations of the given test and to make sure that the result is consistent. This is especially important in the context of the tests described in Section 3, as the buffering testing has historically been the least reliable. The number of iterations SHOULD be explicitly reported. The relative standard deviation SHOULD be below 10%.

2. Line-Rate Testing

2.1. Objective

The objective of this test is to provide a "maximum rate" test for the performance values for throughput, latency, and jitter. It is meant to provide (1) the tests to perform and (2) methodology for verifying that a DUT is capable of forwarding packets at line rate under non-congested conditions.

2.2. Methodology

A traffic generator SHOULD be connected to all ports on the DUT. Two tests MUST be conducted: (1) a port-pair test [RFC2544] [RFC3918] and (2) a test using a full-mesh DUT [RFC2889] [RFC3918].

For all tests, the traffic generator’s sending rate MUST be less than or equal to 99.98% of the nominal value of the line rate (with no further Parts Per Million (PPM) adjustment to account for interface clock tolerances), to ensure stressing of the DUT in reasonable

worst-case conditions (see [RFC8238], Section 5 for more details). Test results at a lower rate MAY be provided for better understanding of performance increase in terms of latency and jitter when the rate is lower than 99.98%. The receiving rate of the traffic SHOULD be captured during this test as a percentage of line rate.

The test MUST provide the statistics of minimum, average, and maximum of the latency distribution, for the exact same iteration of the test.

The test MUST provide the statistics of minimum, average, and maximum of the jitter distribution, for the exact same iteration of the test.

Alternatively, when a traffic generator cannot be connected to all ports on the DUT, a snake test MUST be used for line-rate testing, excluding latency and jitter, as those would become irrelevant. The snake test is performed as follows:

- Connect the first and last port of the DUT to a traffic generator.

- Connect, back to back and sequentially, all the ports in between: port 2 to port 3, port 4 to port 5, etc., to port N-2 to port N-1, where N is the total number of ports of the DUT.

- Configure port 1 and port 2 in the same VLAN X, port 3 and port 4 in the same VLAN Y, etc., and port N-1 and port N in the same VLAN Z.

This snake test provides the capability to test line rate for Layer 2 and Layer 3 [RFC2544] [RFC3918] in instances where a traffic generator with only two ports is available. Latency and jitter are not to be considered for this test.

2.3. Reporting Format

The report MUST include the following:

- Physical-layer calibration information, as defined in [RFC8238], Section 4.

- Number of ports used.

- Reading for "throughput received as a percentage of bandwidth", while sending 99.98% of the nominal value of the line rate on each port, for each packet size from 64 bytes to 9216 bytes. As guidance, with a packet-size increment of 64 bytes between each iteration being ideal, 256-byte and 512-byte packets are also

often used. The most common packet-size ordering for the report is 64 bytes, 128 bytes, 256 bytes, 512 bytes, 1024 bytes, 1518 bytes, 4096 bytes, 8000 bytes, and 9216 bytes.

The pattern for testing can be expressed using [RFC6985].

- Throughput needs to be expressed as a percentage of total transmitted frames.

- Packet drops MUST be expressed as a count of packets and SHOULD be expressed as a percentage of line rate.

- For latency and jitter, values are expressed in units of time (usually microseconds or nanoseconds), reading across packet sizes from 64 bytes to 9216 bytes.

- For latency and jitter, provide minimum, average, and maximum values. If different iterations are done to gather the minimum, average, and maximum values, this SHOULD be specified in the report, along with a justification for why the information could not have been gathered in the same test iteration.

- For jitter, a histogram describing the population of packets measured per latency or latency buckets is RECOMMENDED.

- The tests for throughput, latency, and jitter MAY be conducted as individual independent trials, with proper documentation provided in the report, but SHOULD be conducted at the same time.

- The methodology assumes that the DUT has at least nine ports, as certain methodologies require nine or more ports.

3. Buffering Testing

3.1. Objective

The objective of this test is to measure the size of the buffer of a DUT under typical/many/multiple conditions. Buffer architectures between multiple DUTs can differ and include egress buffering, shared egress buffering SoC (Switch-on-Chip), ingress buffering, or a combination thereof. The test methodology covers the buffer measurement, regardless of buffer architecture used in the DUT.

3.2. Methodology

A traffic generator MUST be connected to all ports on the DUT. The methodology for measuring buffering for a data center switch is based on using known congestion of known fixed packet size, along with maximum latency value measurements. The maximum latency will increase until the first packet drop occurs. At this point, the maximum latency value will remain constant. This is the point of inflection of this maximum latency change to a constant value. There MUST be multiple ingress ports receiving a known amount of frames at a known fixed size, destined for the same egress port in order to create a known congestion condition. The total amount of packets sent from the oversubscribed port minus one, multiplied by the packet size, represents the maximum port buffer size at the measured inflection point.

Note that the tests described in procedures 1), 2), 3), and 4) in this section have iterations called "first iteration", "second iteration", and "last iteration". The idea is to show the first two iterations so the reader understands the logic of how to keep incrementing the iterations. The last iteration shows the end state of the variables.

1) Measure the highest buffer efficiency.

- First iteration: Ingress port 1 sending 64-byte packets at line rate to egress port 2, while port 3 is sending a known low amount of oversubscription traffic (1% recommended) with the same packet size of 64 bytes to egress port 2. Measure the buffer size value of the number of frames sent from the port sending the oversubscribed traffic up to the inflection point multiplied by the frame size.

- Second iteration: Ingress port 1 sending 65-byte packets at line rate to egress port 2, while port 3 is sending a known low amount of oversubscription traffic (1% recommended) with the same packet size of 65 bytes to egress port 2. Measure the buffer size value of the number of frames sent from the port sending the oversubscribed traffic up to the inflection point multiplied by the frame size.

- Last iteration: Ingress port 1 sending packets of size B bytes at line rate to egress port 2, while port 3 is sending a known low amount of oversubscription traffic (1% recommended) with the same packet size of B bytes to egress port 2. Measure the buffer size value of the number of frames sent from the port sending the oversubscribed traffic up to the inflection point multiplied by the frame size.

When the B value is found to provide the largest buffer size, then size B allows the highest buffer efficiency.

2) Measure maximum port buffer size.

At fixed packet size B as determined in procedure 1), for a fixed default Differentiated Services Code Point (DSCP) / Class of Service (CoS) value of 0 and for unicast traffic, proceed with the following:

- First iteration: Ingress port 1 sending line rate to egress port 2, while port 3 is sending a known low amount of oversubscription traffic (1% recommended) with the same packet size to egress port 2. Measure the buffer size value by multiplying the number of extra frames sent by the frame size.

- Second iteration: Ingress port 2 sending line rate to egress port 3, while port 4 is sending a known low amount of oversubscription traffic (1% recommended) with the same packet size to egress port 3. Measure the buffer size value by multiplying the number of extra frames sent by the frame size.

- Last iteration: Ingress port N-2 sending line rate to egress port N-1, while port N is sending a known low amount of oversubscription traffic (1% recommended) with the same packet size to egress port N. Measure the buffer size value by multiplying the number of extra frames sent by the frame size.

This test series MAY be repeated using all different DSCP/CoS values of traffic, and then using multicast traffic, in order to find out if there is any DSCP/CoS impact on the buffer size.

3) Measure maximum port pair buffer sizes.

- First iteration: Ingress port 1 sending line rate to egress port 2, ingress port 3 sending line rate to egress port 4, etc. Ingress port N-1 and port N will oversubscribe, at 1% of line rate, egress port 2 and port 3, respectively. Measure the buffer size value by multiplying the number of extra frames sent by the frame size for each egress port.

- Second iteration: Ingress port 1 sending line rate to egress port 2, ingress port 3 sending line rate to egress port 4, etc. Ingress port N-1 and port N will oversubscribe, at 1% of line rate, egress port 4 and port 5, respectively. Measure the buffer size value by multiplying the number of extra frames sent by the frame size for each egress port.

- Last iteration: Ingress port 1 sending line rate to egress port 2, ingress port 3 sending line rate to egress port 4, etc. Ingress port N-1 and port N will oversubscribe, at 1% of line rate, egress port N-3 and port N-2, respectively. Measure the buffer size value by multiplying the number of extra frames sent by the frame size for each egress port.

This test series MAY be repeated using all different DSCP/CoS values of traffic and then using multicast traffic.

4) Measure maximum DUT buffer size with many-to-one ports.

- First iteration: Ingress ports 1,2,... N-1 each sending [(1/[N-1])*99.98]+[1/[N-1]] % of line rate per port to egress port N.

- Second iteration: Ingress ports 2,... N each sending [(1/[N-1])*99.98]+[1/[N-1]] % of line rate per port to egress port 1.

- Last iteration: Ingress ports N,1,2...N-2 each sending [(1/[N-1])*99.98]+[1/[N-1]] % of line rate per port to egress port N-1.

This test series MAY be repeated using all different CoS values of traffic and then using multicast traffic.

Unicast traffic, and then multicast traffic, SHOULD be used in order to determine the proportion of buffer for the documented selection of tests. Also, the CoS value for the packets SHOULD be provided for each test iteration, as the buffer allocation size MAY differ per CoS value. It is RECOMMENDED that the ingress and egress ports be varied in a random but documented fashion in multiple tests in order to measure the buffer size for each port of the DUT.

3.3. Reporting Format

The report MUST include the following:

- The packet size used for the most efficient buffer used, along with the DSCP/CoS value.

- The maximum port buffer size for each port.

- The maximum DUT buffer size.

- The packet size used in the test.

- The amount of oversubscription, if different than 1%.

- The number of ingress and egress ports, along with their location on the DUT.

- The repeatability of the test needs to be indicated: the number of iterations of the same test and the percentage of variation between results for each of the tests (min, max, avg).

The percentage of variation is a metric providing a sense of how big the difference is between the measured value and the previous values.

For example, for a latency test where the minimum latency is measured, the percentage of variation (PV) of the minimum latency will indicate by how much this value has varied between the current test executed and the previous one.

PV = ((x2-x1)/x1)*100, where x2 is the minimum latency value in the current test and x1 is the minimum latency value obtained in the previous test.

The same formula is used for maximum and average variations measured.

4. Microburst Testing

4.1. Objective

The objective of this test is to find the maximum amount of packet bursts that a DUT can sustain under various configurations.

This test provides additional methodology that supplements the tests described in [RFC1242], [RFC2432], [RFC2544], [RFC2889], and [RFC3918].

- All bursts should be sent with 100% intensity. Note: "Intensity" is defined in [RFC8238], Section 6.1.1.

- All ports of the DUT must be used for this test.

- It is recommended that all ports be tested simultaneously.

4.2. Methodology

A traffic generator MUST be connected to all ports on the DUT. In order to cause congestion, two or more ingress ports MUST send bursts of packets destined for the same egress port. The simplest of the setups would be two ingress ports and one egress port (2 to 1).

The burst MUST be sent with an intensity (as defined in [RFC8238], Section 6.1.1) of 100%, meaning that the burst of packets will be sent with a minimum interpacket gap. The amount of packets contained in the burst will be trial variable and increase until there is a non-zero packet loss measured. The aggregate amount of packets from all the senders will be used to calculate the maximum microburst amount that the DUT can sustain.

It is RECOMMENDED that the ingress and egress ports be varied in multiple tests in order to measure the maximum microburst capacity.

The intensity of a microburst (see [RFC8238], Section 6.1.1) MAY be varied in order to obtain the microburst capacity at various ingress rates.

It is RECOMMENDED that all ports on the DUT be tested simultaneously, and in various configurations, in order to understand all the combinations of ingress ports, egress ports, and intensities.

An example would be:

- First iteration: N-1 ingress ports sending to one egress port.
- Second iteration: N-2 ingress ports sending to two egress ports.
- Last iteration: Two ingress ports sending to N-2 egress ports.

4.3. Reporting Format

The report MUST include the following:

- The maximum number of packets received per ingress port with the maximum burst size obtained with zero packet loss.
- The packet size used in the test.
- The number of ingress and egress ports, along with their location on the DUT.
- The repeatability of the test needs to be indicated: the number of iterations of the same test and the percentage of variation between results (min, max, avg).

5. Head-of-Line Blocking

5.1. Objective

Head-of-line blocking (HOLB) is a performance-limiting phenomenon that occurs when packets are held up by the first packet ahead waiting to be transmitted to a different output port. This is defined in RFC 2889, Section 5.5 ("Congestion Control"). This section expands on RFC 2889 in the context of data center benchmarking.

The objective of this test is to understand the DUT's behavior in the HOLB scenario and measure the packet loss.

The differences between this HOLB test and RFC 2889 are as follows:

- This HOLB test starts with eight ports in two groups of four ports each, instead of four ports (as compared with Section 5.5 of RFC 2889).

- This HOLB test shifts all the port numbers by one in a second iteration of the test; this is new, as compared to the HOLB test described in RFC 2889. The shifting port numbers continue until all ports are the first in the group; the purpose of this is to make sure that all permutations are tested in order to cover differences in behavior in the SoC of the DUT.

- Another test within this HOLB test expands the group of ports, such that traffic is divided among four ports instead of two (25% instead of 50% per port).

- Section 5.3 lists requirements that supplement the requirements listed in RFC 2889, Section 5.5.

5.2. Methodology

In order to cause congestion in the form of HOLB, groups of four ports are used. A group has two ingress ports and two egress ports. The first ingress port MUST have two flows configured, each going to a different egress port. The second ingress port will congest the second egress port by sending line rate. The goal is to measure if there is loss on the flow for the first egress port, which is not oversubscribed.

A traffic generator MUST be connected to at least eight ports on the DUT and SHOULD be connected using all the DUT ports.

Note that the tests described in procedures 1) and 2) in this section have iterations called "first iteration", "second iteration", and "last iteration". The idea is to show the first two iterations so the reader understands the logic of how to keep incrementing the iterations. The last iteration shows the end state of the variables.

1) Measure two groups with eight DUT ports.

- First iteration: Measure the packet loss for two groups with consecutive ports.

  The composition of the first group is as follows:

  Ingress port 1 sending 50% of traffic to egress port 3
  and ingress port 1 sending 50% of traffic to egress port 4.
  Ingress port 2 sending line rate to egress port 4.
  Measure the amount of traffic loss for the traffic from ingress port 1 to egress port 3.

  The composition of the second group is as follows:

  Ingress port 5 sending 50% of traffic to egress port 7
  and ingress port 5 sending 50% of traffic to egress port 8.
  Ingress port 6 sending line rate to egress port 8.
  Measure the amount of traffic loss for the traffic from ingress port 5 to egress port 7.

- Second iteration: Repeat the first iteration by shifting all the ports from N to N+1.

  The composition of the first group is as follows:

  Ingress port 2 sending 50% of traffic to egress port 4
  and ingress port 2 sending 50% of traffic to egress port 5.
  Ingress port 3 sending line rate to egress port 5.
  Measure the amount of traffic loss for the traffic from ingress port 2 to egress port 4.

  The composition of the second group is as follows:

  Ingress port 6 sending 50% of traffic to egress port 8
  and ingress port 6 sending 50% of traffic to egress port 9.
  Ingress port 7 sending line rate to egress port 9.
  Measure the amount of traffic loss for the traffic from ingress port 6 to egress port 8.

- Last iteration: When the first port of the first group is connected to the last DUT port and the last port of the second group is connected to the seventh port of the DUT.

Measure the amount of traffic loss for the traffic from ingress port N to egress port 2 and from ingress port 4 to egress port 6.

2) Measure with N/4 groups with N DUT ports.

The traffic from the ingress port is split across four egress ports (100/4 = 25%).

- First iteration: Expand to fully utilize all the DUT ports in increments of four. Repeat the methodology of procedure 1) with all the groups of ports possible to achieve on the device, and measure the amount of traffic loss for each port group.

- Second iteration: Shift by +1 the start of each consecutive port of the port groups.

- Last iteration: Shift by N-1 the start of each consecutive port of the port groups, and measure the amount of traffic loss for each port group.

5.3. Reporting Format

For each test, the report MUST include the following:

- The port configuration, including the number and location of ingress and egress ports located on the DUT.

- If HOLB was observed in accordance with the HOLB test described in Section 5.

- Percent of traffic loss.

- The repeatability of the test needs to be indicated: the number of iterations of the same test and the percentage of variation between results (min, max, avg).

6. Incast Stateful and Stateless Traffic

6.1. Objective

The objective of this test is to measure the values for TCP Goodput [TCP-INCAST] and latency with a mix of large and small flows. The test is designed to simulate a mixed environment of stateful flows that require high rates of goodput and stateless flows that require low latency. Stateful flows are created by generating TCP traffic, and stateless flows are created using UDP traffic.

6.2. Methodology

In order to simulate the effects of stateless and stateful traffic on the DUT, there MUST be multiple ingress ports receiving traffic destined for the same egress port. There also MAY be a mix of stateful and stateless traffic arriving on a single ingress port. The simplest setup would be two ingress ports receiving traffic destined to the same egress port.

One ingress port MUST maintain a TCP connection through the ingress port to a receiver connected to an egress port. Traffic in the TCP stream MUST be sent at the maximum rate allowed by the traffic generator. At the same time, the TCP traffic is flowing through the DUT, and the stateless traffic is sent destined to a receiver on the same egress port. The stateless traffic MUST be a microburst of 100% intensity.

It is RECOMMENDED that the ingress and egress ports be varied in multiple tests in order to measure the maximum microburst capacity.

The intensity of a microburst MAY be varied in order to obtain the microburst capacity at various ingress rates.

It is RECOMMENDED that all ports on the DUT be used in the test.

The tests described below have iterations called "first iteration", "second iteration", and "last iteration". The idea is to show the first two iterations so the reader understands the logic of how to keep incrementing the iterations. The last iteration shows the end state of the variables.

For example:

Stateful traffic port variation (TCP traffic):

TCP traffic needs to be generated for this test. During the iterations, the number of egress ports MAY vary as well.

- First iteration: One ingress port receiving stateful TCP traffic and one ingress port receiving stateless traffic destined to one egress port.
- Second iteration: Two ingress ports receiving stateful TCP traffic and one ingress port receiving stateless traffic destined to one egress port.
- Last iteration: N-2 ingress ports receiving stateful TCP traffic and one ingress port receiving stateless traffic destined to one egress port.

Stateless traffic port variation (UDP traffic):

UDP traffic needs to be generated for this test. During the iterations, the number of egress ports MAY vary as well.

- First iteration: One ingress port receiving stateful TCP traffic and one ingress port receiving stateless traffic destined to one egress port.
- Second iteration: One ingress port receiving stateful TCP traffic and two ingress ports receiving stateless traffic destined to one egress port.
- Last iteration: One ingress port receiving stateful TCP traffic and N-2 ingress ports receiving stateless traffic destined to one egress port.

6.3. Reporting Format

The report MUST include the following:

- Number of ingress and egress ports, along with designation of stateful or stateless flow assignment.

- Stateful flow goodput.

- Stateless flow latency.

- The repeatability of the test needs to be indicated: the number of iterations of the same test and the percentage of variation between results (min, max, avg).

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 constraints specified in the sections above.

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 DUT.

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

8. IANA Considerations

This document does not require any IANA actions.

9. References

9.1. 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>.

[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>.

[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>.

[RFC8238] Avramov, L. and J. Rapp, "Data Center Benchmarking Terminology", RFC 8238, DOI 10.17487/RFC8238, August 2017, <https://www.rfc-editor.org/info/rfc8238>.

9.2. Informative References

[RFC2432] Dubray, K., "Terminology for IP Multicast Benchmarking", RFC 2432, DOI 10.17487/RFC2432, October 1998, <https://www.rfc-editor.org/info/rfc2432>.

[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>.

[RFC3918] Stopp, D. and B. Hickman, "Methodology for IP Multicast Benchmarking", RFC 3918, DOI 10.17487/RFC3918, October 2004, <https://www.rfc-editor.org/info/rfc3918>.

[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>.

[TCP-INCAST] Chen, Y., Griffith, R., Zats, D., Joseph, A., and R. Katz, "Understanding TCP Incast and Its Implications for Big Data Workloads", April 2012, <http://yanpeichen.com/professional/usenixLoginIncastReady.pdf>.

Acknowledgments

The authors would like to thank Al Morton and Scott Bradner for their reviews and feedback.

Authors' Addresses

Lucien Avramov
Google
1600 Amphitheatre Parkway
Mountain View, CA 94043
United States of America

Email: lucien.avramov@gmail.com

Jacob Rapp
VMware
3401 Hillview Ave.
Palo Alto, CA 94304
United States of America

Email: jhrapp@gmail.com