

Network Working Group
Internet-Draft
Intended status: Standards Track
Expires: October 2, 2021

A. Morton
AT&T Labs
R. Geib
Deutsche Telekom
L. Ciavattone
AT&T Labs
March 31, 2021

Metrics and Methods for One-way IP Capacity
draft-ietf-ippm-capacity-metric-method-09

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.

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 October 2, 2021.

Copyright Notice

Copyright (c) 2021 IETF Trust and the persons identified as the document authors. All rights reserved.

This document is subject to [BCP 78](#) and the IETF Trust's Legal Provisions Relating to IETF Documents (<https://trustee.ietf.org/license-info>) in effect on the date of publication of this document. Please review these documents carefully, as they describe your rights and restrictions with respect to this document. Code Components extracted from this document must include Simplified BSD License text as described in [Section 4.e](#) of

Internet-Draft

IP Capacity Metrics/Methods

March 2021

the Trust Legal Provisions and are provided without warranty as described in the Simplified BSD License.

Table of Contents

1.	Introduction	3
1.1.	Requirements Language	4
2.	Scope, Goals, and Applicability	4
3.	Motivation	5
4.	General Parameters and Definitions	6
5.	IP-Layer Capacity Singleton Metric Definitions	7
5.1.	Formal Name	7
5.2.	Parameters	7
5.3.	Metric Definitions	8
5.4.	Related Round-Trip Delay and One-way Loss Definitions	9
5.5.	Discussion	10
5.6.	Reporting the Metric	10
6.	Maximum IP-Layer Capacity Metric Definitions (Statistic)	10
6.1.	Formal Name	10
6.2.	Parameters	10
6.3.	Metric Definitions	10
6.4.	Related Round-Trip Delay and One-way Loss Definitions	12
6.5.	Discussion	12
6.6.	Reporting the Metric	13
7.	IP-Layer Sender Bit Rate Singleton Metric Definitions	13
7.1.	Formal Name	13
7.2.	Parameters	14
7.3.	Metric Definition	14
7.4.	Discussion	15
7.5.	Reporting the Metric	15
8.	Method of Measurement	15
8.1.	Load Rate Adjustment Algorithm	15
8.2.	Measurement Qualification or Verification	20
8.3.	Measurement Considerations	22
8.4.	Running Code	23
9.	Reporting Formats	24
9.1.	Configuration and Reporting Data Formats	26
10.	Security Considerations	26
11.	IANA Considerations	27
12.	Acknowledgments	27
13.	Appendix A - Load Rate Adjustment Pseudo Code	27
14.	Appendix B - RFC 8085 UDP Guidelines Check	28
14.1.	Assessment of Mandatory Requirements	28

14.2.	Assessment of Recommendations	30
15.	References	33
15.1.	Normative References	33
15.2.	Informative References	34
	Authors' Addresses	36

[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 [[RFC3148](#)] 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 revisits the problem of Network Capacity metrics examined first in [[RFC3148](#)] and later in [[RFC5136](#)]. Maximum IP-Layer Capacity and [[RFC3148](#)] Bulk Transfer Capacity (goodput) are different metrics. Maximum IP-layer Capacity is like the theoretical goal for goodput. There are many metrics in [[RFC5136](#)], such as Available Capacity. Measurements depend on the network path under test and the use case. Here, the main use case is to assess the maximum capacity of the access network, with specific performance criteria used in the measurement.

This memo recognizes the importance of a definition of a Maximum IP-layer Capacity Metric at a time when access speeds have increased dramatically; a definition that is both practical and effective for the performance community's needs, including Internet users. The metric definition is intended to use Active Methods of Measurement [[RFC7799](#)], and a method of measurement is included.

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 offers the most direct assessment possibility, and in the [[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 [[LS-SG12-A](#)] [[LS-SG12-B](#)], 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 [\[RFC8468\]](#) with IPv4, IPv6, and IPv4-IPv6 Coexistence Updates.

[Appendix A](#) describes the load rate adjustment algorithm in pseudo-code. [Appendix B](#) discusses the algorithm's compliance with [\[RFC8085\]](#).

[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.

[2.](#) Scope, Goals, and Applicability

The scope of this memo is to define a metric and corresponding method to unambiguously perform Active measurements of Maximum IP-Layer Capacity, along with related metrics and methods.

Another goal is to harmonize the specified metric and method across the industry, and this memo is the vehicle that captures IETF consensus, possibly resulting in changes to the specifications of other Standards Development Organizations (SDO) (through each SDO's normal contribution process, or through liaison exchange).

A local goal is to aid efficient test procedures where possible, and to recommend reporting with additional interpretation of the results. Fostering the development of protocol support for this metric and method of measurement is also a goal of this memo (all active testing protocols currently defined by the IPPM WG are UDP-based, meeting a key requirement of these methods). The supporting protocol development to measure this metric according to the specified method is a key future contribution to Internet measurement.

The load rate adjustment algorithm's scope is limited to helping determine the Maximum IP-Layer Capacity in the context of an infrequent, diagnostic, short term measurement. It is RECOMMENDED to discontinue non-measurement traffic that shares a subscriber's dedicated resources while testing: measurements may not be accurate and throughput of competing elastic traffic may be greatly reduced.

The primary application of the metric and method of measurement described here is the same as in [Section 2 of \[RFC7479\]](#) where:

- o The access portion of the network is the focus of this problem statement. The user typically subscribes to a service with bidirectional access partly described by rates in bits per second.

In addition, the use of the load rate adjustment algorithm described in [section 8.1](#) has the following additional applicability limitations:

- MUST only be used in the application of diagnostic and operations measurements as described in this memo

- MUST only be used in circumstances consistent with [Section 10](#), Security Considerations

- If a network operator is certain of the access capacity to be validated, then testing MAY start with a fixed rate test at the access capacity and avoid activating the load adjustment algorithm. However, the stimulus for a diagnostic test (such as a subscriber request) strongly implies that there is no certainty and the load adjustment algorithm will be needed.

Further, the metric and method of measurement are intended for use where specific exact path information is unknown within a range of possible values:

- the subscriber's exact Maximum IP-Layer Capacity is unknown (which is sometimes the case; service rates can be increased due to upgrades without a subscriber's request, or to provide a surplus to compensate for possible underestimates of TCP-based testing.

- the size of the access bottleneck buffer is unknown.

Finally, the measurement system's load rate adjustment algorithm SHALL NOT be provided with the exact capacity value to be validated a priori. This restriction fosters a fair result, and removes an opportunity for bad actors to operate with knowledge of the "right answer".

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 2013-2019 time frame, 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 transfer rate and latency are important to user's satisfaction.
3. UDP's growing role in Transport, in areas where TCP once dominated.

Morton, et al.

Expires October 2, 2021

[Page 5]

Internet-Draft

IP Capacity Metrics/Methods

March 2021

4. Content and applications are moving physically closer to users.
5. There is less emphasis on ISP gateway measurements, possibly due to less traffic crossing ISP gateways in future.

4. General Parameters and Definitions

This section lists the REQUIRED input factors to specify a Sender or Receiver 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 MaxHops, 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 (implemented in the TTL or Hop Limit).
- o T0, the time at the start of measurement interval, when packets are first transmitted from the Source.
- o I, the nominal duration of a measurement interval at the destination (default 10 sec)
- o dt, the nominal duration of m equal sub-intervals in I at the destination (default 1 sec)
- o dtn, the beginning boundary of a specific sub-interval, n, one of m sub-intervals in I
- o FT, the feedback time interval between status feedback messages communicating measurement results, sent from the receiver to control the sender. The results are evaluated throughout the test to determine how to adjust the current offered load rate at the sender (default 50ms)
- 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 (default 1 flow)
- 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.

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

packets specified below. Type-P is a parallel concept to "population of interest" defined in clause 6.1.1 of[Y.1540].

- 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).

A non-Parameter which is required for several metrics is defined below:

- o T, the host time of the *first* test packet's *arrival* as measured at the destination Measurement Point, or MP(Dst). There may be other packets sent between source and destination hosts that are excluded, so this is the time of arrival of the first packet used for measurement of the metric.

Note that time stamp format and resolution, sequence numbers, etc. will be established by the chosen test protocol standard or implementation.

[5.](#) IP-Layer Capacity Singleton Metric Definitions

This section sets requirements for the singleton metric that supports the Maximum IP-layer Capacity Metric definition in [Section 6](#).

[5.1.](#) Formal Name

Type-P-One-way-IP-Capacity, or informally called IP-layer Capacity.

Note that Type-P depends on the chosen method.

[5.2.](#) Parameters

This section lists the REQUIRED input factors to specify the metric, beyond those listed in [Section 4](#).

No additional Parameters are needed.

[5.3.](#) Metric Definitions

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

Define the IP-layer capacity, $C(T,dt,PM)$, to be the number of IP-layer bits (including header and data fields) in packets that can be transmitted from the Src host and correctly received by the Dst host during one contiguous sub-interval, dt in length. The IP-layer capacity depends on the Src and Dst hosts, the host addresses, and the path between the hosts.

The number of these IP-layer bits is designated $n0[dt_n,dt_{n+1}]$ for a specific dt .

When the packet size is known and of fixed size, the packet count during a single sub-interval dt multiplied by the total bits in IP header and data fields is equal to $n0[dt_n,dt_{n+1}]$.

Anticipating a Sample of Singletons, the number of sub-intervals with duration dt MUST be set to a natural number m , so that $T+I = T + m*dt$ with $dt_{n+1} - dt_n = dt$ for $1 \leq n \leq m$.

Parameter PM represents other performance metrics [see [section 5.4](#) below]; their measurement results SHALL be collected during measurement of IP-layer Capacity and associated with the corresponding dt_n for further evaluation and reporting. Users SHALL specify the parameter Tmax as required by each metric's reference definition.

Mathematically, this definition is represented as (for each n):

$$C(T,dt,PM) = \frac{(n0[dt_n,dt_{n+1}])}{dt}$$

Equation for IP-Layer Capacity

and:

- o $n0$ is the total number of IP-layer header and payload bits that can be transmitted in standard-formed packets [[RFC8468](#)] 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 $C(T, dt, PM)$ the IP-Layer Capacity, corresponds to the value of n_0 measured in any sub-interval beginning at dt_n , divided by the length of sub-interval, dt .
- o PM represents other performance metrics [see [section 5.4](#) below]; their measurement results SHALL be collected during measurement of IP-layer Capacity and associated with the corresponding dt_n for further evaluation and reporting.
- o all sub-intervals MUST be of equal duration. Choosing dt as non-overlapping consecutive time intervals allows for a simple implementation.
- o The bit rate of the physical interface of the measurement devices must be higher than the smallest of the links on the path whose $C(T, I, PM)$ is to be measured (the bottleneck link).

Measurements according to these definitions SHALL use the UDP transport layer. Standard-formed packets are specified in [Section 5 of \[RFC8468\]](#). The measurement SHOULD use a randomized source port or equivalent technique, and SHOULD send responses from the source address matching the test packet destination address.

Some compression affects on measurement are discussed in [Section 6 of \[RFC8468\]](#).

[5.4](#). Related Round-Trip Delay and One-way Loss Definitions

$RTD[dt_n, dt_n+1]$ is defined as a sample of the [\[RFC2681\]](#) Round-trip Delay between the Src host and the Dst host over the interval $[T, T+I]$ (that contains equal non-overlapping intervals of dt). The "reasonable period of time" in [\[RFC2681\]](#) is the parameter T_{max} in this memo. The statistics used to summarize $RTD[dt_n, dt_n+1]$ MAY include the minimum, maximum, median, and mean, and the range = (maximum - minimum) is referred to below in [Section 8.1](#) for load adjustment purposes.

$OWL[dt_n, dt_n+1]$ is defined as a sample of the [\[RFC7680\]](#) One-way Loss between the Src host and the Dst host over the interval $[T, T+I]$ (that contains equal non-overlapping intervals of dt). The statistics used to summarize $OWL[dt_n, dt_n+1]$ MAY include the lost packet count and the lost packet ratio.

Other metrics MAY be measured: one-way reordering, duplication, and delay variation.

[5.5.](#) Discussion

See the corresponding section for Maximum IP-Layer Capacity.

[5.6.](#) Reporting the Metric

The IP-Layer Capacity SHOULD be reported with at least single Megabit resolution, in units of Megabits per second (Mbps), (which is 1,000,000 bits per second to avoid any confusion).

The Related Round Trip Delay and/or Loss metric measurements for the same Singleton SHALL be reported, also with meaningful resolution for the values measured.

Individual Capacity measurements MAY be reported in a manner consistent with the Maximum IP-Layer Capacity, see [Section 9](#).

[6.](#) Maximum IP-Layer Capacity Metric Definitions (Statistic)

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

[6.1.](#) Formal Name

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

Note that Type-P depends on the chosen method.

[6.2.](#) Parameters

This section lists the REQUIRED input factors to specify the metric, beyond those listed in [Section 4](#).

No additional Parameters or definitions are needed.

[6.3.](#) Metric Definitions

This section defines the REQUIRED aspects of the Maximum IP-layer

Capacity metric (unless otherwise indicated) for measurements between specified Source and Destination hosts:

Define the Maximum IP-layer capacity, $\text{Maximum_C}(T,I,PM)$, to be the maximum number of IP-layer bits $n0[dt_n, dt_{n+1}]$ divided by dt that can be transmitted in packets from the Src host and correctly received by the Dst host, over all dt length intervals in $[T, T+I]$, and meeting the PM criteria. Equivalently the Maximum of a Sample of size m of

Morton, et al.

Expires October 2, 2021

[Page 10]

Internet-Draft

IP Capacity Metrics/Methods

March 2021

$C(T,I,PM)$ collected during the interval $[T, T+I]$ and meeting the PM criteria.

The number of sub-intervals with duration dt MUST be set to a natural number m , so that $T+I = T + m*dt$ with $dt_{n+1} - dt_n = dt$ for $1 \leq n \leq m$.

Parameter PM represents the other performance metrics (see [Section 6.4](#) below) and their measurement results for the maximum IP-layer capacity. At least one target performance threshold (PM criterion) MUST be defined. If more than one metric and target performance threshold are defined, then the sub-interval with maximum number of bits transmitted MUST meet all the target performance thresholds. Users SHALL specify the parameter Tmax as required by each metric's reference definition.

Mathematically, this definition can be represented as:

$$\text{Maximum_C}(T,I,PM) = \frac{\max_{[T,T+I]} (n0[dt_n, dt_{n+1}])}{dt}$$

where:

T $T+I$
 $\text{dt}_n=1 \quad 2 \quad 3 \quad 4 \quad 5 \quad 6 \quad 7 \quad 8 \quad 9 \quad 10 \quad n+1$
 $n=m$

Equation for Maximum Capacity

and:

- o n_0 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 $\text{Maximum_C}(T, I, PM)$ the Maximum IP-Layer Capacity, corresponds to the maximum value of n_0 measured in any sub-interval beginning at dtn , divided by the constant length of all sub-intervals, dt .
- o PM represents the other performance metrics (see [Section 5.4](#)) and their measurement results for the maximum IP-layer capacity. At least one target performance threshold (PM criterion) MUST be defined.

- o all sub-intervals MUST be of equal duration. Choosing dt as non-overlapping consecutive time intervals allows for a simple implementation.
- o The bit rate of the physical interface of the measurement systems must be higher than the smallest of the links on the path whose $\text{Maximum_C}(T, I, PM)$ is to be measured (the bottleneck link).

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 n_0 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 n_0 and metric(s) in the PM list over all dt -length intervals in I .

Measurements according to these definitions SHALL use the UDP transport layer.

[6.4](#). Related Round-Trip Delay and One-way Loss Definitions

$\text{RTD}[dtn, dtn+1]$ and $\text{OWL}[dtn, dtn+1]$ are defined in [Section 5.4](#). Here, the test intervals are increased to match the capacity samples, $\text{RTD}[T, I]$ and $\text{OWL}[T, I]$.

The interval dt_n, dt_{n+1} where $Maximum_C(T, I, PM)$ occurs is the reporting sub-interval within $RTD[T, I]$ and $OWL[T, I]$.

Other metrics MAY be measured: one-way reordering, duplication, and delay variation.

[6.5.](#) Discussion

If traffic conditioning (e.g., shaping, policing) 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 $[T, T+I]$. Each duration dt SHOULD be chosen so that it is an integer multiple of increasing values k times serialization 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.

A $Maximum_C(T, I, PM)$ without any indication of bottleneck congestion, be that an increasing latency, packet loss or ECN marks during a measurement interval I , is likely to underestimate $Maximum_C(T, I, PM)$.

[6.6.](#) Reporting the Metric

The IP-Layer Capacity SHOULD be reported with at least single Megabit resolution, in units of Megabits per second (Mbps) (which is 1,000,000 bits per second to avoid any confusion).

The Related Round Trip Delay and/or Loss metric measurements for the same Singleton SHALL be reported, also with meaningful resolution for the values measured.

When there are demonstrated and repeatable Capacity modes in the Sample, then the Maximum IP-Layer Capacity SHALL be reported for each mode, along with the relative time from the beginning of the stream that the mode was observed to be present. Bimodal Maxima have been observed with some services, sometimes called a "turbo mode" intending to deliver short transfers more quickly, or reduce the initial buffering time for some video streams. Note that modes lasting less than dt duration will not be detected.

Some transmission technologies have multiple methods of operation that may be activated when channel conditions degrade or improve, and these transmission methods may determine the Maximum IP-Layer Capacity. Examples include line-of-sight microwave modulator constellations, or cellular modem technologies where the changes may be initiated by a user moving from one coverage area to another. Operation in the different transmission methods may be observed over time, but the modes of Maximum IP-Layer Capacity will not be activated deterministically as with the "turbo mode" described in the paragraph above.

[7.](#) IP-Layer Sender Bit Rate Singleton Metric Definitions

This section sets requirements for the following components to support the IP-layer Sender Bitrate Metric. This metric helps to check that the sender actually generated the desired rates during a test, and measurement takes place at the Src host to network path interface (or as close as practical within the Src host). It is not a metric for path performance.

[7.1.](#) Formal Name

Type-P-IP-Sender-Bit-Rate, or informally called IP-layer Sender Bitrate.

Note that Type-P depends on the chosen method.

[7.2.](#) Parameters

This section lists the REQUIRED input factors to specify the metric, beyond those listed in [Section 4](#).

- o S, the duration of the measurement interval at the Source
- o st, the nominal duration of N sub-intervals in S (default st = 0.05 seconds)
- o stn, the beginning boundary of a specific sub-interval, n, one of

N sub-intervals in S

S SHALL be longer than I, primarily to account for on-demand activation of the path, or any preamble to testing required, and the delay of the path.

st SHOULD be much smaller than the sub-interval dt and on the same order as FT, otherwise the rate measurement will include many rate adjustments and include more time smoothing, thus missing the maximum rate. The st parameter does not have relevance when the Source is transmitting at a fixed rate throughout S.

[7.3.](#) Metric Definition

This section defines the REQUIRED aspects of the IP-layer Sender Bitrate metric (unless otherwise indicated) for measurements at the specified Source on packets addressed for the intended Destination host and matching the required Type-P:

Define the IP-layer Sender Bit Rate, $B(S, st)$, to be the number of IP-layer bits (including header and data fields) that are transmitted from the Source with address pair Src and Dst during one contiguous sub-interval, st, during the test interval S (where S SHALL be longer than I), and where the fixed-size packet count during that single sub-interval st also provides the number of IP-layer bits in any interval, $[st_n, st_{n+1}]$.

Measurements according to these definitions SHALL use the UDP transport layer. Any feedback from Dst host to Src host received by Src host during an interval $[st_n, st_{n+1}]$ SHOULD NOT result in an adaptation of the Src host traffic conditioning during this interval (rate adjustment occurs on st interval boundaries).

[7.4.](#) Discussion

Both the Sender and Receiver or (source and destination) bit rates SHOULD be assessed as part of an IP-layer Capacity measurement.

Otherwise, an unexpected sending rate limitation could produce an erroneous Maximum IP-Layer Capacity measurement.

[7.5.](#) Reporting the Metric

The IP-Layer Sender Bit Rate SHALL be reported with meaningful resolution, in units of Megabits per second (which is 1,000,000 bits per second to avoid any confusion).

Individual IP-Layer Sender Bit Rate measurements are discussed further in [Section 9](#).

[8.](#) Method of Measurement

The architecture of the method REQUIRES two cooperating hosts operating in the roles of Src (test packet sender) and Dst (receiver), with a measured path and return path between them.

The duration of a test, parameter 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.

[8.1.](#) Load Rate Adjustment Algorithm

The algorithm described in this section MUST NOT be used as a as a general Congestion Control Algorithm (CCA). As stated in the Scope [Section 2](#), the load rate adjustment algorithm's goal is to help determine the Maximum IP-Layer Capacity in the context of an infrequent, diagnostic, short term measurement. There is a tradeoff between test duration (also the test data volume) and algorithm aggressiveness (speed of ramp-up and down to the Maximum Capacity). The parameter values chosen below strike a well-tested balance among these factors.

A table SHALL be pre-built defining all the offered load rates that will be supported (R1 through Rn, in ascending order, corresponding to indexed rows in the table). It is RECOMMENDED that rates begin with 0.5 Mbps at index zero, use 1 Mbps at index one, and then continue in 1 Mbps increments to 1 Gbps. Above 1 Gbps, and up to 10 Gbps, it is RECOMMENDED that 100 Mbps increments be used. Above 10 Gbps, increments of 1 Gbps are RECOMMENDED. A higher starting rate might be configured when the test operator is certain that the Maximum is well-above the starting rate and factors such as test duration and total test traffic play an important role.

Each rate is defined as datagrams of size ss , sent as a burst of count cc , each time interval tt (default for tt is 1ms, a likely system tick-interval). 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. Selection of a new rate is indicated by a calculation on the current row, Rx . For example:

" $Rx+1$ ": the sender uses the next higher rate in the table.

" $Rx-10$ ": the sender uses the rate 10 rows lower in the table.

At the beginning of a test, the sender begins sending at rate $R1$ and the receiver starts a feedback timer of duration FT (while awaiting inbound datagrams). As datagrams are received they are checked for sequence number anomalies (loss, out-of-order, duplication, etc.) and the delay range is measured (one-way or round-trip). This information is accumulated until the feedback timer FT 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 no sequence number anomalies were detected AND the delay range 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 ($Rx+1$). However, if a rate threshold between high and very high sending rates (such as 1Gbps) is exceeded, the offered load rate is only increased by one ($Rx+1$) above the rate threshold in any congestion state.

If the feedback indicates that sequence number anomalies were detected OR the delay range was above the upper threshold, the offered load rate is decreased. The RECOMMENDED values are 0 for sequence number gaps and 30-90 ms for lower and upper delay thresholds, respectively. Also, if congestion is now confirmed for the first time by the current feedback message being processed, then the offered load rate is decreased by more than one rate (e.g., $Rx-30$). This one-time reduction is intended to compensate for the fast initial ramp-up. In all other cases, 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 range 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 inferring congestion is that there were sequence number anomalies AND/OR the delay range was above the upper threshold for two consecutive feedback intervals. The algorithm described above is also illustrated in ITU-T Rec. Y.1540, 2020 version[Y.1540], in Annex B, and implemented in the Appendix on Load Rate Adjustment Pseudo Code in this memo.

The Load Rate Adjustment Algorithm MUST include timers that stop the test when received packet streams cease unexpectedly. The timeout thresholds are provided in the table below, along with values for all other parameters and variables described in this section. Operation of non-obvious parameters appear below:

load packet timeout Operation: The load packet timeout SHALL be reset to the configured value each time a load packet received. If the timeout expires, the receiver SHALL be closed and no further feedback sent.

feedback message timeout Operation: The feedback message timeout SHALL be reset to the configured value each time a feedback message is received. If the timeout expires, the sender SHALL be closed and no further Load packets sent.

Parameter	Default	Tested Range or values	Expected Safe Range (not entirely tested, other values NOT RECOMMENDED)
FT, feedback time interval	50ms	20ms, 50ms, 100ms	20ms <= FT <= 250ms Larger values may slow the rate increase and fail to

			find the max
Feedback message timeout (stop test)	L*FT, L=20 (1sec with FT=50ms)	L=100 with FT=50ms (5sec)	0.5sec <= L*FT <= 30sec Upper limit for very unreliable test paths only
load packet	1sec	5sec	0.250sec - 30sec

Morton, et al.

Expires October 2, 2021

[Page 17]

Internet-Draft

IP Capacity Metrics/Methods

March 2021

timeout (stop test)			Upper limit for very unreliable test paths only
table index 0	0.5Mbps	0.5Mbps	when testing <=10Gbps
table index 1	1Mbps	1Mbps	when testing <=10Gbps
table index (step) size	1Mbps	1Mbps<=rate<= 1Gbps	same as tested
table index (step) size, rate>1Gbps	100Mbps	1Gbps<=rate<= 10Gbps	same as tested
table index (step) size, rate>10Gbps	1Gbps	untested	>10Gbps
ss, UDP payload size, bytes	none	<=1222	Recommend max at largest value that avoids fragmentation; use of too-small payload size might result in unexpected sender limitations.
cc, burst count	none	1<=cc<= 100	same as tested. Vary cc as needed to

			create the desired maximum sending rate. Sender buffer size may limit cc in implementation.
tt, burst interval	100microsec	100microsec, 1msec	available range of "tick" values (HZ param)
low delay range threshold	30ms	5ms, 30ms	same as tested
high delay	90ms	10ms, 90ms	same as tested

Morton, et al.

Expires October 2, 2021

[Page 18]

Internet-Draft

IP Capacity Metrics/Methods

March 2021

range threshold			
sequence error threshold	0	0, 100	same as tested
consecutive errored status report threshold	2	2	Use values >1 to avoid misinterpreting transient loss
Fast mode increase, in table index steps	10	10	$2 \leq \text{steps} \leq 30$
Fast mode decrease, in table index steps	3 * Fast mode increase	3 * Fast mode increase	same as tested

Parameters for Load Rate Adjustment Algorithm

As a consequence of default parameterization, the Number of table steps in total for rates <10Gbps is 2000 (excluding index 0).

A related sender backoff response to network conditions occurs when one or more status feedback messages fail to arrive at the sender.

If no status feedback messages arrive at the sender for the interval greater than the Lost Status Backoff timeout:

$$\text{HDRT} + (2+w)*\text{FT} = \text{Lost Status Backoff timeout}$$

where:

HDRT = high delay range threshold (default 90ms)

FT = feedback time interval (default 50ms)

w = number of repeated timeouts (w=0 initially, w++ on each timeout, and reset to 0 when a message is received)

beginning when the last message (of any type) was successfully received at the sender:

Then the offered load SHALL be decreased, following the same process as when the feedback indicates presence of one or more sequence

number anomalies OR the delay range was above the upper threshold (as described above), with the same Load Rate Adjustment algorithm variables in their current state. This means that rate reduction and congestion confirmation can result from a three-way OR that includes lost status feedback messages, sequence errors, or delay variation.

The RECOMMENDED initial value for w is 0, taking Round Trip Time (RTT) less than FT into account. A test with RTT longer than FT is a valid reason to increase the initial value of w appropriately. Variable w SHALL be incremented by 1 whenever the Lost Status Backoff timeout is exceeded. So with FT = 50ms and HDRT = 90ms, a status feedback message loss would be declared at 190ms following a successful message, again at 50ms after that (240ms total), and so on.

Also, if congestion is now confirmed for the first time by a Lost Status Backoff timeout, then the offered load rate is decreased by more than one rate (e.g., Rx-30). This one-time reduction is intended to compensate for the fast initial ramp-up. In all other

cases, the offered load rate is only decreased by one (Rx-1).

[Appendix B](#) discusses compliance with the applicable mandatory requirements of [\[RFC8085\]](#), consistent with the goals of the IP-Layer Capacity Metric and Method, including the load rate adjustment algorithm described in this section.

[8.2.](#) Measurement Qualification or Verification

It is of course necessary to calibrate the equipment performing the IP-layer Capacity measurement, to ensure that the expected capacity can be measured accurately, and that equipment choices (processing speed, interface bandwidth, etc.) are suitably matched to the measurement range.

When assessing a Maximum rate as the metric specifies, artificially high (optimistic) values might be measured until some buffer on the path is filled. Other causes include bursts of back-to-back packets with idle intervals delivered by a path, while the measurement interval (dt) is small and aligned with the bursts. The artificial values might result in an un-sustainable Maximum Capacity observed when the method of measurement is searching for the Maximum, and that would not do. This situation is different from the bi-modal service rates (discussed under Reporting), which are characterized by a multi-second duration (much longer than the measured RTT) and repeatable behavior.

There are many ways that the Method of Measurement could handle this false-max issue. The default value for measurement of singletons (dt

= 1 second) has proven to be of practical value during tests of this method, allows the bimodal service rates to be characterized, and it has an obvious alignment with the reporting units (Mbps).

Another approach comes from [Section 24 of RFC 2544](#)[\[RFC2544\]](#) and its discussion of Trial duration, where relatively short trials conducted as part of the search are followed by longer trials to make the final determination. In the production network, measurements of singletons and samples (the terms for trials and tests of Lab Benchmarking) must be limited in duration because they may be service-affecting. But there is sufficient value in repeating a sample with a fixed sending rate determined by the previous search for the Max IP-layer Capacity,

to qualify the result in terms of the other performance metrics measured at the same time.

A qualification measurement for the search result is a subsequent measurement, sending at a fixed 99.x % of the Max IP-layer Capacity for I, or an indefinite period. The same Max Capacity Metric is applied, and the Qualification for the result is a sample without packet loss or a growing minimum delay trend in subsequent singletons (or each dt of the measurement interval, I). Samples exhibiting losses or increasing queue occupation require a repeated search and/or test at reduced fixed sender rate for qualification.

Here, as with any Active Capacity test, the test duration must be kept short. 10 second tests for each direction of transmission are common today. The default measurement interval specified here is $I = 10$ seconds. The combination of a fast and congestion-aware search method and user-network coordination make a unique contribution to production testing. The Max IP Capacity metric and method for assessing performance is very different from classic [\[RFC2544\]](#) Throughput metric and methods : it uses near-real-time load adjustments that are sensitive to loss and delay, similar to other congestion control algorithms used on the Internet every day, along with limited duration. On the other hand, [\[RFC2544\]](#) Throughput measurements can produce sustained overload conditions for extended periods of time. Individual trials in a test governed by a binary search can last 60 seconds for each step, and the final confirmation trial may be even longer. This is very different from "normal" traffic levels, but overload conditions are not a concern in the isolated test environment. The concerns raised in [\[RFC6815\]](#) were that [\[RFC2544\]](#) methods would be let loose on production networks, and instead the authors challenged the standards community to develop metrics and methods like those described in this memo.

[8.3.](#) Measurement Considerations

In general, the wide-spread measurements that this memo encourages will encounter wide-spread behaviors. The bimodal IP Capacity behaviors already discussed in [Section 6.6](#) are good examples.

In general, it is RECOMMENDED to locate test endpoints as close to the intended measured link(s) as practical (this is not always possible for reasons of scale; there is a limit on number of test endpoints coming from many perspectives, management and measurement traffic for example). The testing operator MUST set a value for the MaxHops parameter, based on the expected path length. This parameter can keep measurement traffic from straying too far beyond the intended path.

The path measured may be state-full based on many factors, and the Parameter "Time of day" when a test starts may not be enough information. Repeatable testing may require the time from the beginning of a measured flow, and how the flow is constructed including how much traffic has already been sent on that flow when a state-change is observed, because the state-change may be based on time or bytes sent or both. Both load packets and status feedback messages MUST contain sequence numbers, which helps with measurements based on those packets.

Many different traffic shapers and on-demand access technologies may be encountered, as anticipated in [\[RFC7312\]](#), and play a key role in measurement results. Methods MUST be prepared to provide a short preamble transmission to activate on-demand access, and to discard the preamble from subsequent test results.

Conditions which might be encountered during measurement, where packet losses may occur independently from the measurement sending rate:

1. Congestion of an interconnection or backbone interface may appear as packet losses distributed over time in the test stream, due to much higher rate interfaces in the backbone.
2. Packet loss due to use of Random Early Detection (RED) or other active queue management may or may not affect the measurement flow if competing background traffic (other flows) are simultaneously present.
3. There may be only small delay variation independent of sending rate under these conditions, too.

4. Persistent competing traffic on measurement paths that include shared transmission media may cause random packet losses in the test stream.

It is possible to mitigate these conditions using the flexibility of the load-rate adjusting algorithm described in [Section 8.1](#) above (tuning specific parameters).

If the measurement flow burst duration happens to be on the order of or smaller than the burst size of a shaper or a policer in the path, then the line rate might be measured rather than the bandwidth limit imposed by the shaper or policer. If this condition is suspected, alternate configurations SHOULD be used.

In general, results depend on the sending stream characteristics; the measurement community has known this for a long time, and needs to keep it front of mind. Although the default is a single flow ($F=1$) for testing, use of multiple flows may be advantageous for the following reasons:

1. the test hosts may be able to create higher load than with a single flow, or parallel test hosts may be used to generate 1 flow each.
2. there may be link aggregation present (flow-based load balancing) and multiple flows are needed to occupy each member of the aggregate.
3. access policies may limit the IP-Layer Capacity depending on the Type-P of packets, possibly reserving capacity for various stream types.

Each flow would be controlled using its own implementation of the Load Adjustment (Search) Algorithm.

As testing continues, implementers should expect some evolution in the methods. The ITU-T has published a Supplement (60) to the Y-series of Recommendations, "Interpreting ITU-T Y.1540 maximum IP-layer capacity measurements", [[Y.Sup60](#)], which is the result of continued testing with the metric, and those results have improved the method described here.

[8.4](#). Running Code

This section is for the benefit of the Document Shepherd's form, and will be deleted prior to final review.

Much of the development of the method and comparisons with existing methods conducted at IETF Hackathons and elsewhere have been based on the example udpst Linux measurement tool (which is a working reference for further development) [[udpst](#)]. The current project:

- o is a utility that can function as a client or server daemon
- o requires a successful client-initiated setup handshake between cooperating hosts and allows firewalls to control inbound unsolicited UDP which either go to a control port [expected and w/ authentication] or to ephemeral ports that are only created as needed. Firewalls protecting each host can both continue to do their job normally. This aspect is similar to many other test utilities available.
- o is written in C, and built with gcc (release 9.3) and its standard run-time libraries
- o allows configuration of most of the parameters described in Sections [4](#) and [7](#).
- o supports IPv4 and IPv6 address families.
- o supports IP-layer packet marking.

[9](#). Reporting Formats

The singleton IP-Layer Capacity results SHOULD be accompanied by the context under which they were measured.

- o timestamp (especially the time when the maximum was observed in dtn)
- o source and destination (by IP or other meaningful ID)
- o other inner parameters of the test case ([Section 4](#))
- o outer parameters, such as "test conducted in motion" or other factors belonging to the context of the measurement
- o result validity (indicating cases where the process was somehow interrupted or the attempt failed)

- o a field where unusual circumstances could be documented, and another one for "ignore/mask out" purposes in further processing

The Maximum IP-Layer Capacity results SHOULD be reported in the format of a table with a row for each of the test Phases and Number

of Flows. There SHOULD be columns for the phases with number of flows, and for the resultant Maximum IP-Layer Capacity results for the aggregate and each flow tested.

As mentioned in [Section 6.6](#), bi-modal (or multi-modal) maxima SHALL be reported for each mode separately.

Phase, # Flows	Max IP-Layer Capacity, Mbps	Loss Ratio	RTT min, max, msec
Search,1	967.31	0.0002	30, 58
Verify,1	966.00	0.0000	30, 38

Maximum IP-layer Capacity Results

Static and configuration parameters:

The sub-interval time, dt, MUST accompany a report of Maximum IP-Layer Capacity results, and the remaining Parameters from [Section 4](#), General Parameters.

The PM list metrics corresponding to the sub-interval where the Maximum Capacity occurred MUST accompany a report of Maximum IP-Layer Capacity results, for each test phase.

The IP-Layer Sender Bit rate results SHOULD be reported in the format of a table with a row for each of the test Phases, sub-intervals (st) and Number of Flows. There SHOULD be columns for the phases with number of flows, and for the resultant IP-Layer Sender Bit rate results for the aggregate and each flow tested.

Phase, Flow or Aggregate	st, sec	Sender Bit Rate, Mbps
Search,1	0.00 - 0.05	345
Search,2	0.00 - 0.05	289
Search,Agg	0.00 - 0.05	634

IP-layer Sender Bit Rate Results

Static and configuration parameters:

Morton, et al.

Expires October 2, 2021

[Page 25]

Internet-Draft

IP Capacity Metrics/Methods

March 2021

The subinterval time, st, MUST accompany a report of Sender IP-Layer Bit Rate results.

Also, the values of the remaining Parameters from [Section 4](#), General Parameters, MUST be reported.

[9.1](#). Configuration and Reporting Data Formats

As a part of the multi-Standards Development Organization (SDO) harmonization of this metric and method of measurement, one of the areas where the Broadband Forum (BBF) contributed its expertise was in the definition of an information model and data model for configuration and reporting. These models are consistent with the metric parameters and default values specified as lists in this memo. [\[TR-471\]](#) provides the Information model that was used to prepare a full data model in related BBF work. The BBF has also carefully considered topics within its purview, such as placement of measurement systems within the access architecture. For example, timestamp resolution requirements that influence the choice of the test protocol are provided in Table 2 of [\[TR-471\]](#).

[10](#). Security Considerations

Active metrics and measurements have a long history of security considerations. The security considerations that apply to any active measurement of live paths are relevant here. See [\[RFC4656\]](#) and [\[RFC5357\]](#).

When considering privacy of those involved in measurement or those whose traffic is measured, the sensitive information available to potential observers is greatly reduced when using active techniques which are within this scope of work. Passive observations of user traffic for measurement purposes raise many privacy issues. We refer the reader to the privacy considerations described in the Large Scale Measurement of Broadband Performance (LMAP) Framework [[RFC7594](#)], which covers active and passive techniques.

There are some new considerations for Capacity measurement as described in this memo.

1. Cooperating source and destination hosts and agreements to test the path between the hosts are REQUIRED. Hosts perform in either the Src or Dst roles.
2. It is REQUIRED to have a user client-initiated setup handshake between cooperating hosts that allows firewalls to control inbound unsolicited UDP traffic which either goes to a control port [expected and w/authentication] or to ephemeral ports that

are only created as needed. Firewalls protecting each host can both continue to do their job normally.

3. Client-server authentication and integrity protection for feedback messages conveying measurements is RECOMMENDED.
4. Hosts MUST limit the number of simultaneous tests to avoid resource exhaustion and inaccurate results.
5. Senders MUST be rate-limited. This can be accomplished using a pre-built table defining all the offered load rates that will be supported ([Section 8.1](#)). The recommended load-control search algorithm results in "ramp up" from the lowest rate in the table.
6. Service subscribers with limited data volumes who conduct extensive capacity testing might experience the effects of Service Provider controls on their service. Testing with the Service Provider's measurement hosts SHOULD be limited in frequency and/or overall volume of test traffic (for example, the range of I duration values SHOULD be limited).

The exact specification of these features is left for the future protocol development.

[11.](#) IANA Considerations

This memo makes no requests of IANA.

[12.](#) Acknowledgments

Thanks to Joachim Fabini, Matt Mathis, J. Ignacio Alvarez-Hamelin, Wolfgang Balzer, Frank Brockners, Greg Mirsky, Martin Duke, Murray Kucherawy, and Benjamin Kaduk for their extensive comments on the memo and related topics.

[13.](#) [Appendix A](#) - Load Rate Adjustment Pseudo Code

The following is a pseudo-code implementation of the algorithm described in [Section 8.1](#).

```
Rx = 0 # The current sending rate (equivalent to a row of the table)
seqErr = 0 # Measured count of any of Loss or Reordering impairments
delay = 0 # Measured Range of Round Trip Time, RTT, ms
lowThresh = 30 # Low threshold on the Range of RTT, ms
upperThresh = 90 # Upper threshold on the Range of RTT, ms
hSpeedTresh = 1Gbps # Threshold for transition between sending rate step
  sizes (such as 1 Mbps and 100 Mbps)
slowAdjCount = 0 # Measured Number of consecutive status reports
  indicating loss and/or delay variation above upperThresh
slowAdjThresh = 2 # Threshold on slowAdjCount used to infer congestion.
  Use values >1 to avoid misinterpreting transient loss
highSpeedDelta = 10 # The number of rows to move in a single adjustment
  when initially increasing offered load (to ramp-up quickly)
maxLoadRates = 2000 # Maximum table index (rows)
```

```

if ( seqErr == 0 && delay < lowThresh ) {
    if ( Rx < hSpeedTresh && slowAdjCount < slowAdjThresh ) {
        Rx += highSpeedDelta;
        slowAdjCount = 0;
    } else {
        if ( Rx < maxLoadRates - 1 )
            Rx++;
    }
} else if ( seqErr > 0 || delay > upperThresh ) {
    slowAdjCount++;
    if ( Rx < hSpeedTresh && slowAdjCount == slowAdjThresh ) {
        if ( Rx > highSpeedDelta * 3 )
            Rx -= highSpeedDelta * 3;
        else
            Rx = 0;
    } else {
        if ( Rx > 0 )
            Rx--;
    }
}

```

[14. Appendix B - RFC 8085](#) UDP Guidelines Check

The BCP on UDP usage guidelines [[RFC8085](#)] focuses primarily on congestion control in [section 3.1](#). The Guidelines appear in mandatory (MUST) and recommendation (SHOULD) categories.

[14.1](#). Assessment of Mandatory Requirements

The mandatory requirements in [Section 3 of \[RFC8085\]](#) include:

Internet paths can have widely varying characteristics, ... Consequently, applications that may be used on the Internet MUST NOT make assumptions about specific path characteristics. They MUST instead use mechanisms that let them operate safely under very different path conditions. Typically, this requires conservatively probing the current conditions of the Internet path they communicate over to establish a transmission behavior that it

can sustain and that is reasonably fair to other traffic sharing the path.

The purpose of the load rate adjustment algorithm in [Section 8.1](#) is to probe the network and enable Maximum IP-Layer Capacity measurements with as few assumptions about the measured path as possible, and within the range application described in [Section 2](#). The degree of probing conservatism is in tension with the need to minimize both the traffic dedicated to testing (especially with Gigabit rate measurements) and the duration of the test (which is one contributing factor to the overall algorithm fairness).

The text of [Section 3 of \[RFC8085\]](#) goes on to recommend alternatives to UDP to meet the mandatory requirements, but none are suitable for the scope and purpose of the metrics and methods in this memo. In fact, ad hoc TCP-based methods fail to achieve the measurement accuracy repeatedly proven in comparison measurements with the running code [\[LS-SG12-A\]](#) [\[LS-SG12-B\]](#) [\[Y.Sup60\]](#). Also, the UDP aspect of these methods is present primarily to support modern Internet transmission where a transport protocol is required [\[copycat\]](#); the metric is based on the IP-layer and UDP allows simple correlation to the IP-layer.

[Section 3.1.1 of \[RFC8085\]](#) discusses protocol timer guidelines:

Latency samples MUST NOT be derived from ambiguous transactions. The canonical example is in a protocol that retransmits data, but subsequently cannot determine which copy is being acknowledged.

Both load packets and status feedback messages MUST contain sequence numbers, which helps with measurements based on those packets, and there are no retransmissions needed.

When a latency estimate is used to arm a timer that provides loss detection -- with or without retransmission -- expiry of the timer MUST be interpreted as an indication of congestion in the network, causing the sending rate to be adapted to a safe conservative rate...

The method described in this memo uses timers for sending rate backoff when status feedback messages are lost (Lost Status Backoff

timeout), and for stopping a test when connectivity is lost for an longer interval (Feedback message or load packet timeouts).

There is no specific benefit foreseen by using Explicit Congestion Notification (ECN) in this memo.

[Section 3.2 of \[RFC8085\]](#) discusses message size guidelines:

To determine an appropriate UDP payload size, applications MUST subtract the size of the IP header (which includes any IPv4 optional headers or IPv6 extension headers) as well as the length of the UDP header (8 bytes) from the PMTU size.

The method uses a sending rate table with a maximum UDP payload size that anticipates significant header overhead and avoids fragmentation.

[Section 3.3 of \[RFC8085\]](#) provides reliability guidelines:

Applications that do require reliable message delivery MUST implement an appropriate mechanism themselves.

The IP-Layer Capacity Metric and Method do not require reliable delivery.

Applications that require ordered delivery MUST reestablish datagram ordering themselves.

The IP-Layer Capacity Metric and Method does not need to reestablish packet order; it is preferred to measure packet reordering if it occurs [\[RFC4737\]](#).

[14.2.](#) Assessment of Recommendations

The load rate adjustment algorithm's goal is to determine the Maximum IP-Layer Capacity in the context of an infrequent, diagnostic, short term measurement. This goal is a global exception to many [\[RFC8085\]](#) SHOULD-level requirements, of which many are intended for long-lived flows that must coexist with other traffic in more-or-less fair way. However, the algorithm (as specified in [Section 8.1](#) and [Appendix A](#) above) reacts to indications of congestion in clearly defined ways.

A specific recommendation is provided as an example. [Section 3.1.5 of \[RFC8085\]](#) on implications of RTT and Loss Measurements on Congestion Control says:

A congestion control designed for UDP SHOULD respond as quickly as possible when it experiences congestion, and it SHOULD take into

Internet-Draft

IP Capacity Metrics/Methods

March 2021

account both the loss rate and the response time when choosing a new rate.

The load rate adjustment algorithm responds to loss and RTT measurements with a clear and concise rate reduction when warranted, and the response makes use of direct measurements (more exact than can be inferred from TCP ACKs).

[Section 3.1.5 of \[RFC8085\]](#) goes on to specify:

The implemented congestion control scheme SHOULD result in bandwidth (capacity) use that is comparable to that of TCP within an order of magnitude, so that it does not starve other flows sharing a common bottleneck.

This is a requirement for coexistent streams, and not for diagnostic and infrequent measurements using short durations. The rate oscillations during short tests allow other packets to pass, and don't starve other flows.

Ironically, ad hoc TCP-based measurements of "Internet Speed" are also designed to work around this SHOULD-level requirement, by launching many flows (9, for example) to increase the outstanding data dedicated to testing.

The load rate adjustment algorithm cannot become a TCP-like congestion control, or it will have the same weaknesses of TCP when trying to make a Maximum IP-Layer Capacity measurement, and will not achieve the goal. The results of the referenced testing [[LS-SG12-A](#)] [[LS-SG12-B](#)] [[Y.Sup60](#)] supported this statement hundreds of times, with comparisons to multi-connection TCP-based measurements.

A brief review of some of the other SHOULD-level requirements follows (Yes or Not applicable = NA) :

Y?	RFC 8085 Recommendation	Section
Yes	MUST tolerate a wide range of Internet path conditions	3
NA	SHOULD use a full-featured transport (e.g., TCP)	
Yes	SHOULD control rate of transmission	3.1
NA	SHOULD perform congestion control over all traffic	

	for bulk transfers,	3.1.2
NA	SHOULD consider implementing TFRC	
NA	else, SHOULD in other ways use bandwidth similar to TCP	

Internet-Draft	IP Capacity Metrics/Methods	March 2021
----------------	-----------------------------	------------

	for non-bulk transfers,	3.1.3
NA	SHOULD measure RTT and transmit max. 1 datagram/RTT	3.1.1
NA	else, SHOULD send at most 1 datagram every 3 seconds	
NA	SHOULD back-off retransmission timers following loss	
Yes	SHOULD provide mechanisms to regulate the bursts of transmission	3.1.6
NA	MAY implement ECN; a specific set of application mechanisms are REQUIRED if ECN is used.	3.1.7
Yes	for DiffServ, SHOULD NOT rely on implementation of PHBs	3.1.8
Yes	for QoS-enabled paths, MAY choose not to use CC	3.1.9
Yes	SHOULD NOT rely solely on QoS for their capacity non-CC controlled flows SHOULD implement a transport circuit breaker MAY implement a circuit breaker for other applications	3.1.10
	for tunnels carrying IP traffic,	3.1.11
NA	SHOULD NOT perform congestion control	
NA	MUST correctly process the IP ECN field	
	for non-IP tunnels or rate not determined by traffic,	
NA	SHOULD perform CC or use circuit breaker	3.1.11
NA	SHOULD restrict types of traffic transported by the tunnel	
Yes	SHOULD NOT send datagrams that exceed the PMTU, i.e.,	3.2
Yes	SHOULD discover PMTU or send datagrams < minimum PMTU;	
NA	Specific application mechanisms are REQUIRED if PLPMTUD is used.	
Yes	SHOULD handle datagram loss, duplication, reordering	3.3
NA	SHOULD be robust to delivery delays up to 2 minutes	

Yes	SHOULD enable IPv4 UDP checksum	3.4
Yes	SHOULD enable IPv6 UDP checksum; Specific application mechanisms are REQUIRED if a zero IPv6 UDP checksum is used.	3.4.1
NA	SHOULD provide protection from off-path attacks else, MAY use UDP-Lite with suitable checksum coverage	5.1 3.4.2
NA	SHOULD NOT always send middlebox keep-alive messages	3.5
NA	MAY use keep-alives when needed (min. interval 15 sec)	

Internet-Draft	IP Capacity Metrics/Methods	March 2021
Yes	Applications specified for use in limited use (or controlled environments) SHOULD identify equivalent mechanisms and describe their use case.	3.6
NA	Bulk-multicast apps SHOULD implement congestion control	4.1.1
NA	Low volume multicast apps SHOULD implement congestion control	4.1.2
NA	Multicast apps SHOULD use a safe PMTU	4.2
Yes	SHOULD avoid using multiple ports	5.1.2
Yes	MUST check received IP source address	
NA	SHOULD validate payload in ICMP messages	5.2
Yes	SHOULD use a randomized source port or equivalent technique, and, for client/server applications, SHOULD send responses from source address matching request	6
NA	SHOULD use standard IETF security protocols when needed	6

[15.](#) References

[15.1.](#) 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>>.

- [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>>.
- [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>>.
- [RFC4656] Shalunov, S., Teitelbaum, B., Karp, A., Boote, J., and M. Zekauskas, "A One-way Active Measurement Protocol (OWAMP)", [RFC 4656](#), DOI 10.17487/RFC4656, September 2006, <<https://www.rfc-editor.org/info/rfc4656>>.

Morton, et al.

Expires October 2, 2021

[Page 33]

Internet-Draft

IP Capacity Metrics/Methods

March 2021

- [RFC4737] Morton, A., Ciavattone, L., Ramachandran, G., Shalunov, S., and J. Perser, "Packet Reordering Metrics", [RFC 4737](#), DOI 10.17487/RFC4737, November 2006, <<https://www.rfc-editor.org/info/rfc4737>>.
- [RFC5357] Hedayat, K., Krzanowski, R., Morton, A., Yum, K., and J. Babiarz, "A Two-Way Active Measurement Protocol (TWAMP)", [RFC 5357](#), DOI 10.17487/RFC5357, October 2008, <<https://www.rfc-editor.org/info/rfc5357>>.
- [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>>.
- [RFC7479] Moonesamy, S., "Using Ed25519 in SSHFP Resource Records", [RFC 7479](#), DOI 10.17487/RFC7479, March 2015, <<https://www.rfc-editor.org/info/rfc7479>>.
- [RFC7680] Almes, G., Kalidindi, S., Zekauskas, M., and A. Morton, Ed., "A One-Way Loss Metric for IP Performance Metrics (IPPM)", STD 82, [RFC 7680](#), DOI 10.17487/RFC7680, January

2016, <<https://www.rfc-editor.org/info/rfc7680>>.

- [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>>.
- [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>>.

15.2. 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>>.
- [LS-SG12-A] 12, I. S., "LS - Harmonization of IP Capacity and Latency Parameters: Revision of Draft Rec. Y.1540 on IP packet transfer performance parameters and New Annex A with Lab Evaluation Plan", May 2019, <<https://datatracker.ietf.org/liaison/1632/>>.

Morton, et al.

Expires October 2, 2021

[Page 34]

Internet-Draft

IP Capacity Metrics/Methods

March 2021

- [LS-SG12-B] 12, I. S., "LS on harmonization of IP Capacity and Latency Parameters: Consent of Draft Rec. Y.1540 on IP packet transfer performance parameters and New Annex A with Lab & Field Evaluation Plans", March 2019, <<https://datatracker.ietf.org/liaison/1645/>>.
- [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>>.
- [RFC3148] Mathis, M. and M. Allman, "A Framework for Defining Empirical Bulk Transfer Capacity Metrics", [RFC 3148](#), DOI 10.17487/RFC3148, July 2001,

<<https://www.rfc-editor.org/info/rfc3148>>.

- [RFC5136] Chimento, P. and J. Ishac, "Defining Network Capacity", [RFC 5136](#), DOI 10.17487/RFC5136, February 2008, <<https://www.rfc-editor.org/info/rfc5136>>.
- [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>>.
- [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>>.
- [RFC7594] Eardley, P., Morton, A., Bagnulo, M., Burbridge, T., Aitken, P., and A. Akhter, "A Framework for Large-Scale Measurement of Broadband Performance (LMAP)", [RFC 7594](#), DOI 10.17487/RFC7594, September 2015, <<https://www.rfc-editor.org/info/rfc7594>>.
- [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>>.
- [RFC8085] Eggert, L., Fairhurst, G., and G. Shepherd, "UDP Usage Guidelines", [BCP 145](#), [RFC 8085](#), DOI 10.17487/RFC8085, March 2017, <<https://www.rfc-editor.org/info/rfc8085>>.

- [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>>.
- [TR-471] Morton, A., "Broadband Forum TR-471: IP Layer Capacity Metrics and Measurement", July 2020, <<https://www.broadband-forum.org/technical/download/TR-471.pdf>>.

- [udpst] udpst Project Collaborators, "UDP Speed Test Open Broadband project", December 2020,
<<https://github.com/BroadbandForum/obudpst>>.
- [Y.1540] Y.1540, I. R., "Internet protocol data communication service – IP packet transfer and availability performance parameters", December 2019,
<<https://www.itu.int/rec/T-REC-Y.1540-201912-I/en>>.
- [Y.Sup60] Morton, A., "Recommendation Y.Sup60, (09/20) Interpreting ITU-T Y.1540 maximum IP-layer capacity measurements, and Errata", September 2020,
<<https://www.itu.int/rec/T-REC-Y.Sup60/en>>.

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

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

Email: lencia@att.com