                   Metrics and Methods for IP Capacity
                 draft-morton-ippm-capcity-metric-method-00

Abstract

   This memo revisits the problem of Network Capacity metrics first
   examined in RFC 5136. The memo specifies a more practical Maximum
   IP-layer Capacity metric definition catering for measurement
   purposes, and outlines the corresponding methods of measurement.

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.

Table of Contents

1. Introduction

   The IETF's efforts to define Network and Bulk Transport Capacity have
   been chartered and progressed for over twenty years. Over that time,
   the performance community has seen development of Informative
   definitions in RFC 3148 for Framework for Bulk Transport Capacity
   (BTC), RFC 5136 for Network Capacity and Maximum IP-layer Capacity,
   and the Experimental metric definitions and methods in [RFC8337],
   Model-Based Metrics for BTC.

This memo recognizes the importance of a definition of a Maximum IP-
layer Capacity Metric at this time, a definition that is both
practical and effective for the performance community's needs,
including Internet users. The metric definition is intended for
Active Methods of Measurement [RFC7799], and a method of measurement
is included here.

The most direct active measurement of IP-layer Capacity would use IP
packets, but in practice a transport header is needed to traverse
address and port translators. UDP offeres the most direct assessment
possibility, and in the [copycat][copycat] measurement study to
investigate whether UDP is viable as a general Internet transport
protocol, the authors found that a high percentage of paths tested
support UDP transport. A number of liaisons have been exchanged on
this topic [refs to ITU-T SG12, ETSI STQ, BBF liaisons], discussing
the laboratory and field tests that support the UDP-based approach to
IP-layer Capacity measurement.

This memo also recognizes the many updates to the IP Performance
Metrics Framework [RFC2330] published over twenty years, and makes
use of [RFC7312] for Advanced Stream and Sampling Framework, and RFC
8468 [RFC8468]IPv4, IPv6, and IPv4-IPv6 Coexistence Updates.

NOTE: The text contains Author comments, in brackets [RG: , acm:]

2. Scope and Goals

The scope of this memo is to define a metric and corresponding method
to unambiguously perform Active measurements of Maximum IP-Layer
Capacity.

The main goal is to harmonize the specified metric and method across
the industry, and this memo is the vehicle through which working
group (and eventually, IETF) consensus will be captured and
communicated to achieve broad agreement, and possibly changes in
other Standards Development Organizations (SDO).

A local goal is provide efficient test procedures where possible, and
to recommend reporting with additional interpretation of the results.
Also, to foster the development of protocol support for this metric
and method of measurement.

3. Motivation

As with any problem that has been worked for many years in various
SDOs without any special attempts at coordination, various solutions
for metrics and methods have emerged.

There are five factors that have changed (or begun to change) in the last 5 years, and the presence of any one of them on the path requires features in the measurement design to account for the changes:

1. Internet access is no longer the bottleneck for many users

2. Both speed and latency are important to user's satisfaction

3. UDP's growing role in Transport, in areas where TCP once dominated.

4. Content and applications moving closer to users.

5. Possibly less traffic crossing ISP gateways in future.

4. Metric Definitions

This section sets requirements for the following components to support the Maximum IP-layer Capacity Metric.

4.1. Formal Name

Type-P-Max-IP-Capacity, or informally called Maximum IP-layer Capacity.

Note that Type-P depends on the chosen method.

4.2. Parameters

This section lists the REQUIRED input factors to specify a Route metric.

o Src, the address of a host (such as the globally routable IP address).

o Dst, the address of a host (such as the globally routable IP address).

o i, the limit on the number of Hops a specific packet may visit as it traverses from the host at Src to the host at Dst (such as the TTL or Hop Limit).

o MaxHops, the maximum value of i used, (i=1,2,3,...MaxHops).

o T, the time at the start of measurement interval

o I, the duration of measurement interval

o dt, the duration of N equal sub-intervals in I

o Ts, the host time of a transmitted test packet as measured at
   MP(Src), meaning Measurement Point at the Source.

o Ta, the host time of a test packet's *arrival* as measured at
   MP(Dst), assigned to packets that arrive within a "reasonable"
   time (see parameter below).

o Tmax, a maximum waiting time for test packets to arrive at the
   destination, set sufficiently long to disambiguate packets with
   long delays from packets that are discarded (lost), such that the
   distribution of one-way delay is not truncated.

o F, the number of different flows synthesized by the method.

o flow, the stream of packets with the same n-tuple of designated
   header fields that (when held constant) result in identical
   treatment in a multi-path decision (such as the decision taken in
   load balancing). Note: The IPv6 flow label MAY be included in the
   flow definition when routers have complied with [RFC6438]
   guidelines at the Tunnel End Points (TEP), and the source of the
   measurement is a TEP.

o Type-P, the complete description of the packets for which this
   assessment applies (including the flow-defining fields). Note
   that the UDP transport layer is one requirement specified below.

o PM, a list of fundamental metrics, such as loss, delay, and
   reordering, and corresponding Target performance threshold. At
   least one fundamental metric and Target performance threshold MUST
   be supplied (such as One-way IP Packet Loss [RFC7680] equal to
   zero).

4.3. Metric Definitions

This section defines the REQUIRED aspects of the measureable Maximum
IP-layer Capacity metric (unless otherwise indicated) for
measurements between specified Source and Destination hosts:

Define the IP-layer capacity, Maximum_C(T,I,PM), to be the maximum
number of IP-layer bits (including header and data fields) that can
be transmitted from the Src host and correctly received by the Dst
host during one contiguous sub-interval, dt in length, during the
interval [T, T+I], and where the packet count of that single sub-
interval dtn in [T, T+I] indicates the maximum number of IP-layer
bits n0[dtn-1,dtn] which was captured as part of all packet counts n
for all dt in [T, T+I]. The interval dt SHOULD be set to a natural

number m so that T+I = T + m*dt with dtn - dtn-1 = dt and with 0 < n
<= m. Parameter PM represents the other performance metrics [see
section Related Round-Trip Delay and Loss Definitions below] and
their measurement results for the maximum IP-layer capacity. At
least one target performance threshold MUST be defined. If more than
one target performance threshold is defined, then the sub-interval
with maximum number of bits transmitted MUST meet all the target
performance thresholds.

[RG: this requires more explanation. Do you mean that all results
must hit the target performance level? Or is a one / two / k times
hit out of x trials [T, T+I] a criterium indicating that a target has
been reached? Or do you look for the maximum capacity without packet
loss and queuing or added RTD latency, respectively? acm: I think
I've clarified this in the text above...]

Mathematically, this definition can be represented as:

```
                         max  ( n0[dtn-1,dtn] )
                        [T,T+I]
        Maximum_C(T,I,PM) = -------------------------
                                      dt
          where:
            T                                          T+I

          _____
           |   |   |   |   |   |   |   |   |   |   |
          dtn=0   1   2   3   4   5   6   7   8   9   m=10
```

Equation for Maximum Capacity

and:

o n0 is the total number of IP-layer header and payload bits that
   can be transmitted in Standard Formed packets from the Src host
   and correctly received by the Dst host during one contiguous sub-
   interval, dt in length, during the interval [T, T+I],

o Maximum _C(T,I,PM) corresponds to the maximum value of n0 measured
   in any sub-interval ending at dtn (meaning T + n*dt), divided by
   the constant length of all sub-intervals, dt.

o all sub-intervals SHOULD be of equal duration. Choosing dt as
   non-overlapping consecutive time intervals allows for a simple
   implementation. [RG: a sliding window of dt has its charme too,
   why exclude it? acm: seems less practical for real-time feedback.]

o [acm: I think this is a discussion point, not essential to the
   definition] If traffic conditioning applies along a path for which
   Maximum _C(T,I,PM) is to be determined, different values for dt
   SHOULD be picked and measurements be executed during multiple
   intervals [T, T+I]. Any single interval dt SHOULD be chosen so
   that is an integer multiple of increasing values k times
   serialisation delay of a path MTU at the physical interface speed
   where traffic conditioning is expected. This should avoid taking
   configured burst tolerance singletons as a valid Maximum
   _C(T,I,PM) result.

o The bandwidth of the physical interface of the measurement device
   must be higher than that of the interface whose Maximum _C(T,I,PM)
   is to be measured.

In this definition, the m sub-intervals can be viewed as trials when
the Src host varies the transmitted packet rate, searching for the
maximum n0 that meets the PM criteria measured at the Dst host in a
test of duration, I. When the transmitted packet rate is held
constant at the Src host, the m sub-intervals may also be viewed as
trials to evaluate the stability of n0 and metric(s) in the PM list
over all dt-length intervals in I.

Measurements according to these definitions SHALL use UDP transport
layer. [RG: don't we need loss free transmission without added
latency as criteria and add that UDP without closed loop flow control
needs to be applied ? acm: I don't think we should require loss-free
transmission, because most networks allow a small loss ratio which
would likely appear to be zero loss in most trials. Methods may use
feedback, let's talk about how to diffentiate from flow-control]

4.4. Related Round-Trip Delay and Loss Definitions

RTD[dtn-1,dtn] is defined as a sample of the [RFC2681] Round-trip
Delay between the Src host and the Dst host over the interval
[T,T+I]. The statistics used to to summarize RTD[dtn-1,dtn] MAY
include the minimum, maximum, and mean.

RTL[dtn-1,dtn] is defined as a sample of the [RFC6673] Round-trip
Loss between the Src host and the Dst host over the interval [T,T+I].
The statistics used to to summarize RTL[dtn-1,dtn] MAY include the
lost packet count and the lost packet ratio.

4.5. Discussion

Depending on the

4.6. Reporting the Metric

   @@@@ not yet addressed,

5. Method of Measurement

   This section needs development, based on Annex A of Y.1540.

   The duration of a test, I, MUST be constrained in a production
   network, since this is an active test method and it will likely cause
   congestion on the Src to Dst host path during a test.

   Additional Test methods and configurations may be provided in this
   section, after review.

5.1. Load Rate Adjustment Algorithm (from udpst)

   A table is pre-built defining all the offered load rates that will be
   supported (R1 - Rn, in ascending order). Each rate is defined as
   datagrams of size S, sent as a burst of count C, every time interval
   T. While it is advantageous to use datagrams of as large a size as
   possible, it may be prudent to use a slightly smaller maximum that
   allows for secondary protocol headers and/or tunneling without
   resulting in IP-layer fragmentation.

   At the beginning of a test, the sender begins sending at rate R1 and
   the receiver starts a feedback timer at interval F (while awaiting
   inbound datagrams). As datagrams are received they are checked for
   sequence number anomalies (loss, out-of-order, duplication, etc.) and
   the delay variation is measured (one-way or round-trip). This
   information is accumulated until the feedback timer F expires and a
   status feedback message is sent from the receiver back to the sender,
   to communicate this information. The accumulated statistics are then
   reset by the receiver for the next feedback interval. As feedback
   messages are received back at the sender, they are evaluated to
   determine how to adjust the current offered load rate (Rx).

   If the feedback indicates that there were no sequence number
   anomalies AND the delay variation was below the lower threshold, the
   offered load rate is increased. If congestion has not been confirmed
   up to this point, the offered load rate is increased by more than one
   rate (e.g., Rx+10). This allows the offered load to quickly reach a
   near-maximum rate. Conversely, if congestion has been previously
   confirmed, the offered load rate is only increased by one (Rx+1).

   If the feedback indicates that sequence number anomalies were
   detected OR the delay variation was above the upper threshold, the
   offered load rate is decreased. If congestion has not been confirmed

up to this point, the offered load rate is decreased by more than one rate (e.g., Rx-30). This allows the offered load to back off enough to compensate for the fast initial ramp-up. Conversely, if congestion has been previously confirmed, the offered load rate is only decreased by one (Rx-1).

If the feedback indicates that there were no sequence number anomalies AND the delay variation was above the lower threshold, but below the upper threshold, the offered load rate is not changed. This allows time for recent changes in the offered load rate to stabilize, and the feedback to represent current conditions more accurately.

Lastly, the method for confirming congestion is that there were sequence number anomalies OR the delay variation was above the upper threshold for two consecutive feedback intervals.

6. Reporting

This section needs development...

The following text TO BE REPLACED !!!!

======================================

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,       | Max IP-Layer      | Min,Ave,StdDev | Time dt, |
| octets + # Flows  | Capacity, bps     |                | Sec      |
+-------------------+-------------------+----------------+----------+
| 64,100            | 26000             | 25500,27000,20 | 0.00004  |
+-------------------+-------------------+----------------+----------+
```

Maximum IP-layer Capacity Results

   Static and configuration parameters:

   Number of test repetitions, N

   Minimum Step Size (during searches), in frames.

7. Security Considerations

   Active metrics and measurements have a long history of security
   considerations [add references to LMAP Framework, etc.].

   <There are certainly some new ones for Capacity testing>

8. IANA Considerations

   This memo makes no requests of IANA.

9. Acknowledgements

   Thanks to

10. References

10.1. 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>.

   [RFC2330] Paxson, V., Almes, G., Mahdavi, J., and M. Mathis,
              "Framework for IP Performance Metrics", RFC 2330,
              DOI 10.17487/RFC2330, May 1998,
              <https://www.rfc-editor.org/info/rfc2330>.

   [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>.

   [RFC2681] Almes, G., Kalidindi, S., and M. Zekauskas, "A Round-trip
              Delay Metric for IPPM", RFC 2681, DOI 10.17487/RFC2681,
              September 1999, <https://www.rfc-editor.org/info/rfc2681>.

   [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>.

   [RFC5136] Chimento, P. and J. Ishac, "Defining Network Capacity",
              RFC 5136, DOI 10.17487/RFC5136, February 2008,
              <https://www.rfc-editor.org/info/rfc5136>.

   [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>.

   [RFC6438] Carpenter, B. and S. Amante, "Using the IPv6 Flow Label
              for Equal Cost Multipath Routing and Link Aggregation in
              Tunnels", RFC 6438, DOI 10.17487/RFC6438, November 2011,
              <https://www.rfc-editor.org/info/rfc6438>.

   [RFC6673] Morton, A., "Round-Trip Packet Loss Metrics", RFC 6673,
              DOI 10.17487/RFC6673, August 2012,
              <https://www.rfc-editor.org/info/rfc6673>.

   [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>.

   [RFC7312] Fabini, J. and A. Morton, "Advanced Stream and Sampling
              Framework for IP Performance Metrics (IPPM)", RFC 7312,
              DOI 10.17487/RFC7312, August 2014,
              <https://www.rfc-editor.org/info/rfc7312>.

   [RFC7799] Morton, A., "Active and Passive Metrics and Methods (with
              Hybrid Types In-Between)", RFC 7799, DOI 10.17487/RFC7799,
              May 2016, <https://www.rfc-editor.org/info/rfc7799>.

   [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>.

   [RFC8337] Mathis, M. and A. Morton, "Model-Based Metrics for Bulk
              Transport Capacity", RFC 8337, DOI 10.17487/RFC8337, March
              2018, <https://www.rfc-editor.org/info/rfc8337>.

   [RFC8468] Morton, A., Fabini, J., Elkins, N., Ackermann, M., and V.
              Hegde, "IPv4, IPv6, and IPv4-IPv6 Coexistence: Updates for
              the IP Performance Metrics (IPPM) Framework", RFC 8468,
              DOI 10.17487/RFC8468, November 2018,
              <https://www.rfc-editor.org/info/rfc8468>.

10.2. Informative References

   [copycat] Edleine, K., Kuhlewind, K., Trammell, B., and B. Donnet,
              "copycat: Testing Differential Treatment of New Transport
              Protocols in the Wild (ANRW '17)", July 2017,
              <https://irtf.org/anrw/2017/anrw17-final5.pdf>.

   [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>.

   [TST009]   Morton, R. A., "ETSI GS NFV-TST 009 V3.1.1 (2018-10),
              "Network Functions Virtualisation (NFV) Release 3;
              Testing; Specification of Networking Benchmarks and
              Measurement Methods for NFVI"", October 2018,
              <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>.

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

   Ruediger Geib
   Deutsche Telekom
   Heinrich Hertz Str. 3-7
   Darmstadt 64295
   Germany

   Phone: +49 6151 5812747
   Email: Ruediger.Geib@telekom.de

   Len Ciavattone
   AT&T Labs
   200 Laurel Avenue South
   Middletown,, NJ 07748
   USA

   Email: lencia@att.com