Stream: Internet Engineering Task Force (IETF)
RFC: 9097
Category: Standards Track
Published: November 2021
ISSN: 2070-1721
Authors: A. Morton, *AT&T Labs*; R. Geib, *Deutsche Telekom*; L. Ciavattone, *AT&T Labs*

RFC 9097
Metrics and Methods for One-Way IP Capacity

Abstract

This memo revisits the problem of Network Capacity Metrics first examined in RFC 5136. This memo specifies a more practical Maximum IP-Layer Capacity Metric definition catering to measurement and outlines the corresponding Methods of Measurement.

Status of This Memo

This is an Internet Standards Track document.

This document is a product of the Internet Engineering Task Force (IETF). It represents the consensus of the IETF community. It has received public review and has been approved for publication by the Internet Engineering Steering Group (IESG). Further information on Internet Standards is available in Section 2 of RFC 7841.

Information about the current status of this document, any errata, and how to provide feedback on it may be obtained at https://www.rfc-editor.org/info/rfc9097.

Copyright Notice

Table of Contents

1. Introduction

The IETF's efforts to define Network Capacity and Bulk Transport Capacity (BTC) have been chartered and progressed for over twenty years. Over that time, the performance community has seen the development of Informative definitions in [RFC3148] for the Framework for Bulk Transport Capacity, [RFC5136] for Network Capacity and Maximum IP-Layer Capacity, and the Experimental metric definitions and methods in "Model-Based Metrics for Bulk Transport Capacity" [RFC8337].

This memo revisits the problem of Network Capacity Metrics examined first in [RFC3148] and later in [RFC5136]. Maximum IP-Layer Capacity and Bulk Transfer Capacity [RFC3148] (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 one or more networks where the subscriber receives specific performance assurances, sometimes referred to as Internet access, or where a limit of the technology used on a path is being tested. For example, when a user subscribes to a 1 Gbps service, then the user, the Service Provider, and possibly other parties want to assure that the specified performance level is delivered. When a test confirms the subscribed performance level, a tester can seek the location of a bottleneck elsewhere.

This memo recognizes the importance of a definition of a Maximum IP-Layer Capacity Metric at a time when Internet subscription speeds have increased dramatically -- a definition that is both practical and effective for the performance community's needs, including Internet users. The metric definitions are intended to use Active Methods of Measurement [RFC7799], and a Method of Measurement is included for each metric.

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 measurement study to investigate whether UDP is viable as a general Internet transport protocol [copycat], the authors found that a high percentage of paths tested support UDP transport. A number of liaison statements 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 updates to the IP Performance Metrics (IPPM) Framework [RFC2330] that have been published since 1998. In particular, it makes use of [RFC7312] for the Advanced Stream and Sampling Framework and [RFC8468] for its IPv4, IPv6, and IPv4-IPv6 Coexistence Updates.

Appendix A describes the load rate adjustment algorithm, using pseudocode. 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 Active Measurement metrics and corresponding methods to unambiguously determine Maximum IP-Layer Capacity and useful secondary metrics.

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 (SDOs) (through each SDO's normal contribution process or through liaison exchange).

Secondary goals are to add considerations for test procedures and to provide interpretation of the Maximum IP-Layer Capacity results (to identify cases where more testing is warranted, possibly with alternate configurations). 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 Metrics and Methods of Measurement described here is the same as what is described in Section 2 of [RFC7497], where:

> The access portion of the network is the focus of this problem statement. The user typically subscribes to a service with bidirectional [Internet] 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:

- It **MUST** only be used in the application of diagnostic and operations measurements as described in this memo.
- It **MUST** only be used in circumstances consistent with Section 10 ("Security Considerations").
- If a network operator is certain of the IP-Layer Capacity to be validated, then testing **MAY** start with a fixed-rate test at the IP-Layer 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 is **RECOMMENDED**.

Further, the Metrics and Methods 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 increased to provide a surplus to compensate for possible underestimates of TCP-based testing).
- The size of the 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 nefarious operation enabled by knowledge of the correct answer.

3. Motivation

As with any problem that has been worked on 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 began 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 (but subscribers expect network providers to honor contracted performance).
2. Both transfer rate and latency are important to a user's satisfaction.
3. UDP's role in transport is growing in areas where TCP once dominated.
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 the future.

4. General Parameters and Definitions

This section lists the **REQUIRED** input factors to specify a Sender or Receiver metric.

Src: One of the addresses of a host (such as a globally routable IP address).

Dst: One of the addresses of a host (such as a globally routable IP address).

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).

T0: The time at the start of a measurement interval, when packets are first transmitted from the Source.

I: The nominal duration of a measurement interval at the Destination (default 10 sec).

dt: The nominal duration of m equal sub-intervals in I at the Destination (default 1 sec).

dtn: The beginning boundary of a specific sub-interval, n, one of m sub-intervals in I.

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 50 msec).

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.

F: The number of different flows synthesized by the method (default one flow).

Flow: The stream of packets with the same n-tuple of designated header fields that (when held constant) result in identical treatment in a multipath decision (such as the decision taken in load balancing). Note: The IPv6 flow label **SHOULD** be included in the flow definition when routers have complied with the guidelines provided in [RFC6438].

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 concept parallel to "population of interest" as defined in Clause 6.1.1 of [Y.1540].

Payload Content: An aspect of the Type-P Parameter that can help to improve measurement determinism. Specifying packet payload content helps to ensure IPPM Framework-conforming Metrics and Methods. If there is payload compression in the path and tests intend to characterize a possible advantage due to compression, then payload content **SHOULD** be supplied by a pseudorandom sequence generator, by using part of a compressed file, or by other means. See Section 3.1.2 of [RFC7312].

PM: A list of fundamental metrics, such as loss, delay, and reordering, and corresponding target performance threshold(s). 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 that is required for several metrics is defined below:

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 timestamp 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 definitions in Section 6.

5.1. Formal Name

"Type-P-One-way-IP-Capacity" is the formal name; it is 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[dtn,dtn+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[dtn,dtn+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 dtn+1 - dtn = dt for 1 <= n <= 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 dtn 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):

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

Figure 1: Equation for IP-Layer Capacity

and:

- 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].
- C(T,dt,PM), the IP-Layer Capacity, corresponds to the value of n0 measured in any sub-interval beginning at dtn, divided by the length of the sub-interval, dt.
- 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 dtn for further evaluation and reporting.
- All sub-intervals **MUST** be of equal duration. Choosing dt as non-overlapping consecutive time intervals allows for a simple implementation.
- 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 this definition **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 effects of compression on measurement are discussed in Section 6 of [RFC8468].

5.4. Related Round-Trip Delay and One-Way Loss Definitions

RTD[dtn,dtn+1] is defined as a Sample of the Round-Trip Delay [RFC2681] between the Src host and the Dst host during the interval [T,T+I] (that contains equal non-overlapping intervals of dt). The "reasonable period of time" mentioned in [RFC2681] is the Parameter Tmax in this memo. The statistics used to summarize RTD[dtn,dtn+1] **MAY** include the minimum, maximum, median, mean, and the range = (maximum - minimum). Some of these statistics are needed for load adjustment purposes (Section 8.1), measurement qualification (Section 8.2), and reporting (Section 9).

OWL[dtn,dtn+1] is defined as a Sample of the One-Way Loss [RFC7680] between the Src host and the Dst host during the interval [T,T+I] (that contains equal non-overlapping intervals of dt). The statistics used to summarize OWL[dtn,dtn+1] **MAY** include the count of lost packets and the ratio of lost packets.

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

5.5. Discussion

See the corresponding section for Maximum IP-Layer Capacity (Section 6.5).

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, to avoid any confusion, is 1,000,000 bits per second).

The related One-Way Loss metric and Round-Trip Delay 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 (Statistics)

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" is the formal name; it is 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, Maximum_C(T,I,PM), to be the maximum number of IP-Layer bits n0[dtn,dtn+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. An equivalent definition would be the maximum of a Sample of size m of Singletons 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 dtn+1 - dtn = dt for 1 <= n <= 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 is defined, then the sub-interval with the 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:

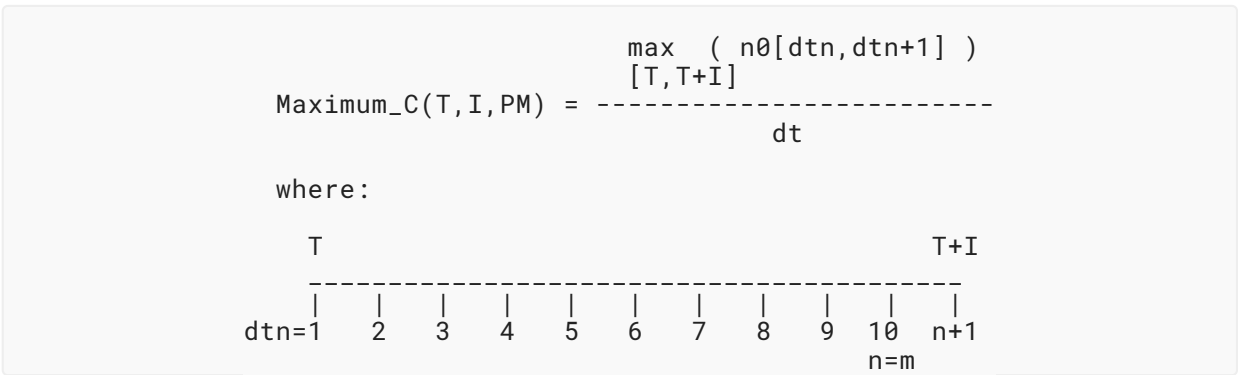

Figure 2: Equation for Maximum Capacity

and:

- 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].
- Maximum_C(T,I,PM), the Maximum IP-Layer Capacity, corresponds to the maximum value of n0 measured in any sub-interval beginning at dtn, divided by the constant length of all sub-intervals, dt.

- PM represents the other performance metrics (see Section 6.4) and their measurement results for the Maximum IP-Layer Capacity. At least one target performance threshold (PM criterion) **MUST** be defined.
- All sub-intervals **MUST** be of equal duration. Choosing dt as non-overlapping consecutive time intervals allows for a simple implementation.
- The bit rate of the physical interface of the measurement systems **MUST** be higher than the smallest of the links on the path whose 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 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 the UDP transport layer.

6.4. Related Round-Trip Delay and One-Way Loss Definitions

RTD[dtn,dtn+1] and OWL[dtn,dtn+1] are defined in Section 5.4. Here, the test intervals are increased to match the capacity Samples, RTD[T,I] and OWL[T,I].

The interval dtn,dtn+1 where Maximum_C(T,I,PM) occurs is the reporting sub-interval for RTD[dtn,dtn+1] and OWL[dtn,dtn+1] 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 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 (PMTU) 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 increased latency, packet loss, or Explicit Congestion Notification (ECN) marks during a measurement interval, I, is likely an underestimate of 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, to avoid any confusion, is 1,000,000 bits per second).

The related One-Way Loss metric and Round-Trip Delay 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, 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 Maximum IP-Layer Capacities 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 duration dt 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 Bit Rate Metric. This metric helps to check that the Sender actually generated the desired rates during a test, and measurement takes place at the interface between the Src host and the network path (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" is the formal name; it is informally called the "IP-Layer Sender Bit Rate".

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.

S: The duration of the measurement interval at the Source.

st: The nominal duration of N sub-intervals in S (default st = 0.05 seconds).

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, possibly smoothing the interval that contains the Maximum IP-Layer Capacity (and therefore losing relevance). 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 Bit Rate 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, [stn,stn+1].

Measurements according to this definition **SHALL** use the UDP transport layer. Any feedback from the Dst host to the Src host received by the Src host during an interval [stn,stn+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, to avoid any confusion, is 1,000,000 bits per second).

Individual IP-Layer Sender Bit Rate measurements are discussed further in Section 9.

8. Method of Measurement

It is **REQUIRED** per the architecture of the method that two cooperating hosts operate 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 path from the Src host to the Dst host during a test.

8.1. Load Rate Adjustment Algorithm

The algorithm described in this section **MUST NOT** be used as a general Congestion Control Algorithm (CCA). As stated in Section 2 ("Scope, Goals, and Applicability"), 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 trade-off between test duration

(also the test data volume) and algorithm aggressiveness (speed of ramp-up and ramp-down to the Maximum IP-Layer Capacity). The Parameter values chosen below strike a well-tested balance among these factors.

A table **SHALL** be pre-built (by the test administrator), 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 initial IP-Layer Sender Bit Rate might be configured when the test operator is certain that the Maximum IP-Layer Capacity is well above the initial IP-Layer Sender Bit Rate and factors such as test duration and total test traffic play an important role. The sending rate table **SHOULD** bracket the Maximum Capacity where it will make measurements, including constrained rates less than 500 kbps if applicable.

Each rate is defined as datagrams of size ss, sent as a burst of count cc, each time interval tt (the default for tt is 100 microsec, 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 (see below for the method for declaring congestion), the offered load rate is increased by more than one rate setting (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 above a high sending rate (such as 1 Gbps) is exceeded, the offered load rate is only increased by one (Rx+1) 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** threshold values are 10 for sequence number gaps and 30 msec for lower and 90 msec for 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 setting (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 for 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 three consecutive feedback intervals. The algorithm described above is also illustrated in Annex B of ITU-T Recommendation Y.1540, 2020 version [Y.1540] and is implemented in Appendix A ("Load Rate Adjustment Pseudocode") 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 Table 1, along with values for all other Parameters and variables described in this section. Operations of non-obvious Parameters appear below:

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

feedback message timeout:
: 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	50 msec	20 msec, 50 msec, 100 msec	20 msec <= FT <= 250 msec; larger values may slow the rate increase and fail to find the max
Feedback message timeout (stop test)	L*FT, L=20 (1 sec with FT=50 msec)	L=100 with FT=50 msec (5 sec)	0.5 sec <= L*FT <= 30 sec; upper limit for very unreliable test paths only
Load packet timeout (stop test)	1 sec	5 sec	0.250-30 sec; upper limit for very unreliable test paths only

Parameter	Default	Tested Range or Values	Expected Safe Range (not entirely tested, other values NOT RECOMMENDED)
Table index 0	0.5 Mbps	0.5 Mbps	When testing <= 10 Gbps
Table index 1	1 Mbps	1 Mbps	When testing <= 10 Gbps
Table index (step) size	1 Mbps	1 Mbps <= rate <= 1 Gbps	Same as tested
Table index (step) size, rate > 1 Gbps	100 Mbps	1 Gbps <= rate <= 10 Gbps	Same as tested
Table index (step) size, rate > 10 Gbps	1 Gbps	Untested	>10 Gbps
ss, UDP payload size, bytes	None	<=1222	Recommend max at largest value that avoids fragmentation; using a payload size that is too small 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 the implementation
tt, burst interval	100 microsec	100 microsec, 1 msec	Available range of "tick" values (HZ param)
Low delay range threshold	30 msec	5 msec, 30 msec	Same as tested
High delay range threshold	90 msec	10 msec, 90 msec	Same as tested
Sequence error threshold	10	0, 1, 5, 10, 100	Same as tested
Consecutive errored status report threshold	3	2, 3, 4, 5	Use values >1 to avoid misinterpreting transient loss

Parameter	Default	Tested Range or Values	Expected Safe Range (not entirely tested, other values NOT RECOMMENDED)
Fast mode increase, in table index steps	10	10	2 <= steps <= 30
Fast mode decrease, in table index steps	3 * Fast mode increase	3 * Fast mode increase	Same as tested

Table 1: Parameters for Load Rate Adjustment Algorithm

As a consequence of default parameterization, the Number of table steps in total for rates less than 10 Gbps is 1090 (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:

```
          UDRT + (2+w)*FT = Lost Status Backoff timeout

   where:

   UDRT = upper delay range threshold (default 90 msec)
   FT   = feedback time interval (default 50 msec)
   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:

The offered load **SHALL** then be decreased, following the same process as when the feedback indicates the 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 lost status feedback messages OR sequence errors OR delay variation can result in rate reduction and congestion confirmation.

The **RECOMMENDED** initial value for w is 0, taking a Round-Trip Time (RTT) of less than FT into account. A test with an RTT longer than FT is a valid reason to increase the initial value of w appropriately. Variable w **SHALL** be incremented by one whenever the Lost Status Backoff timeout is exceeded. So, with FT = 50 msec and UDRT = 90 msec, a status feedback message loss would be declared at 190 msec following a successful message, again at 50 msec after that (240 msec 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 setting (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 unsustainable Maximum Capacity observed when the Method of Measurement is searching for the maximum, and that would not do. This situation is different from the bimodal service rates (discussed in "Reporting the Metric", Section 6.6), 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 has an obvious alignment with the reporting units (Mbps).

Another approach comes from Section 24 of [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 affect service. But there is sufficient value in repeating a Sample with a fixed sending rate determined by the previous search for the Maximum 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 percent of the Maximum IP-Layer Capacity for I, or an indefinite period. The same Maximum Capacity Metric is applied, and the Qualification for the result is a Sample without supra-threshold packet losses or a growing minimum delay trend in subsequent Singletons (or each dt of the measurement interval, I). Samples exhibiting supra-threshold packet losses or increasing queue occupation require a repeated search and/or test at a reduced fixed Sender rate for Qualification.

Here, as with any Active Capacity test, the test duration must be kept short. Ten-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 makes a unique contribution to production testing. The Maximum IP Capacity Metric and Method for assessing performance is very different from the classic Throughput Metric and Methods provided in [RFC2544]: 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, Throughput measurements [RFC2544] 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 the methods discussed in [RFC2544] 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 widespread measurements that this memo encourages will encounter widespread 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 (for reasons of scale, this is not always possible; there is a limit on the number of test endpoints coming from many perspectives -- for example, management and measurement traffic). 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 measured path may be stateful based on many factors, and the Parameter "Time of day" when a test starts may not be enough information. Repeatable testing may require knowledge of 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, bytes sent, or both. Both load packets and status feedback messages **MUST** contain sequence numbers; this helps with measurements based on those packets.

Many different types of traffic shapers and on-demand communications 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 communications access and to discard the preamble from subsequent test results.

The following conditions might be encountered during measurement, where packet losses may occur independently of 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 the 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) is simultaneously present.
3. There may be only a small delay variation independent of the sending rate under these conditions as well.
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 adjustment 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's characteristics; the measurement community has known this for a long time and needs to keep it foremost in mind. Although the default is a single flow (F=1) for testing, the use of multiple flows may be advantageous for the following reasons:

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

Each flow would be controlled using its own implementation of the load rate adjustment (search) algorithm.

It is obviously counterproductive to run more than one independent and concurrent test (regardless of the number of flows in the test stream) attempting to measure the **maximum** capacity on a single path. The number of concurrent, independent tests of a path **SHALL** be limited to one.

Tests of a v4-v6 transition mechanism might well be the intended subject of a capacity test. As long as both IPv4 packets and IPv6 packets sent/received are standard-formed, this should be allowed (and the change in header size easily accounted for on a per-packet basis).

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

9. Reporting Formats

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

- Timestamp (especially the time when the maximum was observed in dtn).
- Source and Destination (by IP or other meaningful ID).
- Other inner Parameters of the test case (Section 4).
- Outer Parameters, such as "test conducted in motion" or other factors belonging to the context of the measurement.
- Result validity (indicating cases where the process was somehow interrupted or the attempt failed).
- 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 tabular format. There **SHOULD** be a column that identifies the test Phase. There **SHOULD** be a column listing the number of flows used in that Phase. The remaining columns **SHOULD** report the following results for the aggregate of all flows, including the Maximum IP-Layer Capacity, the Loss Ratio, the RTT minimum, RTT maximum, and other metrics tested having similar relevance.

As mentioned in Section 6.6, bimodal (or multi-modal) maxima **SHALL** be reported for each mode separately.

Phase	Number of Flows	Maximum IP-Layer Capacity (Mbps)	Loss Ratio	RTT min (msec)	RTT max (msec)
Search	1	967.31	0.0002	30	58
Verify	1	966.00	0.0000	30	38

Table 2: Maximum IP-Layer Capacity Results

Static and configuration Parameters:

The sub-interval time, dt, **MUST** accompany a report of Maximum IP-Layer Capacity results, as well as the remaining Parameters from Section 4 ("General Parameters and Definitions").

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 tabular format. There **SHOULD** be a column that identifies the test Phase. There **SHOULD** be a column listing each individual (numbered) flow used in that Phase, or the aggregate of flows in that Phase. A corresponding column **SHOULD** identify the specific sending rate sub-interval, stn, for each flow and aggregate. A final column **SHOULD** report the IP-Layer Sender Bit Rate results for each flow used, or the aggregate of all flows.

Phase	Flow Number or Aggregate	stn (sec)	Sender Bit Rate (Mbps)
Search	1	0.00	345
Search	2	0.00	289
Search	Agg	0.00	634
Search	1	0.05	499
Search	...	0.05	...

Table 3: IP-Layer Sender Bit Rate Results (Example with Two Flows and st = 0.05 (sec))

Static and configuration Parameters:

The sub-interval duration, st, **MUST** accompany a report of Sender IP-Layer Bit Rate results.

Also, the values of the remaining Parameters from Section 4 ("General Parameters and Definitions") **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 the placement of measurement systems within the Internet 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 Active 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 the 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 that 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 role or the Dst role.
2. It is **REQUIRED** to have a user client-initiated setup handshake between cooperating hosts that allows firewalls to control inbound unsolicited UDP traffic that goes to either a control port (expected and with authentication) or 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 are **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 duration values, I, **SHOULD** be limited).

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

11. IANA Considerations

This document has no IANA actions.

12. References

12.1. 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>.

[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>.

[RFC7497] Morton, A., "Rate Measurement Test Protocol Problem Statement and Requirements", RFC 7497, DOI 10.17487/RFC7497, April 2015, <https://www.rfc-editor.org/info/rfc7497>.

[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>.

12.2. Informative References

[copycat] Edeline, K., Kühlewind, M., Trammell, B., and B. Donnet, "copycat: Testing Differential Treatment of New Transport Protocols in the Wild", ANRW '17, DOI 10.1145/3106328.3106330, July 2017, <https://irtf.org/anrw/2017/anrw17-final5.pdf>.

[LS-SG12-A] "Liaison statement: 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", From ITU-T SG 12, March 2019, <https://datatracker.ietf.org/liaison/1632/>.

[LS-SG12-B] "Liaison statement: 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", From ITU-T-SG-12, May 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., "Maximum IP-Layer Capacity Metric, Related Metrics, and Measurements", Broadband Forum TR-471, July 2020, <https://www.broadband-forum.org/technical/download/TR-471.pdf>.

[Y.1540] ITU-T, "Internet protocol data communication service - IP packet transfer and availability performance parameters", ITU-T Recommendation Y.1540, December 2019, <https://www.itu.int/rec/T-REC-Y.1540-201912-I/en>.

[Y.Sup60] ITU-T, "Interpreting ITU-T Y.1540 maximum IP-layer capacity measurements", ITU-T Recommendation Y.Sup60, October 2021, <https://www.itu.int/rec/T-REC-Y.Sup60/en>.

Appendix A. Load Rate Adjustment Pseudocode

This appendix provides a pseudocode 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 that includes Loss or Reordering
                    # or Duplication impairments (all appear
                    # initially as errors in the packet sequence
                    # numbering)

seqErrThresh = 10   # Threshold on seqErr count that includes Loss or
                    # Reordering or Duplication impairments (all
                    # appear initially as errors in the packet
                    # sequence numbering)

delay = 0           # Measured Range of Round Trip Delay (RTD), msec

lowThresh = 30      # Low threshold on the Range of RTD, msec

upperThresh = 90    # Upper threshold on the Range of RTD, msec

hSpeedThresh = 1    # Threshold for transition between sending rate
                    # step sizes (such as 1 Mbps and 100 Mbps), Gbps

slowAdjCount = 0    # Measured Number of consecutive status reports
                    # indicating loss and/or delay variation above
                    # upperThresh

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

Appendix B. RFC 8085 UDP Guidelines Check

Section 3.1 of [RFC8085] (BCP 145), which provides UDP usage guidelines, focuses primarily on congestion control. The guidelines appear in mandatory (**MUST**) and recommendation (**SHOULD**) categories.

B.1. Assessment of Mandatory Requirements

The mandatory requirements in Section 3 of [RFC8085] include the following:

> 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 described 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 of applications described in Section 2. There is tension between the goals of probing conservatism and minimization of 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; this 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 methods described in this memo use 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 a longer interval (feedback message or load packet timeouts).

This memo does not foresee any specific benefit of using Explicit Congestion Notification (ECN).

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 Metrics and Methods do not require reliable delivery.

> Applications that require ordered delivery **MUST** reestablish datagram ordering themselves.

The IP-Layer Capacity Metrics and Methods do not need to reestablish packet order; it is preferable to measure packet reordering if it occurs [RFC4737].

B.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 **SHOULD**-level requirements as listed in [RFC8085], of which many are intended for long-lived flows that must coexist with other traffic in a 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] (regarding the implications of RTT and loss measurements on congestion control) says:

> A congestion control [algorithm] designed for UDP **SHOULD** respond as quickly as possible when it experiences congestion, and it **SHOULD** take into 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 following:

> 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 requirements from [RFC8085] follows (marked "Yes" when this memo is compliant, or "NA" (Not Applicable)):

Yes?	Recommendation in RFC 8085	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	

Yes?	Recommendation in RFC 8085	Section
	For bulk transfers,	**3.1.2**
NA	**SHOULD** consider implementing TFRC	
NA	else, **SHOULD** in other ways use bandwidth similar to TCP	
	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	3.1.10
NA	non-CC controlled flows **SHOULD** implement a transport circuit breaker	
Yes	**MAY** implement a circuit breaker for other applications	
	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,	**3.1.11**
NA	**SHOULD** perform CC or use circuit breaker	
NA	**SHOULD** restrict types of traffic transported by the tunnel	

Yes?	Recommendation in RFC 8085	Section
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	5.1
	else, **MAY** use UDP-Lite with suitable checksum coverage	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)	
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?	Recommendation in RFC 8085	Section
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

Table 4: Summary of Key Guidelines from RFC 8085

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 this memo and related topics. In a second round of reviews, we acknowledge Magnus Westerlund, Lars Eggert, and Zaheduzzaman Sarker.

Authors' Addresses

Al Morton
AT&T Labs
200 Laurel Avenue South
Middletown, NJ 07748
United States of America
Phone: +1 732 420 1571
Email: acm@research.att.com

Rüdiger Geib
Deutsche Telekom
Heinrich Hertz Str. 3-7
64295 Darmstadt
Germany
Phone: +49 6151 5812747
Email: Ruediger.Geib@telekom.de

Len Ciavattone
AT&T Labs
200 Laurel Avenue South
Middletown, NJ 07748
United States of America
Phone: +1 732 420 1239
Email: lencia@att.com